Thèse de doctorat

de l'Université Sorbonne Paris Cité

Préparée à l'Université Paris Diderot

Ecole doctorale HEMATOLOGIE ONCOGENESE ET BIOTHERAPIES

(**HOB – ED561**)

Unité Epigénétique et Destin Cellulaire UMR7216

The ICF syndrome and emergent players in DNA methylation and development: when studying a rare genetic disease sheds new light on an "old" field

Syndrome ICF et acteurs émergents dans la méthylation de L'ADN et le développement: l'étude d'une maladie génétique rare apporte un regard nouveau sur un « ancien » domaine

Par **Giacomo GRILLO**

Thèse de doctorat de Hématologie et Oncologie

Dirigée par le Dr. Claire FRANCASTEL et le Dr. Guillaume VELASCO

Présentée et soutenue publiquement à Paris le 06 Juillet 2017

Président du jury :	VEITIA, Reiner / Professeur / Université Paris Diderot
Rapporteurs :	BOURC'HIS, Deborah / Docteur / Institut Curie
	PETERS, Antoine / Professeur/ FMI (Suisse)
Examinateurs :	GENDREL, Anne-Valerie / Docteur / Institut Curie
	CAVAILLE, Jérôme / Docteur / Université Paul Sabatier Toulouse III
Directrice de thèse :	FRANCASTEL, Claire / Docteur / Université Paris Diderot
Co-directeur de thèse :	VELASCO, Guillaume / Docteur / Université Paris Diderot

To all the people who supported me during all these years, even for a single moment…

"Joy in looking and comprehending is nature's most beautiful gift"
(Albert Einstein)

"Se oggi capissi tutto, domani non avrei nulla da fare"
(mamma)

ACKNOWLEDGEMENTS

These four years of PhD have been a great adventure. I am very happy to start to say “thank you”. I want to thank all the people who supported me and did a part of this great voyage with me.

First of all, I want to thank all the members of my jury for accepting to evaluate my work. Many thanks to my rapporteurs Dr Deborah Bourc’his, who accepted despite a very bad mistake in my very first mail, and Pr Antoine Peters for accepting to revise my heavy manuscript and for their interest in my work. Many thanks to the examiteurs Dr Anne-Valerie Gendrel and Dr Jerome Cavaille for the interest they showed in my work. Many thanks also to the president of my jury, Pr Reiner Veitia for your suggestions and help.

Many many thanks to “la chef” Dr Claire Francastel for everything. I arrived in your team for my M2 without any experience in a research lab, without knowing anything but very interested in all the things that you could teach me. Thank you for all the time you gave me, in the lab, at home working on my abstracts, posters, fourth year projects, presentations and in particular on my thesis. Thank you for the time we spent together outside the lab, drinking beers or having barbeques. Thank you because you always push me to do my best, to give everything I could give to a project, for a result, for an intuition, for a collaboration and to keep my mind open and not closing myself on my project. As we discussed once after the submission of a paper: you taught me a lot and now I am ready for the step forward. Thank you very much also because you helped me to defend my thesis before summer, although it was not your idea. THANKS CHEF!

I am very grateful to great equipe Fracasse! All and each of you, past and actual members. Many thanks to my PhD co-director Guillaume Lelasco. No, it is not a typo. I wrote it with the “L” as it is on your access badge! I am still laughing! Thank you very much for everything! Thank you because you contributed to lead me during this adventure. You taught me a lot about techniques, way of thinking, how to risk on an experiment, how to deal with collaborators, how to organize my work, my thoughts and how to give everything the best I had. I am grateful because during my PhD you have been a great co-director, a friend (teaching me hilarious French way of saying, crazy movies, crazy situations in serious movies), a brother and sometimes even a father. Thank you for your jokes (you are famous in all the unit for these!), games with words and for sharing with me your passions, like climbing (and not only…). Florentino!!! Many thanks! You really contributed to my formation, you helped me a lot in some critical moments of my project, especially at the beginning and you gradually become more than a simple colleague. I was very surprised, grateful and moved (and I am still now) when you gave me as present for my birthday a cup saying “Il faut toujours viser la lune, car même en cas d’échec, on atterrit dans les étoiles” saying “you ARE this sentence! Do not forget it! And, in particular, manage to arrive on the

moon!". It is crazy to think that my parents grew me saying the same thing. Thank you because you shared with me many beers, many thoughts, many doubts and also the art of barbecue. THANK YOU! Many thanks to Ivanna! Also in this case it is not a typo but the way I am used to call you, causing your laugh! Many thanks for your help during these years in the lab, for your sincere friendship and because you shared with me thoughts and things that were happening to you. Thank you very much! You have been very patient and brave to come to have lunch with me Guillaume, Miguel, Tharvesh and the others listening to our high-level discussions…I am happy that sometimes you have even participated! Many thanks also to the last arrived Baptiste for your joyful way of working, your way of explaining experiments and results during lab meeting. You are in the perfect place to grow up and have a great PhD. Good luck! Many thanks to all the past members of the equipe Fracasse. Thanks to Alain for sharing protocols, for teaching me to work with radioactivity, for great discussion about science and also about travels. Many thanks also to the M2 students Thomas and Nadiia because somehow working with you I learnt something about myself, about my way of working and about how I organize my time and my work. Many many thanks to Emma, who taught me how to work during the first steps of my M2, about ES cells cultivation and all the basic techniques of the equipe Fracasse. Last, but not least of course, many many thanks to Sardine (not a typo!)! I am very grateful to have worked with you on your project, for the discussion about our projects, about life, about difficulties and about future. Thank you very much because you showed me great places in Paris, not only for eating but also places out of my knowledge to relax and enjoy Paris. I am very grateful to you who has become an elder sister.

I am grateful to all the great UMR7216. It has been a real pleasure to work with you, to share with you weeks and weekends of work. Thanks to Jonny for the quick discussions we had in the hall of the unit. Many thanks to Slimane and his team, past and actual members. Sharing the cell culture room has been the occasion of great discussion about science and life. Many thanks to the "crew of the cantine" Miguel, Christophe and Tharvesh and in particular to Miguel and Tharvesh because you really take care of me asking what was wrong or good depending on the situation and on the color of my face! Thank you very much! Many thanks to all the PhD students of the unit for discussion bout everything and for sharing the great and tough time of PhD. A special thanks to Robi for sharing ideas, protocols, ideas, evenings, beers and so on. Thank you very much for your friendship and your endless moral support! Many thanks to Florent LP, who started a new adventure for his friendship and his support.

I am grateful to the Ministère de la recherche et de l'enseignement superieur as well as the Fondation ARC pour la recherche sur le cancer for believing in me and my project, allowing me to work for four intense years on this project. Many thanks to all the collaborators, who worked with us and share with us projects, techniques, skills and results.

RINGRAZIAMENTI

Un enorme grazie a tutti gli amici che mi hanno accompagnato durante questa avventura parigina della tesi, ognuno con il suo carattere e temperamento. Grazie a tutti quelli che da Milano non hanno mai fatto mancare il loro sostegno e il loro affetto. Grazie a Vitto, Elisa, Tommo, Alice, Elenina, Pizzo, Tia, Merik, Zanche…purtroppo non riesco a nominarvi tutti. Grazie a tutti quelli che diverse parti nel mondo mi hanno accompagnato e sostenuto durante il dottorato. Grazie infinite a tutti gli amici a Parigi e più in generale in Francia. Un grazie speciale a Clara e Virgilio per la loro sempre calorosa amicizia e attenzione. Grazie di cuore ai miei coinquilini Andre, Frarro e Deca (anche se non più coinquilino sempre presente e sempre signore!). Ne abbiamo passate di tutti i colori insieme in appartamento e fuori, da 30 metri quadri a 100, serate, cene, fughe d'acqua…grazie perché mi avete fatto sentire a casa e mi avete aiutato a trattare l'appartamento e me stesso con dignità. Siete un pezzo di famiglia che ho avuto la fortuna di trovare lungo la mia strada. Grazie davvero! Grazie mille a tutti gli amici di Comunione e Liberazione, in particolare Cri, Luis, Natalie, la Stef (si ci vuole l'articolo!), i Turi, Roberta, Cate cuoca, Cate Vaglio, Valmo, Pit, Leo e tutti quanti voi. Grazie mille a Mimmo per averci dato un posto dove vivere e per i suoi saluti e il suo modo di parlare sempre pieno di attenzione ed entusiasmo. Grazie di vero cuore per tutto! Grazie a tutti coloro, che anche per poco, hanno condiviso un pezzo di vita con me.

Grazie dal profondo del mio cuore ai compagni del master, poi diventati compagni del dottorato e compagni di vita. In particolare grazie a Teo, Madda, Ale, Giulia e Lara. Grazie delle serate, delle risate, delle discussioni e di tutto tutto. Un grazie particolare a Madda, a tratti coinquilina aggiunta quando tu e Franci mi avete ospitato per la scrittura della tesi e non solo. Grazie di cuore per la tua amicizia! Un altro grazie particolare a Teo per le grasse risate, le serate, le bevute e anche per il lavoro fatto insieme. Grazie di tutto cuore per la tua generosità e amicizia! Cerca per favore di non uscire dal lavoro prima delle 16 però! Ahahah!

Un grazie tutto particolare va a Bubba, Tommi e Bella per le continue provocazioni sul vivere alla grande e sul non lasciar perdere nulla. Vi sento fratelli aggiunti, siete stati spesso determinanti e sono grato e onorato di camminare verso una decisione molto importante per la mia vita con voi accanto. Grazie di cuore a tutta la mia famiglia. Grazie a mia mamma, mio papà e ai miei fratelli Gabri e Gio È difficilissimo esprimere quello che provo…grazie perché mi avete sempre sostenuto, avete gioito e sofferto con me perché non avete mai smesso di mostrarmi grande affetto e stima e mi avete sempre dato grande energia per non abbattermi mai. Lontani ma mai distanti. Da grande vorrei tanto avere una famiglia così! Grazie ai cari nonni, zie e zii, cugini e cugine e tutti gli amici di famiglia per il loro affetto, il loro supporto e a volte per la loro sorpresa nel vedermi a Milano. Grazie di cuore! Grazie di cuore alla famiglia di Francesca per il loro

sostegno, la loro fiducia, la loro pazienza e la loro stima nei miei confronti. Grazie perché siete diventati famiglia con una grande semplicità. Grazie di cuore!

Dulcis in fundo grazie alla paziente, incredibile, sempre presente, santa Francesca. Sei sempre stata accanto a me per aiutarmi, insultarmi (quando ci vuole, ci vuole…e spesso ci vuole!), per sostenermi, per consolarmi, per spronarmi a non mollare mai e se mai mollavo a rialzarmi in fretta. Grazie perché mi sopporti e hai voluto accettare di camminare con me per l'eternità. Chissà cosa ci riserverà la vita…sono molto curioso di scoprirlo con te che sei casa.

Ultimo, ma non meno importante, grazie di cuore a Chi ha concesso tutto questo, che mi ha accompagnato sempre senza mai lasciarmi, facendosi vivo in tutti i nomi che ho citato e che, aihmè ho dimenticato.

L'avventura continua…

TABLE OF CONTENTS

ABBREVIATIONS **9**
ABSTRACT **11**
RESUME **12**
PROLOGUE **13**
PREAMBULE **15**
PROLOGO **17**

INTRODUCTION 19
Chapter I: Genomic features of mammalian genomes **20**
1. Historical overview 20
2. ENCODE 2012: Over 80% of the human genome might be “functional” 22
3. Interspersed DNA repeats: transposable elements 24
3.1. Genomic distribution of TEs 26
3.2. Transposable elements as “functional elements” 27
4. Tandem repeats: centromeres, pericentromeric regions and telomeres 28
Chapter II: Epigenetics and chromatin **31**
1. Historical overview of epigenetics 31
2. Chromatin 32
2.1. Historical overview of chromatin 32
2.2. Chromatin structure and compaction 34
2.3. Histone modifications 36
2.4. Histone variants, chromatin identity and dynamic 38
2.4.1. Histone variants in transcription and DNA repair 38
2.4.2. Histone variants that define specialized chromosomal regions 39
2.5. Histone writers, readers and erasers 41
2.6. The ENCODE project 43
3. The special case of heterochromatin 44
3.1. Facultative heterochromatin: H3K27me3 and Polycomb 44
3.2. Constitutive heterochromatin: H3K9me3 and H3K20me3 46
3.3. Non-coding RNA and assembly of heterochromatin 48
3.3.1. The long non-coding RNA XIST (X-inactive specific transcript) and inactive X heterochromatin 48
3.3.2. (peri)centromeric transcripts and heterochromatin assembly 49

3.3.3. Telomeric repeat-containing RNA (TERRA) and telomeric heterochromatin 51
4. DNA methylation 52
4.1. Techniques to study DNA methylation 52
4.1.1. Restriction-based 53
4.1.2. Fluorescence-based methods 53
4.1.3. Affinity-based methods 54
4.1.4. DNA methylation arrays 54
4.1.5. Sequencing-based methods 54
4.2. CpG context of DNA methylation 55
4.3. Genomic context of CpG methylation 56
4.3.1. Methylation at promoters 56
4.3.2. Methylation at gene bodies 57
4.3.3. Methylation at DNA repeats 57
4.3.4. Methylation at monoallelically expressed regions: the case of genomic imprinting and X chromosome inactivation 58
4.4. Dynamics of DNA methylation and its players 59
4.4.1. Reprogramming of DNA methylation during development 59
4.4.2. DNA methylation activities: *de novo* and maintenance 59
4.4.3. Mechanisms of DNA demethylation 62
4.4.4. 5hmC: just an intermediate of DNA demethylation? 63
4.5. The impact of DNA methylation in the different genomic contexts 64
4.5.1. On promoters and transcription start sites 64
4.5.2. At gene bodies 65
4.5.3. At tandem and interspersed DNA repeats 65
4.5.4. At Imprinted Control Regions (ICRs) 66
4.6. Readers of DNA methylation 67
4.6.1. 5mC readers 67
4.6.2. Methylation-sensitive transcription factors 68
4.6.3. Protein binding both methylated and unmethylated DNA 69
4.6.4. 5hmC, 5fC and 5caC readers 69
4.7. Interplay between DNA methylation and histone methylation 70
4.7.1. DNA methylation and H3K4 methylation at promoters 70
4.7.2. H3K36 methylation and DNA methylation in gene bodies 71
4.7.3. DNA methylation and H3K9me3: old good friends 72

4.7.4. Mutual exclusion between DNA methylation and H3K27me3 75
4.8. DNA methylation in normal development 76
4.8.1. Analyses of DNA methylation *in vivo* 76
4.8.2. Analyses of DNA methylation *ex vivo* in the mouse – culture systems . 78
Chapter III: DNA methylation in physiopathology 84
1. DNA methylation in physiology 84
1.1. DNA methylation and aging 84
1.2. DNA methylation and environment 86
1.2.1. Benzo[a]pyrene (BaP) and DNA methylation 86
1.2.2. Alimentation and DNA methylation: the case of folate 87
1.2.3. Chemotherapy: 5-azacytidine, what about normal cells? 89
2. DNA methylation in pathology 90
2.1. Somatic alterations 91
2.1.1. Autoimmune diseases 91
2.1.2. Cancer 92
2.1.3. Stress: a model to study the causal link between DNA methylation, transcription of pericentromeric and centromeric repeats and perturbed phenotypes .. 96
2.2. Germline mutations 97
2.2.1. Mutations in DNMT3A cause an "Overgrowth syndrome with intellectual disability"…. 98
2.2.2. Mutations in DNMT1 cause an "Hereditary sensory neuropathy with dementia and hearing loss" 98
2.2.3. Mutations in MECP2: Rett and Angelman syndromes 99
2.2.4. The Immunodeficiency, Centromeric instability and Facial anomalies syndrome……..... 100
2.2.4.1. Historical overview 100
2.2.4.2. Diagnosis, clinical and molecular features 101
2.2.4.3. Mutations in DNMT3B cause ICF1 syndrome 102
2.2.4.4. Mutations in the zinc finger ZBTB24 genes cause ICF2 syndrome……..... 105
OBJECTIVES 108

RESULTS 109

ARTICLE 1 Mutations in CDCA7 and HELLS cause immunodeficiency–centromeric instability–facial anomalies syndrome 110

ARTICLE 2 Comparative methylome analysis of ICF patients: ZBTB24, CDCA7 and HELLS are required for DNA methylation at heterochromatic loci..... 128

UNPUBLISHED RESULTS..... 230

1. ZBTB24 is essential for embryonic development but is dispensable for ES cell differentiation..... 231

2. Loss of ZBTB24 functions impairs DNA methylation at tandem and interspersed DNA repeats..... 233

3. ZBTB24 function is important for the recruitment of DNMTs..... 236

4. ZBTB24 controls DNA methylation at interspersed DNA repeats LINE1 and IAP by promoting H3K9me3 establishment..... 237

MATERIALS AND METHODS (Unpublished results)..... 259

DISCUSSION AND PERSPECTIVES..... 269

DISCUSSION..... 270

PERSPECTIVES..... 280

REFERENCES..... 282

ANNEXES..... 340

ANNEXE 1 341

ANNEXE 2 386

ANNEXE 3 413

ABBREVIATIONS

5caC – 5-carboxyl-cytosine

5fC – 5-formyl-cytosine

5hmC – 5-hydroxy-methyl-cytosine

5hmU – 5-hydroxy-methyl-uracil

5mC – 5-methyl-cytosine

AVE – Anterior Visceral Endoderm

AZA – 5-azacytidine

BAC – Bacterial Artificial Chromosome

BaP – Benzo[a]pyrene

BER – Base Excision Repair

BPA – Bisphenol-A

BS – Bisulfite Sequencing

CENP-A – Centromeric protein A

CFP1 – CXXC Finger Protein 1

CGI – CpG Island

ChIP – Chromatin Immuno-Precipitation

COBRA – Combined Bisulfite Restriction Analysis

CTCF – CCCTC-binding factor

DNMT – DNA methyltransferase

Dpc – days post coitum

DVE – Distal Visceral Endoderm

ENCODE – Encyclopedia of DNA Elements

EPC – Ectoplacental Cone

Epi – epiblast cells

EpiSC – Epiblast Stem Cells

ERV – Endogenous Retroviral Elements

ES – Embryonic Stem

ExE – extra-embryonic ectoderm

FISH – Fluorescence In Situ Hybridization

HAT – Histone Acetyl-Transferase

HCP – High CpG promoter

HDAC – Histone De-Acetylase

hESC – human Embryonic Stem Cell

HM – Human Methylation BeadChip

HMT – Histone Methyl-Transferase

HP1 – Heterochromatin Protein 1

IAP – Intracisternal A-Particle

ICF – Immunodeficiency Centromeric instability and Facial anomalies

ICM – Inner Cell Mass

ICP – Intermediate CpG promoter

ICR – Imprinting Control Region

Kb – Kilobase

KMD – Lysine Methyl-Demethylase

KMT – Lysine Methyl-transferase

KRAB – Kruppel-Associated box

LCP – Low CpG Promoter

LIF – Leukemia Inhibitory Factor

LINE – Long Interspersed Nuclear Elements

LTR – Long Terminal Repeat

MBD – Methyl-Binding Protein

MeCP – Methyl-CpG-binding protein

MeDIP – Methylated DNA Immuno-Precipitation

MEF – Mouse Embryonic Fibroblast

mRNA – messenger RNA

MSRE – Methylation-Sensitive Restriction Enzymes

ncRNA – non-coding RNA

NER – Nucleotide Excision Repair

NuRD – Nuclear Remodelling and Deacetylation Complex

ORF – Open Reading Frame

PRC – Polycomb Repressive Complex

PrE – Primitive Endoderm

REST - RE1-silencing transcription factor

RITS – RNA-Induced Transcriptional Silencing

RMT – Arginine Methyla-Transferase

RNA Pol – RNA Polymerase

RNAi – RNA interference

RNP – ribonucleoprotein

RRBS – Reduced-Resolution Bisulfite Sequencing

rRNA – ribosomal RNA

SAM – S-Adenosyl Methionine

Sat – Satellite repeats

SINE – Short Interspersed Nuclear Elements

SRP – signal recognition particle ribonucleoprotein complex

SUV39HDN – SUV39H1/2 Double knock-out

TDG – Thymine DNA Glycosylases

TE – Transposable Elements

TEm - Trophectoderm

TERRA – telomeric repeat-containing RNA

TET – Ten-Eleven Translocation enzymes

TKO – DNMT triple knock-out

TLC – telocentric tandem repeats

TPRT – Target-Primed Reverse Transcription

TRD – Transcriptional-Repression Domain

tRNA – tranfer RNA

TSA – Trichostatin A

UTR- UnTranslated Region

WGBS – Whole Genome Bisulfite Sequencing

XIST – X-Inactive Specific Transcript

ZBTB – zinc finger and Broad complex, Tramtrack, and Bric-a`-brac

ABSTRACT

DNA methylation is an essential process for the development of mammals. Its abnormal distribution, particularly at the level of the repeated regions of the genome, is a pathological signature. The discovery of hereditary diseases affecting DNA methylation and the stability of the genome allowed a considerable progress in the identification of their actors and mechanisms. We chose to study the ICF (Immunodeficiency, Centromeric Instability and Facial Abnormalities) syndrome, the first genetic disorder identified with defects in the distribution of DNA methylation, linked to chromosomal instability. When I started my PhD, mutations in two genes had been described to cause the ICF syndrome: DNMT3B and ZBTB24. However, the genetic origin of a subset of ICF patients remained unknown. We identified mutations in CDCA7 and HELLS as causative of the ICF syndrome. I showed that their loss of function in somatic cells results in the loss of DNA methylation at centromeric repeats, strongly suggestive of a role DNA methylation maintenance. Hence, the study of the aetiology of a genetic disease provided new candidate "guardians" of DNA repeats and genome stability, with virtually unknown functions but with exciting potential roles in the DNA methylation machinery and in development. During my PhD, I established methylation maps in ICF patients cells to identify common and distinct targets of these factors, as well as their genomic and epigenomic characteristics. In contrast to DNMT3B mutations, those in ZBTB24, CDCA7 and HELLS affect methylation at CpG-poor regions in intergenic genomic locations and at interspersed DNA repeats, and more generally, at genomic locations with heterochromatic features. Their integrity is required for the methylated status of coding and non-coding clusters of genes, some of which are expressed in a monoallelic manner. To better characterize the role of ZBTB24 in development and DNA methylation pathways, we generated a mouse model carrying mutations in ZBTB24. We showed that ZBTB24 is essential for early development, while it seemed to be dispensable for *in vitro* differentiation of murine ES cells. We implicated ZBTB24 in the establishment of DNA methylation at DNA repeats, both in tandem or interspersed, in differentiating ES cells. Interestingly, ZBTB24 seems to be also implicated in the establishment of the repressive mark H3K9me3 suggesting that ZBTB24 may indirectly control DNA methylation through an interplay with histone marks. As a whole, our work sheds light on how DNA methylation and heterochromatin marks are established and maintained at unique genes and DNA repeats, and provides new actors and mechanisms to consider in studies of the maintenance of genome stability.

RESUME

La méthylation de l'ADN est un processus vital pour le développement des mammifères. Sa distribution anormale, notamment au niveau des régions répétées du génome, est une signature pathologique. La découverte de maladies héréditaires touchant la stabilité du génome a permis des avancées considérables dans l'identification des acteurs et des mécanismes. Nous avons choisi d'étudier le syndrome ICF (Immunodéficience, instabilité Centromérique et anomalies Faciales), première maladie génétique identifiée avec des défauts de la méthylation de l'ADN, liés à une instabilité chromosomique. Lorsque j'ai commencé ma thèse, des mutations dans les gènes DNMT3B et ZBTB24 avaient été décrites comme causes génétiques du syndrome. Cependant, d'autres causes génétiques restaient inconnues. Nos travaux ont permis d'identifier deux nouveaux gènes, CDCA7et HELLS, dont les mutations sont responsables du syndrome. J'ai montré que leur perte de fonction dans les cellules somatiques entraîne un défaut de méthylation des répétitions centromériques, suggérant leur rôle dans le maintien de la méthylation de l'ADN. Par conséquent, l'étude de l'étiologie d'une maladie génétique rare a permis d'identifier de nouveaux «gardiens» de la stabilité du génome, avec des fonctions jusqu'alors insoupçonnées dans les processus de méthylation de l'ADN et dans le développement. Au cours de mon doctorat, j'ai établi des cartes de méthylation des cellules de patients ICF afin d'identifier les cibles communes et distinctes de ces facteurs, ainsi que leurs caractéristiques génomiques et épigénomiques. Contrairement aux mutations de DNMT3B, celles de ZBTB24, CDCA7 et HELLS affectent la méthylation dans des régions pauvres en CpG, dans des régions intergéniques et dans des répétitions d'ADN intercalées. Plus généralement, ce sont les régions d'hétérochromatine qui sont les plus touchées et en particulier des clusters des gènes codants et non codants, dont certains sont exprimés de manière monoallélique. Pour mieux caractériser le rôle de ZBTB24 dans le développement et la méthylation de l'ADN, nous avons généré un modèle murin mutant qui nous a permis de monter que ZBTB24 était essentielle pour le développement embryonnaire précoce. De plus, ZBTB24 jouerait un rôle dans l'établissement de la méthylation des séquences répétées de l'ADN, à la fois en tandem ou intercalé. Fait intéressant, ZBTB24 semble être également impliqué dans l'établissement de la marque répressive H3K9me3, suggérant un rôle de la protéine dans le "dialogue" entre la méthylation de l'ADN et celle des histones. Dans l'ensemble, mon travail met l'accent sur la façon dont la méthylation de l'ADN et les marques d'hétérochromatine sont établies et maintenues à des gènes uniques et des répétitions de l'ADN, et fournit de nouveaux acteurs et mécanismes à considérer dans les études sur le maintien de la stabilité du génome

PROLOGUE

Our genome contains all the information necessary for the development of a healthy subject. Maintenance of its integrity is fundamental for survival and normal functioning of the whole organism and protection from the emergence of disease. Still, our genome is in constant interaction with external and internal environments, which is required for normal development and health maintenance, but may also jeopardize its normal functions.

Mutations in our genome have different impact depending on which region and which cell type are affected. Mutations arising in somatic cells can be acquired during an individual's life, are present only in certain cells, may lead to sporadic or acquired diseases and are not passed on to the next generation. In contrast, the so-called "hereditary mutations" are present in the parent's egg or sperm cells, which are also termed germ cells. When an egg and a sperm cell meet, the resulting fertilized egg cell receives the genome of both parents. If the genomic DNA contains a mutation, the child who grows from the fertilized egg will carry the mutation in each cell of his body. It is possible then to inherit mutations that alter a protein function, leading to disruption of normal development or causing disease. A famous example is the inheritance of mutations leading to the loss of function of the Dystrophin gene implicated in muscle contraction, that cause the emergence of the Duchenne Myotonic Dystrophy, characterized by progressive muscle wasting and weakness.

The correct exploitation of the information carried by our genome is insured by chemical modifications that are collectively defined as the epigenome, literally meaning "above genes". As a consequence, mutations in genes coding for factors implicated in the regulation of the epigenome lead to more pleiotropic effects. The first of these epigenetic modifications to have been described in 1950 is the methylation of DNA. DNA methylation is essential for human development and alterations of this modification characterizes many diseases like cancer, which in general arises from somatic mutations, or like the inherited ICF syndrome (Immunodeficiency, Centromeric instability and Facial anomalies) caused by germ line mutations that affect DNA methylation landscapes. ICF patients show developmental delay and a primary immunodeficiency that can cause the death of patients during childhood, and a typical chromosomal instability that is also a hallmark of cancer cells.

Hence, the interest of studying such a rare syndrome is at least two-fold: first, it allows to better characterize the disease and to provide new biomarkers with diagnosis value. Second, the study of monogenic diseases provides a unique opportunity to identify new factors and mechanisms implicated in DNA methylation with added relevance for other pathological cases characterized by alteration of DNA methylation patterns. As a whole, studying the aetiology of an

extremely rare disease like the ICF syndrome shed new light on the "old" field of DNA methylation. As Marcel Proust said:

"Le véritable voyage de découverte ne consiste pas à chercher de nouveaux paysage,
mais à avoir de nouveaux yeux
(The real voyage of discovery does not consist in seeking new landscapes,
but in having new eyes)"
Marcel Proust

During my PhD, I did not look for new landscapes because DNA methylation has been studied for almost 60 years now. But, the study of the ICF syndrome provided me with the "new eyes" to tackle some unknowns that remain in the field.

In the following introduction, I will briefly expose the main features of the mammalian genome, the history of the study of epigenetic modifications with a particular attention to DNA methylation, to describe its distribution, its impact on the different genomic regions where it occurs and its importance in a normal physiological context. Then, I will describe how DNA methylation profiles can be altered by the effects of the environment or the alterations that characterize human pathologies like cancer, autoimmune diseases and the ICF syndrome.

PREAMBULE

Notre génome contient l'information nécessaire à la formation, au développement et au bon fonctionnement de notre organisme. Il est donc essentiel d'empêcher que l'intégrité de cette information soit compromise au risque et de voir émerger des pathologies.

L'altération de cette information par des mutations peut être héritée des parents et transmise à la descendance via les gamètes, ou peut être acquise au cours de la vie sous l'effet de facteurs environnementaux auxquels notre organisme est constamment soumis et doit faire face. Qu'elles soient transmises ou acquises, ces mutations peuvent être délétères si elles touchent des gènes essentiels aux grandes fonctions de notre organisme et à l'homéostasie cellulaire, conduisant ainsi à des pathologies dont les cancers sont un des exemples les plus connus et étudiés.

L'étude des causes génétiques des pathologies, et en particulier des maladies monogéniques, permet d'identifier les gènes dont le rôle est essentiel au bon fonctionnement de l'organisme. La Dystrophie de Duchenne, caractérisée par une perte progressive des fonctions musculaires, est un exemple célèbre de pathologie causée par la transmission de mutations délétères et la perte de fonction de la protéine Dystrophine impliquée dans le processus de contraction musculaire.

L'exploitation correcte de l'information portée par notre génome est assurée par des modifications chimiques qui le décorent et régulent sont expression. L'ensemble de ces modifications est défini comme l'épigenome, signifiant littéralement « au-dessus des gènes ». Il en découle que les mutations des gènes codant pour les acteurs de la régulation de l'épigénome conduisent à des effets pléiotropiques. La première des modifications épigénétiques à avoir été décrite en 1950, et la plus étudiée à ce jour, est la méthylation de l'ADN. La méthylation de l'ADN est essentielle pour le développement des mammifères et des altérations de cette modification sont décrites dans de nombreuses maladies comme les cancers, ou des maladies monogéniques comme le syndrome héréditaire ICF (Immunodéficience, instabilité Centromérique et anomalies Faciales). Ce syndrome est causé par des mutations causant la perte de fonction d'acteurs de la methylation de l'ADN. Les patients ICF présentent un retard de développement et une immunodéficience primaire qui peuvent causer la mort des patients pendant la petite enfance. Les cellules des patients présentent une instabilité chromosomique typique due à des défauts de méthylation de l'ADN, également caractéristique des les cellules cancéreuses.

Par conséquent, l'intérêt d'étudier un syndrome aussi rare est au moins double : d'abord, pour mieux comprendre les bases moléculaires de la maladie et ainsi fournir de nouveaux biomarqueurs pertinents pour le diagnostic. Deuxièmement, comme évoqué précédemment, cette maladie monogénique offre une occasion unique d'identifier de nouveaux acteurs et mécanismes impliqués

dans la méthylation de l'ADN, ce qui est pertinent pour d'autres contextes pathologiques comme les cancers en général ou certains troubles immunologiques et mentaux.

Dans l'ensemble, l'étude de l'étiologie d'une maladie extrêmement rare comme le syndrome ICF apporte une lumière nouvelle sur le « vieux » domaine de la méthylation de l'ADN. Comme le disait Marcel Proust :

"Le véritable voyage de découverte ne consiste pas à chercher de nouveaux paysage, mais à avoir de nouveaux yeux"
Marcel Proust

Au cours de mon doctorat, je n'ai pas cherché de « nouveaux paysages », car la méthylation de l'ADN est étudiée depuis près de 60 ans maintenant. Mais l'étude du syndrome ICF m'a apporté de « nouveaux yeux » qui m'ont permis d'accroitre notre compréhension des bases moléculaires du maintien de l'intégrité du génome humain.

Dans l'introduction qui suit, j'exposerai brièvement les principales caractéristiques du génome des mammifères, l'historique des études des modifications épigénétiques avec une attention particulière sur la méthylation de l'ADN pour décrire sa répartition, son impact sur les différentes régions génomiques concernées ainsi que son importance dans un contexte physiologique normal. Ensuite, je décrirai comment les profils de méthylation de l'ADN peuvent être affectés par les effets de l'environnement ou les altérations génétiques caractéristiques de maladies humaines comme les cancers, les maladies auto-immunes et le syndrome ICF.

PROLOGO

Il nostro genoma contiene tutte le informazioni necessarie per la formazione, lo sviluppo e il buon funzionamento di un individuo. È quindi necessario evitare che l'integrità di questa informazione sia compromessa, rischiando di provocare l'emergenza di malattie.

L'alteramento di questa informazione a causa di mutazioni può essere ereditata dai genitori e trasmessa alle generazioni successive tramite i gameti, oppure può essere acquisita durante il corso della vita, sotto l'effetto di fattori ambientali che il nostro organismo deve costantemente affrontare. Sia che le mutazioni siano trasmettibili o acquisite, queste mutazioni possono essere deleterie se colpiscono gene essenziali per le funzioni del nostro organismo o per l'omeostasi cellulare, conducendo alla formazione di malattie come il cancro, che rimane il caso più noto e studiato.

Lo studio delle cause di malattie genetiche, e in particolare di malattie monogeniche, permette di identificare i geni la cui funzione è essenziale per il buon funzionamento del nostro organismo. La distrofia di Duchenne, caratterizzata da una progressiva perdita delle funzioni muscolari, è un famoso esempio di una patologia causata dalla trasmissione di mutazioni deleterie che provocano la perdita delle funzioni della proteina chiamata Distrofina, implicata nel processo di contrazione muscolare.

La corretta interpretazione dell'informazione contenuta nel nostro genoma è garantita da delle modificazioni chimiche che lo decorano e che garantiscono la corretta espressione. L'insieme di queste modificazioni è definito come epigenoma, che significa letteralmente "sopra i geni". Ne deriva che mutazioni in geni che codificano per attori implicati nella regolazione dell'epigenoma hanno effetti pleiotropici. La prima modificazione epigenetica ad essere stata descritta nel 1950, e la più studiata ad oggi, è la metilazione del DNA. L metilazione del DNA è essenziale per lo sviluppo dei mammiferi ed alterazioni di questa modificazione caratterizzano numerose malattie come il cancro o malattie monogeniche come la sindrome ereditaria nota come ICF (Immunodeficienza, instabilità del Centromero e anomalie Facciali). Questa sindrome è causata da mutazioni che causano la perdita della funzione di fattori implicati nel processo di metilazione del DNA. I pazienti affetti dalla sindrome ICF presentano un ritardo di sviluppo un'immunodeficienza primaria che può causare la morte dei pazienti durante l'infanzia. Le cellule di questi pazienti presentano un'instabilità cromosomica particolare, causata da difetti di metilazione del DNA, che caratterizzano anche le cellule cancerose.

Di conseguenza, l'interesse dello studio di una sindrome rara come l'ICF è almeno a due livelli: primo, permette di meglio comprendere le basi molecolari della malattia per poter fornire nuovi biomarkers utilizzabili a scopo diagnostico. Secondo, come precedentemente accennato,

questa malattia monogenica offre un'occasione unica per l'identificazione di nuovi fattori e meccanismi implicati nella metilazione del DNA, che può essere pertinente anche per altri contesti patologici come il cancro o, in generale, altre malattie caratterizzate da alterazioni immunologiche o mentali.

In generale, lo studio dell'eziologia di una rarissima malattia come la sindrome ICF ha permesso di meglio chiarire il "vecchio" ambito della metilazione del DNA. Come diceva Marcel Proust:

"Le véritable voyage de découverte ne consiste pas à chercher de nouveaux paysage,
mais à avoir de nouveaux yeux
(Il vero viaggio della scoperta non consiste nello scoprire nuovi paesaggi,
ma nell'avere occhi nuovi)"
Marcel Proust

Durante il mio dottorato, io non ho cercato "nuovi paesaggi", dato che la metilazione del DNA viene studiata da ormai 60 anni, ma lo studio della sindrome ICF ha costituito i miei "occhi nuovi" per accrescere la nostra comprensione delle basi molecolari del mantenimento dell'integrità del genoma umano.

Nell'introduzione che segue, presenterò le principali caratteristiche del genoma dei mammiferi, la storia delle modificazioni epigenetiche, con una particolare attenzione per la metilazione del DNA, la sua distribuzione nel genoma, il suo impatto a seconda delle regioni genomiche in cui viene deposta e la sua importanza in condizioni fisiologiche. Successivamente, descriverò come i profili di metilazione del DNA posso essere alterati dall'effetto dell'ambiente e le alterazioni che caratterizzano malattie come il cancro, malattie autoimmuni e la sindrome ICF.

INTRODUCTION

Chapter I: Genomic features of mammalian genomes

1. Historical overview

The genome is defined as the genetic material of an organism. Each genome contains all the information needed to build and maintain that organism. In mammals, a copy of the entire genome is contained in all the cells that have a nucleus.

Nowadays it is known that the genome is composed of discrete units, called genes. Their existence was first suggested in 1866 by Mendel. He was studying the inheritance of distinct traits from parent to the offspring in edible pea plants and proposed that the inheritance of discrete units could give rise to observable physical characteristics. Its theories were largely unnoticed until the end of the 19th century, when, in 1889, de Vries published his book, *Intracellular Pangenesis*, containing very similar results to Mendel's studies (Stamhuis, 2003). The term gene was then coined and defined as the "physical and functional unit of heredity", in 1905, by Eduard Strasburger.

A great advance in the study of genes came in 1928, when Frederick Griffith showed that DNA was the molecule of inheritance (as it will be better explained in **Chapter II**). In 1941, follow-up studies by Beadel and Tatum showed that genes coded for specific proteins which were implicated in metabolic mechanisms (Beadle and Tatum, 1941). These results led to the "one gene, one enzyme" hypothesis. From 1940 a series of great discoveries contribute to better understand the structure of DNA and the composition and the organisation of the genome. A part of these events, characterising the "DNA-era" are summarised in **Table 1**.

Since proteins were known to play a direct role in processes, like the metabolic processes, the protein-coding genes firmly took centre stage as the mean by which genetic information is trans-acted. Thus, any DNA sequence deemed to lie outside of protein-coding genes, or promoter and enhancers that control gene expression and hence the protein repertoire of the cell, have been largely considered as less interesting, or even useless, and classified as "junk". Some opposition to this view came from the Nobel Laureate Barbara McClintock who wrote in a letter in 1950 "are we letting the philosophy of a protein-coding gene control (our) reasoning? What then is the philosophy of the gene?". However, her insights were largely ignored in the decades that followed.

Between 1950 and 1960 the identification of functional RNA that do not code for polypeptides, termed non-coding RNA (ncRNA), including ribosomal RNA (rRNA) and transfer RNA (tRNA) (Hoagland et al., 1957, 1958; Palade, 1955) highlighted the existence of genomic regions that did not code for proteins, but that were far from being useless. However, since these ncRNA were involved in messenger RNA (mRNA) translation to ensure the expression of protein-

YEAR	Discoveries
1941	Edward Lawrie Tatum and George Wells Beadle show that genes code for proteins
1944	The Avery–MacLeod–McCarty experiment isolates DNA as the genetic material (at that time called transforming principle)
1947	1947: Salvador Luria discovers reactivation of irradiated phage, stimulating numerous further studies of DNA repair processes in humans
1948	Barbara McClintock discovers transposons in maize
1950	Erwin Chargaff determined the pairing method of nitrogenous bases. Chargaff and his team studied the DNA from multiple organisms and found that i) the concentration of the pyrimidines (guanine and adenine) are always found in the same amount as one another, ii) the concentration of purines (cytosine and thymidine) are also always the same and that iii) the proportion of pyrimidines and purines correspond each other
1952	an X-ray diffraction image of DNA taken by Raymond Gosling, a student supervised by Rosalind Franklin
1953	The DNA structure is resolved to be a double helix by James Watson and Francis Crick
1955	Joe Hin Tjio, while working in Albert Levan's lab, determined the number of chromosomes in humans to be of 46
1957	Arthur Kornberg with Severo Ochoa discovered the DNA Polimerase I to be responsible of DNA synthesis and synthesised DNA *in vitro*
1958	The Meselson–Stahl experiment demonstrates that DNA is semiconservatively replicated
1958	Central Dogma of Biology
1960 - 1961	Jacob, Monod and collaborators discovered the operon and demostrate the central dogma of biology. They also defined the messenger RNA, as the RNA translated in protein product
1960 - 1969	Combined efforts of scientists "crack" the genetic code, including Marshall Nirenberg, Har Gobind Khorana, Sydney Brenner & Francis Crick.

Table 1. Summary of key discoveries during the "DNA era".

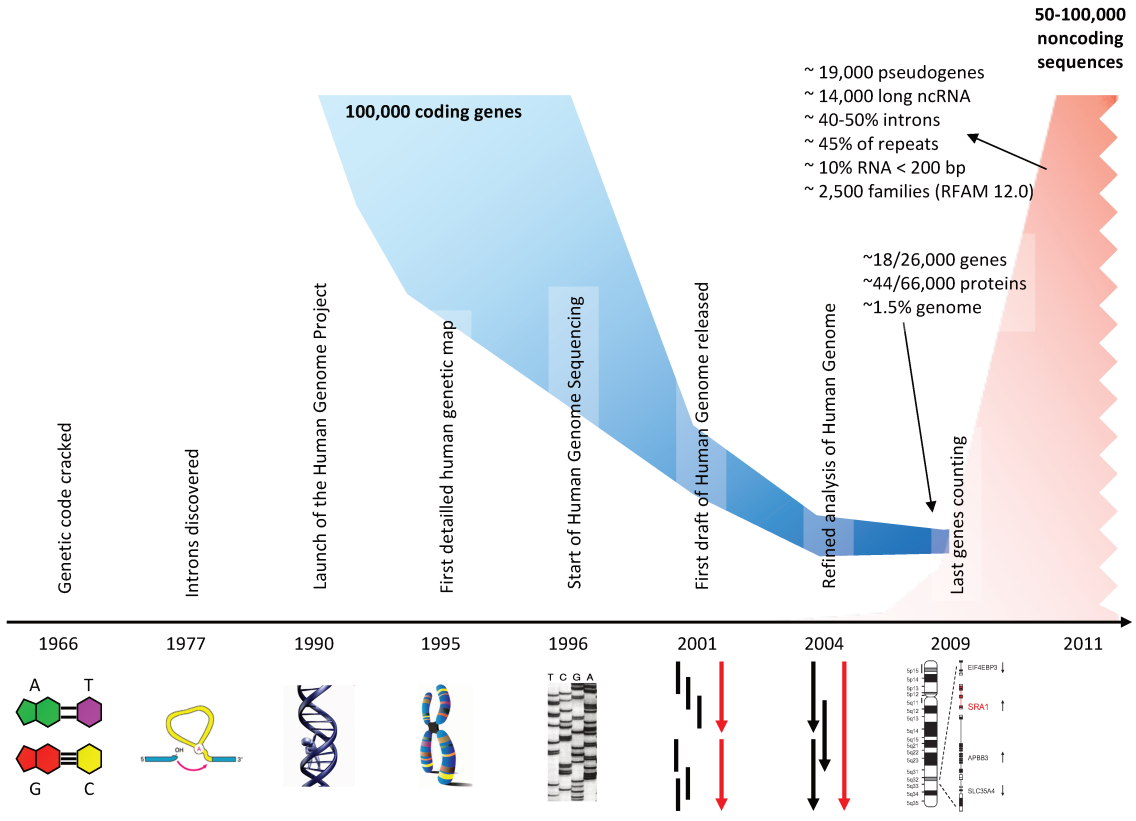


Figure 1. The human genome contains far less genes than expected. The progression of events leading to the counting of the genes contained in the human genome is represented. The numbers on the right correspond to the last gene counting. *Courtesy of Florent Hubé.*

coding genes, although essential, their functionality remained secondary to polypeptides. This view focused on protein-coding genes was reinforced in 1958 by Crick who summarised the flow of genetic information from DNA to protein as passing via an RNA intermediate (Crick, 1958). This has since become known as the Central Dogma of Biology and until recently identified mRNA, serving no other purpose than specifying protein assembly. The experimental demonstration of this dogma came between 1960 and 1961, when Jacob and Monod, studying the Lac operon system in bacteria. They discovered that genomic regions are transcribed to RNA, that they called RNA messenger (mRNA), which was then translated into a protein product (Jacob and Monod, 1961). This protein-coding-centric view even remained relatively unscathed following the discovery of introns and RNA splicing by the Nobel Laureates Roberts and Sharp in 1977 (Berget and Sharp, 1977; Chow et al., 1977). They described introns as segments of a gene between exons, which are transcribed but do not participate in the production of the final protein product as they are removed before translation, allowing joining of exons through a process known as splicing. In the same period, Chambon, Leder and O'Malley separately showed, by comparing DNA and mRNA sequences, that 90% of typical eukaryotic genes was not present in the mRNA (Breathnach et al., 1977; Nishioka and Leder, 1979; Robertson et al., 1979). This period represented also the dawn of the "genomic era" characterized by the development of DNA sequencing, with Sanger's lab being the first to completely sequence the genome of bacteriophage Φ-X174 (Sanger et al., 1977). The development of this technique and its progressive amelioration allowed the launch of the Human Genome project, in 1990, aiming to sequence all the human genome, that was started 6 years later (**Figure 1, bottom**).

In 2001 came the publication of the first draft of the human genome, which brought a first big surprise. The human genome is composed of 18/26,000 protein-coding genes, as opposed to the 100,000 that were announced about two decades ago, corresponding to 1.5% of the genome (Lander et al., 2001) (**Figure 1**). In the following years, high-throughput sequencing technologies allowed scientists to probe the organisational complexity of genomes of many other organisms. The comparison between the sequencing data obtained of the human genome and of other organisms brought a second surprise: the previous assumption that the number and repertoire of protein-coding genes should increase with organism complexity, defined as the number and types of cells, and the degree of cellular organization, was wrong. In contrast, the relative amount of non-coding DNA sequence increases consistently with organism complexity (Taft et al., 2007), a curious observation should this DNA be functionally unimportant. 98% of the human genome was composed of that "junk" or "parasitic" DNA, apparently without any known function (Lander et al., 2001; Mouse Genome Sequencing Consortium et al., 2002).

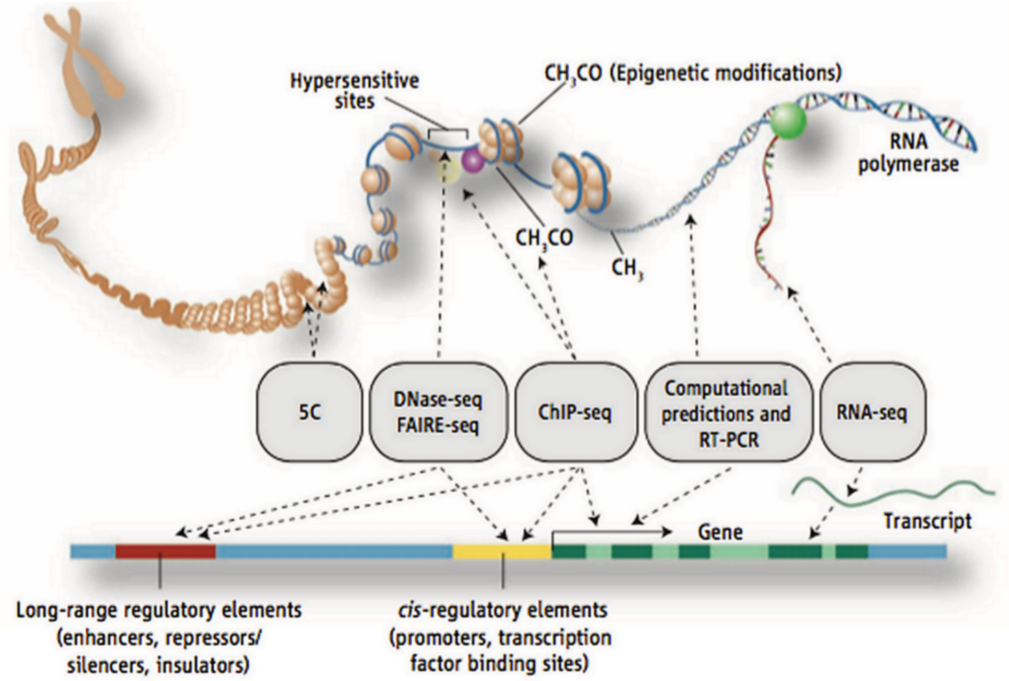


Figure 2. The principle axes of the ENCODE project 2012. Encodes functional tests (gray boxes) reveal "functional elements" along a chromosome. DNase-seq and FAIRE seq identity "open" chromatin via either "hypersensitivity" to Dnase I or reduced efficiency of formaldhyde cross-linking (which is less efficient in nucleosome-depleted regions). 5C maps long range chromosomal interactions. *From E. Pennisi 2012.*

The consideration of this huge non-coding DNA as junk DNA started to be considered as out-dated by the scientific community as a whole, largely thanks to the advent of new sequencing technologies that have allowed the genome-wide analysis. These studies have led to two important conclusions that bring into question our previous assumptions regarding the non-functionality of "junk" DNA:

- The majority of the genome is transcribed
- Many sequences lying outside of traditionally-defined proximal promoters contribute to the regulation of genome output, including elements that establish and maintain specific chromatin contexts (as it will be described in **Chapter II**).

All together, these observations launched the search for "treasures in the junk".

2. ENCODE 2012: Over 80% of the human genome might be "functional"

In 2012, the collaborative effort of 32 research institutions resulted in the culmination of one of the most ambitious research projects of all times: the decade-long Encyclopedia of DNA Elements (ENCODE) project. The goal of ENCODE was to catalogue all "biochemically functional" elements of the human genome, that is to say regulatory elements that control gene expression, sequences that control how DNA is packaged and folded, and regions of the genome that are actually transcribed. The consortium estimates that 80% of the human genome is transcribed or may hold regulatory functions (**Figure 2**). These functional DNA elements are defined on the basis of results obtained through different types of experiments described below, which were carried out in several types of cells and cell lines.

- RNA sequencing (RNA-seq): extraction of total RNA is followed by high-throughput sequencing. ENCODE estimates that up to 80% of the human genome can be transcribed depending on context (ENCODE Project Consortium, 2012), and defined 8,800 small and 9,600 long non-coding RNA molecules (Dunham et al., 2012).
- DNase I hypersensitive sites sequencing (DNase-seq) and Formaldehyde-Assisted Isolation of Regulatory Elements sequencing (FAIRE-seq). These techniques are used to identify the location of regulatory regions, based on the genome-wide sequencing of regions sensitive to cleavage by DNase I (DNases-seq) or accessible to crosslinking of chromatin with formaldehyde (FAIRE-seq). These techniques allow the separation of non-accessible from accessible chromatin fractions that are then sequenced. These techniques mapped 2.89 million sites where the protein-

DNA complexes could unwind, and estimated that across all cell types about 42% of the genome could be accessible to various machineries.

- Chromatin Immuno-Precipitation sequencing (ChIP-seq): this technique allows the identification of genomic regions bound by a factor or occupied by a specific histone modification. To date, ENCODE has examined about 100 of the 1,500 or so known transcription factors and about 20 other DNA-binding proteins, including those involved in modifying histones, and up to 12 modified histones (see also **Chapter II**). Overall, 8% of the genome falls within a binding site of the tested transcription factors.
- Chromosome Conformation Capture Carbon Copy (5C): this technique allows the identification of long-range DNA-DNA interactions and their organisation within the context of the nuclear space. On average, 3.9 distal stretches of DNA are in contact with the transcription start site of a given gene.

ENCODE has come under some criticism from evolutionary biologists because of its loose definition of the word "function" of genomic elements (Graur et al., 2013) and indeed whether these interactions have any functional relevance remains to be shown. However, the ENCODE data provide useful roadmaps for those who may go on to demonstrate such functionality, and arguably, put to rest the idea of a genome overridden with "junk" DNA. The ENCODE project is continuing to contribute to our understanding of the linear and 3D organization of the genome and its relationship with genome functions. The organisation of the human genome is complex and the goal of the following chapters is not to provide an overview of this immense organisational complexity. Rather, I will focus on what is defined as "repetitive DNA", *i.e.* repeated DNA sequences, which may be present at high copy numbers ranging from several hundreds to several thousand per genome and were largely considered for many decades as non-functional DNA. This includes interspersed transposable elements, which have contributed to shape the mammalian genomes throughout evolution through various events like new gene formation, exonization, as well as by the formation of new transcriptional regulatory elements, like promoters (Faulkner and Carninci, 2009; Schmitz and Brosius, 2011) and may act as "functional elements" that can modulate gene expression and chromatin states, and large tandemly repeated arrays of DNA, which organise specialized domains like centromeres and telomeres. In contrast to their positive impact on genome plasticity, transposable elements may also cause insertional mutagenesis that can be deleterious for the cells and can cause diseases, like cancer.

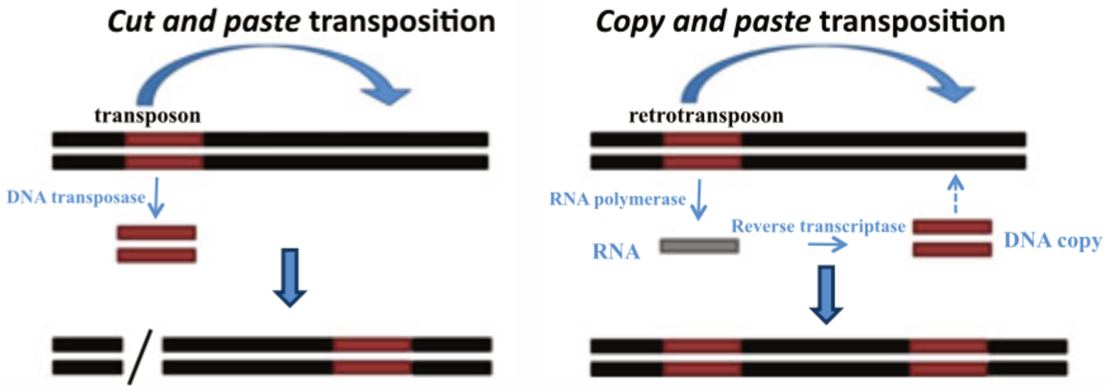


Figure 3. Schematic mechanisms of transposition of TEs. On the left, the cut-and-paste mechanism of DNA transposons mobilisation. On the right, the copy-and-paste mechanism of retrotransposons mobilisation. *From Bonchev and Parisod 2013.*

TE class	Full length repeat unit	Mechanism	Protein coded	Fraction in mammalian genome
DNA transposons	1.5 – 5 kb	Cut-and-paste	Transposase	0,87 – 3%
Non-LTR retrotransposons LINE1	~6kb	Copy-and-paste (Autonomous)	ORF1 ORF2	~20%
Non-LTR retrotransposons SINE	300bp	Copy-and-paste (LINE-dependent)	None	~10%
LTR retrotransposons	~5kb	Copy-and-paste	gag, pro, pol, (env)	8 – 10%

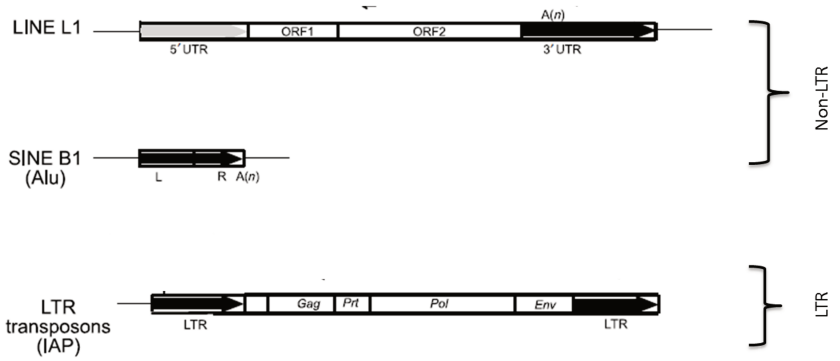


Figure 4. Different families of TE. Top: table summarising TEs families, highlighting the full-length repeat unit, the mechanism of transposition, the protein encoded and the fraction of the mammalian genome they represent. Bottom: "Non-LTR" retrotransposons encompass LINE and SINE elements such as human and mouse LINE L1 (LINE1), mouse SINE B1, and their correspondent human Alu. In the mouse, "LTR" retrotransposons encompass a large family of elements that resemble integrated retroviruses in structure, possessing "gag", "pol", and "env" genes for their autonomous transpostion. The intracisternal A particle (IAP) element is shown as an example. *Adapted from Martens et al., 2005.*

3. Interspersed DNA repeats: transposable elements

The majority of mammalian repeated sequences are transposable elements (TEs), meaning that they can "move" from one genomic location to another. These elements make up approximately 45% of the genome (Deininger and Batzer, 2002). TEs are divided into two categories based on their mode of mobilisation. DNA transposons "move" using a cut-and-paste mechanism, whereas retrotransposons "move" in a copy-and-paste fashion involving an RNA intermediate (**Figure 3**).

In mammals, DNA transposons like Mariner and Tigger are thought to be the inactive remnants of transposons with a mutated DNA transposase (Kazazian, 2004). Retrotransposons are further classified according to their structure and putative viral origin (**Figure 4**). Among two classes can be identified: i) Long Terminal Repeat (LTR) retrotransposons, that are retroviral-like in structure and ii) non-LTR retrotransposons, which include Long Interspersed Nuclear Elements (LINEs) and Short Interspersed Nuclear Elements (SINEs).

In mammals, the most widespread family of TEs capable of autonomous retrotransposition is the Long Interspersed Element L1 (LINE1) element that represents approximately 20% of mammalian genomes (Martens et al., 2005). Full-length LINE1 elements are 7 kb in length and show a specie-specific 5'-untranslated region (5'UTR) that contains regulatory sequences responsible for their transcriptional activity. In mice, the LINE1 5'UTR is composed of distinct classes of 200 bp monomers: F, A, Tf, and Gf. Hence, these sequences divide LINE1 elements into four categories: L1MdF, L1MdA, L1MdTf, and L1MdGf, which appeared in the mouse genome at about 5, 2.5 and 0.5 million years ago, respectively (Jubier-Maurin et al., 1992; Naas et al., 1998; Padgett et al., 1988; Schichman et al., 1993). In addition, LINE1 elements include two conserved open reading frames ORF1 and ORF2, and highly interspecies variable 3'-untranslated regions (3'UTR) (Furano et al., 1988; Loeb et al., 1986; Padgett et al., 1988; Scott et al., 1987). During retrotransposition, LINE1 elements are first transcribed by RNA Polymerase II (RNA Pol II) from their promoter located at the 5'UTR. The resulting full-length sense LINE1 RNA then serves both as the template for making a new copy of LINE1 DNA and as a mRNA that directs the synthesis of ORF1 and ORF2 proteins. ORF1 encodes a 40kDa RNA-binding protein that is implicated in the priming of LINE1 reverse transcription (Kolosha and Martin, 1997). ORF2 encodes a 150kDa protein that possesses endonuclease and reverse transcriptase enzymatic activities (Feng et al., 1996; Mathias et al., 1991). ORF1 and ORF2, together with LINE1 RNA, form a ribonucleoprotein (RNP) particles that are imported into the nucleus. When finding their genomic target, generally an AT-rich region, the endonuclease generates a single-strand cut in DNA and the reverse transcriptase reverse transcribes the LINE1 RNA, leading to the integration

of the LINE1 into a new genomic locus in a process termed Target-Primed Reverse Transcription (TPRT) (Feng et al., 1998; Luan et al., 1993). In human and mouse genomes, more than 500,000 copies of LINE1 elements have been identified (Lander et al., 2001; Waterson et al., 2002). However, the vast majority (>90%) of these elements are truncated in their 5'end or have inactivating deletions or point mutations that prevent their transcription and hence, their transposition. Even among the full length LINE1 elements, only a subset, corresponding to 30-60% of total full-length LINE1 elements, is estimated to be active for retrotransposition (Brouha et al., 2003; Sassaman et al., 1997).

SINE elements are derived from 7SL RNA, the RNA component of the signal recognition particle ribonucleoprotein complex (SRP), transcribed by RNA polymerase III (RNA Pol III) (Okada, 1991; Singer, 1982). Both RNA Pol III genes and SINEs have promoter elements called A- and B-boxes (Schramm and Hernandez, 2002), which are recognised by the multi-subunit RNA Pol III co-factor TFIIIC and subsequent binding of the TFIIIB complex and recruitment of RNA Pol III (Geiduschek and Kassavetis, 2001). SINEs do not encode for any protein for an autonomous mobilisation, and are thus parasites that rely on the reverse transcriptase and endonuclease activity of LINE elements. The 3'UTR regions of LINE and SINE elements usually contain a specific sequence and/or structure that is recognised by the reverse transcriptase (Anzai et al., 2005; Luan and Eickbush, 1995; Osanai et al., 2004). In the case of LINE1, this activity is in the polyA tract (Wei et al., 2001). SINE elements take advantage of this dependency for their transposition owing to their homology with the 3'UTR region of LINE elements (Ohshima et al., 1996). For example, mouse B1 and B2 SINEs possess a polyA tract in their 3' termini so that they can be mobilised by mouse LINE1 reverse transcriptase (Dewannieux and Heidmann, 2005).

Long terminal repeat (LTR) elements are also called Endogenous Retroviral Elements (ERVs) since they are retroviral-like in structure and mechanism of retro-transposition. ERVs constitute a range of elements, owing similarities and differences, with varying copy number, autonomy, and expression patterns, which make up approximately 8% and 10% of human and mouse genomes, respectively (Lander et al., 2001; Waterson et al., 2002). In the mouse, ERV elements include (but are not limited to) the murine endogenous retrovirus-like (MuERV-L), Intracisternal A-Particle (IAP) elements, MusD, and Mammalian apparent LTR-retrotransposons (MalR). Autonomous ERVs, such as full-length IAP elements, possess GAG, PRO, POL, and ENV genes that encode proteins that facilitate the viral cycle, flanked by 5' and 3' LTRs that contain promoter and transcription termination signals. GAG encodes a polyprotein with a capsid and a nucleocapsid domain (Sandmeyer and Clemens, 2010). These proteins form virus-like particles in the cytoplasm, inside which reverse-transcription occurs. The POL gene produces three proteins: a protease (PRO), a reverse-transcriptase endowed with an RT (reverse-transcriptase)

and an RNAse-H domain, and an integrase (IN) (Wicker et al., 2007). The lack of a functional viral envelope gene (ENV), distinguishes ERVs elements from retroviruses (Boeke and Stoye, 1997). Although the majority of IAP elements have defective genes, a few have only defective or deleted ENV genes. Nonetheless, retro-transposition of IAP elements can occur in *trans* using the proteins encoded by intact IAP elements (Ribet et al., 2008). Because of the higher number of active ERV in the mouse than in humans, mutational insertions account for up to 10% of all mutations in mice (Maksakova et al., 2006), compared to only about 0,1% of mutations in humans.

3.1. Genomic distribution of TEs

LINE1-encoded ORF2 protein is an endonuclease that does not cut DNA at random sites, but instead shows specificity for consensus sequences, for example 5'-TTAAAA-3' for human LINE1 (Jurka, 1997). Given that SINEs are dependent on LINE-encoded endonucleases for their retrotransposition, it follows that these elements should be inserted in similar genomic sequences. However, they occupy distinct regions of the genome and their densities are negatively correlated. Early DNA-FISH and chromosome banding experiments revealed that LINE1 elements reside in so-called G-dark bands, which are AT-rich, comprise facultative heterochromatin and tissue-specific genes that are transcribed only in selected cell types. In contrast, SINE elements predominantly reside in G-light bands (or R bands), which contain housekeeping genes, which are early replicating and actively transcribed in almost all cells (Boyle et al., 1990; Korenberg and Rykowski, 1988).

ERVs are not known to have a consensus site of insertion, although they have preferences for certain genomic regions. Active and evolutionary young human ERVs were shown to integrate into gene-rich regions, particularly in gene bodies (Brady et al., 2009; Mitchell et al., 2004). In contrast, evolutionary ancient ERV elements are generally found in gene-poor regions (Medstrand et al., 2002). Gene-rich regions represent a more permissive chromatin context (chromatin structure will be described in the next sections of this manuscript) that could facilitate the insertion of TEs. However, some TEs are found in gene-poor regions, suggesting that a natural selection may be responsible for the control of the maintenance of certain TEs in gene-poor regions, and that their maintenance may confer some selective advantage.

In mammalian genomes, the TE content is higher on the X chromosome compared to the other chromosomes. This bias mainly relies on the enrichment in LINE1 elements, which occupy large domains of the X chromosome. In humans, there is a portion of around 200kb on the Xp11 arm which is composed at 99% of TEs, half of it being composed of LINE1 elements (Lander et al., 2001). The analysis of the Human Genome Project confirmed this enrichment, especially at the Xq13–q21 region that contains the X Inactivation Centre (XIC) that controls the inactivation

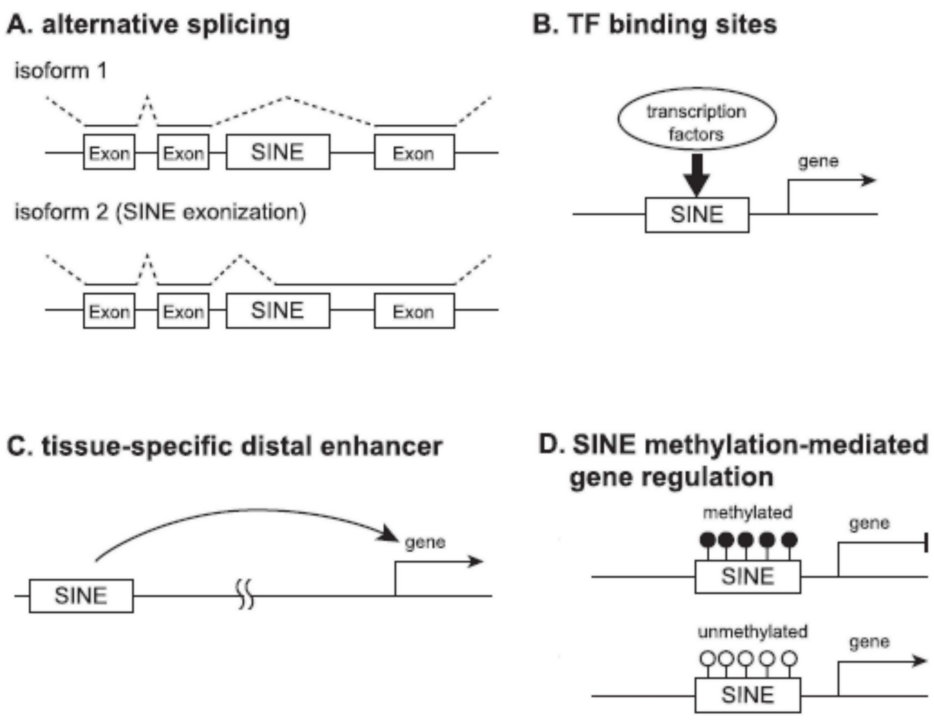


Figure 5. Possible functional roles of TEs as controlling elements as exemplified by SINEs. *Adapted from Ichiyanagi 2013.*

of one X chromosome in females (the mechanism will be explained in the following chapters) (Bailey et al., 2000). This non-random distribution of LINE1 elements on the X chromosome led Bailey and colleagues to propose that LINE1 could be key elements for the mechanism of X inactivation (Bailey et al., 2000; Lyon, 2000).

There are also other regions in which TEs are particularly enriched, such as centromeric, pericentromeric and telomeric domains (Bartolomé et al., 2002; Wright et al., 2003).

3.2. Transposable elements as "functional elements"

Far from their initial definition as parasitic DNA, TEs have the potential to influence the evolutionary trajectory of their host in three distinct ways: i) via alterations of gene function through insertion, ii) through the induction of chromosomal rearrangements or iii) as a source of coding and noncoding material that allows for the emergence of genetic novelty, such as new genes and regulatory sequences (Feschotte and Pritham, 2007). In particular, the insertion of TEs within or close to genes has been shown to affect gene activity in various ways: i) by regulating patterns of alternative splicing (Sorek et al., 2002), creating binding sites for transcription factors and even acting as alternative promoters (Macfarlan et al., 2012; Oei et al., 2004; Peaston et al., 2004; Vansant and Reynolds, 1995), ii) acting as tissue-specific distal enhancers (Bejerano et al., 2006), and iii) controlling the transcription of nearby genes via DNA methylation, since, as I will describe in the following chapter, TEs are known to be enriched in this epigenetic mark (see also **Chapter II**) (Cooney et al., 2002) (**Figure 5**).

TEs are emerging as key regulators in the acquisition of species-specific transcriptional networks in embryonic stem cells, mediating the creation of species-specific binding sites for pluripotency factors (Kunarso et al., 2010) as well as at early stages of pre-implantation development (Xie et al., 2010). Moreover, TEs were shown to even contribute to the formation of a new gene regulatory networks dedicated to pregnancy in placental mammals (Lynch et al., 2011). Moreover, the transposition of TEs in somatic cells, although it is generally associated with insertional mutagenesis and tumour formation, may have a beneficial effect on the host organism, like in the brain where recent studies have showed that mobile elements actively retrotranspose during neurogenesis, thereby creating genomic diversity between neurons, contributing to neuroplasticity of the brain (Reilly et al., 2013).

4. Tandem repeats: centromeres, pericentromeric regions and telomeres

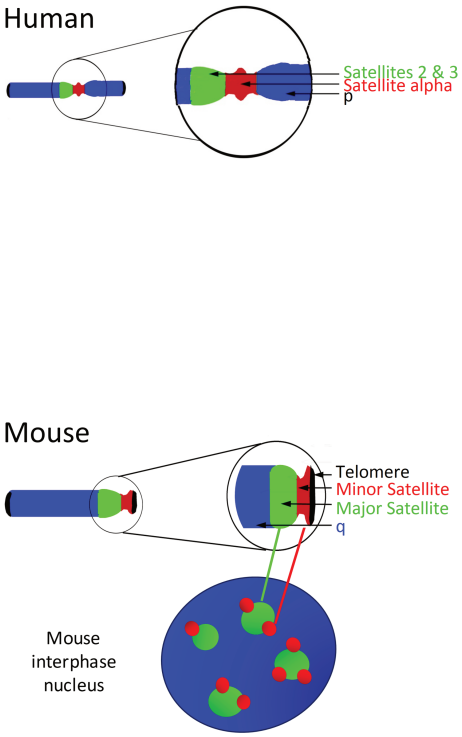
In higher eukaryotes, the clustering of tandem arrays of repetitive elements in large segments, called satellite repeat sequences, contributes to the formation of specialised structures, notably telomeres, which are regions that protect the end of the chromosomes from deterioration or fusion with other chromosomes, and centromeres, which are essential for faithful chromosome segregation during cell division.

Surprisingly, no sequence conservation can be found at centromeric and pericentromeric repeats sequences across species. Among species, the primary determinant for centromere identity is the presence of nucleosomes containing a specific variant of histone H3, called CENP-A (see also **Chapter II**). However, in most higher eukaryotes, repetitive elements around centromere are assembled on large arrays of tandemly repeated DNA sequences, which can be divided into two distinct functional domains (Guenatri et al., 2004): the centromere *per se*, that serves as the point of assembly for the kinetochore, and flanking heterochromatin domains, thought to protect the centromere and ensure sister chromatid cohesion (Yamagishi et al., 2008).

In humans and mice, these regions are characterized by the presence of a particular class of DNA repeats, called satellite repeats. This name comes from the identification of these DNA repeats, by gradient density centrifugation. Satellite DNA repeats were observed to give a sharp "satellite" band in a density gradient, which can be resolved from the broader main band of DNA in the gradient (Cooney and Matthews, 1985). The particular density of these sequence was shown to rely on their sequence, which, as I will explain later, is AT-rich (Cooney and Matthews, 1985)

In humans, centromeric regions contain 171-bp AT-rich α-satellite repeats motifs. However, each human centromere shows specific sequence polymorphisms as well as a variable number of α-satellite repeats (Schueler and Sullivan, 2006). The size and structure of pericentromeric regions varies between chromosomes, and three types of satellite repeats account for about 4%–5% of the human genome. Satellite repeats type I are short AT-rich sequences found at pericentromeric regions of most chromosomes. Type II and type III satellite repeats are made of a 5bp GGAAT repeat unit possibly found on all chromosomes but unevenly distributed over up to several megabases, where they form heterochromatin blocks in pericentromeric regions of certain chromosomes like chromosomes 1, 9, 16 and heterochromatin of chromosome Y (Vourc'h and Biamonti, 2011).

In the mouse, pericentromeric domains are made up of A/T rich 234 bp major satellite repeats, and adjacent centromeric regions comprise 120 bp minor satellite repeats (Choo, 1997). The contribution of these sequences to the genome is substantial, with major satellite repeats representing approximately 3% of the mouse genome and minor satellite repeats representing



Repeat type	Repeat consensus sequence	Chromosomal location
Satellite alpha repeats **HUMAN**	CATTCTCAGAAACTTCTTTGTGATGTATACATTCAACT CACAGAGTTGAACCTTCCTTTTCATAGAGCAGTTTTG AAACACTCTTTTTGTAGAATCTGCAAGTGGATATTTG GACCGCTTTGAGGCCTTGGTTGGAAACGGGAATATC TTCATATAAAAACTAGACAGAAG **(171bps)**	**Centromere** all chromosomes
Satellite II repeats **HUMAN**	(ATTCCATTCG)2 + $(ATG)_{1or2}$	**Pericentromere** Most chromosomes, large blocks on chr1 and 16
Satellite III repeats **HUMAN**	(ATTCC) + or – (AT/GTCGGGTTG)	**Pericentromere** Most chromosomes, large block on chr9
Minor satellite repeats **MOUSE**	GGAAAATGATAAAAACCACACTGTAGAACCTTTTAG AGGAGTGAGTTACACTGAAAAACACATTCGTTGGAA ACGGGATTTGTAGAACAGTGTATATCAATGAGTTAC AATGAGAAACAT (**120bps**)	**Centromere** all chromosomes
Major Satellite repeats **MOUSE**	GGACCTGGAATATGGCGAGAAAACTGAAAATCACG GAAAATGAGAAATACACACTTTAGGACGTG AAATATGGCGAGGAAAACTGAAAAAGGTGGAAAAT TTAGAAATGTCCACTGTAGGACGTGGAATATGGCAA GAAAACTGAAAATCATGGAAAATGAGAACATCCACT TGACGACTTGAAAAATGACGAAATCACTAAAAAACG TGAAAAATGAGAAATGCACACTGAA (234bps)	**Pericentromere** all chromosomes

Figure 6. Structural organisation of pericentromeric and centromeric repeats in the human and mouse. (Left panel) Model of a human chromosome (top) and mouse acrocentric chromosome (bottom) depicting the pericentric satellite repeats, (green), flanking the inner centromere region, (red). Telomeres are depicted in black. In interphase mouse nuclei, pericentric domains from different chromosomes cluster into chromocenters, while the minor satellite repeats remain as separate entities. (Right panel) Table showing the various classes of human and mouse centromeric satellite repeats along with their sequence composition and chromosomal location.

approximately 0.45% of the genome (Martens et al., 2005). In the mouse, this discrete organisation can be easily visualised in interphase nuclei. Pericentromeric regions from several chromosomes cluster together to form chromocentres that are densely stained with the DNA dye 4',6-diamidino-2-phenylindole (DAPI). Centromeric regions remain closely associated but separate entities (Guenatri et al., 2004) (**Figure 6**).

Despite the differences in the DNA sequence and 3D organisation, and although have previously been considered tightly compacted and transcriptionally inert regions, transcripts emanating from both the pericentromere and centromere core have been identified in a multitude of organisms (Eymery et al., 2009; Stimpson and Sullivan, 2010). These transcripts play a critical role in the organization and function of centromeric and pericentromeric regions, as it will be described in **Chapter II**. The transcription of these regions is tightly regulated during the cell cycle, since accumulation of centromeric and pericentromeric transcripts were observed in several types of cancer (Ting et al., 2011). The fact that both human and mouse centromeric and pericentromeric repeats contain CpG dinucleotides, which represents the main context for DNA methylation, suggests a potential regulation of the transcription of these regions by DNA methylation (see also **Chapter II**).

Telomeres are specialized chromatin structures at the end of eukaryotic chromosomes that prevent the chromosome ends from being recognized as a DNA break. In all species, telomeres are composed of tandem repeats of the TTAGGG sequence that are generated by a cellular reverse transcriptase, known as telomerase, that prevents erosion of chromosome ends after replication (Osterhage and Friedman, 2009). The subtelomeric regions are located adjacent to telomeric regions and contain other families of repetitive elements whose sequence varies among mammals and from one chromosome to the other (Riethman et al., 2005). Telomeric repeats are bound by the SHELTERIN protein complex that has a role in chromosome protection and in the maintenance of telomere length (Xin et al., 2008), which is crucial to safeguard chromosome integrity. The length of human telomeres is between 5 and 15 kb, although differences between individuals are observed (Moyzis et al., 1988; Zhu et al., 1998). A progressive reduction of telomere length, up to a critical size of 3 kb, was observed in aging cells *in vivo,* and shown *in vitro* to trigger entry of the cells into senescence, an irreversible cell cycle arrest in which cells are refractory to growth factor stimulation (Allsopp and Harley, 1995; Blasco, 2007).

Similarly to centromeric and pericentromeric repeats, also telomeric repeats are transcribed and the transcripts originated from these regions are essential for the organization of the telomeric heterochromatin (as I will describe in **Chapter II**). Although CpG dinucleotides are absent from the telomeric TTAGGG repeat unit, human subtelomeric regions are rich in CpG sites (Brock et

al., 1999; Cross et al., 1990; de Lange et al., 1990) and hence, similarly to centromeric and pericentromeric repeats, DNA methylation could play a role in the regulation of their transcription.

In the mouse, the length of telomeres was shown to be highly heterogeneous from one chromosome to another (Zijlmans et al., 1997) and even among mouse strains. *Mus musculus* chromosomes have incredibly long telomeres, exceeding 25kb. In contrast, *Mus spretus* shows a length between 5 and 15 kb, which is closer to that of human telomeres (Zhu et al., 1998). Hence, *Mus spretus* is a model system used to study the relationship between telomere shortening and senescence in mice (Coviello-McLaughlin and Prowse, 1997).

In contrast to human chromosomes, mouse chromosomes were described as being acrocentric, *i.e.* having its centromere located at one end of the chromosome and thus lacking a cytologically detectable short arm (**Figure 6**). Interestingly, highly conserved truncated LINE1 (tL1) elements and varying amounts of a new type of telocentric tandem repeats, called TLC, were found between centromeric and telomeric regions (Kalitsis et al., 2006). To date, it is unclear what function the tL1 and TLC sequences might have. It is possible that their presence serves to demarcate the telomeric and centric regions (Kalitsis et al., 2006).

Repetitive elements represent a substantial burden to genome stability since, on one hand, transposable elements can cause new insertions in germ cells resulting in germ line mutations (Maksakova et al., 2006), whereas these events in somatic cells have detrimental consequences for a number of reasons, including genomic instability, accumulation of deleterious mutations, toxic accumulation of protein or RNA products, and DNA-damage that may cause malignant transformations and are involved in numerous human diseases (Kaer and Speek, 2013; Reilly et al., 2013). On the other hand, aberrant transcription originating from tandem repeats, in particular from centromeric and pericentromeric repeats, was associated with chromosomal instability and perturbed phenotypes, such as cancer (Bouzinba-Segard et al., 2006; Ting et al., 2011). Thus, the transcription of these elements must be tightly controlled in order to avoid potentially aberrant events that may lead to disease. This regulation is ensured by the so-called epi-genetic modifications, since they occur “on the top” of the DNA sequence (from the Greek “epi” means “above”).

Chapter II: Epigenetics and chromatin

1. Historical overview of epigenetics

Embryologists were the first to question the nature and location of the components that allow to start and regulate developmental programs (Felsenfeld, 2014). They were divided in two groups: one supporting the "preformationism" suggesting that each cell contains preformed elements that spread during development, and the other supporting the "epigenesis" along which nothing is preformed but the new organism starts as an undifferentiated mass and develops through several steps called "morphogenesis" (Aulie, 1961). Debates between these groups already started in the 17th and 18th centuries. In the 19th century, the explosion of the experimental embryology led "epigenesists", who followed the theory elaborated by Aristotle (384-322 before Christ), to suggest that development processes involve chemical reactions between soluble components (Ribatti, 2014). The genetic contribution to development was debated inside the group of "epigenesists". One of them, Weismann thought that development proceeds with the progressive loss of genetic material and retention of the material necessary to perform specific functions in a given cell (Weismann et al., 1893). In contrast, the so-called new geneticists believed that the inherited genetic information was the same in each cell but that it was somehow differentially exploited. The formal demonstration that the genetic material is conserved during differentiation, and that development relies on changes in gene expression, was given by the works of Briggs and King in 1952 and Gurdon in 1962. Briggs and King were the first to establish the procedure of transplantation of the nucleus of a living cell to an egg or oocyte (Briggs and King, 1952). They transplanted the nucleus of a frog blastula cell in an enucleated frog egg, and observed that a new frog was formed. Gurdon, ten years later, performed the same experiment but he transplanted the nucleus of a frog differentiated intestine cell in an enucleated frog egg and obtained the same result (Gurdon, 1962a, 1962b, 1962c). All together, these findings suggested that the genetic information is the same in all the cells, independently of their differentiation stage, but that it is differentially exploited to form the different cells of an organism.

Since then, the question of how the genetic information is differentially exploited during cellular differentiation is still a very current issue. Moreover, differentiated cells, like fibroblast, neurons or lymphocytes, when extracted from their physiological environment and transferred to the artificial environment of cell culture, maintain their phenotypes throughout cell divisions. These findings suggest that this exploitation can be maintained and transmitted to daughter cells. All the phenomena involved in the regulation of the processes essential for cell differentiation and, more globally, organism development, without changing the genomic information, can be

collectively grouped under the definition of "epigenetics". This term was coined by Waddington in 1942 and defined as "the study of the mechanisms of development through which genes bring about phenotypic effects" (Waddington, 2012). In the last 50 years, researchers used the term "epigenetics" in different ways and many tried to give their personal definition. The most common definition heard nowadays is the one proposed by Holliday in 1994 and Riggs in 1996: "the study of mitotically or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence" (Holliday, 2006; Riggs, 1975). Among the various definitions, I want to propose the more "poetic" and intriguing by Barlow, less heard and less used in the scientific field: epigenetics is "all weird and wonderful things that cannot be explained by genetics" (Denise Barlow, Epigenome Network of Excellence website).

2. Chromatin

2.1. Historical overview of chromatin

Whatever definition we choose, epigenetics is tightly related to the genome that is not just floating inside the cell nucleus but is well organized and packaged. Eukaryotic cells are faced with the dual challenge of organizing the incredibly long linear DNA (two meters-long in the case of humans) in order to pack it into the restricted nuclear volume (around 10μm in diameter) and to regulate genome functions. Miescher was the first, in 1871, to describe a novel kind of cellular substance in leukocytes and then in salmon sperm, that unexpected properties different to those of the other proteins (Dahm, 2005). He called this mysterious structure "nuclein", because he believed it had come from the cell nucleus. Unbeknown to him, Johann had discovered the molecular basis of all life: DNA. At the time, the vast majority of scientists were convinced that the more complex proteins must be the carriers of genetic information. Many scientists started to get interested in Miescher's discovery, mainly chemists. Among them, Kossel, in 1880s, discovered that nuclein was comprised of four bases and sugar molecules (nucleotides) and renamed "nuclein" to "nucleic acid". In addition, Kossel proposed that nucleic acids might bind to a new class of proteins that he called histones (Kossel, 2009). One of the fathers of chromatin was Flemming, who, in the 1870s, described the morphology and behaviour of the chromosomes during mitosis (Paweletz, 2001), coining both the terms "chromatin" and "mitosis". In particular, Flemming explored the fibrous network that could be easily stained with aniline dyes, that he called chromatin ("chromo" stands for "colour" in ancient Greek), and he thought that nuclein and chromatin could be the same thing. Flemming observed also some coloured bodies, the chromosomes (literal translation of coloured bodies). In the late 1880s and 1890s, Boveri began

advancing the theory that the chromosomes not only harbour the genetic information in a cell, but that individual chromosomes carry different parts of the hereditary material (Maderspacher, 2008).

In 1941, although it was accepted that chromatin and chromosomes were the structural basis of heredity and that genes were harboured on chromosomes, proteins were still supposed to be responsible for the genetic information (Schultz, 1941). In 1943, Avery, MacLeod and McCarty injected different macromolecules which compose the pneumococcus *Streptococcus pneumonia*, namely proteins or nucleic acids, in mice. The death of the mice was observed only after injecting nucleic acids, reproducing the effects of the complete pneumococcus. The molecule with a lethal power had the chemical composition of DNA and could be destroyed by a depolymerase (deoxyribonuclease I). These findings showed that DNA, and not proteins, is the molecule of heredity (Avery et al., 1999). Almost ten years later, the resolution of the DNA structure provided the molecular basis for copying and possibly transmitting the genetic material (Watson and Crick, 1953). Five years later, Crick proposed that, in general, the transfer of information is performed from DNA, through RNA to proteins, with what is known today as the Central Dogma of Biology (Crick, 1958). This dogma was then validated by the work of Jacob and Monod in 1961 who explored the Lac operon system in bacteria and discovered that a genomic region was transcribed into RNA, that they called messenger, which was then translated into a protein product (Jacob and Monod, 1961). This great discovery showed that there were genetic determinants, called regulators, that could control the expression of a given gene.

The relationship between chromatin and gene expression has been of great interest since the before mentioned discovery of histones and the realization that DNA is not naked in the nucleus. In 1950, the histones were proposed as repressor of gene expression (Stedman, 1950). It was common thought that, in order to induce transcription, it was necessary to strip off the histones and once that was done, it was thought transcription would proceed pretty much as it did in prokaryotes ((Felsenfeld, 2014) and described in the next chapter). A great contribution in this field was given by Allfrey and Mirsky, who proposed that chemical modification of histones could be important to regulate transcription, more than strip histones away of chromatin (Allfrey and Mirsky, 1964). The modifications of histones will be described in the following part of my manuscript. In the same year, X-ray diffraction studies showed that histones played a crucial role in the organization of chromatin. In particular, Vittorio Luzzati noticed that chromatin had a repeating pattern with intervals of about 100 angstroms (Luzzati et al., 1964). Aaron Klug confirmed the presence of these repetition in the structure of chromatin, suggesting that histones may paly a role in "packaging" chromatin in discrete units. Support to the hypothesis of packaging came in 1973, when Kornberg's lab isolated the enzyme DNA nuclease from rat liver cells that was then use by Kornberg and Klug to digest chromatin. They electrophoresed the digested

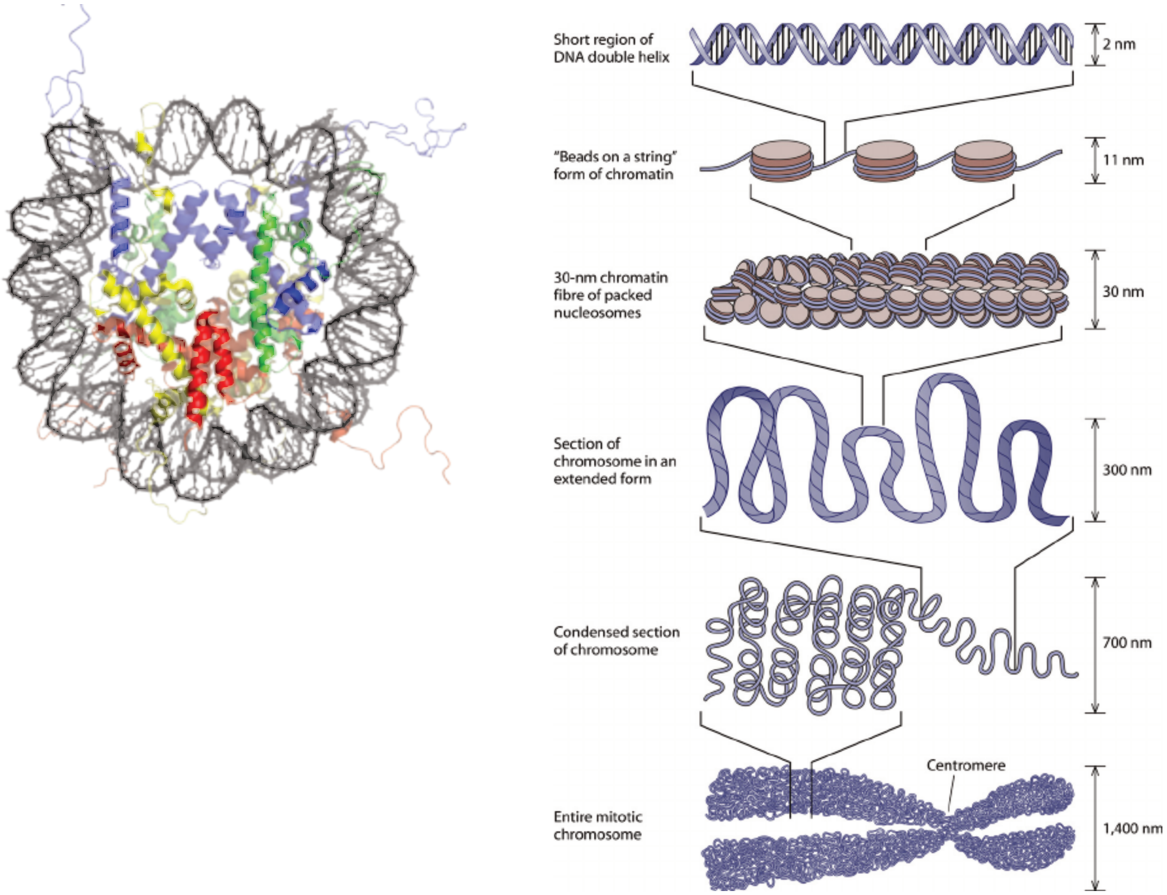


Figure 7. Multiple levels of chromatin structure. On the left the crystal structure of the nucleosome, composed of H2A, H2B, H3 and H4 core histones and DNA. The view is from the top. *From Luger et al., 1997*. On the right: progressive packing of DNA into chromatin via into association with histone proteins to give rise to highly condensed mitotic chromosomes. *From Verstrepen, 2011.*

chromatin material and saw a regular pattern of bands on the gel, defined as basic chromatin subunits, consisting of ~200bp of DNA and two each of histones H2A, H2B, H3 and H4, called the nucleosome (Kornberg, 1974). Then, in 1976, Finch and Klug proposed the "solenoidal" model for winding nucleosomes into the next level of organization and showed that histone H1 is necessary for this (Finch and Klug, 1976). Crystals of the nucleosome core (Finch et al., 1977) were gradually improved (Klug et al., 1980) and the group, then including Tim Richmond, solved its structure to 7 A˚ resolution in 1984 (Richmond et al., 1984). Follow-up studies by electron and atomic force microscopy have revealed the familiar "beads-on-a-string" structure of chromatin, leading to possible more detailed description of chromatin structure (Graziano et al., 1994).

As a whole, it is fascinating to realize the progression of events and discoveries that contributed to the description of the chromatin structure, that is now analysed at the atomic level (Luger et al., 1997) (**Figure 7 left panel**). In the same years a lot of attention has been put on the modifications of chromatin (at the level of DNA and/or at the level of histones) and the relationship between these modifications and gene expression (Bascom and Schlick, 2017; Brien et al., 2016; Johnson and Dent, 2013). The following chapter of this manuscript aims to give an overview of the structure and compaction of the chromatin, the modifications of histones and the impact of these modifications on transcription.

2.2. Chromatin structure and compaction

Today it is known that DNA is tightly packaged into nucleosomes that form the basic building blocks of chromatin (**Figure 7**). In each nucleosome, approximately two super-helical turns of DNA (146 bp) wrap around an octamer of core histone proteins consisting of a central $(H3/H4)_2$ tetramer flanked on each side by two H2A/H2B dimers. Core histones also contain N-terminal and C-terminal unfolded tails that extend radially from the nucleosome core and are subjected to many different post-translational modifications (Luger et al., 1997). A fifth family of histones, the H1 family, called linker histones owing to their association with linker DNA between two neighbouring nucleosomes, promotes the organisation of nucleosomes into a helical array (Hergeth and Schneider, 2015; Thomas, 1999) to create a 30nm fibre that is found in both interphase chromatin and mitotic chromosomes. This fibre is further packaged to give an overall DNA packing ratio of around 10,000 (van Holde and Zlatanova, 1995).

Intuitively, this compaction of the genome represents a serious obstacle to DNA-templated processes such as transcription, replication, and DNA repair (Bell et al., 2011). Indeed, DNA templates were assembled *in vitro*, in presence or absence of one nucleosome, and then incubated with RNA polymerase. In the absence of the nucleosome, transcription proceeded, whereas the presence of a nucleosome prevented transcription (Knezetic and Luse, 1986; Lorch et al., 1987;

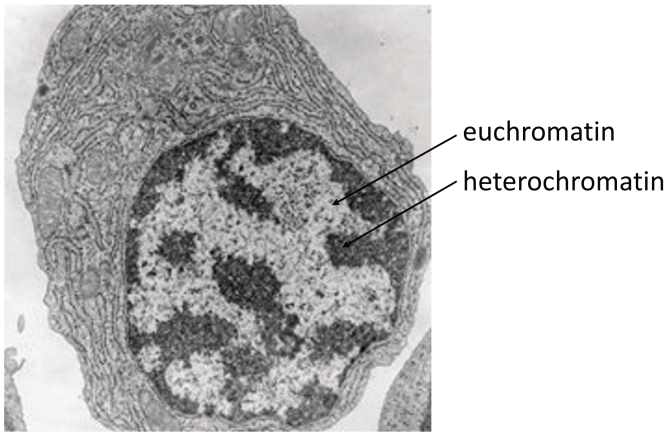


Figure 8. Euchromatin and heterochromatin in an interphase nucleus. Chromatin presents different layers of compaction. Euchromatin and heterochromatin correspond to light and dark regions respectively, as indicated on the right. (Euchromatin Dr. Jastrow's electron microscopy atlas).

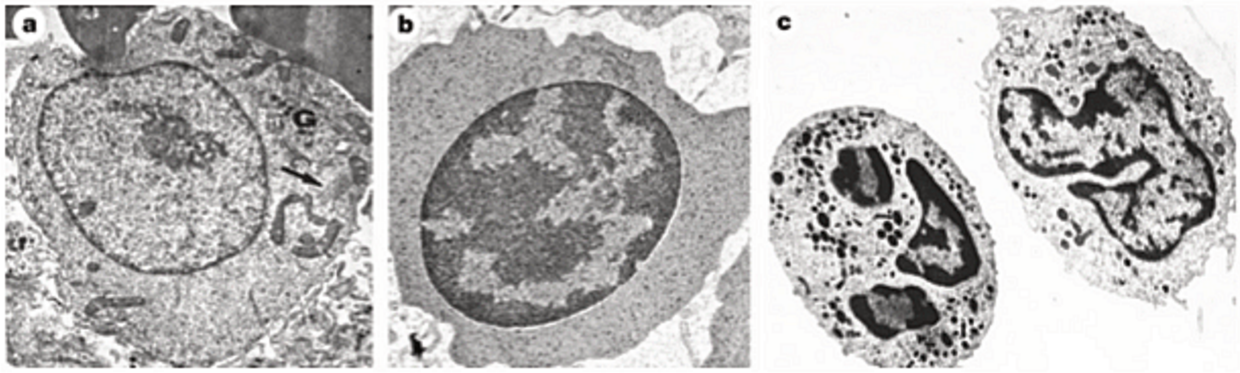


Figure 9. Patterns of chromatin condensation in haematopoietic cells. The electron micrographs show haematopoietic cells from normal human bone marrow, and illustrate the distinct patterns of chromatin condensation in distinct terminally differentiated cells. a) Proerythroblast (immature red blood cell). The nucleus contains almost no heterochromatin. b) Late erythroblast. Heterochromatin in the nucleus is predominant and appears as darkly staining regions. c) The nucleus of a monocyte (right) appears less condensed than that of a granulocyte (left). *From Francastel et al., 2000.*

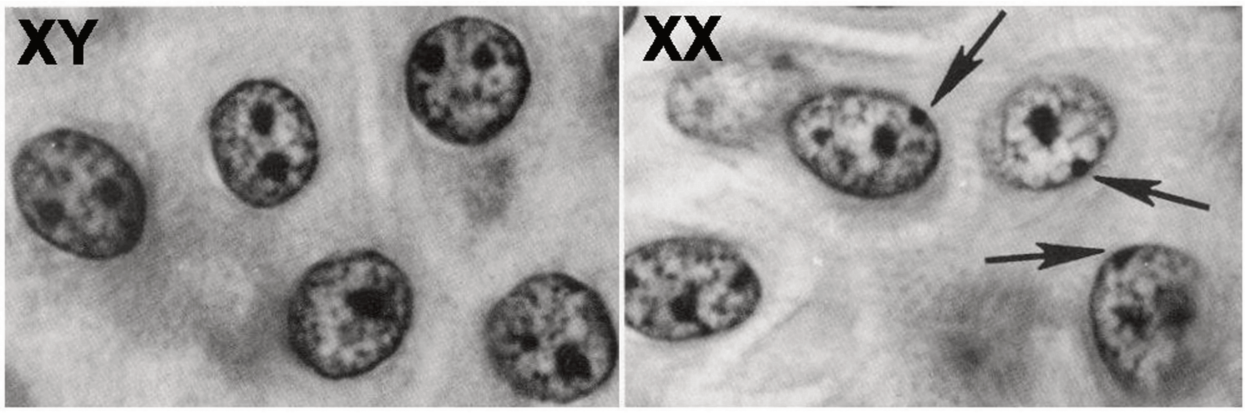


Figure 10. Female nuclei show an additional block of heterochromatin. On the left: nuclei of male cells (XY) are characterized by the presence of 1 or 2 dark dots, corresponding to heterochromatin. On the right: nuclei of female cells (XX) present an additional dot, an additional block of heterochromatin localizing at the periphery of the nucleus. This dot has been then "Barr body" and correspond to the inactive chromosome X. *From Bertram and Barr, 1949.*

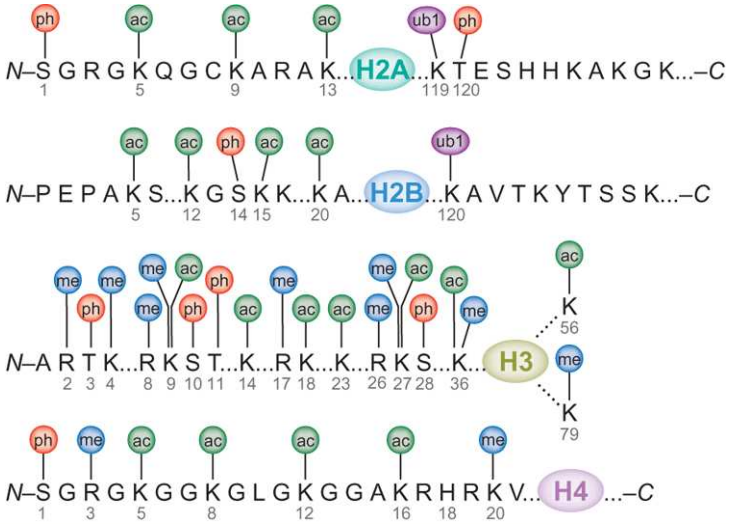
Wasylyk and Chambon, 1979). Moreover, in 1988, Han and Grunstein used a yeast strain carrying the gene coding for the histone H4 under the control of a GAL1 promoter, to activate H4 mRNA synthesis in presence of galactose or its repression with glucose (Kim et al., 1988). The inhibition of histone H4 expression induced the expression reporter genes, confirming that histones could inhibit transcription (Han and Grunstein, 1988).

However, examples that simple presence of histones could not inhibit transcription have previously been proposed in human cells, with Weintraub and Groudine showing that actively transcribed genes have histones in the promoter region but are in an altered conformation. In particular, the authors found that the globin locus, expressed specifically during erythroblasts differentiation (Groudine et al., 1974), were more sensible to deoxyribonuclease I in human red cells, compared to fibroblasts. This results suggested that a more compacted chromatin of a genomic region was associated with a lower expression of the same region (Weintraub and Groudine, 1976). Thus, chromatin was started to be seen as a dynamic structure, whose structure could change from a cellular type to another.

Chemical and biochemical observations have been complementary to microscopy analysis. In particular, in 1928, Emil Heitz was the first to observe that chromatin is not homogeneously organized along chromosomes. He defined two forms of chromatin as: "heterochromatin refers to a part of a chromosome that remains heteropyknotic after telophase and thus behaves opposite to euchromatin" (Passarge, 1979). Using electron microscopy, it was then possible to visualize the two forms of chromatin in the interphase nucleus: heterochromatin, a more condensed conformation, corresponding to the dark regions in the nucleus, and euchromatin, a less condensed conformation which corresponds to light regions (**Figure 8**).

Moreover, evidence of the dynamic nature of chromatin came from the observation that chromatin organization and compaction change upon differentiation. In general, the level of chromatin compaction increases during differentiation and accompanies the reduction of the differentiation potential and the fraction of the genome that remains transcriptionally active. It also varies with cell type, suggesting that it may have a great impact on the regulation of tissue-specific genome expression (Francastel et al., 2000) (**Figure 9**).

Since this first description, a great deal of studies focused on the characterization of the differences between heterochromatin and euchromatin. Heterochromatin is highly condensed, generally poor in genes and replicates late in S phase. In contrast, euchromatin is less condensed, gene-rich and replicates early in S phase (Goldman et al., 1984). However, the observation of nuclei from male and female mouse cells revealed additional differences in nuclear chromatin domains, with female nuclei presenting an additional heterochromatin region compared to male nuclei, which was called the Barr body (Barr and Bertram, 1949) (**Figure 10**). The link between



Modification	Histone	Residue	Known function
Acetylation	H2A	K5	Transcriptional activation
	H2B	K5, K12, K15, K20	Transcriptional activation
	H3	K4, K27, K36	Transcriptional activation
		K9	Transcriptional activation, histone deposition
		K14, K23	Transcriptional activation, histone deposition, DNA repeair
		K18	Transcriptional activation, DNA repair, DNA replication
	H4	K5, K12	Transcriptional activation, histone deposition, DNA repeair
		K8, K16	Transcriptional activation, DNA repair,
Methylation	H3	R2, K9, K27	Transcriptional repression
		R8	Transcriptional control?
		R17, R26, K4, K79	Transcriptional activation
	H4	R3	Transcriptional control?
		K20	Transcriptional repression
Phosphorylation	H2A	S1, T119	mitosis
	H2A X	S139	DNA repair
	H2B	S14	DNA repair, apoptosis
	H3	T2, S10, T11, S28	mitosis
Ubiquitylation	H2A	K119	Transcriptional repression
	H2B	K120	Transcriptional activation, meiosis

Figure 11. N-terminal and C-terminal histone modifications. Position of the major histone modifications: acetylation (ac), methylation (me), phosphorylation (ph) and ubiquitylation (ub) are shown. A table containing the known histone modifications and the putative function is presented underneath. *From Bhaumik et al., 2007 and HIstome: the Histone Infobase* http://www.iiserpune.ac.in/~coee/histome/index.php

the “Barr” body and a condensed X chromosome in female cells was made only later by Susumu Ohno (Ohno et al., 1959) and was critical for the formulation by Mary Lyon of the X-inactivation theory in 1961 (Lyon, 1961).

These studies contributed to divide heterochromatin into two different classes constitutive and facultative. Constitutive heterochromatin regions form at the same genomic regions in every cell type and at all stages of differentiation and development. The bulk of constitutive heterochromatin forms at pericentromeric regions and at telomeres. In contrast, facultative heterochromatin regions can alternate between different states in a development-specific manner, and usually contains genes that must be kept silent upon developmental cues. The most prominent example of facultative heterochromatin is the inactive X chromosome in females (Brown, 1966; Saksouk et al., 2015).

Subsequent studies described the DNA sequences associated with these two groups of heterochromatin. It is now assumed that constitutive heterochromatin is composed of tandem repetitive elements, while facultative heterochromatin does not show preferential sequences but is referred to regions which can change their chromatin state in response to cellular stimuli (Yunis and Yasmineh, 1971).

In the next paragraph, I will give an overview of the epigenetic characteristics of these regions, with a special emphasis on DNA methylation.

2.3. Histone modifications

The identification of the nucleosome structure showed the presence of unfolded tails that extend radially from the nucleosome core that can be the target of different post-translational modifications (Luger et al., 1997). The possibility that the modifications of histone tails could have an effect on transcription was first proposed almost 50 years ago (Allfrey and Mirsky, 1964; Murray, 1964). Since then, much research focused on the identification of the various histone modifications and their impact on transcription. Histone tails are the target of a wide range of post-translational modifications, including methylation, acetylation, phosphorylation, ubiquitylation and the latest ADP-ribosylation and SUMOylation (Kleinsmith et al., 1966; Ord and Stocken, 1967; Shiio and Eisenman, 2003; Stanley et al., 2001; Ueda et al., 1975). Each type of histone modification occurs on a particular residue, with a specific outcome in terms of chromatin compaction and biological function (summarised in **Figure 11**).

Histone modifications were first reported as early as 50 years ago (Allfrey and Mirsky, 1964) with Allfrey and colleagues realising the potential of such modifications “to affect the capacity of ribonucleic acid synthesis *in vivo*”. Histone acetylation and methylation were the first modifications to be discovered (Allfrey and Mirsky, 1964). The authors described the isolation of

acetylated and methylated histones and proposed that the acetyl-group, by neutralizing the positive charge of histones, would weaken their interaction with the negatively charged DNA backbone and make chromatin more amenable to gene activation and expression, but it was not the case for methylated histones. Acetylation was the first modification shown to correlate with gene active transcription (Allfrey and Mirsky, 1964) and, since then, it generated great interest. The correlation between histone acetylation and transcription was then shown by Allis and Schreiber teams. They confirmed that acetylation could reduce affinity between histones and DNA, or could be recognized by others factors, mediating an easier access to DNA sequence for transcription factors (Brownell et al., 1996; Taunton et al., 1996).

Methylation occurs on arginines or lysines. It is the only histone modification that can be added more than once on the same residue. Lysines can be either mono-, di- or tri-methylated. Arginines can be either mono- or di-methylated, which can be symmetric or asymmetric, depending if the modification occurs on the same nitrogen or on different nitrogen atoms. Histone methylation was originally associated with inactive genes suggesting that this mark could block the transcription by condensing nucleosomes (DeLange and Smith, 1971). The inhibition was not related to the neutralization of the positive charge of histone, as it was proposed for acetylation, but the underlying mechanisms was largely unknown. Now it is known that it can be associated with both activation and repression depending on the residue that is modified (Kouzarides, 2007).

Phosphorylation is found on serines, threonines and tyrosines. Like acetylation, it could counteract the positive charge of histones, leading to the disruption of interactions between histones and DNA. For this reason, histone phosphorylation was also associated with transcriptional activation, as it is the case for H3S10 phosphorylation in interphase (Kouzarides, 2007). However, counter-intuitively, this mark has been associated with chromatin condensation during mitosis (Wei et al., 1998).

Ubiquitylation occurs on lysines through the action of three enzymes that activate, conjugate and ligate ubiquitin groups to lysine residues (E1, E2 and E3 enzymes, respectively) (Hershko and Ciechanover, 1998). H2A and H2B histones present two well-characterized examples of ubiquitylation. Ubiquitylation of histone H2A on lysine 119 has been involved in transcriptional repression of coding genes as well as at centromeric repeats (Wang et al., 2004; Zhu et al., 2011). In contrast, ubiquitylation on lysine 123 of histone H2B is associated with transcription, by promoting the efficient reassembly of nucleosomes during the elongation of transcription by RNA Pol II (Pavri et al., 2006).

Less is known about ADP-ribosylation and SUMOylation of histones. ADP-ribosylation is catalysed by a family of enzymes called ADP-ribosyltransferases (ART) that promotes the addiction of ADP-ribose mark mainly on lysisnes. This mark is mainly associated with

Histone	Variant	Chaperone	Process
H2A	**H2AZ**	**ANP32E**	Gene activation, gene silencing and chromosome segregation
	H2AX	**(?)**	DNA repair
	macroH2A.1/2	**APLF**	Gene silencing (inactive X chromsome)
H3	**H3.1**	**CAF-1**	DNA replication, DNA repair
	H3.2	**CAF-1 (?)**	DNA replication, DNA repair
	H3.3	**HIRA** **DAXX**	Active transcription Deposition and telomeres and mouse pericentromeric regions
	CENP-A	**HJURP**	Centromere-specific histone variant

Table 2. Summary of H2A and H3 histone variants. For each histone variant, the corresponding chaperone (if known) and the process in which the histone variant is implicated are shown.

transcriptional activation and this could result from a more relaxed conformation of chromatin, increasing the accessibility to DNA (Hottiger, 2011).

SUMOylation occurs mainly on lysines through an enzymatic cascade very similar to ubiquitylation. This mark is supposed to antagonize histone acetylation. Hence, SUMOylation is thought to play a role in transcriptional silencing (Cubeñas-Potts and Matunis, 2013).

2.4. Histone variants, chromatin identity and dynamic

An additional level of complexity came from the identification of histone variants. With the exception of histone H4, all core histone proteins have variant counterparts, which often differ by surprisingly few amino acids (Hake and Allis, 2006).

Substitutions of the individual components of the histone core with structurally distinct histone variants provide a versatile mechanism through which epigenetic changes can be incorporated into the chromatin fibre itself. Histone variants allow to define specific chromosome regions, like centromere or inactive X chromosome, and participate in the dynamics of chromatin during transcription, DNA repair or DNA replication and can function as temporary place-holders for other histones (Buschbeck and Hake, 2017; Gurard-Levin and Almouzni, 2014) (**Table 2**).

2.4.1. Histone variants in transcription and DNA repair

Affinity purification of H3 variants-associated complexes in human cells, followed by mass spectrometry identification, shows that distinct chromatin assembly factors specifically associate with each histone variant *in vivo*: the H3.1 complex contains the CAF-1 histone chaperone, which stimulates its deposition in a DNA-synthesis-dependent manner, whereas the H3.3 complex contains the HIRA histone chaperone, which promotes the deposition of the replacement H3.3 variant independently of DNA synthesis (Ray-Gallet et al., 2002; Tagami et al., 2004). The replication-independent deposition of H3.3 in chromatin is linked to transcriptional activation. The deposition of H3.3 is not restricted to the promoter region but extends throughout the transcribed region and correlates with the presence of transcription-competent RNA Pol II. Moreover, it was proposed that the deposition of H3.3 could be a mechanism to remove inactivating histone marks, like H3K9me3, from gene promoters (Daury et al., 2006). The nucleosomes containing these histone variants may mark sites for active transcription after mitosis, constituting a putative bookmark for gene activation (Ahmad and Henikoff, 2002).

H2A.Z is hypothesized to affect the conformation/stability of the nucleosome subunit as well as that of the higher levels of folding of the chromatin fibre (Redon et al., 2002). Biochemical studies suggest that nucleosomal arrays containing H2A.Z may be only partially compact, highly

susceptible to chromatin disassembly, thereby facilitating transcription (Redon et al., 2002). Studies in yeast showed that H2A.Z incorporation into chromatin is achieved through an ATP-dependent chromatin remodelling complex, called SWR1 that promote the exchange of canonical H2A with H2A.Z variant (Wu et al., 2005). Furthermore, at least in yeast, H2A.Z has been shown to prevent the binding of silencing proteins and thus the spreading of heterochromatin into neighbouring euchromatin regions (Meneghini et al., 2003).

Histone variants are also crucial genome integrity sensors (Li and Fang, 2015). H2A.X was identified as one of the key histones to undergo various post-translational modifications that include acetylation, phosphorylation and ubiquitination, in response to DNA double-strand breaks (DSBs). H2A.X marks the damaged DNA double helix and is thought to facilitate local recruitment and retention of DNA repair proteins and chromatin remodelling factors to restore genomic integrity (Rogakou and Sekeri–Pataryas, 1999).

A dynamic replacement between histone variants was observed in 1984 by Grove and Zweidler in mammals, along which H3.1 and H3.2 are the replicative histones deposed during the S phase, while H3.3 is the replacement histone deposed throughout the cell cycle (Grove and Zweidler, 1984).

2.4.2. Histone variants that define specialized chromosomal regions

Centromeres are a paradigm for epigenetically-defined chromosomal domains. Indeed, the primary determinant for centromeric identity is the presence of nucleosomes containing a specific variant of histone H3, termed CENP-A in mammals. The protein was discovered as one of the epitopes recognized by anti-centromeres antibodies found in patients suffering from a form of scleroderma known as the "CREST syndrome" (*Calcinosis, Raynaud's syndrome, Esophageal dismotility, Sclerodactyly and Telangiectasia*), and was given its name (*CENtromeric Protein A*) due to its exclusive centromeric localization (Earnshaw and Rothfield, 1985). Later studies demonstrated that CENP-A is *a bona fide* histone variant, able to substitute the canonical histone H3 inside a nucleosome (Palmer et al., 1987, 1991). Since then, CENP-A has been shown to be universal in eukaryotes with orthologs identified in mouse, Drosophila (Cid), S. pombe (Cnp1), C. elegans (HCP-3), and S. cerevisiae (Cse4) (Buchwitz et al., 1999; Henikoff, 2000; Kalitsis et al., 1998; Stoler et al., 1995; Takahashi et al., 2000). CENP-A is the anchor point for the formation of the kinetochore, a large protein assembly that connects chromosomes to microtubules of the mitotic and meiotic spindles in order to distribute the replicated genome from a mother cell to its daughters (Musacchio and Desai, 2017). CENP-A depletion results in the mislocalization of most kinetochore proteins, whereas depletion of most kinetochore proteins has no effect on CENP-A localization (Fachinetti et al., 2013). Moreover, depletion of CENP-A by targeted gene disruption,

antibody microinjection, or RNA-mediated interference confers a lethal phenotype in every organism tested to date (Howman et al., 2000). Conversely, overexpression of CENP-A results in its mislocalization to non-centromeric regions and the formation of ectopic kinetochores, suggesting that CENP-A is both a structural and functional foundation of the kinetochore (Heun et al., 2006; Lacoste et al., 2014). Localization studies have indicated that the presence of CENP-A onto centromeric DNA is an early event in kinetochore assembly.

CENP-A is highly expressed in the G2 phase but is loaded into centromeres only in G1 phase through its chaperon HJURP, hence at each round of replication, CENP-A decreases by 50% (Jansen et al., 2007). This observation led to ask if there is one particular histone which working as placeholder for following CENP-A incorporation. Indeed, H3.1 and H3.3 are deposited at centromeres during S phase, but only H3.3 is then replaced by CENP-A suggesting that H3.3 plays as placeholder for CENP-A (Dunleavy et al., 2011).

Other histone variants which identify specific chromosomal regions are macroH2A1 and macroH2A2. MacroH2A variants have an N-terminal domain with high sequence homology with the canonical H2A histone, but also contain a 25kDa C-terminal non-histone macro-domain of unknown function (Aravind, 2001). This macro-domain constitutes two thirds of the molecular mass of the protein and shows sequence similarity with viral proteins that interact with RNA-binding proteins (Pehrson and Fuji, 1998). MacroH2A variants are generally considered transcriptionally repressive in nature due to their association with forms of condensed chromatin such as the inactive X chromosome (Xi) (Costanzi et al., 2000; Costanzi and Pehrson, 1998; Csankovszki et al., 1999; Heard and Disteche, 2006; Hernández-Muñoz et al., 2005) and inactive genes (Buschbeck et al., 2009; Changolkar et al., 2010; Gamble and Kraus, 2010; Ratnakumar et al., 2012). A kinetics of events that lead to chromosome X inactivation was proposed to start with the transcription of the long non-coding RNA XIST (X-inactive specific transcript) that is followed by a depletion in transcriptional activating histone modifications, silencing by the repressive mark H3K27me3, and a subsequent incorporation of macroH2A variants (Heard, 2004). Macro-H2As were found to interact directly with XIST RNA, confirming its RNA-binding ability (Gilbert et al., 2000). The Deletion of XIST in mouse embryonic fibroblasts disrupts the localization of macro-H2A variants at inactive X chromosome, without altering the maintenance of inactivation (Csankovszki et al., 1999). This suggests that the role of macroH2As in X inactivation needs further elucidations (Csankovszki et al., 2001). The mechanism of inactivation of X chromosome mediated by XIST will be described more in details in a section dedicated to the role of non-coding RNA and heterochromatin formation.

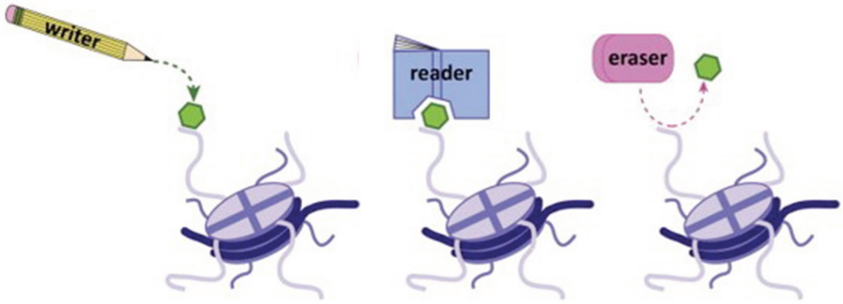


Figure 12. Establishment, reading and erasure of histone modifications. Histone modifications are established by writers (pencil) as HAT or HMT, in the case of acetylation or methylation respectively. The modification can be read by readers (book) via specific domains. Readers may be themselves effectors or may serve as platform for other chromatin effectors. Histone marks are not irreversible; they can be removed by erasers (rubber). *From Gardner et al., 2011.*

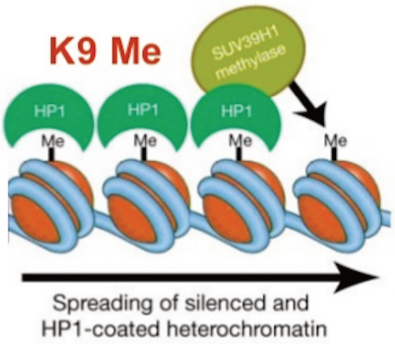


Figure 13. Heterochromatin formation promoted by HP1 reader. HP1 reader recognizes H3K9me3 via its chromodomain. The interaction with this mark allows the recruitment of SUV39H1 and the increasing deposition of H3K9me3 and heterochromatin formation. *Adapted from Bannister et al., 2001.*

Histone mofication	Writers	Readers (domains)	Erasers
Acetylation	HAT1 (KAT1) CBP/P300 (KAT3B) GCN5/PCAF (KAT2A/KAT2B) TIP60 (KAT5) HB01 (KAT7)	Bromo-domain	HDAC1-11 SIRT1-3 and 6
Lysine Methylation	SUV39H1 (KMT1A) SUV39H2 (KMT1B) G9a (KMT1C) ESET/SETDB1 (KMT1E) EuHMTase/GLP (KMT1D) CLL8 (KMT1F) MLL1-5 (KMT2A-E) SET1A (KMT2F) SET1B (KMT2G) ASH1 (KMT2H) SET2 (KMT3A) NSD1 (KMT3B) SYMD1-3 (KMT3C-E) DOT1 (KMT4) Pr-SET 7/8 (KMT5A) SUV420H1 (KMT5B) SUV420H2 (KMT 5C) EZH2 (KMT6A) RIZ1 (KMT8)	Ankyrin domain Chromo-domain MBT domain PHD domain PWWP domain Tudor-domain WD domain	LSD1/BHC110 (KDM1) JHDM1a (KDM2A) JHDM1b (KDM2B) JHDM2a (KDM3A) JHDM2b (KDM3B) JMJD2A/JHDM3A (KDM4A) JMJD2B (KDM4B) JMJD2C/GASC1 (KDM4C) JMJD2D (KDM4D)

Table 3. Main histone acetyltranferases and lysine methyltranferases. Human acetyltransferases and lysisne methyltranferases. *Adapted from Kouzarides 2007, HIstome: the Histone Infobase* http://www.iiserpune.ac.in/~coee/histome/index.php *and Rockland website:* http://www.rockland-inc.com/epigenetic-tools.aspx

2.5. Histone writers, readers and erasers

All the histone modifications outlined above are reversible and are dynamically regulated through the dual action of enzymes that have become known as "chromatin writers" and "chromatin erasers" (**Figure 12**). Writers are enzymes that establish histone modifications. Histone acetylation, for example, is mediated by histone acetyltranferases (HATs). Histone methyltransferases (HMTs) are named according to their specificity towards lysine (KMTs) or arginine (RMTs). In the reverse reaction, chromatin erasers, such as histone deacetylases (HDACs) and histone lysine demethylases (KMDs), catalyse the removal of histone marks. Within these classes of chromatin writers, distinct enzymes may show a substrate specificity for a particular position on the histone tail.

As mentioned earlier, histone acetylation and phosphorylation could directly impact chromatin compaction, although it is not always the case and the question of how they influence chromatin structure remains. One hypothesis is that they are "read" by effector proteins (Yun et al., 2011) (**Figure 12**). A number of proteins termed "chromatin readers" have been described to integrate a given chromatin mark into an information for an appropriate response. The interaction between the reader and a given chromatin mark is mediated by a specific domain.

The first of such domains that has been identified was the bromo-domain that preferentially binds to acetylated histone tails (Dhalluin et al., 1999; Winston and Allis, 1999). For example, GCN5 was first discovered in *Tetrahymena thermophila* to contain a bromo-domain and to function as a platform for the histone acetylase HAT1, although it was then found to be a HAT itself (Brownell et al., 1996) able to promote gene activation through its interaction with adaptors of the transcriptional machinery (Sterner and Berger, 2000).

Subsequent studies showed that histone methylation could be recognized by several domains, and more precisely PWWP, PHD, chromo-domain or Tudor-domain (Kouzarides, 2007). For example, H3K4me3 is recognized by readers with a PHD domain. Nucleosome Remodelling Factors (NURF) is able to recognize this mark and to promote chromatin disruption at the promoter regions and activate transcription (Wysocka et al., 2006). Another PHD finger showing highly selective binding for H3K4me3 was also found within TAF3, a component of the basal transcription factor TFIID, suggesting that this mark may help to position the basal transcription machinery (Vermeulen et al., 2007). Alternatively, heterochromatin Proteins 1 (HP1) contain a chromo-domain which allows the interaction with H3K9me3 and the formation of heterochromatin (Lachner et al., 2001) (**Figure 13**). The mechanisms of heterochromatin formation will be described in the following chapters.

In **Table 3**, I summarised the proteins working as writers, the main domains contained in readers and erasers for histone acetylation and lysine methylation.

Chromatin readers can have themselves an enzymatic activity or they can function as platforms. They can mediate the recruitment of another protein or a complex of proteins in order to remodel chromatin and orchestrate a response to a given stimulus. An appropriate signal can then induce the recruitment of chromatin erasers that catalyse the removal of their respective marks, disabling the interaction between the histone modification and the readers.

An additional level of complexity is suggested by the cross-talk between different histone modifications, that probably helps the general control. This cross-talk can follow different forms: i) competition between different modifications for the same site. For example, a lysine can be either acetylated or methylated and the two modifications are mutually exclusive; ii) the establishment of a modification may depend on the presence of another modification. For example, ubiquitylation of H2B is required for H3K4 di- and trimethylation during the activation of transcription (Kim et al., 2009); iii) the presence of one modification may inhibit the interaction of a chromatin reader and adjacent marks. This is the case of phosphorylation of H3S10 that disrupts the interaction between H3K9me3 and its reader HP1 during mitosis (Fischle et al., 2005; Hirota et al., 2005).

Generally active and repressive histone marks are not found in the same region, although, in mouse ES cells, many key developmental genes are marked by the dual presence of activating H3K4me3 and repressive H3K27me3 in what have been referred to as “bivalent”, or “poised” chromatin domains (Azuara et al., 2006; B. E. Bernstein et al., 2006). H3K27me3 is dominant as the underlying gene remains silent, but the presence of H3K4me3 has been proposed to ‘poise’ the underlying gene for transcriptional activation. These bivalent domains are progressively resolved during differentiation by loss of either activating or repressive mark (B. E. Bernstein et al., 2006; Mikkelsen et al., 2007).

The evidence presented so far suggests that the transcriptional state of genomic loci results from the deposition of a combination of different histone modifications. Moreover, the discovery of chromatin readers implied an additional layer of information “encoded” by a diverse set of histone modifications. This observation led Strahl and Allis to propose the existence of a combination of histone modifications, on one or more histone tails, called “histone code” which may be read by proteins promoting downstream events (Strahl and Allis, 2000). A similar “epigenetic code” involving histone acetylation was proposed by Turner in the same year (Turner, 2000).

The fact that histone modifications could have strong consequences on transcription needed further investigations. This was one of the reasons why, in September 2003, the National Human Genome Research Institute started the Encyclopedia of DNA Elements project (ENCODE project). The aim of the project was “to build a comprehensive list of functional elements in the

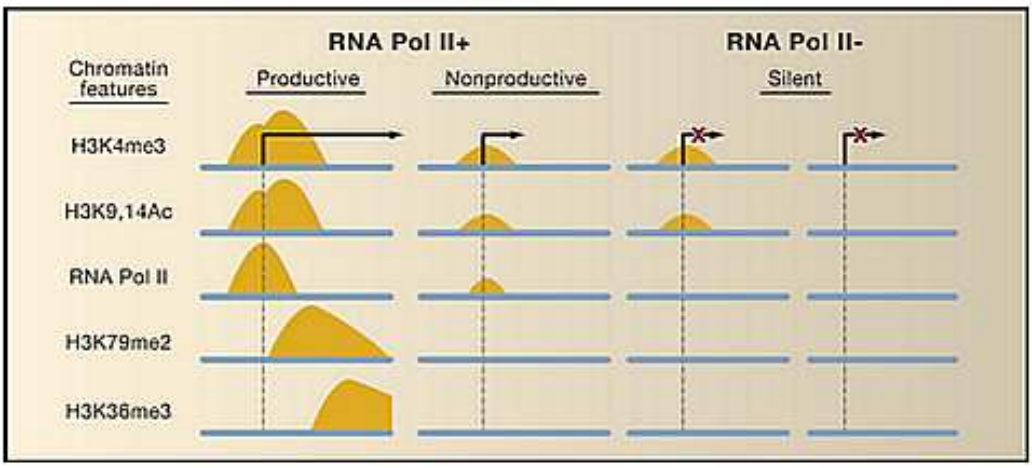


Figure 14. Active chromatin marks, RNA polymerase and transcription. Two classes of genes transcribed by RNA polymerase II (RNA Pol II+) are present in the human genome, defined by the presence of covalent modifications of histone H3. Both classes are marked around the transcription start site with H3K4me3 and H3K9,14Ac. However, only one class (productive) is also marked downstream of the transcription start site by H3K36me3 and H3K79me2, hallmarks of ongoing transcriptional elongation. Genes lacking these latter marks initiate transcription at a relatively low efficiency but do not proceed to the elongation stage of transcription (non-productive). A third class of genes (RNA Pol II−, left) harbours H3K4me3 and H3K9,14Ac around the transcription start site but is not simultaneously bound by Pol II. A fourth class of genes (RNA Pol II−, right) is enriched for neither Pol II nor any of the above-mentioned histone marks. The promoter region of this class of genes may adopt a chromatin structure that is refractory to recruitment of Pol II and associated histone-modifying enzymes. *From Lorincz and Schubeler 2007.*

human genome and regulatory elements that control cells and circumstances in which a gene is active" (ENCODE project website: www.encodeproject.or. The ENCODE project is still ongoing today, maintaining the original aim. The consortium is not limiting its analysis to the human genome and extended genome annotation to mice, flies, worms. In January 2017 three hundred ninety-three new datasets were released, including new ChIP-Seq and RNA-Seq data obtained in human, mouse, worm and fly (www.encodeproject.org).

2.6. The ENCODE project

In the **Chapter I**, I mentioned that the ENCODE project had been crucial to identify "biochemically functional" elements of the human genome and the elements controlling the packaging and folding of DNA (**Figure 2**).

Among the several approaches used by ENCODE, the consortium assessed the chromosomal locations of up to 12 histone modifications and variants in 46 cell types. The goal was to associate specific peaks of enrichment in every histone modification/histone variants with gene landmarks and transcriptional states.

Active promoters were shown to be marked by either histone acetylation, H3K9Ac, H3K14Ac or H3K27Ac, or histone H3 methylation of K4, in particular H3K4me2 or H3K4me3 (Bell et al., 2011) (**Figure 14**). In contrast, the repressive marks, H3K9me3 and H3K27me3, were observed at promoters of inactive genes. However, H3K9me3 or H3K27me3 are generally mutually exclusive, and the presence of one mark or the other depends on the cell line analysed and the genomic location (ENCODE 2012).

The presence of H3K36me3 and H3K79me2 was found in gene bodies of actively transcribed genes. In particular, H3K79me2 is preferentially found at the 5'UTR of gene bodies while H3K36me3 is mostly found at the 3' region along the gene body (Huff et al., 2010). H3K36 methylation, in turn, was shown to be recognized by histone deacetylases and to promote propagation of a permissive chromatin environment and transcriptional elongation by RNA Pol II (Joshi and Struhl, 2005).

The distribution of histone marks associated with active transcription was compared to sites of RNA Pol II occupancy, confirming the earlier observations of Guenther and colleagues who first described an overlap between "permissive" histone marks like H3K4me3 and H3K14Ac and loading of RNA Pol II (Guenther et al., 2007). However, only a subset of these promoters also present hallmarks of ongoing transcriptional elongation, like H3K36me3 and H3K79me2. In the absence of these marks, transcription is initiated but does not proceed to productive elongation, characterized by a poised RNA Polymerase (Guenther et al., 2007; Lorincz and Schübeler, 2007) (**Figure 14**). Moreover, active marks are also found at promoters devoid of RNA Pol II. The

deposition of these marks may depend on sporadic recruitment of polymerase (Guenther et al., 2007; Lorincz and Schübeler, 2007).

ENCODE also reported the great enrichment of H3K9me3, corresponding to large blocks more than discrete peaks at repetitive elements, reinforcing the findings described in previous works in yeast, *Drosophila* and mouse (Elgin and Grewal, 2003; Guenatri et al., 2004; Hall et al., 2002; Henikoff, 2000; Lehnertz et al., 2003). This observation confirmed that H3K9me3 is the histone modification which characterizes constitutive heterochromatin, together with H4K20me3 (Schotta et al., 2004).

3. The special case of heterochromatin

As I mentioned before, heterochromatin has been divided into two groups: facultative and constitutive heterochromatin. In general, heterochromatin is characterized by hypoacetylation of histones and high levels of repressive marks, like DNA methylation, H3K9me3, H3K27me3 and H4K20me3. DNA methylation and H3K9me3 are generally associated with constitutive heterochromatin, while H3K27me3 is the mark of facultative heterochromatin (Schotta et al., 2004; Trojer and Reinberg, 2007). This discrimination in terms of chromatin marks relies on the fact that DNA methylation and H3K9me3 may be present at the same genomic regions while DNA methylation and H3K27me3 are mutually exclusive (Lindroth et al., 2008).

Less is known about H4K20me3, but it was found at pericentromeric and telomeric heterochromatin, suggesting that this mark may contribute to the formation of constitutive heterochromatin (Jørgensen et al., 2013).

DNA methylation and the interplay with histone modifications in the special case of heterochromatin will be described in the following chapter, but I will now focus on the histone modifications that distinguish facultative and constitutive heterochromatin.

3.1. Facultative heterochromatin: H3K27me3 and Polycomb

The most prominent example of facultative heterochromatin is the inactive X chromosome in female cells. Other examples include developmentally regulated genes, like the HOX genes. They are both characterized by the presence of H3K27me3 (E. Bernstein et al., 2006; Heard, 2004; Mohn et al., 2008). H3K27me3 is established and maintained by Polycomb group proteins (PcG) that were originally identified as important regulators of developmentally relevant homeotic (HOX) genes in Drosophila and epigenetic cellular memory (Ringrose and Paro, 2004).

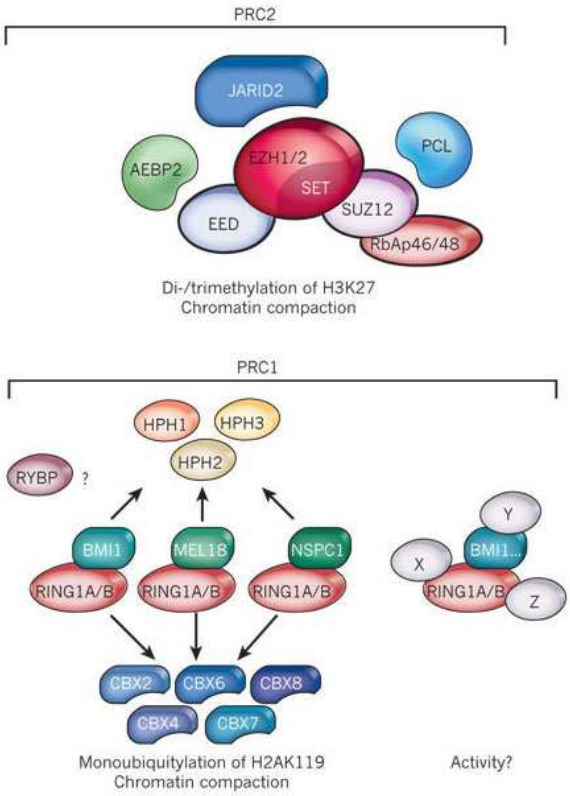


Figure 15. Composition of PRC2 and PRC1. In PRC1, the diagrams shown on the left correspond to the classical PRC1 complexes, whereas those on the right correspond to the so-called PRC1-like complexes. The "pocket" shape of the CBX proteins represents the chromodomain that specifically recognized H3K9/27me3. *Adapted from Margueron and Reinberg 2011.*

Progress over the past decade has defined two main protein complexes, Polycomb Repressive Complex 1 (PRC1) and PRC2, with fundamental roles in PcG-mediated gene silencing. PRC2 catalyses H3K27me3, EZH2 being the HMT. This mark is in turn recognised by chromodomain-containing proteins of PRC1 (CBX family of proteins), whose recruitment is thought to lead to transcriptional silencing (Simon and Kingston, 2009) (**Figure 15**).

PRC2 is composed of different proteins with different proprieties. EZH1 or 2 are the proteins responsible of H3K27me3 deposition. In the PRC2 complex there are co-factors that have been shown to be important for the stabilization of PRC2, including JARID2 or PCL, which are involved in the recruitment of PRC2 complex to chromatin (Margueron and Reinberg, 2010). In particular, JARID2 is a DNA-binding protein and is thought to be required to recruit PRC2 complex to CG-rich regions (Li et al., 2010; Peng et al., 2009). Among all these components, SUZ12 and RBAP46 have been shown to anchor the complex to chromatin *via* their binding to unmodified H3. This allows the catalytic components, EZH1 and EZH2, to catalyse the di- and tri-methylation of H3K27 from a mono-methylated substrate (Nekrasov et al., 2005). Alternatively, PRC2 can also be recruited to H3K27me3 through the C-terminal domain of EED, reading H3K27me3 through its WD-40 domain (Margueron et al., 2009). The resolution of the molecular structure of PRC2 suggested that EED binds to one nucleosome and places EZH2 close to the following nucleosome, suggesting a mechanism for H3K27me3 spreading (Ciferri et al., 2012). The enzyme responsible for H3K27me1 remains unknown. Using *in vitro* studies, Tachibana and colleagues showed that G9A could methylate H3K27 (Tachibana et al., 2001). However, following studies showed that, although H3K27me1 was reduced in G9A knock-out cells, H3K27me2 and H3K27me3 were not altered. This observation strongly suggested that G9A was not a good candidate (Wu et al., 2011).

PRC2 or H3K27me3 do not seem to play a direct role in chromatin condensation. However, PRC2 and H3K27me3 are thought to mediate gene repression by recruiting other factors like PRC1.

The core of PRC1 complex is composed of two E3 ubiquitin-ligase called RING1A and RING1B. These enzymes catalyse the last step of the addition of an ubiquitin group from a donor, called E2, to its targets (Berndsen and Wolberger, 2014). The PRC1 complex is defined then by the mutually exclusive presence of one of the six Polycomb group RING finger proteins (PCGF 1-6) proteins. The association of RING1A/B with a specific PCGF protein defines six distinct PRC1 complexes that can be referred as PRC1.1 to PRC1.6. The six PRC1 complexes have distinct protein partners and are functionally different (Gao et al., 2012). These modules can associate with the chromo-domain-containing CBX proteins, which are small proteins that can interact only with PCGF2 and 4 and are crucial for the recruitment of PRC1 on PRC2 targets, by binding H3K27me3

(E. Bernstein et al., 2006). This recruitment allows then the mono-ubiquitylation of H2AK119 that is thought to repress transcription by interfering with polymerase elongation (Shao et al., 1999; Stock et al., 2007; Wang et al., 2004). The mechanism underlying this transcriptional repression is still unknown, but we can not exclude that could rely on steric hindrance caused by the PRC1 complex, or the recruitment of a still unknown reader of H2AK119 mono-ubiquitylation. It was shown that PRC1 components can promote condensation of nucleosome arrays *in vitro* and also long-range interactions, suggesting a possible role for PRC1 in higher-order repressive chromatin structure (Francis et al., 2004; Lanzuolo et al., 2007; Tiwari et al., 2008). A very interesting work by Eskeland and colleagues showed another possible scenario. They observed that RING1B component was able to promote chromatin condensation independently of its ubiquitin-ligase activity, suggesting that PRC1 complex could create a physical barrier that interferes with the binding of the transcriptional machinery (Eskeland et al., 2010).

The molecular mechanism of silencing by PRCs remains unclear. Additional efforts will be required to understand whether Polycomb repression is based either on chromatin condensation or on the creation of a physical barrier to transcriptional machinery or probably a combination of the two mechanisms.

3.2. Constitutive heterochromatin: H3K9me3 and H3K20me3

Constitutive heterochromatin, as outlined above, assembles mainly on large blocks of tandem repeats at pericentromeres, telomeres as well as families of interspersed DNA repeats. The fundamental mechanism leading to the formation of constitutive heterochromatin is the deposition of H3K9me3 (Nakayama et al., 2001). Two of the enzymes that are responsible for the establishment of this mark are SUV39H1/2 (Melcher et al., 2000). Mice knock-out for SUV39H1/2 (SUV39HDN) are viable, although they born under sub-Mendelian ratios, are sterile and present impaired formation of heterochromatin associated with genomic instability. SUV39H1/2 were shown to be responsible for the establishment of H3K9me3 at pericentromeric repeats as well as at telomeric, subtelomeric regions and at LINE elements (Bulut-Karslioglu et al., 2014; Lehnertz et al., 2003). At pericentromeric repeats, SUV39H1/2 are able to recognize H3K9me1 through their chromo-domain and to specifically catalyse the H3K9me3 (Peters et al., 2001; Rea et al., 2000). H3K9me1 is established specifically by PRDM3 and PRDM16 (Positive Regulatory Domain zinc-finger protein 3 and 16). The enzymatic reaction takes place in the cytoplasmic pool of histones (Black et al., 2012) but the mechanism bringing the modified histones to pericentromeres is still unknown. Interestingly, the absence of PRDM3 and PRDM16 leads to the disappearance of H3K9me3 and the disruption of DAPI-dense heterochromatin foci, which is not the case in SUV39HDN (Pinheiro et al., 2012). One of the mechanisms through which

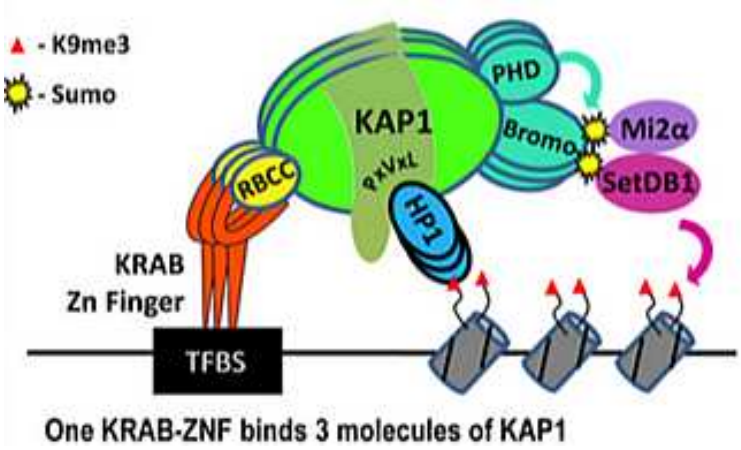


Figure 16. Recruitment of KAP1 on the genome and recruitment of SETDB1. The recruitment depends on interaction of the RBCC domain of KAP1 with a KRAB-ZNF that is bound to its recognition motif (indicated as TFBS); 3 molecules of KAP1 interact with a KRAB-ZNF. The PHD domain sumoylates the bromo-domain, leading to recruitment of SETDB1 and Mi2 and creation of the H3K9me3 mark on nearby nucleosomes. HP1 can bind to KAP1 at the PxVxL motif and also to H3K9me3, stabilizing the bound KAP1-containing complex. *From Iyengar and Farnham 2011.*

H3K9me3 is the building block of heterochromatin is through the recruitment of HP1 proteins family. HP1, in turn, is involved in the recruitment of SUV39H1/2, creating a self-reinforcing loop and leading to the spreading of H3K9m3 and HP1 on heterochromatin (Lachner et al., 2001).

The other enzyme which is implicated in establishing H3K9me3 is ESET/SETDB1 (Greer and Shi, 2012). The deletion of SETDB1 gene leads to peri-implantation lethality between 3.5 and 5.5 days post coitum (dpc), preventing the correct formation of the inner cell mass (ICM) and establishment of ES cells (Dodge et al., 2004; Matsui et al., 2010). SETDB1 can successively catalyse the mono-, di- and tri-methylation of H3K9 (Schultz et al., 2002; Wang et al., 2003). SETDB1-mediated H3K9me3 deposition seems to be crucial for the formation of heterochromatin at interspersed repeats, in particular at retrotransposons of classes I and II in ES cells but not in somatic cells (Matsui et al., 2010). The mechanisms of heterochromatin formation at retrotransposons is complex, because it varies with the class of elements and with the developmental stage. However, it was shown that SETDB1 could play a crucial role in deposing H3K9me3 at ERV during the developmental stages characterized by absence of DNA methylation, such as early embryonic and germ line development (Wu and Zhang, 2010). Moreover, recent studies proposed that the formation of heterochromatin at retrotransposons as well as at telomeric regions may depend on the deposition of H3.3 histone variant *via* a DAXX-dependent mechanism. It seemed to be contradictory with other studies that showed H3.3 enrichment in transcribed regions, although, as we will see in the next section, transcription of DNA repeats is required for assembly of heterochromatin (Nishibuchi and Déjardin, 2017).

SETDB1 is tightly associated with KAP1 protein, also known as TIF1β or TRIM28 (Nielsen et al., 1999; Ryan et al., 1999). KAP1 is a multi-domain protein defined as a platform for the recruitment of repressive enzymes. It contains an N-terminal Ring finger and a coiled coil domain (RBCC domain), a C-terminal PHD-bromo domain and a central HP1 binding domain (Iyengar and Farnham, 2011). The RBCC domain is crucial for the interaction with the N-terminal part of KRAB-Zinc finger proteins (KRAB stands for Kruppel-Associated box), a family of proteins able to bind DNA through their C2H2 zinc finger domain. Around half of annotated transcription factors belong to the zinc finger family and, among them, 40% show a KRAB domain (Bellefroid et al., 1991). The interaction of KAP1 with KRAB-zinc finger proteins allows the recruitment of KAP1 at specific loci. KAP1 then recruits SETDB1 (**Figure 16**) as well as the catalytic subunit of the Nuclear Remodelling and Deacetylation Complex (NuRD complex), promoting the compaction of chromatin (Schultz et al., 2001; Underhill et al., 2000).

Despite the fact that the molecular function of H4K20me3 in heterochromatin formation is still unknown, it is important to mention that this mark is found at pericentromeric heterochromatin as well as at telomeric regions (Saksouk et al., 2015; Schotta et al., 2004). The deposition of this

mark is catalysed by the histone methyltransferase SUV4-20H1/2 which specifically catalyses the trimethylation of H4K20. In the absence of H3K9me3 at pericentromeres, H4K20me3 is lost, suggesting a tight dependency of this mark on H3K9me3. Indeed, it was observed that HP1 can interact with SUV4-20H1/2, leading to propose that, once H3K9me3 is established, HP1 recognizes this mark and recruits SUV4-20H1/2 to establish H4K20me3 at H3K9me3-rich regions (Schotta et al., 2008). However, a very recent paper showed that SMYD5 can also establish H4K20me3 at heterochromatin regions, like at pericentromeric repeats and retrotransposons. Deletion of SMYD5 in mouse ES cells leads to the loss of H4K20me3 as well as the redistribution of H3K9me3 and HP1, suggesting that the deposition of H4K20me3 also influences the H3K9me3 patterns (Kidder et al., 2017).

Another particularly interesting mechanism in the formation of both facultative and constitutive heterochromatin is through non-coding RNAs. The next paragraph will be dedicated to the description of this mechanism.

3.3. Non-coding RNA and assembly of heterochromatin

3.3.1. The long non-coding RNA XIST (X-inactive specific transcript) and inactive X heterochromatin

The inactive X chromosome was the first example of facultative heterochromatin described more than 60 years ago, and we are still trying to understand the underlying mechanism. The process of inactivation and "heterochromatisation" of the chromosome depends on the presence of a non-coding RNA, XIST, transcribed from a particular locus of the X chromosome called the X-chromosome inactivation centre (Avner and Heard, 2001). The transcription of XIST and the remarkable coating of the future inactive X by this 17 kb-long non-coding RNA were shown to be the triggers of the inactivation process (Chaumeil et al., 2002; Heard et al., 2001; Okamoto et al., 2004). At the chromatin level, a global deacetylation of histone H4 occurs at the coated X-chromosome. The mechanism of this deacetylation is still unknown. It could be either a passive histone-loss during replication, or an active removal. This deacetylation is then followed by the loss of RNA Pol II (Chaligné and Heard, 2014). Several repressive histone modifications become enriched on the X chromosome, such as H3K9me2, H3K27me3 and H4K20me1, H2AK119ub being the best documented in the mouse (de Napoles et al., 2004; Fang et al., 2004; Heard et al., 2001; Kohlmaier et al., 2004; Peters et al., 2002; Plath et al., 2003). H3K27me3 precedes the other marks (Okamoto et al., 2004). Indeed, PRC2 can be recruited by XIST on the X chromosome undergoing inactivation, promoting H3K27me3 (Plath et al., 2003). There are two components of PRC2 which can bind RNA: EZH2 and JARID2, although the interaction seems to be more

specific for JARID2 since the depletion of this protein impairs the recruitment of PRC2 at the inactive X chromosome (da Rocha et la., 2014). PRC1 enrichment is then observed, through the recognition of H3K27me3 (Escamilla-Del-Arenal et al., 2011). PRC1 catalyses the mono-ubiquitylation of H2AK119, promoting chromatin condensation. The processes that follow, which are the deposition of the histone variant macroH2A and DNA methylation at the gene promoters, are considered later events and contribute to the maintenance of the inactive state of the X-chromosome (Heard, 2004).

3.3.2. (peri)centromeric transcripts and heterochromatin assembly

The regions around centromere were often viewed as a transcriptionally inactive domains of the chromosome. However, although the sequence of (peri)centromeric DNA repeats is not conserved among species, the transcripts originating from them have been described in a broad range of organisms, including plants (Topp et al., 2004; Zhang et al., 2005), parasites (F. Li et al., 2008), yeast (Volpe et al., 2002), insects (Grewal and Elgin, 2007; Pezer and Ugarković, 2008), amphibians (Diaz et al., 1981; Varley et al., 1980), and mammals (Bouzinba-Segard et al., 2006; Carone et al., 2009; Ferri et al., 2009; Martens et al., 2005; Nakano et al., 2003; Rudert et al., 1995).

The importance of pericentromeric transcripts in the assembly is best described in the fission yeast *Saccharomyces pombe*. The current model states that the mechanisms of pericentromeric repeat transcription and RNAi-mediated heterochromatin formation at pericentromeric regions are under cell cycle control (Chen et al., 2008). Pericentromeric repeats are transcribed by RNA Pol II at the S phase of the cell cycle (Kato et al., 2005; Kloc et al., 2008). dsRNAs are generated either via bidirectional transcription of the repeated sequences of the centromeres, the RNA molecule folding back on itself, or by the activity of RNA-dependent RNA polymerases on centromere-derived RNA templates. The double stranded precursor RNA is recognized and processed by DICER into small RNAs, approx. 25 nt long, which are incorporated into the RNA-Induced Transcriptional Silencing (RITS) complex *via* AGO1. The dsRNA is then sliced into siRNA and a passenger strand that is degraded. RITS complex may serve as a docking platform for the recruitment of chromatin modifying factors, such as histone H3K9 methyltransferase and histone deacetylases (Chen et al., 2008). H3K9me3 favours the binding of SWI6 (homolog of mammalian HP1) and, in turn, the establishment of H3K9me3 in a self-reinforcing loop (Chen et al., 2008). Inversely, the deletion of any component of RNAi system, such as DICER, RNA-dependent RNA polymerase or components of RITS, leads to loss of pericentromeric heterochromatin, with no accumulation of methylated histone H3K9 or binding of SWI6. In addition, long ncRNAs homologous to the centromeric repeats, were found to accumulate in all these mutant cells, but not

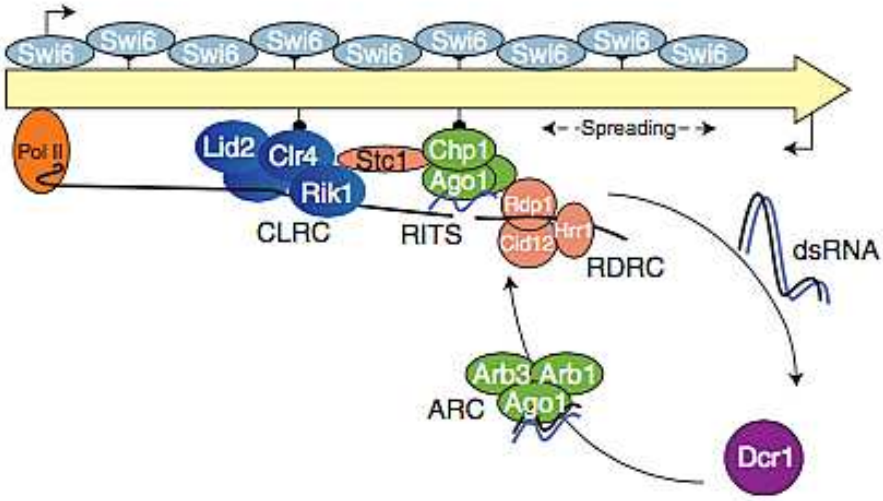


Figure 17. Model for heterochromatin assembly and spreading at pericentromeric repeats in *S. pombe*. Pericentromeric regions (yellow arrow) are transcribed by the RNA Pol II from bidirectional promoters (black arrows). dsRNAs are then processed by DICER (Dcr1) to generate siRNA that are incorporated in the RITS complex via AGO1. Association of RITS with centromere heterochromatin is strengthened by binding of Chp1 to H3mK9. RITS activity can recruit both CLRC, via interactions with Stc1, and RDRC resulting in spreading of H3mK9 and amplification of siRNAs, respectively. *From Volpe and Martienssen 2011.*

in wild-type cells, whereas siRNAs with centromeric sequence were described in normal cells (Hall et al., 2002; Volpe et al., 2002) (**Figure 17**).

The requirement for RNAi in the assembly of pericentromeric heterochromatin has been well established in fission yeast, and thus, it was hypothesized that RNAi would play a pivotal role also in mammals.

Evidence for RNAi-directed centromere regulation was demonstrated in chicken-human hybrid DT40 cells containing human chromosomes, in which DICER was inactivated by homologous recombination. These DICER-deficient cells have defects in sister-chromatid cohesion, and de-repress pericentromeric heterochromatin, as witnessed by the accumulation of transcripts from the human Satellite repeats sequence. Moreover, HP1 proteins were found to be partially delocalized in the nucleus, suggesting a disruption of heterochromatin by the RNAi machinery (Fukagawa et al., 2004).

During mouse development, RNAs derived from major satellite repeats are ubiquitously distributed in various tissues, with high level expression in cells of the developing central nervous system, whereas in adult tissues major satellite repeat transcripts are detected in liver and testis (Rudert et al., 1995). In addition, in proliferating mouse cells, a heterogeneous population of major satellite RNAs, ranging from 1 kb to more than 8 kb, accumulates during the course of G1 phase, reaching a peak in late G1/early S phase, whereas smaller transcripts of ~200 nt are detected only in mitosis (Lu and Gilbert, 2007). Centromeric RNAs (minor satellite repeats) of 2 kb and 4 kb in size are present in exponentially growing cells but accumulate as ~120 nt molecules during differentiation of two model systems, erythroid and myogenic cells (Ferri et al., 2009). In these conditions, the levels of minor satellite repeats RNA greatly vary during cell cycle progression, peaking in G2/M phase. These non-coding RNAs are low detectable in G1 and S phase. Like in human cells, pericentromere-derived transcripts were up-regulated in DICER-null mouse ES cells. These cells have also reduced levels of H3K9me3 and diffuse staining of HP1 proteins (Cobb et al., 2005; Kanellopoulou et al., 2005). These results resemble to the mislocalization of H3K9me3 and HP1 protein which were observed upon aberrant accumulation of transcripts throughout the cell cycle (Bouzinba-Segard et al., 2006).

Transcription and/or remodelling of the nucleosomes at the centromeres may also be important for the deposition of CENP-A at the centromere (Nakano et al., 2003; Topp et al., 2004).

In the fission yeast *S. pombe*, centromeric transcription may be induced by the GATA-type transcription factor AMS2, which is able on one hand to bind to the GATA-core sequence present in the central centromere regions, and, on the other hand, is crucial for the loading of CENP-A at centromeres (Chen et al., 2003). However, a role of AMS2 in the regulation of centromeric transcription has not been clarified yet.

In the mouse, minor satellite repeat RNAs associate with CENP-A-associated chromatin domains, and stabilize centromeric-associated complexes (Ferri et al., 2009). Inversely, aberrant accumulation of transcripts throughout the cell cycle cause delocalization of CENP-A away from centromeres (Hédouin et al., 2017).

In human cells, a long non-coding RNA seems to be implicated in centromere assembly. Quenet and Dalal showed that RNA Pol II is responsible of the transcription of a 1.3 kb transcript that originates from centromeric Satα. The transcription occurs from late mitosis to early G1, concurrent with the deposition timing of CENP-A. Inhibition of RNA Polymerase II-dependent transcription abrogates the recruitment of CENP-A and its chaperone HJURP to native human centromeres, suggesting that the action of transcription is essential for CENP-A targeting at centromeres. Moreover, the authors show that the long non-coding RNA interacts directly with the complex CENP-A/HJURP and its down-regulation leads to gradual loss of CENP-A at centromeres (Quénet and Dalal, 2014).

As a whole, these findings suggest a crucial role of (peri)centromeric transcripts in the maintenance of pericentromeric heterochromatin and in centromere identity and function.

3.3.3. Telomeric repeat-containing RNA (TERRA) and telomeric heterochromatin

Like centromeric and pericentromeric repeats, eukaryotic telomeric repeats are also transcribed by RNA Pol II, leading to the production of telomeric repeat-containing RNA (TERRA), a long non-coding RNA which is an integral component of telomeric heterochromatin (Azzalin et al., 2007; Schoeftner and Blasco, 2008, 2009). TERRA transcription is cell-cycle regulated, showing the highest levels in G1 and the lowest in late S/G2 (Porro et al., 2010). TERRA was found to interact directly with the double-stranded DNA-binding proteins TRF1 and TRF2. These two proteins are components of the SHELTERIN complex, the guardian of the length and integrity of telomeres. All these observations suggest a role of this long non-coding RNA in telomeres structure and integrity (Deng et al., 2009; Maestroni et al., 2017).

The importance of heterochromatin for telomere homeostasis and stability was first suggested before the discovery of TERRA in 2007 (Azzalin et al., 2007), by Garcia-Cao and colleagues. They showed that deletion of SUV39H1/2 caused a loss of heterochromatin features at telomeres, in particular a decrease in H3K9me3 levels at telomeres, associated with loss of chromo-box proteins (HP1 proteins). The loss of these heterochromatin features was related to abnormal elongation of telomeres (García-Cao et al., 2004). Later, Deng and colleagues showed that TERRA interacted directly with H3K9me3. Moreover, depletion of TERRA resulted in decreased H3K9me3 levels at telomeres (Deng et al., 2009). The importance of heterochromatin at telomeres was then reinforced by several studies which underlined an interaction with proteins

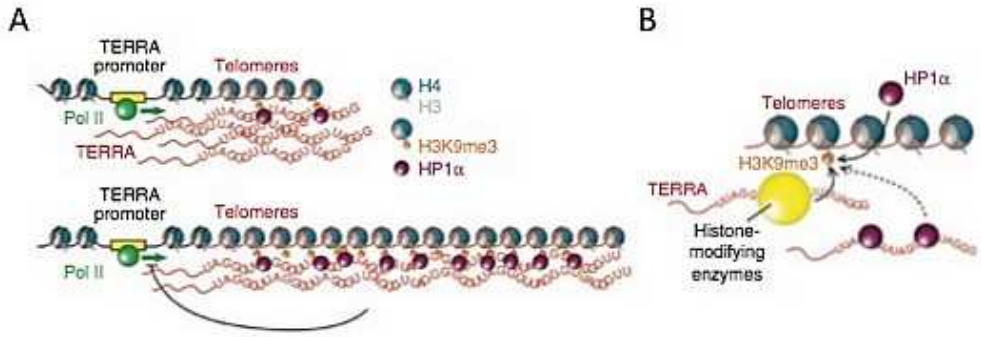


Figure 18. Non-coding RNA TERRA and telomeric heterochromatin. A) TERRA is highly transcribed in case of short telomeres. TERRA interact directly with H3K9me3 and recruit HP1α. The increase in the length of telomeres is associated with an increase in H3K9me3 and HP1α density, which can repress TERRA transcription acting on its promoter. B) The model proposes that TERRA recruits directly histone-modifying enzymes (SUV39h1/2) at telomeres to establish H3K9me3. TERRA recruits HP1α proteins at telomeres, forming a block of telomeric heterochromatin. *From Arnoult et al., 2012.*

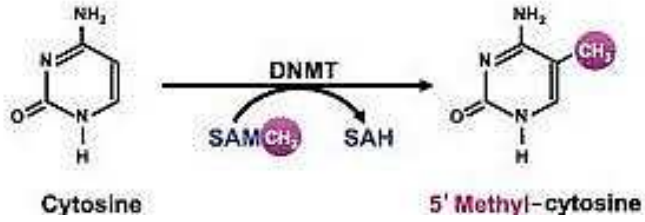


Figure 19. DNA methylation occurring on a cytosine residue. Cytosines can be covalently modified by attachment of a methyl group at the carbon 5 of the pyrimidinie ring. This reaction is performed by DNA methyltransferase enzymes (DNMTs) using S-adenosyl methionine (SAM) as a donor of methyl group. *From Zakhari 2013.*

involved in the heterochromatin machinery, HP1 and SUV39H1 (Deng et al., 2009; López de Silanes et al., 2010; Porro et al., 2014; Scheibe et al., 2013). Arnoult and colleagues showed that TERRA can directly recruit SUV39H1 and HP1 proteins at telomeres, promoting the formation of telomeric heterochromatin and the following down-regulation of TERRA (Arnoult et al., 2012). The model proposed by the authors is shown in **Figure 18**.

4. DNA methylation

DNA methylation is a covalent modification occurring mainly on cytosines, through the addition of a methyl-group at the carbon 5 of the pyrimidine ring by enzymes called DNA methyltransferases (DNMT) that use S-Adenosyl Methionine (SAM) as a methyl-donor (Suzuki and Bird, 2008) (**Figure 19**).

The presence of a 5-methyl-cytosine was first observed in 1925 in the *Tubercle bacillus* and described 25 years later in animals and plants DNA (Johnson and Coghill, 1925; Wyatt, 1950). In 1975, Holliday and Pugh, at the same time as Riggs, proposed that the programmed methylation and demethylation of DNA might be the molecular mechanism behind Waddington's hypothesis of “epi-genetics” to explain how dynamic interactions between the developmental environment and the genome lead to the production of phenotype (Holliday and Pugh, 1975; Riggs, 1975).

I will go over the tremendous progress made in our understanding of the dynamic non-random repartition of DNA methylation on the mammalian genome and how key biological processes like X-chromosome-inactivation, genome imprinting and lineage-specific gene silencing all rely on DNA methylation, but I will start by highlighting the technical improvement that greatly favored analysis of DNA methylation at local to whole genome scale.

4.1. Techniques to study DNA methylation

The techniques used to discover this modification were mainly based on crystallography and chromatography (Johnson and Coghill, 1925; Wyatt, 1950). Today a wide variety of techniques have been developed to allow its study at the level of genes and genome.

The first approaches were based on the use of methylation sensitive restriction enzymes, and were proposed by Bird and Southern in 1978 (Bird and Southern, 1978). They based this tool on the differential sensitivity of some restriction enzymes to cytosine methylation in their recognition sites (Bibikova and Fan, 2009). In particular, the existence of isoschizomers, restriction enzymes that recognize the same sequence but are sensitive or resistant to DNA methylation, is a powerful

tool. The most widely used couple of isoschizomers are HpaII, sensitive to DNA methylation, and MspI, which is not.

Several techniques are based on the conversion of genomic DNA by sodium bisulfite, which mediates the deamination of cytosine into uracil, and these converted residues will be converted to thymine. In contrast, methylated cytosines are resistant to this conversion and, so, will remain as cytosine. Thus, bisulfite conversion allowed to easily distinguish methylated DNA from unmethylated DNA.

4.1.1. Restriction-based

Several and widely-used approaches use this technique, like MSRE (Methylation-Sensitive Restriction Enzymes) analysis and COBRA (COmbined Bisulfite Restriction Analysis) that are coupled to PCR reactions. In MSRE, DNA is digested with a DNA methylation-sensitive restriction enzyme. If the DNA is methylated, methylation-sensitive enzymes cannot cut and PCR-amplification across the restriction site is permitted. In contrast, if genomic DNA is not methylated, these enzymes will cut the sequence leading to a decreased PCR amplification (Melnikov et al., 2005). For COBRA, the principle is similar but genomic DNA is first converted by sodium bisulfite and the restriction enzyme is not necessarily methylation-sensitive. The appearance of digestion fragments indicates DNA methylation, whereas the persistence of an undigested product reveals partial or no methylation. The products can also be quantified, for example by using densitometry, as described in (Xiong and Laird, 1997).

Alternatively, digested genomic DNA is analysed by Southern blot that will show a strong and unique signal if the DNA is methylated, whereas DNA hypomethylation will lead to the formation of multiple fragments. This technique is particularly powerful to study tandem repeats, which gives a ladder-like profile when they are unmethylated. The use of enzyme-based approaches is mostly limited to regions which contain a consensus sequence recognized by methylation-sensitive enzymes and requires to know the restriction map of the region of interest.

4.1.2. Fluorescence-based methods

MethyLight is a sensitive, real-time PCR technique that allows quantifying DNA methylation at a particular locus by using DNA oligonucleotides that differentially anneal to bisulfite-converted DNA depending on the methylation status of the original genomic DNA. The use of three oligonucleotides (forward and reverse primers, and a nested fluorescent probe) in MethyLight, any one or more of which can be used for methylation discrimination, allows for a high degree of specificity, sensitivity, and flexibility in methylation detection (Eads et al., 2000).

The limits of this technique is that it is a candidate approach that requires to design primers for the regions of interest.

4.1.3. Affinity-based methods

In this category, two different approaches are possible. Methylated DNA Immuno-Precipitation (MeDIP) uses antibodies against 5-methyl-cytosine (5mC) to precipitate methylated DNA after fragmentation (Mohn et al., 2009; Weber et al., 2005). Methylated CGI Recovery Assay (MIRA) takes advantage of the high affinity of the Methyl-CpG-Binding Domain (MBD) protein MBD2 for methylated CpG dinucleotides, which is strongly enhanced by MDB3L1. Fragmented genomic DNA is incubated with the MBD2B/MBD3L1 high affinity protein complex. Unlike the MeDIP technique, which requires single-stranded DNA for antibody recognition, MIRA works on normal double-stranded DNA. The CpG-methylated DNA is easily captured from the binding reaction via the GST-tagged MBD2B and glutathione beads (Rauch and Pfeifer, 2009). Both techniques can be coupled to PCR reactions or to next generation sequencing (NGS). The limitation of these approaches is the low efficiency in precipitating regions with low methylation density.

4.1.4. DNA methylation arrays

Bisulfite-converted genomic DNA is hybridized to arrays that contain predesigned probes or BAC libraries, to distinguish methylated from unmethylated cytosines converted to uracil. A single-base PCR extension step incorporates a fluorescently labelled nucleotide, a guanine or an adenine complementary to “methylated” cytosine or unmethylated” thymine respectively. Scanning of the array detects the ratio of fluorescent signal from the unmethylated probe compared to that of the methylated probe, allowing the levels of methylation to be determined (Stirzaker et al., 2014). This category of assays groups for example all the evolutions of BeadChips by Illumina, from the Human BeadChip HM27K and HM450K to the most recent EpicBeadChip 850K (the numbers correspond to the number of cytosines that are analysed). On the HM450K, for example, 99% of RefSeq genes and 96% of CGI are covered. However, genomic regions outside these categories are poorly represented (Stirzaker et al., 2014).

4.1.5. Sequencing-based methods

This is the type of technique that saw huge improvements in the last years. Three different techniques can be grouped in this category: Bisulfite-Sequencing (BS) is based on the clonal sequencing of bisulfite-converted PCR products from a single genomic region, which is widely

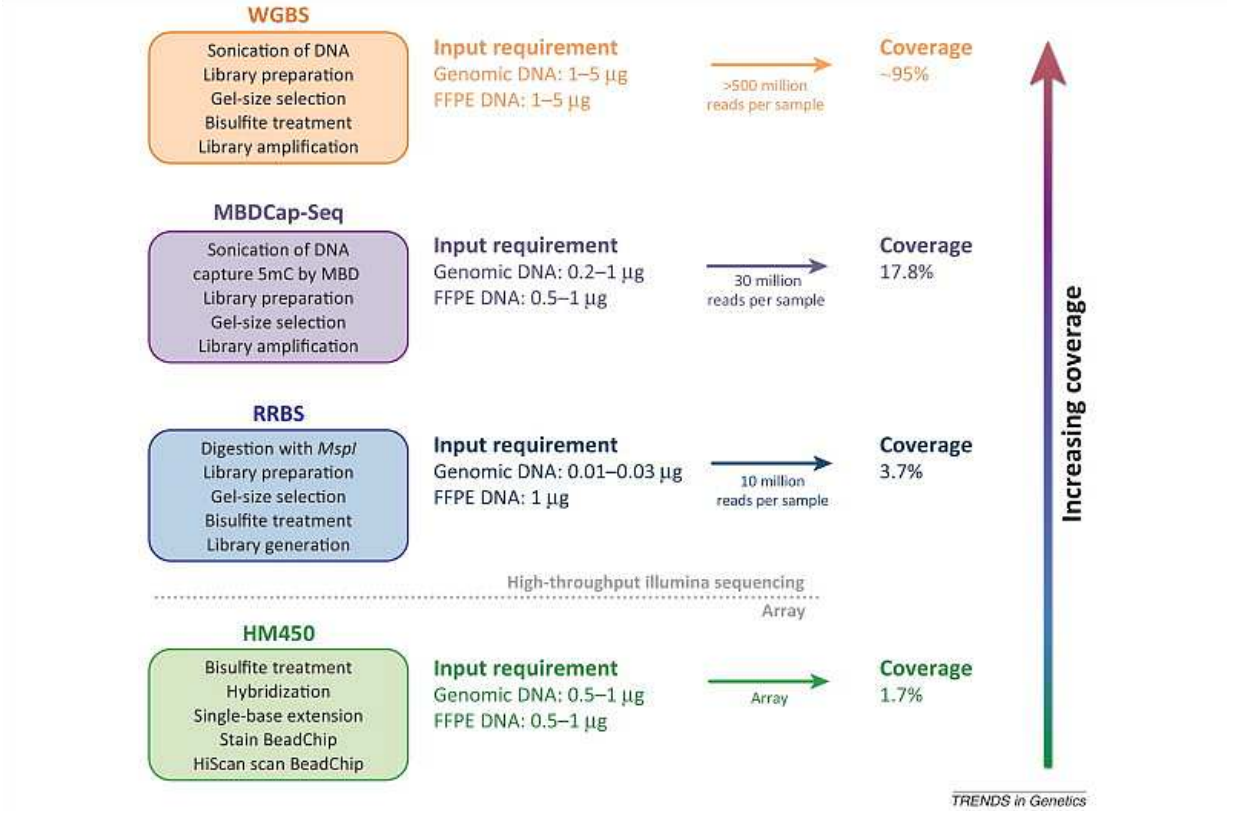


Figure 20. Summary of techniques for genome-wide analysis. The summary compares the maximum coverage of Whole Genome Bisulfite Sequencing (WGBS) with Methyl-Binding Domain Capture coupled with sequencing (MBDCap-Seq), Reduced Resolution Bisulfite Sequencing (RRBS) and HumanMethylaition 450K BeadChip (HM450). The workflow is summarized underneath each technique. The amount of genomic DNA or the formalin-fixed paraffin-embedded tissue (FFPE) needed for each technique is indicated. The minimum number of reads per sample is indicated under the arrows, except for the HM450 that the number is fixed to 482,422 cytosines. *From Stirzaker et al., 2014.*

used to discriminate allele-specific methylation. Whole Genome Bisulfite Sequencing (WGBS) is performed on genomic DNA ligated to methylated adaptors before bisulfite conversion, followed by library construction and high-throughput sequencing. Reduced Resolution Bisulfite Sequencing (RRBS) starts with digestion using a methylation-insensitive restriction enzyme (like MspI). After end-repair, A-tailing and ligation to methylated Illumina adapters, the CpG-rich DNA fragments (generally 40–220 bp) are size selected, subjected to bisulfite conversion, PCR amplified and end sequenced. The use of RRBS data is restricted to regions with moderate to high CpG density, and hence, enriched for promoter-associated CGI. Although RRBS interrogates only less than 4% of the approximately 28 million CpG dinucleotides distributed throughout the human genome, the lack of coverage at intergenic and distal regulatory elements is a potential drawback of the method (Stirzaker et al., 2014).

A summary of the genome-wide techniques and the percentage of the human genome covered are shown in **Figure 20**.

4.2. CpG context of DNA methylation

In mammals, DNA methylation occurs mainly on cytosines in the context of CpG dinucleotides (Doskocil and Sorm 1962). CpG dinucleotides are globally depleted in mammalian genomes, which is thought to reflect a high mutation rate of methylated cytosines that can spontaneously deaminate into thymine (Coulondre et al., 1978), probably by the action of the Base-Excision Repair (BER) pathway. Yet, the genome is characterized by the presence of stretches of DNA with a high CpG density, called CpG islands (CGI), which, by definition, cover at least 200 bases pairs over less than 1 kb, a CG content of greater than 50% and a CpG frequency (observed/expected) of 0.6 (Illingworth and Bird, 2009). The human genome contains around 25,000 CGI outside of DNA repeats (Lander et al., 2001). Other genomic annotations based on density in CpG dinucleotides have been defined: CpG “shores” as the 2kb regions near CGIs, CpG “shelves” as the additional flanking 2 kb from CGIs, and “open sea” as isolated CpGs (Irizarry et al., 2009; Sandoval et al., 2011).

Several studies showed that only a small fraction of CGIs seems to be methylated, even in terminally differentiated cells (Rollins et al., 2006; Weber et al., 2007). In contrast, regions with lower density of CpGs show high levels of DNA methylation (Weber et al., 2007).

In certain cellular contexts, DNA methylation has been found on cytosines in a non-CpG context. Indeed, in human embryonic stem cells (hESCs) around 25% of the DNA methylation occurs in CpHpG (around 17.5%) and CpHpH (7.5%), where H corresponds to Adenine, Thymine or Cytosine. In contrast, in human fibroblasts, the levels of non-CpG methylation were evaluated to represent less than 1% of the global DNA methylation. This suggests that non-CpG methylation

decreases upon differentiation (Lister et al., 2009), although it has been described in somatic tissues, particularly in the brain where CpH methylation increases during synaptogenesis and can reach up to 53% of total DNA methylation in neurons (Lister et al., 2013).

4.3. Genomic context of CpG methylation

As mentioned above, DNA methylation is not randomly distributed on the genome (Varriale and Bernardi, 2010), with the bulk of the genome being methylated with the exception of CGI (Weber and Schübeler, 2007).

4.3.1. Methylation at promoters

Around 60% of gene promoters are characterized by the presence of a CGI and, inversely, 50% of CGI map to gene promoters (Illingworth et al., 2010). The great majority of these promoters are unmethylated at all stages of development and in all types of tissues (Antequera and Bird, 1993).

A more detailed description of promoters methylation was performed by Weber and colleagues in 2007. They analysed 16,000 human promoters classified based on their CpG content, CpG density and the length of the CpG-rich region. High CpG promoters (HCPs) and low CpG promoters (LCPs) formed two distinct groups, corresponding to CGI promoters and clearly non-CGI promoters respectively. Promoters with an intermediate level of CpGs were called intermediate CpG promoters (ICPs). DNA methylation levels at these promoters was tested in different somatic cells (primary fibroblast, primary lung cells, primary kidney and colon tissues) and in sperm samples using MeDIP followed by hybridization on microarrays analysing 24,134 putative human promoters. In somatic cells only 3% on HCPs were methylated, whereas 21% of ICPs and the majority of LCPs were methylated. In sperm cells, similar DNA methylation patterns were found at LCPs, whereas most of the methylated HCPs (86%) and half of the ICPs (49%) were unmethylated. Promoters of genes associated with germ cells processes, namely germline genes, were overrepresented within the differentially methylated HCPs between somatic cells and sperm, suggesting a drastic change in the DNA methylation levels of a particular set of CGI promoters from germline to somatic cells (Weber et al., 2007), which is consistent with DNA methylation being the primary mechanism to maintain these genes silent in somatic cells (De Smet et al., 1999; Maatouk et al., 2006; Rodić et al., 2005; Velasco et al., 2010).

4.3.2. Methylation at gene bodies

The interest in DNA methylation of gene bodies came from the observation that half of mammalian CGI lie outside annotated promoters, with 25% of them being found in gene bodies as shown by hybridization of precipitated methylated DNA from human primary cells to a microarray containing a BAC library, representing large parts of chromosomes 3, 7, 12 and 14 showed that DNA methylation occurs in genic regions (Weber et al., 2005). Bisulphite sequencing of 873 genes, on human chromosomes 6, 20 and 22, showed that sequences outside gene promoters, like gene bodies, were highly methylated (Eckhardt et al., 2006), although exons were more methylated than introns. This pattern is likely to be influenced by the higher CG content of exons compared to their surrounding introns, probably caused by the lower selective pressure on introns sequences (Laurent et al., 2010).

4.3.3. Methylation at DNA repeats

As mentioned in the first chapter, 50% of the human and mouse genomes is composed of repetitive elements (Lander et al., 2001). Most of the studies on the DNA methylation status of interspersed and tandem repeats were performed in mice and showed that repetitive elements are highly methylated in somatic tissues (Chapman et al., 1984). From the work of Okano and colleagues, we know that centromeric and pericentromeric repeats are already fully methylated in mouse ES cells, cultivated in Serum + LIF (the different culture conditions will be explained later in this chapter) despite that, at this stage, DNA methylation levels are fairly low (Okano et al., 1999). An in-depth analysis of the methylation status of interspersed DNA repeats at the pre-implantation stages of mouse development, using RRBS, showed that only IAP were methylated, with 60% of CpGs retaining DNA methylation. In contrast, at post-implantation stages, all the families of repetitive elements showed an increase in DNA methylation, reaching 80% of methylated CpGs at 6.5 dpc (Auclair et al., 2014). Unfortunately, minor and major satellite repeats were not analysed. The dynamic of DNA methylation during development will be described in details in the following sections.

DNA methylation of repetitive elements have been analysed in human cells in 2005 by MethylLight technique (Eads et al., 2000). Weisenberger and colleagues analysed the methylation status of Satellite II repeats (SatII) and alpha-satellite repeats (Satα), LINE1 and Alu sequences in peripheral blood cells, and reported that the levels of DNA methylation of all the tested repetitive elements were higher compared to cancer cell lines and sperm samples. These findings are in agreement with what was observed in mice, whereby DNA repeats are globally methylated in somatic cells (Weisenberger, 2005).

4.3.4. Methylation at monoallelically expressed regions: the case of genomic imprinting and X chromosome inactivation

In 1984, Surani and Solter performed experiments of nuclear transplantation in mice to generate androgenotes (embryos with two paternal genomes) or gynogenotes (embryos with two maternal genomes). The zygotes were formed, but neither type was able to develop and generate a viable embryo (McGrath and Solter, 1984; Surani et al., 1984). This was the first evidence that maternal and paternal genomes carry different and complementary information, and that genomic imprinting plays a major role in growth and development (Monk et al., 1987). Imprinted genes are defined as genes that are expressed in a non-random mono-allelic manner, meaning that these genes are expressed exclusively either from the paternal or the maternal allele, suggesting that a mark, called imprint, allows to distinguish between maternal and paternal alleles. The idea that DNA methylation could represent the heritable molecular differences between maternally- and paternally-derived alleles came from studies using transgenes in mice (Reik et al., 1987; Sapienza et al., 1987) and confirmed in DNMT1 knockout mice where imprinting of the H19-IGF2 locus was lost (Li et al., 1993).

In mammals, a small number of genes (97 confirmed in humans and 151 in mice) carries an imprint that allows the differential expression from the paternal or the maternal allele (updated list on the following websites http://www.geneimprint.com/site/genes-by-species and http://www.mousebook.org/mousebook-catalogs/imprinting-resource?catalog=imprinting). Most of these genes are organised in clusters in the mammalian genomes (W. Reik, Dean, and Walter 2001; Ferguson-Smith and Surani 2001), and most of the imprinted clusters host non-coding RNA clusters, like the six well-studied clusters IGF2R, KCNQ1, Prader-willi syndrome – Angelman syndrome, IGF2 – H19 and DLK1 – DIO3 loci (Koerner et al., 2009). Several studies underlined that each imprinted cluster contains a region called Imprinted Control Region (ICR) which corresponds, in most of the cases, to a Differentially Methylated Region (DMR). The DMRs are differentially methylated on the two parent alleles, leading to allelic distinction. The mechanisms of establishment of DNA methylation on the bulk of the genome will be described in the next chapter, but it is important to mention here that sex-specific DNA methylation of ICRs is established during gametogenesis and that their methylation persists during development in most of the cases (Weaver and Bartolomei, 2014). The KRAB-Zinc finger protein ZFP57 was shown to be crucial for the establishment of imprinting at one locus (SNRPN locus), but is required mainly for the maintenance of all maternal imprints (Hirasawa and Feil, 2008; X. Li et al., 2008).

DNA methylation is also important for X chromosome inactivation in female cells, which represents a remarkable case of imprinting since it affects a whole chromosome. The first steps of this process have been described in the section 3.3 of this chapter, dedicated to the function of the

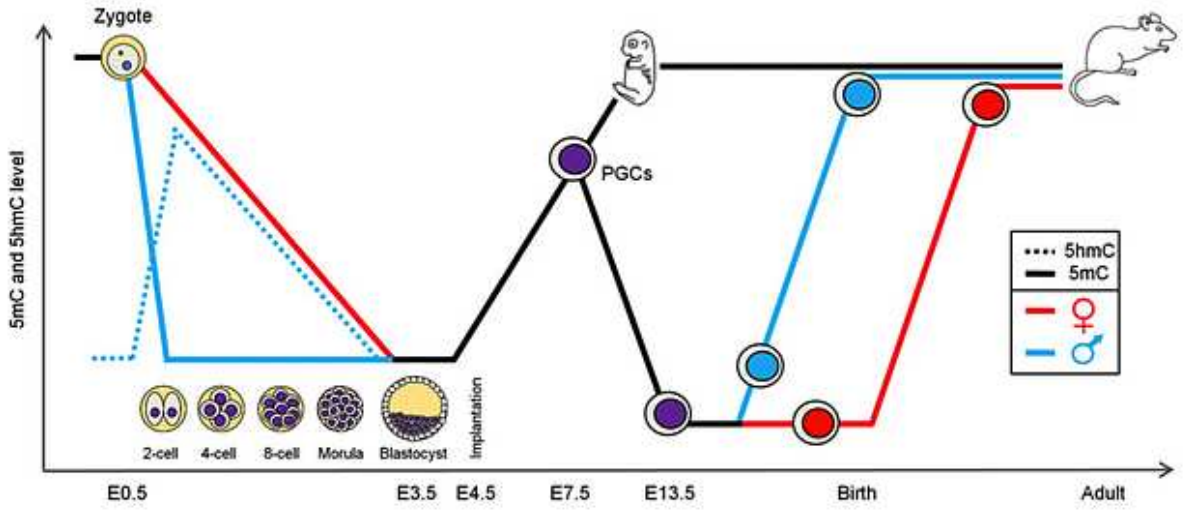


Figure 21. Waves of DNA methylation reprograming during development. After fertilization, levels of DNA methylation of the maternal and paternal genomes globally decrease with different kinetics, until the blastocyst stage. The paternal genome (blue) is demethylated by an active process, concomitant with the apparition of 5-hydroxymethylation (blue dotted line). Conversely, the maternal genome (red) is thought to be lost by a passive, replication-dependent process. Note that many sequences escape this DNA methylation resetting, like IAP retroelements. New patterns of DNA methylation are established following implantation by the de novo methylation machinery. After implantation, epiblast cells generate all the cells of the embryo, including the primordial germ cells (PGCs). These cells undergo a second wave of DNA demethylation that is completed by day 13.5 after fertilization, after which PGCs enter in meiosis in females and mitotic arrest in males. *From Auclair and Weber 2012.*

long non-coding RNA XIST. In contrast to XIST, which is dispensable for maintenance of X inactivation (Csankovszki et al., 2001; Hernández-Muñoz et al., 2005), DNA methylation is a rather "late" event in the cascade and contributes to the maintenance of the silent state of the inactive X chromosome (Csankovszki et al., 2001; Heard, 2004).

4.4. Dynamics of DNA methylation and its players

4.4.1. Reprogramming of DNA methylation during development

DNA methylation is a key epigenetic modification, highly dynamic during development (**Figure 21**). After the formation of the zygote, the genome undergoes a genome-wide demethylation. Demethylation of paternal and maternal genomes proceeds following different kinetics. The maternal genome is demethylated *via* a replication-dependent process, meaning that DNA methylation is not maintained upon DNA replication, in what is called a "passive" demethylation. In contrast, the paternal genome is actively demethylated more rapidly than the maternal genome (Mayer et al., 2000; Oswald et al., 2000). When the embryo reaches the blastocyst stage, DNA presents the lowest levels of methylation (Auclair et al., 2014; Borgel et al., 2010; Smallwood et al., 2011; Smith et al., 2012). Like ICRs, the IAP family of DNA repeats does not undergo complete demethylation, keeping around 60% of DNA methylation at the blastocyst stage (Auclair et al., 2014). It has been proposed that these regions are "protected" from DNA demethylation (Sanz et al., 2010) and a possible mechanism for this protection will be described in the section 4.5.4.

It was thought for several years that the genome-wide methylation erasure was important to reset the epigenome after gametogenesis. However, the fact that the global methylation levels in ES cells and sperm is not so different, except at ICRs and IAP elements, seems to discredit this hypothesis (Farthing et al., 2008). However, this reprograming might be necessary for the reactivation of the pluripotency program in the early embryo. Indeed, Farthing and colleagues showed that promoters of the key pluripotency factors, like NANOG, were methylated in the sperm and un-methylated in ES cells (Farthing et al., 2008).

4.4.2. DNA methylation activities: *de novo* and maintenance

From the blastocyst stage, when the parental profiles of DNA methylation have been erased, new patterns of DNA methylation have to be established in a genome-wide and tissue-specific manner and maintained throughout cell divisions, suggesting distinct activities.

All mammals possess four catalytically active DNA methyltransferases (DNMTs) (**Figure 22A**). DNMT1 was the first DNMT to be identified and characterized (Bestor et al., 1988). It was

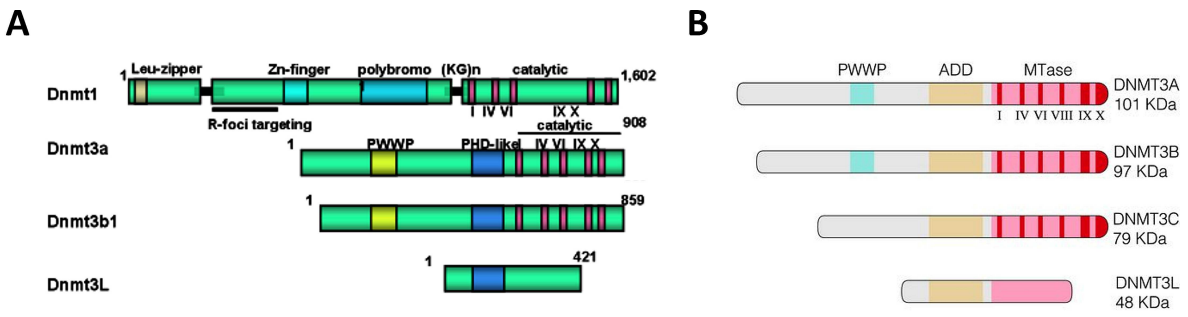


Figure 22. Schematic overview of DNA methyltransferases. The catalytic domains of the enzymes are well conserved, but the similarity is very low at the N-terminal regulatory elements. A) DNMT3A and DNMT3B as shown as well as the catalytically inactive DNMT3L which function as cofactor for the *de novo* methyltransferases are shown. *Adapted from Shoji Tajima*. B) The structure of the newly identified DNMT3C compared to the DNMT3 enzymes and DNMT3L. Up to know this new enzyme seems to be present only in rodents. *From Barau et al., 2016.*

then showed to be essential for life in mice (Li et al., 1992). Around ten years later, DNMT3A and DNMT3B were then identified (Okano et al., 1998). Deletion of DNMT3B cause embryonic lethality around 13.5 dpc, while deletion of DNMT3A causes the formation of smaller mice which die after four weeks of life (Okano et al., 1999). All these enzymes show enzymatic activity on DNA, in contrast with DNMT2 and DNMT3L (Aapola et al., 2000; Yoder and Bestor, 1998). DNMT2 has a catalytic activity but on RNA (Goll and Bestor, 2005), whereas DNMT3L does not show any enzymatic activity, but contributes to the activity of DNMT3A and DNMT3B. Mice with deletion of DNMT3L are viable and show a normal phenotype, but they are sterile (Bourc'his et al., 2001). The recently identified DNMT3C adds a new player in the establishment of DNA methylation at the promoter of young retrotransposons. The gene coding for this enzyme was previously annotated as a pseudogene of DNMT3B. DNMT3C showed characteristics of DNMT3 enzymes: a catalytic domain in the C-terminal part of the protein and an ADD domain, although it is missing the PWWP domain (**Figure 22B**). This enzyme is specific of male germ cells in rodents. DNMT3C has two isoforms, a long and a short one which are differentially expressed during germ cells development. The long isoform, coding for the functional enzyme, is expressed before birth, during the prospermatogonia phase, while the short non-coding isoform is expressed in post-natal testes (Barau et al., 2016).

DNA methylation is established on unmethylated DNA through the synergistic action of DNMT3A and DNMT3B enzymes, together with their enzymatically inactive co-factor DNMT3L. The patterns of methylation that are established are then maintained by DNMT1 which is able to copy these patterns from a hemi-methylated template during DNA replication (Li and Zhang, 2014). In agreement with this model, DNMT3B is highly expressed when DNA establishment takes place *in vivo* immediately after implantation, while DNMT3A, although is expressed during DNA establishment (Auclair et al., 2014), reaches the highest levels at 10.5 dpc (Watanabe et al., 2002). Later in development, both enzymes are gradually down-regulated, whereas DNMT1 is highly expressed in dividing somatic cells. Although the first *in vitro* tests did not show that DNMT3 enzymes have a higher affinity for unmethylated DNA compared to hemi-methylated DNA (Okano et al., 1998), a following *in vitro* test showed that DNMT3A has a three-times higher affinity for unmethylated DNA, confirming its role as *de novo* methyltransferase (Yokochi and Robertson, 2002). In contrast, DNMT1 showed a higher affinity for hemi-methylated DNA (Pradhan et al., 1999). DNMT3C was shown to be essential to establish DNA methylation at IAP and LINE1 repeats, whose methylation is established before fertilization, during gametogenesis (Barau et al., 2016). This hypomethylation was also observed in DNMT3L knock-out mice (Bourc'his et al., 2001), suggesting that DNMT3L and DNMT3C may act together in DNA methylation establishment at IAP and LINE1 elements during gametogenesis.

Since DNA methylation is not randomly distributed on the genome, it raises the question of how DNMTs are recruited at specific genomic loci, or excluded from some others. The mechanisms that have been proposed include the interaction with chromatin marks and the recruitment through transcription factors binding to consensus sequences. As I will describe in the following sections, DNMT3A is mainly recruited through chromatin modifications. In contrast, an interaction between DNMT3B and the centromeric protein CENP-C was shown to be important for the recruitment of DNMT3B at centromeric and pericentromeric repeats in human cells (Gopalakrishnan et al., 2009). At the promoter of a subset of germline genes, E2F6 was shown to be required for DNMT3B recruitment (Velasco et al., 2010). For DNMT1, its interaction with the Proliferating Cell Nuclear Antigen (PCNA) and Ubiquitin-like with PHD and Ring Finger domains 1 (UHRF1), which is able to bind hemi-methylated DNA (Bostick et al., 2007; Chuang et al., 1997; Sharif et al., 2007), allows its localization at the replication fork and the inheritance of DNA methylation profiles to the daughter strand

It is important to mention that there is not evidence for a mutual exclusion between *de novo* and maintenance activities. Mouse ES cells depleted in both DNMT3A and DNMT3B show a progressive loss of DNA methylation, in particular at genomic regions with a high CpG density as well as at LINE1 promoters. These findings led Liang and colleagues to propose that DNMT3A and DNMT3B may cooperate with DNMT1 for the maintenance of DNA methylation at CpG-rich regions (Liang et al., 2002). This hypothesis was analysed in a follow-up study which described the DNA methylation levels in mouse ES cells single knock-out for DNMT1, DNMT3A, DNMT3B or both DNMT3A/3B. In the absence of DNMT1, DNA methylation levels were clearly reduced, even if a subset of DNA repeats maintained a hemi- or even a fully-methylated profile, notably major satellite repeats and IAP elements (Arand et al., 2012). The individual affinity of each DNMT for each DNA repeat was calculated and, in contrast to what was proposed by *in vitro* experiments, DNMT1 may work as a *de novo* DNMT *in vivo* at IAP and major satellite repeats. This activity was absent at LINE1 and SINE B1 elements, suggesting that a cooperation between DNMT1 and DNMT3A and/or DNMT3B is required at these regions (Arand et al., 2012).

The potential role of *de novo* DNMTs in the maintenance of DNA methylation was also proposed at the level of particular class of germline genes, where DNA methylation is established by DNMT3B (Borgel et al., 2010; Velasco et al., 2010). My team showed that DNMT3B remains at the promoter of these genes also in somatic cells, suggesting that DNMT3B could participate to maintain DNA methylation in somatic cells (Velasco et al., 2010). Even in absence of DNMT1, these promoters remain methylated, thus suggesting that DNMT3B is essential for DNA methylation maintenance at the promoter of this class of germline genes (Walton et al., 2011).

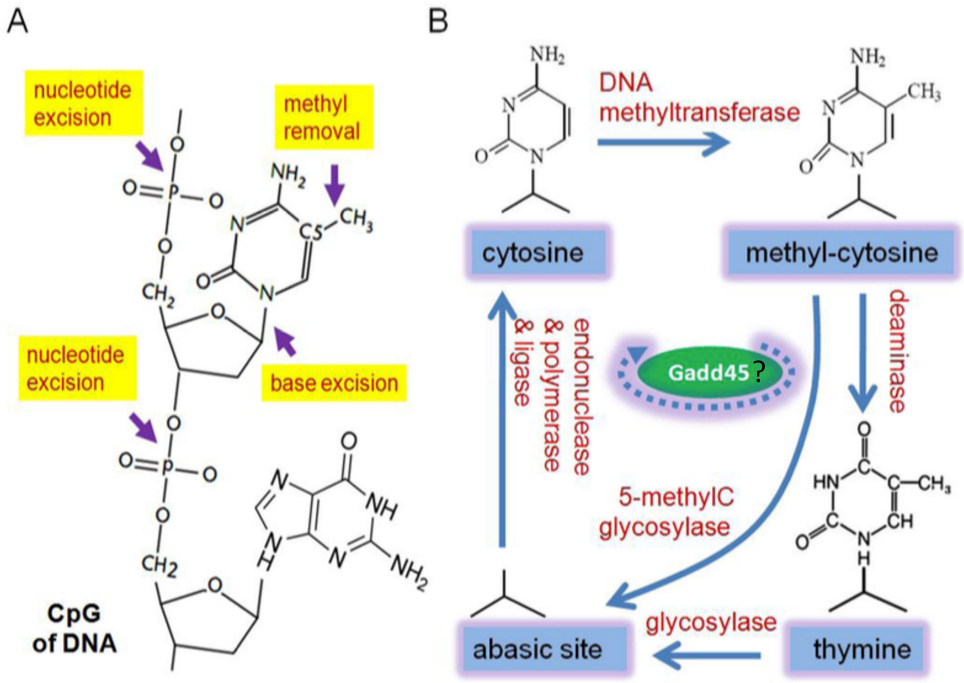


Figure 23. Mechanistic models of active DNA demethylation. A) active demethylation can be potentially carried out by direct removal of methy-group, nucleotide excision (NER) or base excision. (BER). Base and nucleotide excisions are followed by similar subsequent fill-in repair with unmethylated cytosine. B) 5-methylcytosine initially formed by DNA methyltransferases might be directly removed by DNA glycosylases or indirectly by thymine glycosylase after being converted by cytidine deaminases. Abasic sites are then filled-in through DNA polymerases and ligases. GADD45 proteins may promote the coupling of deamination, glycosylation and fill-in repair resulted from both base and nucleotide excisions. *Adapted from Ma et al., 2009.*

4.4.3. Mechanisms of DNA demethylation

Methylation of cytosines was initially thought to be irreversible since no DNA demethylase enzyme was identified. However, both passive and active DNA demethylation were described in pre-implantation embryos, first in the male pronucleus through an active mechanism independent of DNA replication and subsequently in both paternal and maternal chromosomes through a passive process (Sanz et al., 2010) (**Figure 21**).

The first mechanism proposed for DNA demethylation was based on the activity of an enzyme called Thymine DNA Glycosylases (TDGs), first isolated by Jost in 1993 (Jost, 1993). He proposed that these enzymes would provide a direct way for removing the whole 5mC base by breaking the β-N glycosidic bond to create an apurinic/apyrimidinic (AP) site and excision of AP mononucleotide by endonucleases, followed by fill-in of unmethylated nucleotides by DNA polymerase and ligases (Jost, 1993). TDGs are implicated in DNA repair mechanisms, hence DNA demethylation was thought to use a DNA repair-like process to remove the 5-methycytosine-containing base or a stretch of nucleotides rather than the methyl group directly (**Figure 23A**).

Barreto and colleagues then identified GADD45A as a non-enzymatic factor that promotes active DNA demethylation in cultured cells. GADD45 was shown to interact with the endonuclease XPG (Xeroderma Pigmentosum complementation group G), component of the Nucleotide Excision Repair (NER). Hence, it was proposed that 5mC were recognized and removed through GADD45-XPG complex, and the resulting gap in DNA was filled by DNA polymerase δ/ε and ligases with resulting unmethylated CpGs (Barreto et al., 2007). Notably, GADD45 and XPG have also been implicated in Base Excision Repair (BER) that has been shown to excise methylated cytosine suggesting another potential mechanism of DNA demethylation (Ma et al., 2009) (**Figure 23A**).

However, the role of GADD45 in DNA demethylation is controversial. While XPG is highly expressed in oocyte (Menezo et al., 2007), GADD45 was shown to be expressed in pre-implantation embryos, hence after the wave of DNA demethylation. Moreover, GADD45 did not promote DNA demethylation *in vitro* (Jin et al., 2008).

Alternatively, a cytidine deaminase was shown to convert the 5mC into a thymine, followed by glycosylase (TDG) removal and replacement of unmethylated nucleotides (Métivier et al., 2008) (**Figure 23B**). These findings reinforced the role of BER systems in DNA demethylation, since TDG and BER proteins were found to be concomitantly recruited to promote DNA demethylation (Métivier et al., 2008).

The mechanism of DNA demethylation needed still to be understood. For this reason, in the recent years, great efforts have been performed to identify the enzymatic process of DNA demethylation, which was supposed to involve the formation of 5-hydroxy-methyl-cytosine

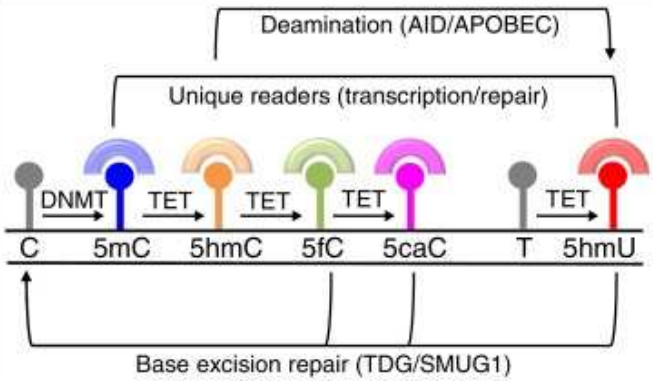


Figure 24. TET-mediated oxidation products in DNA demethylation processes. DNA methyltransferases (DNMT) can add a methyl-group on a cytosine (C) to form 5-methyl-cytosine (5mC). TET enzymes can then modifiy the 5mC, promoting the formation of5-hydroxy-methyl-cytosine (5hmC), 5-formyl-cytosine (5fC), 5-carboxyl-cytosine (5caC). TET proteins can induce the base excision repair system (BER) via DNA glycosylases (TDG/Smug1) that can target 5fC, 5caC or 5-hydroxy-methyl-uracil (5hmU), which is generated by deamination of 5hmC or by oxidation of tymidine (T). Each oxidation product can recruit specific readers (colored semi-circles) that may promote different biological functions. *Adapted from Cimmino and Aifantis 2017.*

(5hmC) observed during the demethylation process of the paternal genome ((Saitou et al., 2012) (**Figure 21**). A great advance came with the discovery of the Ten-Eleven Translocation enzymes (TETs) and their ability to oxidise 5mC to 5hmC (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). This new modification, which was already described in primary tissues since 1972 (Penn et al., 1972), represents an intermediate in the DNA demethylation pathway. The formation of 5hmC results from the activity of three TET enzymes in vertebrates, TET1, TET2 and TET3, but only the deletion of TET3 impairs its formation and the DNA demethylation process during early development (Gu et al., 2011). All TET enzymes catalyse several modifications starting from 5mC: a series of oxidation reactions may form 5hmC, 5-formyl-cytosine (5fC) and 5-carboxyl-cytosine (5caC) (Wu and Zhang, 2010) (**Figure 24**).

5hmC could be recognized by the AID/APOBEC family that can trigger the deamination of 5hmC, forming 5-hydroxy-methyl-uracil (5hmU). 5hmU, as well as 5fC and 5caC, can be recognized by the base excision repair (BER) system, *via* DNA glycosylases (TDG/SMUG1) (**Figure 24**). The action of the TDG/SMUG1 complex can restore an unmodified cytosine, providing an active mechanism of DNA demethylation (Cimmino and Aifantis, 2017).

4.4.4. 5hmC: just an intermediate of DNA demethylation?

A more exciting hypothesis is that 5hmC may represent a new epigenetic modification carrying an information by itself, rather than only an intermediate in the DNA demethylation process. Because of the abundance of this mark in physiological conditions, mainly in pre-implantation embryo and brain, 5hmC may have a role beyond the simple role of an intermediate of DNA demethylation (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). Moreover, 5hmC is not randomly distributed on the genome and its distribution depends on the cell type, developmental stage and genomic loci. Single-nucleotide sequencing showed that 5hmC is highly present in ES cells, in particular at enhancers and CGI, which are free of DNA methylation (Pastor et al., 2013). In neuronal cells, 5hmC is particularly enriched in gene bodies and it was proposed to be also the case in ES cells (Mellén et al., 2012; Pastor et al., 2013). TET enzymes may be important to keep CGI methylation-free in ES cells, while the function in neuronal cells is still unknown (Kriaucionis and Tahiliani, 2014).

Understanding the role of 5hmC requires to develop new techniques to discriminate between 5mC and 5hmC. Bisulfite sequencing is not able to distinguish between the two marks and does not correctly interpret 5fC and 5caC (Pastor et al., 2013). For this reason, it is important to note that bisulphite sequencing must be interpreted very carefully in particular in cellular contexts where 5hmC is highly present. Several techniques for genome-wide analysis were developed to study 5hmC and to distinguish this mark from 5mC (Pastor et al., 2013) The most famous example

is the Oxidative Bisulfite Sequencing (ox-BS Seq) in which 5hmC is oxidized to 5fC using potassium perruthenate. Bisulphite sequencing is performed on the same sample both before and after oxidation. As 5hmC is unmodified by sodium bisulphite treatment, like 5mC, and 5fC is deaminated, the extent of 5hmC at a site can be approximated by subtracting the frequency of deamination after perruthenate oxidation from the frequency of deamination before oxidation. 5hmC is a rare base, so very high sequencing coverage is required to determine the extent of hydroxymethylation (Booth et al., 2012).

4.5. The impact of DNA methylation in the different genomic contexts

4.5.1. On promoters and transcription start sites

Methylation of gene promoters has been associated with gene repression although the link might not be that straightforward. Methylated CGI promoters are usually restricted to genes that require the long-term stabilization of repressed states. Examples include imprinted genes, genes located on the inactive X chromosome and genes that are exclusively expressed in germ cells and whose expression would be inappropriate in somatic cells (Jones, 2012). It has been proposed that the methylation of promoter and the impact on gene expression depend on CpG density within the promoter. Indeed, *in vitro* studies showed that methylation of HCPs inhibits transcriptional initiation, which could not be reverted neither by adding an enhancer region nor by histone acetylation (Boyes and Bird, 1992; Hsieh, 1994).

More recently, Weber and colleagues showed that the methylation of HCPs or ICPs (*Cf* section 4.3.1 of this chapter) was associated with gene repression. They coupled DNA methylation analysis with profiling of RNA Pol II occupancy, and observed a negative correlation. Most of genes whose CGI promoter is methylated belong to the germline-specific set of genes. In contrast, methylation of non-CGI promoters was associated both with transcriptional activity as well as with silencing, suggesting another layer of regulation for low density of CpG promoters (Weber et al., 2007).

Germline genes showed hypomethylated promoters and associated expression in germ cells, whereas they are methylated and silent in somatic cells (Storre et al., 2005; Weber et al., 2007). These findings suggested the key role played by DNA methylation in silencing of germline genes in somatic cells, acting as genome-defence mechanism (Hackett et al., 2012). As I mentioned above, the silencing of a set of these gene promoters is promoted by DNMT3B through its recruitment by the transcriptional repressor E2F6, which establishes (Borgel et al., 2010; Velasco et al., 2010) and even maintains (Walton et al., 2011) the methylated status of these promoters in somatic cells.

4.5.2. At gene bodies

In contrast to promoters and transcription start sites, methylation of gene bodies positively correlates with transcription (Wolf et al., 1984). Genome-wide profiling of DNA methylation in different human cells (lymphocytes, fibroblasts and induced-pluripotent stem cells) confirmed this correlation (Ball et al., 2009). Within gene bodies, introns and exons are not equally methylated: exons are more highly methylated than introns, and a strong transition in DNA methylation occurs at the exon-intron junctions. This observation led to suggest that DNA methylation could impact the splicing of transcripts (Laurent et al., 2010).

Many genes have alternative promoters, classically located upstream of the transcription start site but also commonly present within gene bodies (Kimura et al., 2006). Starting from this observation, Maunakea and colleagues decided to test whether the CGI that were found methylated in human brain corresponded to alternative promoters. They compared their methylation data with Capped Analysis of Gene Expression (CAGE) which identifies the presence of a transcription start site. They observed that a significant proportion of the methylated CGI in the gene body or in the 3'UTR portion of the gene matched with a putative transcription start site. Therefore, they suggested that methylation at CGI in the gene body may help to repress transcription from alternative promoters (Maunakea et al., 2010).

4.5.3. At tandem and interspersed DNA repeats

Transcription of transposable elements must be prevented to avoid potentially aberrant events of retrotransposition. In germ and mouse ES cells, methylation levels at transposable elements are low compared to somatic cells, and correlate with transcriptional activity, with the exception of IAP which present a higher rate of DNA methylation (Efroni et al., 2008; Ergun et al., 2004; Fadloun et al., 2013; Lane et al., 2003; Peaston et al., 2004; Taruscio and Mantovani, 2004; Trelogan and Martin, 1995). In germ cells, new insertions can take place and can result in germ line mutations (Maksakova et al., 2006). The fact that DNA methylation plays a critical role in silencing of transposable elements in somatic as well as in germ cells is supported by evidence obtained in murine cells. The first evidence was given by Gaudet and colleagues who showed that DNMT1 is essential to keep IAP elements silent in pre- and post-implantation embryos (Gaudet et al., 2004). These findings led the authors to propose that IAP transcription could have profound consequences on development (Gaudet et al., 2004). Only two years later, Bourc'his showed hypomethylation of LINE1 and IAP elements in DNMT3L knock-out germ cells compared to wild-type, and correlated transcription of these elements. This aberrant transcription was associated with meiotic catastrophes (Bourc'his and Bestor, 2004).

In addition, a study performed on mouse ES cells knock out for DNMT3A and DNMT3B or DNMT1 suggested another important role for DNA methylation at DNA tandem repeats to prevent aberrant recombination events. Indeed, DNMT3A/DNMT3B double knock-out mouse ES cells showed a reduction in the number of centromeric repeats, which was associated with an increase in mitotic centromeric recombination. Interestingly, this decrease was not observed in DNMT1 knock-out mouse ES cells, despite the fact that both cell lines showed hypomethylation of centromeric and pericentromeric repeats (Jaco et al., 2008). In addition, DNA methylation contributes to prevent aberrant transcription of tandem repeats, that has been associated with chromosomal instability (Bouzinba-Segard et al., 2006).

As a whole, DNA methylation plays a crucial role not only in transcriptional silencing of genomic repeats, but also in maintenance of genome integrity. Its action is synergic with other epigenetic mechanisms that will be described in the following chapters of this manuscript.

4.5.4. At Imprinted Control Regions (ICRs)

DNA methylation has been shown to be essential for establishment and maintenance of imprinting. Although differential DNA methylation at ICRs is present in all somatic lineages of the embryo, the expression of imprinted genes can be tissue-specific (Kacem and Feil, 2009). Around 88% of mouse imprinted genes show a CGI which often corresponds to the ICR (Neumann et al., 1995). H19 - IGF2 is one of the most studied imprinted loci. In physiological conditions, the paternal allele of H19 is methylated and transcriptionally silent (Bartolomei et al., 1991). Indeed, the expression of these genes was analysed in DNMT1 knock-out mice, where DNA methylation at ICR was lost. This loss of DNA methylation caused an increase in the expression in H19, suggesting that the inactive allele started to be transcribed (Li et al., 1993). The follow-up studies focused on the differentially methylated region (DMR) controlling the maternal expression of H19. Bartolomei's lab studied the deletion of this region in mice, reporting that in the case of a paternal inheritance of the deletion there was an increase in the expression of H19 from the paternal allele which, however, did not reach the level of the maternal allele. Surprisingly, they observed that, in the case of a maternal inheritance of the deletion, the expression of H19 decreased while the expression of IGF2 increased. They propose that the deletion of the region which is normally methylated resulted in a competition for the endodermal enhancer, downstream of H19 gene (Thorvaldsen et al., 1998). DNA methylation, hence, allows the correct monoallelic expression of imprinted genes and contributes to the correct organization of imprinted loci. Moreover, mice with aberrant expression of imprinted genes showed increase or decrease in body weight and size, supporting the importance of regulating the expression of these genes for physiological growth and metabolism (Smith et al., 2006).

4.6. Readers of DNA methylation

The impact of DNA methylation on transcription can be explained by the fact that DNA methylation can be read by specific readers, which in turn may recruit co-repressors or co-activators. Alternatively, DNA methylation could also play a role on transcription by preventing the binding of methylation-sensitive transcription factors.

4.6.1. 5mC readers

The family of proteins that can specifically bind to methylated DNA are generally called methyl-CpG-binding proteins (MBD). This family of proteins comprises Methyl-CpG-binding protein 1 and 2 (MECP1 and MECP2), which were the first to be described. MECP1 is known to bind regions dense in CpG methylation whereas MECP2 binds to single methylated CpG (Meehan et al., 1989, 1992). MECP1 is less studied compared to MECP2, because of the implication of MeCP2 in the Rett syndrome (**Chapter III**), and has been found to collaborate with other methyl-binding proteins as we will see later.

MECP2 is known to contribute to gene repression and this is suggested by its interaction with histone deacetylases and histone methyl-transferases like SETDB1 and SUV39H1. Hence, it has been involved in heterochromatin formation and chromatin organization (Jones et al., 1998; Lunyak et al., 2002; Nan et al., 1998). For example, Nan and colleagues showed that MECP2 binds to DNA in a methylation-dependent manner and interacts with histone deacetylase (HDAC) enzymes *via* its Transcriptional-Repression Domain (TRD). The formation of this complex causes the repression of transcription *in vitro* and *in vivo*. The inhibition of histone deacetylates by Trichostatin A (TSA) inhibited the transcriptional repression, without affecting MECP2 binding (Nan et al., 1998).

Other MBD proteins were identified later by sequence homology with MECP2. Up to now, 6 MBD proteins have been identified, named MBD1 to 6 (Du et al., 2015). Among the MBD proteins, only MBD1, 2 and 4 can bind to methylated DNA (Hendrich and Bird, 1998). MBD1 and 2 have been described to promote gene repression by recruiting histone deacetylases as well as the nucleosome remodelling and deacetylation complex (NuRD complex). Ng and colleagues, indeed, showed that MBD2 binds to methylated DNA, interacts with MECP1-NuRD complex mediating gene repression in HeLa cells. MBD2 is thought to increase the affinity of the MECP1 complex for methylated DNA, given the low affinity of MECP1 for methylated DNA (Ng et al., 1999). In contrast, MBD4 seems to be more implicated in DNA repair processes than in repression of transcription (Hendrich and Tweedie, 2003). MBD5 and MBD6 are less-characterized members of this family. Although they contain a MBD domain that allows their localization at

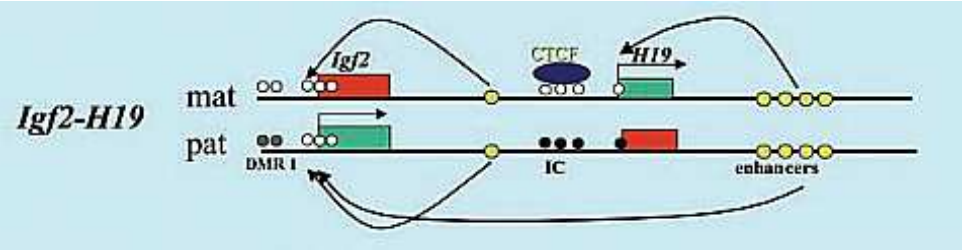


Figure 25. Schematic representation of the role of CTCF in the expression of the imprinted genes H19 and IGF2. CTCF is able to bind to the unmethyalted ICR or IC on the maternal allele. Its binding allows the boundary effect, contributing to the expression of H19 on the maternal allele. On the paternal the ICR or IC is methylated, inhibiting the binding of CTCF, contributing to the expression of IGF2 on the paternal allele. Gray circles indicate partial methylation. Larger yellow circles denote enhancers. Active alleles of genes are noted in green with silent alleles in red. Arrows indicate interactions between cis-elements on the two parental chromosomes. *From Feguson-Smith and Surani 2001.*

heterochromatin regions, Laget and colleagues showed that these proteins bind DNA in a methylation-independent manner *in vitro* (Laget et al., 2010).

4.6.2. Methylation-sensitive transcription factors

In contrast with the previous category of proteins, DNA methylation prevents the binding of this class of factors to their consensus DNA sequence. Among theses factors, several factors have been implicated in the differential expression from imprinted loci. For example, the recruitment of YY1 to DNA is prevented by the methylation of one CpG contained in its consensus binding, GCGCCATCTT (Kim et al., 2003). YY1 is recruited at the promoter of the imprinted gene PEG3, but only on the unmethylated paternal allele, promoting its expression from this allele and not from the methylated maternal allele (Kim et al., 2003).

A protein domain which seems to be sensitive to DNA methylation is the CXXC domain. This domain is known to favour the targeting of proteins to methylation-free regions (Long et al., 2013). For example, CXXC Finger Protein 1 (CFP1) is never found at methylated promoters. Genome-wide analysis showed that it is binds at 80% of the mouse CGI, that are generally unmethylated, and it is contained in the SETD1 H3K4 methylation complex which establishes H3K4 methylation at the promoter region of genes (Lee and Skalnik, 2005; Thomson et al., 2010).

Moreover, there are two proteins that can be classified as methylation-sensitive factors which have been shown not only to bind to unmethylated sites, but also to protect DNA from DNA methylation. These factors are SP1 and CTCF (CCCTC-binding factor). SP1 binds to the consensus motif CCGCCC, which is particularly enriched in CGI (Brandeis et al., 1994; Kriwacki et al., 1992; Macleod et al., 1994). Mutations of SP1 binding sites within the promoter of its target GTF2A1l caused the loss of SP1 recruitment, increase in DNA methylation and loss of gene expression (Lienert et al., 2011). Similarly, 25% of regions bound by CTCF contain a CpG that is not methylated (Feldmann et al., 2013). CTCF has been shown to protect different regions from DNA methylation (Schoenherr et al., 2002; Wang et al., 2012). For example, CTCF plays a crucial role in the complex organization of the H19/IGF2 locus. CTCF binds to unmethylated ICR on the maternal allele, contributing to the expression of H19. At the same time, it contributes to the repression of IGF2 forming a boundary effect which blocks the action of the downstream enhancer (Hark et al., 2000; Lee et al., 2010) (**Figure 25**). Moreover, insertion of mutations in the four CTCF binding sites at this ICR, caused a loss of CTCF recruitment and significant increase in DNA methylation (Schoenherr et al., 2002).

CTCF is not only able to prevent DNA methylation by binding to precise sites, like ICRs, but it was proposed to be implicated in the dynamics of DNA demethylation. Feldmann and colleagues, indeed, compared genome-wide methylation levels at low methylation regions, CTCF

binding and 5hmC levels. They observed that absence of methylation corresponds to binding of CTCF, as previously described, and that high levels of 5hmC were detected at the same regions in ES cells and neuronal progenitors. They finally depleted a factor similar to CTCF, which is essential for ES survival, called REST (RE1-silencing transcription factor) and observed a drastic decrease in 5hmC which corresponds to a strong enrichment in DNA methylation. These findings suggest that this category of transcription factors could be important in the dynamic of DNA demethylation (Feldmann et al., 2013). The precise mechanism implicated in this dynamic still needs further investigations.

4.6.3. Protein binding both methylated and unmethylated DNA

Some proteins with this feature belong to the family of the zinc finger and BTB (Broad complex, Tramtrack, and Bric-a`-brac) proteins (Perez-Torrado et al., 2006), like KAISO and KAISO-like proteins. KAISO, also known as ZBTB33, was discovered as a DNA-binding factor interacting with the adhesion molecule p120 catenin. KAISO was found to repress transcription in a CpG-methylation-dependent manner, taking part to the MECP1 complex (Feng and Zhang, 2001; Prokhortchouk et al., 2001). A homology search of KAISO-like zinc-fingers proteins led to the identification of two proteins closely related to KAISO, ZBTB4 and ZBTB38 (Filion et al., 2006). Both ZBTB4 and ZBTB38 have a preference for DNA sequences containing methylated CpGs, but it has been shown to bind also to unmethylated DNA with a lower affinity. Interestingly, these two proteins bind to methylated DNA in a sequence-independent manner, while fixation at unmethylated DNA occurs at the consensus sequence $C^{C}/_{T}GCCATC$. The consensus sequence was identified by homology with KAISO consensus binding site, corresponding to CTGCNA (Daniel et al., 2002; Sasai et al., 2010).

When these proteins bind to methylated DNA, it can repress transcription through the recruitment of histone deacetylases *in vitro* (Filion et al., 2006). Mice knock-out for ZBTB4 show normal development although mice are smaller that normal littermates. However, cells depleted in ZBTB4 present defects in cell proliferation, in the G2/M checkpoint regulation and increase of apoptosis (Roussel-Gervais et al., 2017). The connection between these anomalies and DNA methylation still needs to be elucidated.

4.6.4. 5hmC, 5fC and 5caC readers

Like the other epigenetic marks, 5hmC, as a candidate epigenetic mark, was thought to be "read" by specific proteins. In 2013, Vermeulen's lab derived a pull-down strategy combined with mass-spectromerty in order to identify new readers of 5mC, 5hmC and all their derivatives. The

number of proteins interacting with 5hmC was lower than those interacting with 5mC in ES cells. Indeed, this study identified 10 proteins that interacted with 5hmC, only 6 of them were specific of this mark and 4 of them included helicases and glycosylases that could be involved in the BER pathway. Surprisingly, UHRF1 was found to interact with 5hmC, which is in contradiction with the inability of the DNA methylation maintenance machinery to recognize this modification. On the other hand, 5fC and 5caC were able to recruit a large number of proteins (84 and 67 respectively). Of particular interest, the tumour suppressor p53 was found to interact with 5fC (Spruijt et al., 2013). In the same paper, the authors showed the results obtained performing the same pull-down strategy in neuronal progenitors, differentiated *in vitro* from mouse ES cells, and adult brain extracts. Interestingly, a large number of neuron-specific proteins were found to interact with 5hmC and, for the majority of these proteins, the explanation could not only result from a differential abundance of the proteins (Spruijt et al., 2013). These last findings suggest that readers of DNA modifications are highly dynamic upon neuronal differentiation (Eberl et al., 2013).

4.7. Interplay between DNA methylation and histone methylation

DNA methylation, in the chromatin context, may be determined by the presence of some chromatin modifications and, inversely, may be mutually exclusive with others (Jin et al., 2011). In this section I will focus on the crosstalk between DNA methylation and histone methylation.

4.7.1. DNA methylation and H3K4 methylation at promoters

H3K4 methylation, and more precisely H3K4me2 and me3, mark active promoters (Kim et al., 2005; Peters and Schübeler, 2005), while DNA methylation occurring at HCPs and ICPs is generally accepted to inhibit gene expression (Jones, 2012). These two observations suggest that the two marks could be mutually exclusive. Indeed, Weber and colleagues showed that unmethylated HCPs were enriched in H3K4me2, independently of the transcriptional state of these genes. Inversely, methylated HCPs did not show any enrichment in H3K4me2. This suggested that the dimethylation of H3K4 is a common feature of CGI promoters and that this modification could prevent the DNA methylation of gene promoters (Weber et al., 2007). This hypothesis is further emphasized by the fact that H3K4 is established by the SETD1 H3K4 complex which is targeted at promoters by CFP1, whose binding to DNA is inhibited by DNA methylation (*Cf* 4.6.2 in this section) (Lee and Skalnik, 2005; Thomson et al., 2010). Moreover, studies in yeast showed that the insertion of mutations in SETD1 H3K4 homolog caused a global increase in ectopic DNA methylation (Rose and Klose, 2014). A possible hypothesis to explain the exclusion of DNMTs

by H3K4 methylation may be the fact that H3K4me3 can act as a docking site for transcriptional factors like TFIID (Vermeulen et al., 2007).

The model of mutual exclusion between these two mark is strengthened by the work of Ooi and colleagues who observed that the unmodified histone tail of the histone H3 co-crystalized with DNMT3L at the level of the lysine 4 and that this interaction was lost if that lysine was methylated. This interaction allows the recruitment of DNMT3A on chromatin and the methylation of DNA (Ooi et al., 2007). Structural studies proposed that this interaction between H3K4 and the complex DNMT3L-DNMT3A could activate DNMT3A in a DNMT3L-dependent manner. Indeed, the interaction between H3K4 and DNMT3L promotes a conformational change of the complex, disrupting the inhibitory interaction between the ADD domain and the catalytic domain of DNMT3A, promoting its activity (Guo et al., 2015).

Although all these observations suggest that H3K4 methylation inhibit DNA methylation, while unmodified H3K4 stimulates it, the mechanism behind the recruitment of SETD1 H3K4 or DNMTs is still under investigation. Chen's lab, indeed, published two papers showing that mutations in the histone demethylases KDM1 and KDM1B, also known as LSD1 and LSD2 respectively, cause loss of DNA methylation. In the case of LSD1, the loss of methylation is related to the degradation of DNMT1 which results no more stable (Wang et al., 2009). In contrast, for LSD2, the loss of DNA methylation at imprinted loci was reminiscent of an imprinting defect, as observed in mice deficient for DNMT3A or DNMT3L (Bourc'his et al., 2001; Hata et al., 2002; Kaneda et al., 2004). These findings suggest that a H3K4 demethylation activity is required to promote DNA methylation at imprinted loci and gene promoters (Ciccone et al., 2009).

4.7.2. H3K36 methylation and DNA methylation in gene bodies

Gene bodies are highly methylated, in particular gene bodies of transcribed genes (Jones, 2012). They are also characterized by the presence of H3K36me3, even if its repartition is not homogeneous. It has been observed, indeed, that there is progressive shift from mono-methylation to trimethylation of H3K36 between the promoters and the 3' part of genes (Hahn et al., 2011). *In vitro* studies suggested that H3K36me3 could interact directly with DNMT3A. By pull-down experiments, Dhayalan and colleagues showed that DNMT3A could be recruited on native nucleosomes from human cells depleted in DNA methylation, through its PWWP domain. This observation suggested that H3K36me3 could guide *de novo* DNA methylation on nucleosomes (Dhayalan et al., 2010). A more recent paper contributed to clarify the mechanism of recruitment of DNMTs at gene bodies through H3K36me3. Indeed, the mapping of DNMT3A and DNMT3B binding, as well as the distribution of DNA methylation, were analysed by reintroducing these DNMTs in TKO mouse ES cells. This experimental procedure allowed Baubec and colleagues to

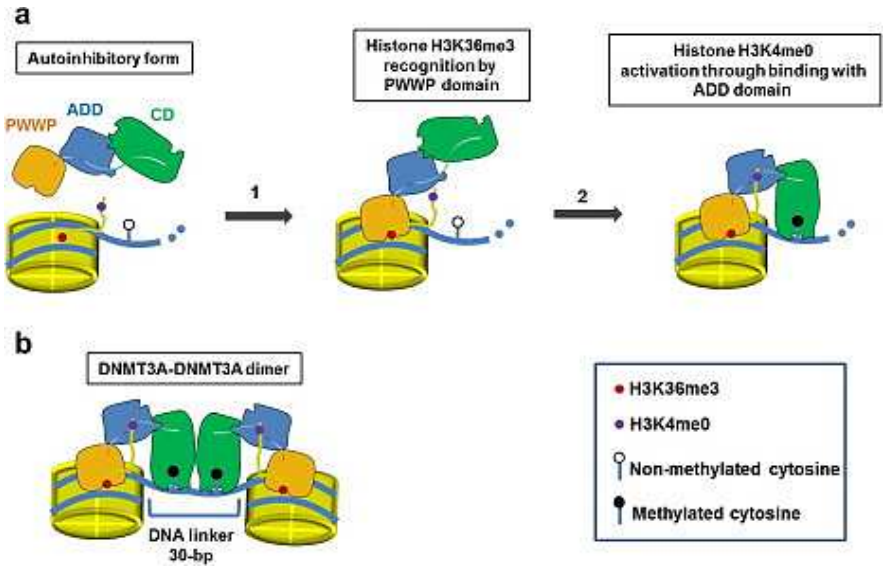


Figure 26. Proposed mechanism for recognition of nucleosome and DNA methylation by DNMT3A (for clarity, the DNMT3L was removed)**.** a) Sequential recognition mechanism of DNMT3A for nucleosome and methylation activation with H3 tail: 1-Nucleosome recognition by PWWP domain through interaction with methylated histone H3 tail (H3K36me3). 2-Histone H3 tail activation (H3K4me0) through binding with ADD domain. ADD domain of DNMT3A interaction with the catalytic domain is disrupted, promoting the interaction with DNA and DNA methyltranferase activity. b) Periodicity in de novo methylation patterns between two CpG sites. *From Rondelet et al., 2016.*

analyse the targeting and the *de novo* DNA methylation activity of DNMT3A and DNMT3B, independently of pre-existing DNA methylation. The global increase in DNA methylation was higher when reintroducing DNMT3A compared to DNMT3B, although DNMT3B preferentially methylated gene bodies of active genes characterized by the presence of H3K36me3 (Baubec et al., 2015). Moreover, the binding of DNMT3B at these gene bodies correlated with an increase in transcription of these genes. Interestingly, H3K36me3 distribution matched perfectly with the binding of DNMT3B. Moreover, cells knock-out for SETD2, the methyltransferase responsible for the trimethylation of H3K36, showed loss of H3K36me3 and impairment of DNMT3B binding. In contrast, the binding of DNMT3A was not affected. The authors observed that mutations in the PWWP domain of DNMT3B resulted in loss of interaction with H3K36me3 and loss of DNA methylation in genes bodies enriched in H3K36me3. All these findings support a H3K36me3-dependent recruitment of DNMT3B at gene bodies of transcribed genes to establish DNA methylation. In contrast with the findings of Dhayalan and colleagues, this way of recruitment was not observed for DNMT3A, suggesting that the preferential recruitment of DNMT3B may be specific for mouse ES cells and that the targeting of DNMT3A could take place in other cell contexts (Baubec et al., 2015).

The possible targeting of DNMT3A and DNMT3B by H3K36me3 was analysed also from a structural point of view. Rondelet and colleagues showed that H3K36me3 is physically associated with DNMT3A and DNMT3B (Rondelet et al., 2016). Moreover, the authors of this work suggest a model for the recruitment of DNMT3A which takes into account also the results of Guo and colleagues, presented in the previous section (Guo et al., 2015). The model in **Figure 26** proposes that DNMT3A can be targeted to genomic regions by H3K36me3 through its PWWP, favouring the formation of a dimer. The interaction with unmodified H3K4 can be recognized by the ADD domain of DNMT3L activating the complex and allowing the DNA methyltransferase activity (Rondelet et al., 2016).

4.7.3. DNA methylation and H3K9me3: old good friends

As the subtitle suggests, DNA methylation and H3K9me3 have always been considered good friends, not only because they are found many times at the same regions, but also because their mechanisms of establishment and maintenance are tightly related. In particular, it is commonly thought that H3K9me3 recruits the DNA methylation machinery (Rose and Klose, 2014). It is important to say that, generally, H3K9me1 and H3K9me2 are established before DNA methylation by a G9A and GLP dependent-mechanism. In the absence of these two enzymes, DNA methylation is not established at pluripotency genes in mouse ES cells (Epsztejn-Litman et al., 2008). This loss of DNA methylation relies on the absence of the recruitment of DNMT3A and

DNMT3B by G9A and GLP directly or indirectly, in mouse ES cells. Indeed, Epsztejn-Litman and colleagues showed that DNMT3A and DNMT3B can be recruited by G9A, independently on its methyltransferase activity, based on its ankyrin repeat domain (ANK domain) (Epsztejn-Litman et al., 2008). Auclair and colleagues observed a global loss of DNA methylation in G9A knock-out ES cells, affecting DNA repeats, a set of pluripotency genes and germline genes. Strikingly, part of these results was not observed in G9A knock-out embryos at 8.5 dpc (Auclair et al., 2016). Indeed, in these embryos, DNA methylation was barely affected at DNA repeats and pluripotency genes. However, germline genes were hypomethylated similarly to DNMT3B knock-out embryos and also the imprinted gene SLC38A4 was hypomethylated. The authors showed that G9A is required to recruit DNMT3B at these germline genes to establish DNA methylation. However, H3K9me2 alone is not sufficient to silence these germline genes, supporting the pivotal role of DNA methylation in the silencing of these genes (Auclair et al., 2016). Nevertheless, the recruitment of DNMT3A and DNMT3B can occur also by an indirect mechanism. Among the readers of H3K9me2, MPP8 protein is known to interact with G9A and to bind H3K9me2 through its chrom-odomain. MPP8 was shown to interact with N-terminal portion of DNMT3A, resulting in its recruitment at H3K9 methylated regions (Chang et al., 2011; Kokura et al., 2010).

H3K9me3 was found to mediate the recruitment of DNMTs. In particular, pull-down experiments suggested that H3K9me3 can be directly recognized by the PWWP domain of DNMT3A, suggesting that the histone mark may be directly implicated in the establishment of DNA methylation at some loci, like at imprinted loci (Dhayalan et al., 2010).

As I described in section dedicated to constitutive heterochromatin, H3K9me3 can be read by the family of HP1 proteins. This factor is crucial for the further recruitment of SUV39H1 and increase in H3K9me3 levels. However, this is not the only function of these proteins. By pull-down experiment HP1 proteins were found to interact with DNMT3A, DNMT3B and DNMT1. These findings suggest that the recognition of H3K9me3 by HP1 proteins may be crucial for the recruitment of *de novo* as well as for the maintenance activity of DNA methylation DNMTs (Fuks, 2003; Lehnertz et al., 2003). More recent evidence for the privileged link between H3K9me3 and DNA methylation maintenance was reinforced. Structural studies showed that one of the main recruiter of DNMT1, UHRF1, co-crystalizes with H3K9me3 through its Tandem Tudor Domain (TTD)-PHD domain (Arita et al., 2012). Further studies then showed that the domain of UHRF1 protein called “Linker”, localized between the TTD and PHD domain, was also crucial for H3K9me3 recognition (Rothbart et al., 2013). Moreover, in the same work the authors tried to rescue the drastic loss of DNA methylation caused by knock-down of UHRF1 in human cells with different truncated forms of the protein. In absence of the TTD domain, the “linker” domain or the PHD domain, DNA methylation was not restored, whereas it was clearly possible with the UHRF1

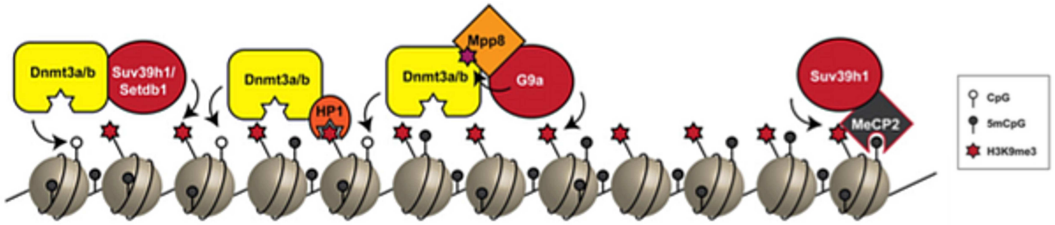


Figure 27. Targeting of DNA methylation and H3K9 methylation to heterochromatin. *Adapted from Rose and Klose 2014.*

wild-type protein. These findings suggest that UHRF1 recognition of H3K9me3 is crucial for the maintenance of DNA methylation at the level of chromatin (Rothbart et al., 2013). This mechanism of DNA methylation maintenance may be particularly important for repeated sequences that resist to DNA methylation reprograming, like ERVs class I and II (IAP). These regions present high levels of DNA methylation and H3K9me3, even in the absence of DNMT3A and DNMT3B. In contrast, in the absence of H3K9me3, caused by deletion of SETDB1, DNA methylation was lost through activation of TETs enzyme and active DNA demethylation (Leung et al., 2014).

SUV39HDN mouse ES cells or SETDB1 knock-out showed DNA methylation defects (Karimi et al., 2011; Lehnertz et al., 2003; Matsui et al., 2010). These findings may suggest that the activity or specific domains of SUV39H1/2 or SETDB1 could be important for the recruitment of DNMTs. ES cells SUV39HDN show hypomethylation of pericentromeric repeats and LINE1 elements (Bulut-Karslioglu et al., 2014; Lehnertz et al., 2003). SUV39H1 interacts with DNMT3A and DNMT1 *in vitro* and *in vivo* in human cells but, unfortunately, DNMT3B was not tested. Anyhow, this experiment strongly suggested that SUV39H1/2 enzymes may provide one of the way of recruitment for *de novo* and maintenance DNMT enzymes (Fuks, 2003). Direct connections between SETDB1 and DNMTs have already been proposed by Li and colleagues who showed direct interaction between SETDB1 and DNMT3A. Indeed, this interaction seemed to promote the recruitment of DNMT3A at gene promoters in the particular case of cancer cells (Li, 2006). More recently, other evidences suggest that the link between SETDB1 and DNMT enzymes could be mostly related to KAP1, one of the closest partner of SETDB1. As I described in the section 3.2 of this chapter, KAP1 proteins is recruited at genomic loci *via* KRAB Zinc-finger proteins (**Figure 16**). KAP1 is recruited to ICRs by the KRAB zinc-finger protein ZFP57. In this context, the KRAB domain is essential for KAP1 recruitment, confirming the necessity of this domain for its genomic targeting. Moreover, the complex ZFP57/KAP1 recruits DNMT3A and DNMT1 at the ICRs to establish and maintain the methylation at imprinted genes. These findings support a strong interconnection between KMT and DNMT recruiters to maintain heterochromatin at different genomic loci (Quenneville et al., 2011; Zuo et al., 2012).

Inversely, it is not excluded that DNA methylation may promote the establishment and/or maintenance of H3K9me3 at genomic loci via DNA methylation readers, suggesting a reciprocal action. Indeed, MECP2 was shown to associate with SUV39H1 at different genomic loci, engaging DNA and histone methylation. These findings suggest the establishment of a self-reinforcing loop in which DNA methylation and H3K9me3 contribute to heterochromatin compaction and gene silencing (Lunyak et al., 2002). The different mechanisms described in this chapter are summarised in **Figure 27**.

4.7.4. Mutual exclusion between DNA methylation and H3K27me3

The mutual exclusion between DNA methylation and H3K27me3 is not so straightforward to explain in a mechanistic point of view and seems to rely on several features. Indeed, the density in CpGs and the differentiation stage play crucial roles in this crosstalk. In human or mouse differentiated cells, H3K27me3 has been shown to spread away from CGI covering around 10% of the genome (Hawkins et al., 2010; Pauler et al., 2008). In these blocks enriched in H3K27me3, DNA methylation was not depleted, showing that DNA methylation and H3K27me3 can overlap. However, methylated CGI remained poorly enriched in H3K27me3 (Brinkman et al., 2012; Statham et al., 2012). In mouse ES cells, this mutual exclusion is even more drastic. Indeed, between 80 and 90% of regions enriched in PRC1 and PRC2 are unmethylated (Ku et al., 2008; Tanay et al., 2007). Several *in vitro* studies provided controversial results, showing on one hand that PRC1 and PRC2 were not able to interact with methylated DNA, and on the other hand that PRC2 was able to methylate H3K27 on methylated DNA, even with a high density of CpG in mouse ES cells (Cooper et al., 2014). A controversial study in cells lines showed that PRC2 complex could directly interact with DNMT enzymes, leading to DNA methylation at some EZH2 targets, like MYT1 or WNT (Viré et al., 2005). This study may suggest that a complex between two exclusive activities may form and act to promote DNA methylation and H3K27me3 depending on the differentiation stage, in particular H3K27me3 in mouse ES cells and DNA methylation in differentiated cells. In contrast, another study showed that a spreading of H3K27me3, like it was observed in somatic cells, was observed only in TKO mouse ES cells (Brinkman et al., 2012). In this context, H3K27me3 is deposited at major satellite repeats together with H3K9me3, but the two marks never overlap (Cooper et al., 2014; Marks et al., 2012; Saksouk et al., 2015). This enrichment results from the recruitment of PRC2 through BEND3, which binds directly to major satellite repeats, probably in a methylation-dependent manner (Dai et al., 2013; Saksouk et al., 2015). The spreading of H3K27me3 at major satellite repeats was even higher in SUV39HDN cells, in absence hence of DNA methylation and H3K9me3, suggesting that H3K27me3 could compensate the absence of these marks (Lehnertz et al., 2003; Peters et al., 2003).

Conversely, absence of H3K27me3 had only a modest effect on DNA methylation. Hagarman and colleagues studied the genome-wide levels of DNA methylation by MeDIP-Seq in mouse ES cells defective for PRC2 complex. They observed that 861 regions, which were marked by H3K27me3, gained DNA methylation, while 552 showed a decrease in DNA methylation. They performed the same experiment also in mouse ES cells knock-out for DNMT1, DNMT3A and DNMT3B (DNMT triple knock-out, TKO) and observed an increase in H3K27 at promoter regions (Hagarman et al., 2013), in agreement with previous results, at other loci (Lehnertz et al., 2003; Peters et al., 2003). In order to show that DNA methylation is antagonizing the placement

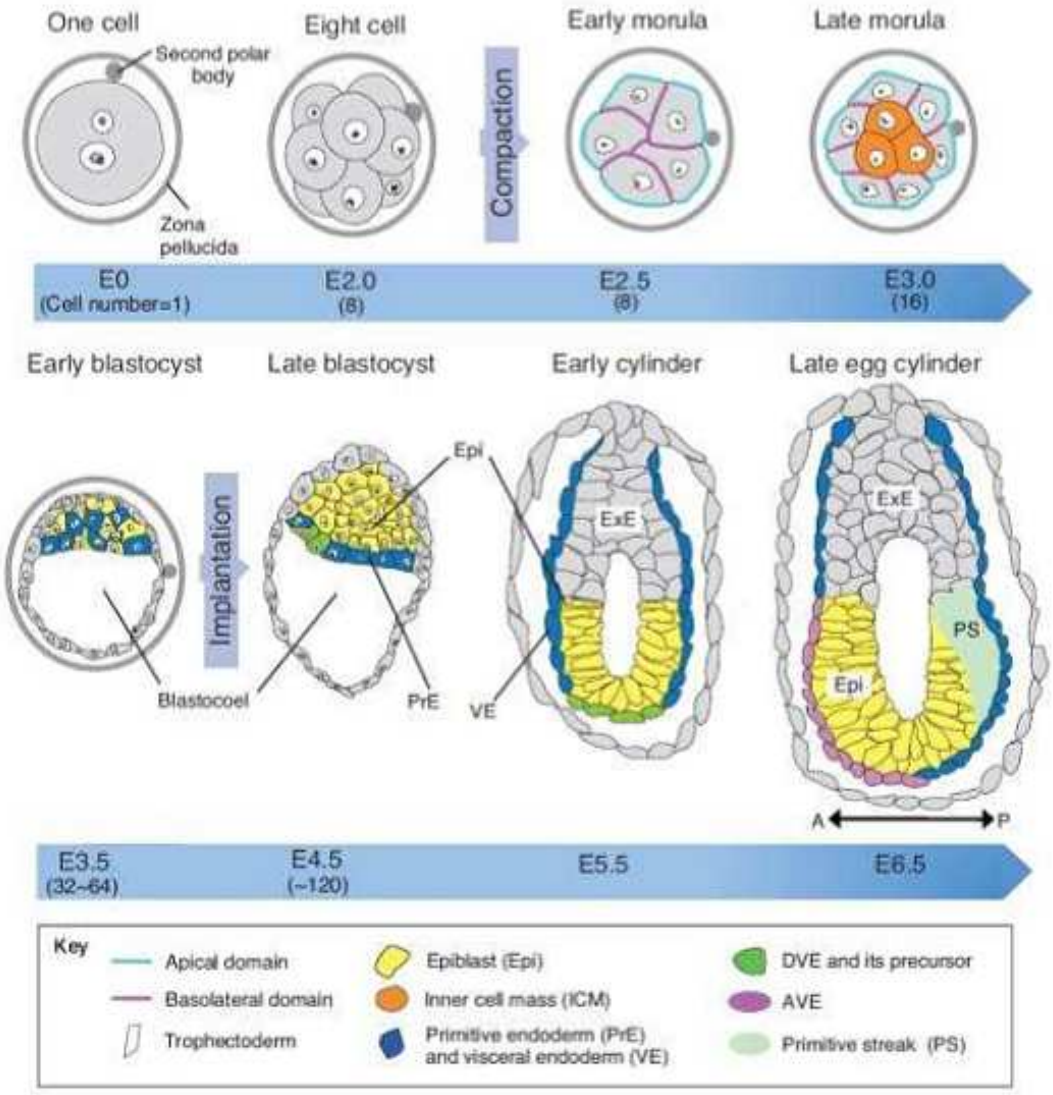


Figure 28. Early mouse development from 1-cell stage to gastrulation. Kinetics of morphological changes and cell-specification from zygote to late egg cylinder. The cell types in the embryos are coded with different colours: epiblast (Epi), primitive ectoderm (PrE), visceral endoderm (VE), extra-embryonic ectoderm (Exe), distal visceral endoderm (DVE), anterior visceral endoderm (AVE), primitive streak (PS). *From Takaoka and Hamada 2012.*

of H3K27me3, they compared promoters which gained H3K27me3 in TKO mouse ES cells to DNA methylation data in wild-type mouse embryonic stem cells. They observed that, in 99% of the cases, these promoters presented DNA methylation in wild-type cells and concluded that DNA methylation is actually antagonizing H3K27me3 (Hagarman et al., 2013). Unfortunately, the authors did not propose a mechanistic explanation for this phenomenon. We can not exclude that the absence of DNA methylation may favour the recruitment of methylation-sensitive protein that could recruit PRC2, mediating H3K27me3 deposition.

As a whole, DNA methylation and H3K27me3 seem to be mutual exclusive, especially at CGI promoters. However, the mechanism underlying this mutual exclusion still needs to be clarified.

4.8. DNA methylation in normal development

4.8.1. Analyses of DNA methylation *in vivo*

As I described in the previous part of this chapter, reprogramming of DNA methylation is required for normal development. As a consequence, perturbations in the process of reprogramming are associated with propensity to postnatal homeostatic instability (Ashe et al., 2008; Mahsoudi et al., 2007; Watkins et al., 2008). However, the description of the events occurring from zygote to epiblastocyst is based on fragmentary pieces of evidences and a mix of results obtained *in vivo* or *in vitro*.

The early mouse embryo development goes from totipotency, through pluripotency to differentiated cells. Totipotency is defined as the ability of a cell to give rise to a fertile adult individual, thus, including formation of the extra-embryonic tissues like placenta or yolk sac. In mouse, totipotency is restricted to the zygote and to the blastomeres of a 2-cells stage embryo. In contrast, pluripotency is defined as the ability of cells to self-renew *in vitro* and to differentiate into the three embryonic lineages, called mesoderm, endoderm and ectoderm, that will give rise to all the tissues of an adult organism, but not to extra-embryonic tissues (Condic, 2014).

Mammalian development starts with the fertilization of the oocyte by a spermatocyte, and the formation of the totipotent zygote (fertilized egg). The first steps before implantation are characterized by proliferation and differentiation into the first two lineages, the pluripotent inner cell mass (ICM) and the trophectoderm (TEm), that lead to the formation of the primitive blastocyst, 3.5 days after fertilization in the mouse (Wang and Dey, 2006). ICM cell populations are not homogeneous, but are formed by pluripotent cells that will form the epiblast cells (Epi), and by more committed cells called primitive endoderm (PrE) cells. By E4.5, prior to implantation, PrE cells form a layer at the surface of the ICM facing the blastocoel cavity (**Figure 28**). At this

stage, the blastocyst invades the uterine tissues and implants. Mural TEm surrounding the blastocyst cavity first makes contacts with maternal tissues and differentiates into primary TEm giant cells that mediate the mother-embryo exchanges. Conversely, polar TEm surrounding the outer face of the ICM proliferates and differentiates in extra-embryonic ectoderm (ExE) and ectoplacental cone (EPC) that will build the proximal half of the egg cylinder and, later on, the placenta (**Figure 28**). After implantation at E5.0, there is a burst of cell proliferation and growth resulting in the expansion of embryonic as well as extra-embryonic lineages into the blastocyst cavity. From the PrE emerge the visceral endoderm cells that cover both Epi and ExE. The elongating egg cylinder grows, reorganizes and forms the primitive streak in order to initiate the gastrulation (Camus et al., 2006). During peri-implantation stages, the Epi reorganises itself from a compact non-polarized ball to a rosette-like organization at the time of implantation, and finally to a cup-shaped polarized epithelium surrounding the pro-amniotic cavity in 24 hours (Bedzhov et al., 2014) (**Figure 28**). The anterior-posterior axis of the mouse embryo is established when distal visceral endoderm (DVE) forms the anterior visceral endoderm (AVE) on the future anterior side of the E6.5 embryo (Perea-Gomez et al., 2002).

Several genes have been shown to be required for the specification of the different cell populations that are formed during development, as exemplified by the famous experiments of reprogramming of somatic cells into pluripotent cells through the ectopic expression of combinations of these factors (Takahashi and Yamanaka, 2006). The expression of REX1, NANOG, OCT4, SOX2, KLF4 and KLF2 characterises ICM pluripotent cells and have been defined as pluripotency factors or pluripotency network. These factors are expressed homogeneously in the nuclei of all pluripotent cells *in vivo* (Avilion et al., 2003; Bourillot and Savatier, 2010; Chambers et al., 2003; Mitsui et al., 2003; Nichols et al., 1998; Rogers et al., 1991). In contrast, PrE cells are characterized by the expression of endoderm markers like GATA4 and GATA6 (Chazaud et al., 2006), whereas TEm cells are characterized by the expression of CDX2 gene, which starts to be expressed at the 8-cells embryo stage (Dietrich and Hiiragi, 2007; Ralston and Rossant, 2008). These markers are crucial for *in vitro* experiments, because they allow analysing the pluripotency states of manipulated cells, as I will describe later.

All these developmental steps are accompanied by changes in DNA methylation and hydroxymethylation, which were nicely documented last year. In agreement with earlier studies (Howlett and Reik, 1991; Monk et al., 1987; Reik et al., 2001), Li and colleagues reported that the lowest levels of DNA methylation are reached at the early blastocyst stage (Li et al., 2016). The authors detected 5mC and 5hmC by immunofluorescence on single blastomeres at each developmental stage. They showed a reorganization of these modifications, with the formation of high-density signal foci starting from the 8-cells stage. At the blastocyst stage, the ICM cells show

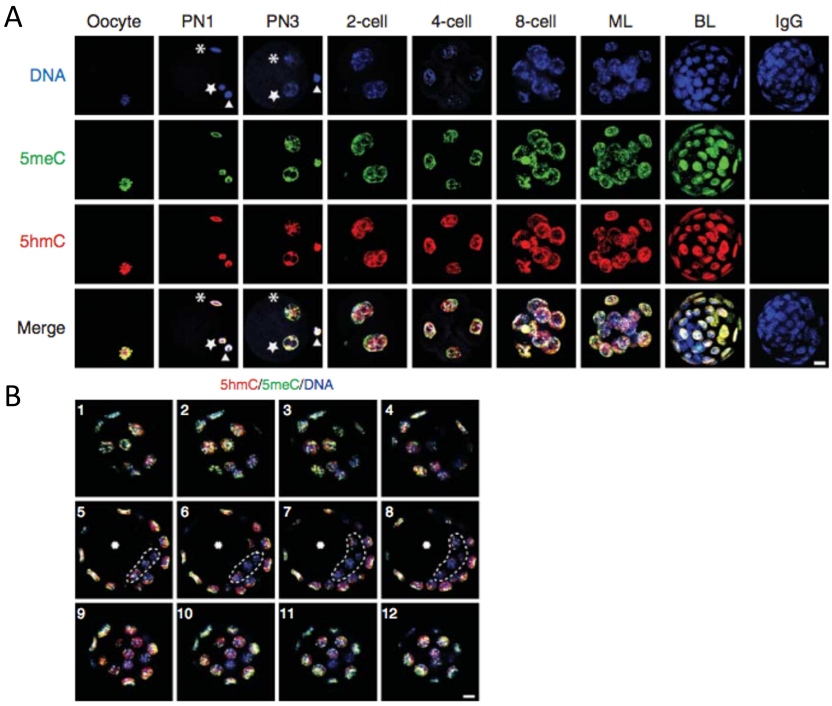


Figure 29. Patterns of 5mC and 5hmC staining in the oocyte and different stages of preimplantation embryo development. A) z-stack projection images of staining for DNA (DAPI, blue), 5meC (green), 5hmC (red), and their merged (merge) images in the oocyte; zygote (PN1 and PN3); 2-cell, 4-cell, and 8-cell embryos; morula (ML); blastocyst (BL); and a BL stained with non-immune control IgG. (*) indicates the decondensing sperm head (in newly fertilized oocyte) or the male pronucleus (in PN3 zygote). (open star) indicates a female pronucleus. (open triangle) indicates the extruded polar body. B) Single confocal optical sections through a blastocyst triple stained for 5meC (green), 5hmC (red), and DNA (blue). The images are sequential sections at 1-mm intervals at the lower pole (1–4, top row), through the equator (5–8, middle row), and upper pole (9–12, bottom row) of the embryo. Cells in the upper and lower poles are TEm cells, the equatorial sections show the ICM (ICM enclosed by dotted line) encircled by the TEm. (*) indicates blastocoel cavity. *Adapted from Li et al., 2016.*

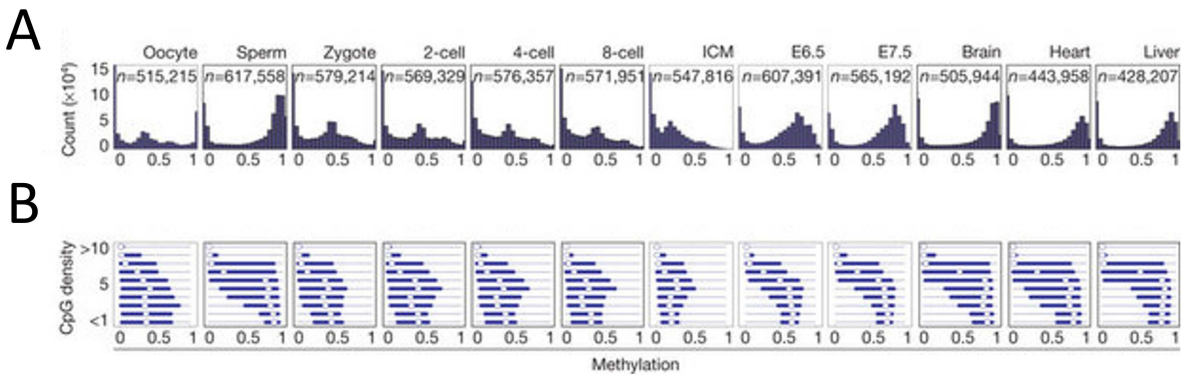


Figure 30. Global CpG methylation dynamics across early murine embryogenesis. A) Histogram of methylation values across 100-bp tiles. n is the number of tiles for each stage. B) Box plots of methylation values across local CpG densities highlight the difference between hypomethylated pre-implantation tissues and the somatic pattern seen in sperm, post-implantation and adult samples. Circle indicates the median, edges the 25th/75th percentile and whiskers the 2.5th/97.5th percentile. *From Smith et al., 2012.*

very low levels of 5mC and 5hmC compared to the surrounding TEm cells. These findings are consistent with the expression profiles of DNMT enzymes, with high levels of DNMT3B and DNMT3L in TEm cells and almost absence of all *de novo* DNMTs in ICM cells (Burton et al., 2013; Li et al., 2016) (**Figure 29**). Interestingly, DNA methylation levels *in vivo* are comparable between the epiblast and ICM cells (Borgel et al., 2010; Veillard et al., 2014), although the expression of DNMT3B increases immediately after implantation, which may be indicative of an increase in DNA methylation levels. The expression of DNMT3A is detected only from E10.5 (Watanabe et al., 2002).

In 2012, Smith and colleagues had performed a quantitative dynamic and genome-wide scale study, performing RRBS on oocytes, sperm, 2 to 8 cells embryos, total ICM, 6.5 and 7.5 dpc embryos and three differentiated tissues (brain, liver and heart). Unfortunately, RRBS analysis does not distinguish between 5mC and 5hmC. The authors "solved" the problem stating that "given this ambiguity, regions of high 5mC or 5hmC methylation, especially those retained over multiple time points, could still be expected to function as if methylated". However, we could argue that to function as if methylated, 5hmC should be read by MBDs and prevent transcription factors binding in the same manner.

Oocyte and zygote showed a greater hypomethylation compared to sperm, suggesting that, starting from the fertilization, a global and genome-wide demethylation takes place on the paternal genome, which is perfectly in agreement with previous studies (Mayer et al., 2000; Oswald et al., 2000). The DNA methylation levels decrease progressively reaching the lowest levels of DNA methylation in ICM cells. From epiblast cells, DNA methylation increases although it does not reach the levels found in differentiated tissues (**Figure 30A**). DNA methylation at all stages occurs mainly at regions of low CpG density, similarly to what was found in somatic tissues *in vitro* (Mohn et al., 2008) and *in vivo* (Smith et al., 2012). Interestingly, the DNA methylation profiles in sperm resemble more that of somatic tissues than pre- or post-implantation embryo (**Figure 30B**) (Smith et al., 2012).

This *in vivo* analysis revealed the presence of regions that resist to the genome-wide demethylation, like ICRs and IAP retrotransposons, which is in agreement with *in vitro* studies (Howlett and Reik, 1991) (as described in section 4.4).

4.8.2. Analyses of DNA methylation *ex vivo* in the mouse – culture systems

The possibility to derivate mouse ES cells from blastocysts at 3.5 dpc and Epiblast Stem Cells (EpiSCs) from epiblast at 5.5 dpc (Brons et al., 2007; Tesar et al., 2007) allows recapitulating the dynamic establishment of DNA methylation *ex vivo*. Smith and Nichols proposed that mouse embryonic pluripotency can be divided into two states: a "naïve" state corresponding to the pre-

Proprety	Naïve State	Primed State
Embryonic tissue	Early balstocyst	Post-implantation epiblast
Culture stem cells	Rodent ES cells	Rodent EpiSCs
Balstocyst chimeras	Yes	No
Teratomas	Yes	Yes
Differentiation bias	None	Variable
Pluripotency factors	OCT3/4, NANOG, SOX2, KLF2, KLF4	OCT3/4, NANOG, SOX2
Naïve markers	REX1, NROB1, FGF4	Absent
Specification markers	None	FGF5, T
Cell signalling dependency self-renewal in culture	LIF/STAT3	FGF/ERK
XX status	XaXi	XaXi

Table 4. Comparison of properties of "Naïve" versus "Primed" pluripotent states. *Adapted from Nichols and Smith 2009.*

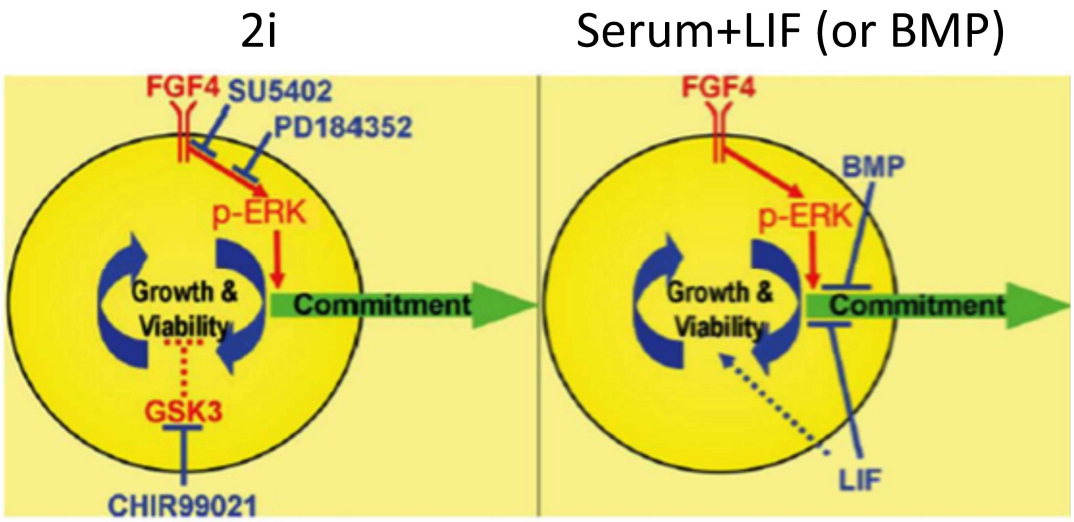


Figure 31. Signalling pathways inhibited under 2i condition. Pro-differentiation signals arising from autocrine FGF4 serum lead to the phosphorylation of ERK (p-ERK). In conditions of serum+LIF (or BMP) (right panel), commitment is inhibited downstream of p-ERK by LIF signalling and BMP. In 2i (left panel), p-ERK is inhibited upstream by chemical antagonists (a single MEK inhibitor PD0325902 has since replaced the two chemical agents depicted). Inhibition of GSK3 serves to relieve inhibition mediated by TCF3 on the pluripotency network. In 2i culture, LIF is usually added as well for robust self-renewal. *Adapted from Ying et al. 2008.*

implantation ICM cells, and a "primed" state corresponding to the post-implantation epiblast (J. Nichols et al., 2009). These states showed particular features that are summarized in **Table 4**.

Mouse ES cells were first isolated and expanded by plating ICM cells onto a layer of mitotically inactivated fibroblasts in the presence of media conditioned by teratocarcinoma stem cells (Martin, 1981). These fibroblasts were thought to provide trophic factors that aided the growth of the ES cells and became known as "feeders". A key advance in the refining of this culture environment came with the finding that pluripotency could be maintained with the addition of the IL-6 cytokine leukaemia inhibitory factor (LIF) (Smith et al., 1988; Williams et al., 1988). LIF has pleitropic effects on ES cell self-renewal, (Hirai et al., 2011), although its major downstream effector is the transcription factor STAT3 (Niwa et al., 1998). ChIP-Seq and ChIP-on-chip experiments performed to identify STAT3 genomic targets revealed that this transcription factor is a core component of the pluripotency network (X. Chen et al., 2008; Kidder et al., 2008). STAT3 binds to the distal enhancer of the genes coding for the core pluripotency transcription factors OCT4, SOX2, and NANOG (Do et al., 2013). Serum can be replaced by Bone Morphogenetic Protein 4 (BMP4), which promotes the expression of the inhibitors of differentiation (Id) proteins (Ying et al., 2003).

In mouse ES cells cultivated in serum+LIF, the expression of several key pluripotency factors, including NANOG and REX1, is heterogeneous (Chambers et al., 2007; Hayashi et al., 2008; Toyooka et al., 2008). The heterogeneous expression of these markers may reflect the coexistence of distinct cell states within the ES cell population, despite their genetic homogeneity and global ability to contribute to a new organism.

The derivation and propagation of ES cells in the absence of exogenously added serum, growth factors, or cytokines, was reported by the lab of Austin Smith (Ying et al., 2008). This system involves the dual inhibition of mitogen-activated protein kinase (MEK) and glycogen synthase kinase-3 (GSK3) by the small chemical inhibitors PD0325902 and CHIR992021, respectively, in what became widely known as "2i" (**Figure 31**).

Inhibition of MEK shields ES cells from inductive differentiation signals, including autocrine FGF4, which are transmitted via the RAS-MEK-ERK intracellular signalling cascade (Kunath et al., 2007). As a result, in 2i conditions, phosphorylation of ERK is inhibited at the top of the cascade. Inhibition of GSK3 results in an increase in the steady-state levels of β-catenin, and stimulates canonical WNT signalling (Ying et al., 2008). Although the precise role of WNT signalling in maintaining pluripotency has been hotly debated, Smith and colleagues show that the responsiveness of self- renewal to GSK3 inhibition is dependent on the interaction of β-catenin with TCF3, which serves to relieve the transcriptional repression promoted by TCF3 at multiple genes of the pluripotency network (Wray et al., 2011). This 2i culture condition, which is usually

implemented in combination with LIF, supports the robust long-term propagation of ES cells in culture, which are able to give rise to chimeras and readily contribute to germ line transmission when injected into blastocysts (Ying et al., 2008). Moreover, these conditions support the derivation of ES cells from mouse strains reputed to be non-permissive (Kiyonari et al., 2010; Jennifer Nichols et al., 2009; Ying et al., 2008), and for the first time from rats (Buehr et al., 2008; P. Li et al., 2008).

Unlike ES cells cultured in serum, the 2i condition promotes the homogenous expression of the pluripotency factors NANOG and REX1 (Wray et al., 2010), suggesting that the fluctuations in transcription factors expression observed with serum is related to cell culture conditions. Furthermore, NANOG is expressed in a monoallelic manner in serum conditions but in a bi-allelical manner in 2i conditions. The bi-allelic expression of NANOG was also observed in a greater proportion of cells from the ICM of the late blastocyst (Miyanari and Torres-Padilla, 2012). These observations, together with a proliferative state with minimal requirements for extrinsic stimuli, support the claims of Smith's lab that ES cells cultured in 2i represent a "ground state" of pluripotency (Wray et al., 2010; Ying et al., 2008).

The fact that ES cells represent a powerful model to study the waves of DNA methylation that occur pre- or post-implantation was recently reinforced by the genome-wide analysis of DNA methylation in the different culture conditions. WGBS was performed in ES cells cultivated in Serum+LIF and reprogrammed by switching to the 2i condition. Strikingly, ES cells cultivated in the standard Serum+LIF condition have heavily methylated genomes, corresponding to around 75% of methylated CpGs (Habibi et al., 2013; Stadler et al., 2011). In contrast, ES cells in 2i have between 20 and 30% of methylated CpGs. These results were compared to RRBS data obtained on pre-implantation ICM and post-implantation embryos. The amount and the distribution of methylated CpGs were very similar between ICM cells and ES cells cultivated in 2i, whereas ES cells in Serum+LIF presented strikingly higher levels of DNA methylation (Habibi et al., 2013). More specifically, ES cells in 2i retained DNA methylation at ICRs, IAP elements and, to a lesser extent, at major satellite repeats and LINE1 elements. Hence, the DNA methylation profiles of genomic repeats and ICRs of ES cells cultivated in 2i is more similar to that of ICM cells, whereas DNA methylation levels in ES in Serum+LIF might reflect the levels observed in the late blastocyst. The similarity between ES cells in 2i and ICM cells of the primitive blastocyst was also confirmed by 5mC and 5hmC immunofluorescence (Habibi et al., 2013). Hypomethylation in 2i conditions correlated with reduced levels of *de novo* DNMTs, thought to be mediated, at least in part, by increased levels of PRDM14 (PR-domain containing protein), a transcriptional repressor of DNMT3B. The expression patterns of *de novo* DNMTs reflect those described for the ICM *in vivo* (Hirasawa et al., 2008; Hirasawa and Sasaki, 2009). Thus, "ground state" pluripotency, as

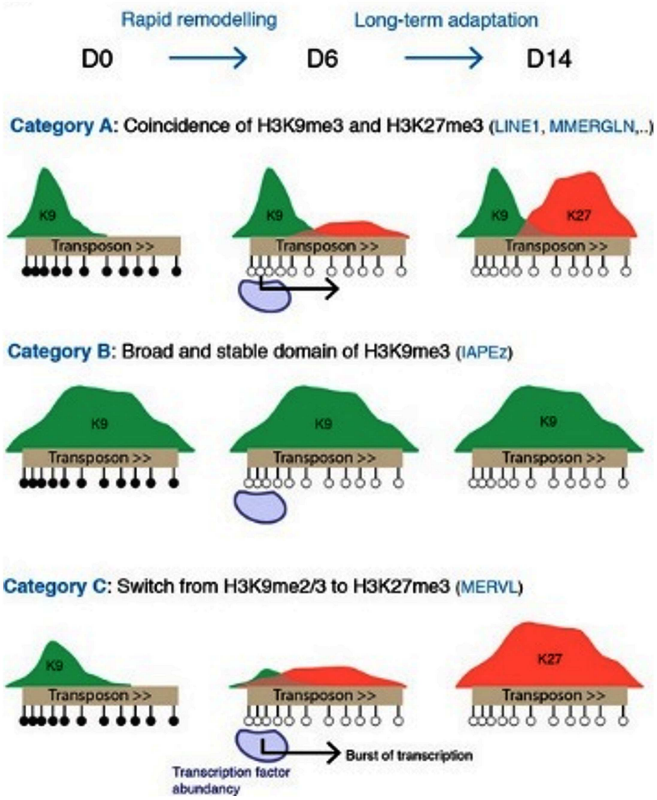


Figure 32. Model for chromatin changes at repetitive elements during genome-wide demethylation. *From Walter et al., 2016.*

recapitulated *in vitro* under 2i conditions, is likely associated with the protection from *de novo* methylation.

Recently, Blaschke and colleagues, testing new ES cells culture conditions containing vitamin C, observed the transcriptional expression of DAZL, a germline gene normally methylated and silent in mouse ES cells. This observation led the authors to test whether vitamin C could induce DNA demethylation. They added vitamin C to 2i medium and observed an increase in 5hmC levels. In order to distinguish between 5mC and 5hmC, they performed immunoprecipitation of 5mC (MeDIP) or 5hmC (hMeDIP) followed by deep-sequencing. The increase in 5hmC was genome-wide, affecting exons, introns and intergenic regions, and it was accompanied by reduced DNA methylation. Notably, the authors observed a global loss of DNA methylation at promoters of other germline genes like ASZ1, which is known to be methylated in 2i ES cells but unmethylated in ICM cells. In contrast, ICRs and IAP elements were resistant to the global demethylation induced by the vitamin C. The increase in 5hmC was rapid, within the first 12 hours of treatment, suggestive of an active mechanism. Vitamin C was then found to be a biochemical activator of TET enzymes, which are responsible for this increase in 5hmC concomitantly with a loss of DNA methylation (Blaschke et al., 2013).

The dynamics of DNA demethylation in the presence of vitamin C resembles the reprogramming of DNA methylation observed *in vivo* between fertilization and blastocyst stage (Smith et al., 2012), suggesting that this *in vitro* system could help to decipher the dynamics of DNA demethylation during early development. Last year, Walter and colleagues took advantage of this tool to study the mechanisms of demethylation and, more generally, the chromatin remodelling at repetitive elements, which is characterized by the presence of DNA methylation and H3K9me3 (as described in the section 4.7.3) (Walter et al., 2016). They confirmed the previously described loss of DNA methylation at repetitive elements occurring after fertilization, with the exception of IAP elements, and classified DNA repeats based on the chromatin changes that take place during DNA demethylation. In the first class (category A), which includes LINE1 elements, they observed that the loss of DNA methylation was associated with a loss of H3K9me2, but not H3K9me3, and replacement by H3K27me3. The second class (category B), that includes IAPEz, showed loss of DNA methylation and maintenance of H3K9me3, without increase in H3K27me3. In the last class (category C), containing MERVL, the loss of DNA methylation was associated with a loss of H3K9me3 and replacement by H3K27me3 on the total length of DNA repeats (**Figure 32**). These findings have important implications for understanding the molecular basis of transposon control in the pluripotent cells of the early mammalian embryo (Walter et al., 2016).

ES cells, cultivated in Serum+LIF or in 2i conditions, can be easily converted *in vitro* in EpiSCs, through adaptation to Activin A and FGF2 culture conditions (Buecker et al., 2014; Orkin and Hochedlinger, 2011; Schulz et al., 2014). Genome-wide analysis showed an increase in DNA methylation at the level of gene promoters. This suggests that a first wave of DNA methylation occurs between ES cells and EpiSCs. However, in the same study, Veillard and colleagues showed that *in vitro* converted EpiSCs presented higher levels of DNA methylation at promoters compared to EpiSCs directly isolated from E6.5 embryos (Veillard et al., 2014). In contrast, centromeric and pericentromeric repeats seemed to be methylated at the same levels in both contexts (Tosolini et al., *under review*). Interestingly, *in vitro* reversion of EpiSCs to ES cells is possible, although very inefficiently, again by switching culture conditions (Bao et al., 2009; Guo et al., 2009). To gain in efficiency, it is necessary to over-express pluripotency factors, like KLF4. In addition, it is necessary to overcome "epigenetic barriers" like X chromosome reactivation and DNA demethylation. This reversion is a crucial criterion to consider an efficient de-differentiation (Bao et al., 2009; Guo et al., 2009; Zhang et al., 2016). Therefore, this cellular system represents a unique tool and model to study the first waves of DNA methylation during development, as well as to identify epigenetic barriers to reprogramming, that could be informative for studies on the reprogramming of induced pluripotent stem cells (iPSCs).

Mouse ES cells can also be differentiated *in vitro*. A widely used method to induce the first steps of differentiation is through the formation of embryoid bodies (EBs) following various protocols (Kurosawa, 2007). Among those, the hanging drop technique (HD) promotes the formation of EBs that can be differentiated into a broad spectrum of cell types, including neuronal cells (He et al., 2006), cardiomyocytes (Takahashi et al., 2003), and even gametes (Geijsen et al., 2004). In general, during the kinetics of differentiation, the pluripotency genes are silenced within three days, while the tissue-specific expression programs are activated. For example, in the case of neuronal differentiation, neuro-specific genes like PAX6 and NESTIN are activated after around six days of differentiation (Mohn et al., 2008).

DNA methylation seemed to be dispensable to maintain mouse ES cells in culture, since mouse ES cells knock-out for DNMT1, DNMT3A or DNMT3B can be maintained in culture. Moreover, in the absence of all three enzymes (DNMT triple knock-out, TKO) and, although complete loss of DNA methylation is observed, mouse ES cells can be maintained in culture. Mouse ES TKO cells could be differentiated *in vitro*, although only towards Extraembryonic Lineages (Sakaue et al., 2010). The promotion of their differentiation towards progenitors of the three germ layers showed normal expression of molecular markers that characterize the three germ layers but high levels of cellular death through apoptosis. These findings suggest that DNA methylation plays a key role in the processes of differentiation. Mohn and colleagues focused on

neuronal differentiation to show that, starting from ES cells cultivated in Serum+LIF, the profiles of DNA methylation at promoters do not change prominently regardless their CpG content, with the exception of ICPs that gain DNA methylation during neuronal differentiation (Mohn et al., 2008). This increase in DNA methylation occurs in the first steps of differentiation, mainly at promoters of pluripotency factors, like NANOG, SOX2 and REX1, and at germline gene promoters, like PAPOLB and DMRTC7, involved in spermatogenesis and male meiosis, respectively (Kashiwabara et al., 2002; Kim et al., 2007; Mohn et al., 2008). Interestingly, DNA demethylation also occurs during neuronal differentiation. Although this is a rare event, occurring at 22 promoters out of the almost 15,000 analysed, the fact that it takes place at the promoter of brain-specific genes, like the gene coding for a neuron-specific transmembrane protein, LRRTM2, is highly interesting because it suggests that a tissue-specific DNA demethylation can occur during differentiation of somatic cells (Mohn et al., 2008).

One study analysed the dynamics of DNA methylation at repeated sequences during ES cells differentiation. Meissner and colleagues described an accumulation of DNA methylation at interspersed repetitive elements, LTR retrotransposons, LINE and SINE elements, showing that DNA methylation increases upon neuronal differentiation (Meissner et al., 2008). Unfortunately, the kinetics of DNA methylation was not performed at tandem repeats. Moreover, the results were obtained from undifferentiated mouse ES cells and in neural precursors, without taking into account a kinetics of differentiation. Such study would clarify if DNA methylation is established at the first steps of *in vitro* differentiation, *i.e.* commitment, as it is the case *in vivo* (Auclair et al., 2014).

In contrast to mouse ES cells, human Embryonic Stem cells (hESCs) knock-out for DNMT1 are not able to survive in culture (Li et al., 1992; Liao et al., 2015), and somatic cells die through a p53-dependent pathway (Jackson-Grusby et al., 2001). In fact, hESCs do not resemble mouse ES cells, but are similar to a “more differentiated” cell state, called epiblast (De Los Angeles et al., 2012). This feature may explain hESC lethality, suggesting that DNMT1 is essential for cell survival starting from epiblast stage in human cells (Liao et al., 2015).

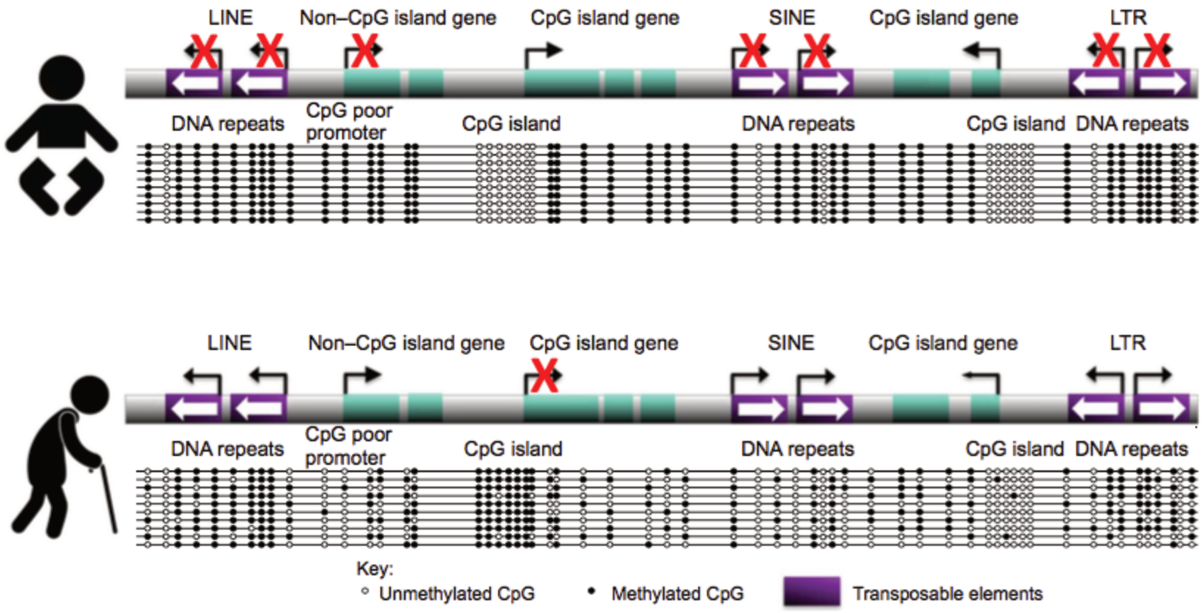


Figure 33. DNAmethylation changes during aging. Young mammalians are characterized by methylation of repetitive elements, like LINE, SINE and transposable elements [long-terminal repeats (LTR)]. In contrast, CGI are mostly unmethylated. During aging, there is a global loss of DNA methylation at DNA repeats, associated with increase in their transcription. Inversely, hypermethylation events are observed at CGI promoters. *From Pal and Tyler 2016.*

Chapter III: DNA methylation in physiopathology

1. DNA methylation in physiology

DNA methylation and its machinery play an essential role during human development and during the lifetime in general. However, the functioning of the machinery responsible for normal DNA methylation profiles has to face the impact of "the passage of time" and the adverse effect of the environment. Indeed, cells are constantly exposed to environmental or endogenous stress that may jeopardize their identity and viability. Agents that inflict damage to the cell's genome or epigenome, whether from UV light, chemicals in food or reactive molecules produced by metabolism, are particularly deleterious since they severely compromise their integrity. Hence, the ability of a cell to safeguard its genome integrity is fundamental for the survival and normal functioning of the whole organism and protection from development of diseases like cancer.

1.1. DNA methylation and aging

Aging is a complex multifactorial biological process shared by all living organisms. It is characterized by a gradual loss of physiological functions in a time-dependent manner (Pal and Tyler, 2016).

DNA methylation is one of the most studied and best-characterized epigenetic modification in the context of aging (Jung and Pfeifer, 2015). A first link between DNA methylation variation and aging came from studies performed in salmon and showed that the highest levels of DNA methylation were reached at the stage of the embryo, followed by a global decrease at later stages (Berdyshev et al., 1967). More recently, studies in rodents confirmed that a decrease in DNA methylation was associated with decreased lifespan and aging (Wilson et al., 1987). Likewise, genome-wide analysis performed on $CD4^+$ T cells from new-borns compared to centenarian individuals showed a loss of DNA methylation during aging, mainly at low density CpG promoters, gene bodies and intergenic regions (H. Heyn et al., 2012). However, the regions that were mostly hypomethylated seemed to be repetitive sequences (**Figure 33)**, particularly pericentromeric SatII repeats (Cruickshanks et al., 2013), in association with their abnormal transcription and a risk of retrotransposition leading to genomic instability (Bollati et al., 2009; Jintaridth and Mutirangura, 2010). Interestingly, this hypomethylation could contribute to the general loss of heterochromatin observed during aging (Tsurumi and Li, 2012).

The loss of DNA methylation, observed during aging, is thought to mainly result from defects in the methylation maintenance activity. Studies in $DNMT1^{+/-}$ mice showed that reduced activity of DNMT1 was associated with decreased DNA methylation levels. Moreover,

behavioural tests underlined a loss of biological functions, characterized by impairment in learning and memory functions in these mice in an age-related manner (Liu et al., 2011). In parallel to the global loss of DNA methylation in aging, hypermethylation events occur at CGI promoters, which are generally unmethylated in both human and murine somatic cells (Jung and Pfeifer, 2015).

Altogether, these results suggest that alterations of DNA methylation patterns can directly contribute to the gradual loss of biological functions observed in aged subjects, which could affect cellular functions in adult stem cell compartments and depletion of the pool of progenitors and impaired tissue homeostasis (Kuilman et al., 2010). This observation is associated with the fact that, in mammals, age-related degeneration gives rise to well-recognized pathologies, such as atherosclerosis and heart failure, osteoporosis, pulmonary insufficiency, renal failure, neurodegeneration, which includes prominent neurodegenerative diseases like Alzheimer's and Parkinson's diseases, and many others (Campisi, 2013). Moreover, in humans, there are also inherited disorders characterized by accelerated aging, or premature aging, which show only certain features of aging that are termed segmental progerias (Scheibye-Knudsen, 2016).

Aging has some similarities with cellular senescence, considered as a sign of aging but also as a safeguard mechanism against oncogenic events (Pal and Tyler, 2016). Senescence is generally defined as an irreversible cell cycle arrest in which cells are refractory to growth factor stimulation. Primary cells in culture initially proliferate but eventually enter a state of permanent growth arrest, called replicative senescence. In physiopathological context, oncogenic and mitogenic stimuli or DNA damage can trigger cellular senescence (Nakagawa and Opitz, 2007). *In vivo*, senescent cells are observed with increased frequency in aging skin for instance, and were found to contribute to the aging process. Therefore, cellular senescence, also defined as cellular aging, represents a powerful cellular model to study the molecular mechanisms of aging, although this conclusion is not commonly accepted by the scientific community (Campisi, 2013; Wagner et al., 2016).

Studies on this in culture system showed that also human senescent cells are characterized by a progressive loss of DNA methylation, in particular at heterochromatin regions (Hänzelmann et al., 2015). In addition, senescent cells showed local hypermethylation events that are better characterized compared to hypomethylation events. Hypermethylation occurs at genomic regions generally enriched in H3K4me1/3 or H3K27me3. Most of the hypermethylated gene promoters lead to transcriptional silencing. The genes linked to the hypermethylated promoters are mainly implicated in process of DNA repair, chromosome organization and DNA replication (Hänzelmann et al., 2015). Moreover, the hypermethylation events might also be associated with the formation of Senescence-Associated Heterochromatin Foci (SAHF). Indeed, Narita and colleagues described the formation of heterochromatin foci in senescent cells, enriched in H3K9me3 and HP1 protein and thought to contribute to cellular senescence (Narita et al., 2003).

The alterations of the DNA methylation profiles observed in senescent cells, including the hypomethylation of pericentromeric SatII repeats, are reminiscent of the alterations observed in cancer cells (Jung and Pfeifer, 2015). This observation suggested that cellular senescence may act as a barrier to cancer formation. Indeed, senescence is triggered by an oncogenic hit (Oncogene-Induced Senescence, or OIS), like the activation of an oncogene or the inactivation of a tumour suppressor gene. In this context, the activation of the p53 pathway was showed to protect from aberrant proliferation and to promote cell cycle arrest (Kuilman et al., 2010; Pal and Tyler, 2016). However, it was shown that senescent cells secrete many factors, including pro-inflammatory factors, which can have pro-oncogenic effects, promoting proliferation, invasion and cellular migration (Kuilman et al., 2010).

1.2. DNA methylation and environment

There is a considerable interest in studying whether and how environmental factors modulate the establishment and the maintenance of DNA methylation patterns, influencing gene expression and cellular phenotypes. Chemical pollutants, dietary components and other external stresses can, indeed, have an impact on chromatin organization in general and, in particular, on DNA methylation. The underlying mechanisms remain unknown, but mechanistic insights are emerging from studies in mouse models (Feil and Fraga, 2012). In the frame of this manuscript, an exhaustive description of the effects of the exposure to all kind of stresses would be very difficult and meaningless. For this reason, I decided to focus on a few examples belonging to everyday life. I will conclude this part on the impact of chemical drugs used for chemotherapy, like 5-azacytidine, on DNA methylation and the often-neglected consequences on normal cells.

1.2.1. Benzo[a]pyrene (BaP) and DNA methylation

Benzo[a]pyrene (BaP) is one of the many polycyclic aromatic hydrocarbons that are by-products of incomplete combustion (Tong et al., 2006). BaP is a ubiquitous environmental contaminant, with major sources including tobacco smoke, automobile exhaust, residential and commercial heating and even charcoal-grilled meat (Chappell et al., 2016; Miousse et al., 2015).

In vitro studies on human and mouse cells showed that exposure to BaP causes a global decrease in DNA methylation (Wilson and Jones, 1983, 1984; Wojciechowski and Meehan, 1984). Indeed, studies on human cell lines showed hypomethylation of repetitive elements like LINE1, SINE and LTR retrotransposons upon exposure to BaP. The transcription state of these repeats was assessed only for LINE1 elements, showing increased transcription (Sadikovic and Rodenhiser, 2006; Stribinskis, 2006). Moreover, the decrease in DNA methylation was showed to

A

Diet compound	Food source
Folate	Leafy vegetables, fruits, fortified cereal
B vitamins (B2, B6, B12)	Meat, nuts, various sources
Methionine	Dairy products, nuts, fish
Choline	Egg, milk, meat sources
Betaine	Spinach, beets, wheat
Phytoestrogen	Soy, legumes, cereal

B

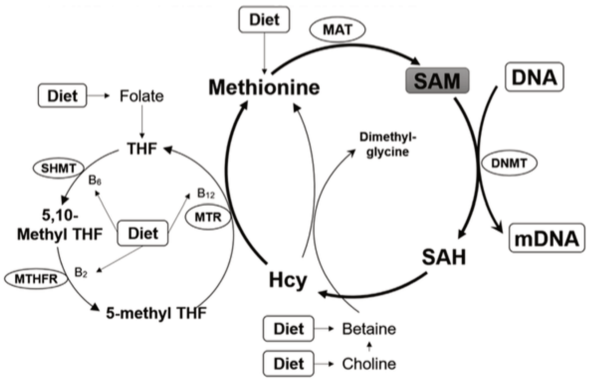


Figure 34. Diet compounds and DNA methylation patterns. A) Diet Compounds with a potential effect on DNA methylation and their food source. B) Possible ways that nutrition influences patterns of DNA methylation (*Adapted from McKay and Mathers 2011*). MAT = methionine adenosyltransferase; SAM = S-adenosyl methionine; SHMT = serine hydroxymethyltransferase; THF = tetrahydrofolate; DNMT = DNA methyltransferase; MTR = 5-methyltetragydrofolate-homo- cysteine methyltransferase; MTHFR = methylentetrahydrofolate reductase; Hcy = homocysteine; SAH = S-adenosylhomocysteine; mDNA = methylated DNA. *Adapted from Cheng and Xu 2010; Zhang 2015.*

promote the enrichment in activating histone marks, like H3K4me3 and H3K9Ac, which may lead to an increase in transcription and possible retrotransposition (Teneng et al., 2011). A possible explanation for this global hypomethylation could be the inhibition of the catalytic activity of DNMT1, which, as a consequence, would be unable to transfer methyl-groups from SAM substrate to DNA. This inhibition was indeed reported by *in vitro* studies, proposing a direct inhibition of DNMTs by BaP-DNA adducts (Wojciechowski and Meehan, 1984). This observation is particularly interesting because, as I will describe in the next chapter, global hypomethylation, in particular at genomic repeats, is a peculiar feature of cancer cells (Ehrlich, 2009). Other studies performed on mice showed that hypomethylation events could also occur at the level of gene promoters, including that of the pluripotency factor NANOG and MAPK3 (ERK1), which were described to promote cancer formation or progression (Tommasi et al., 2014; Wang et al., 2013).

Studies performed directly on blood samples of people exposed to high doses of BaP also showed hypermethylation events. Indeed, the promoter of INFγ gene was hypermethylated in blood samples of women exposed to high doses of hydrocarbons, and the same hypermethylation was found for the promoter of DUSP22 (Dual Specificity Phosphatase 22) in blood samples of fire fighters (Ouyang et al., 2012; Tang et al., 2012). Interestingly, these genes are found hypermethylated and down-regulated in several cancer types (Chappell et al., 2016). Moreover, hypermethylation may increase BaP-induced damage. Indeed, metabolic intermediates were shown to preferentially bind to highly methylated regions, causing conformation changes of the double-helix that may induce DNA damage through interference with the mechanisms of DNA replication and DNA repair. The promoter of the tumor suppressor p53 was found among the regions that are targeted by these events of hypermethylation (Chen et al., 1998; Dong et al., 2004; Tretyakova et al., 2008; Weisenberger and Romano, 1999).

These findings may suggest that overexposure to BaP could modify the DNA methylation patterns, which may predispose to the onset of diseases like cancer.

1.2.2. Alimentation and DNA methylation: the case of folate

Dietary nutrients not only play a direct role in maintaining the epigenome that safeguards normal cell differentiation and growth, but they also contribute to counteract the insults caused by environmental toxicants on the human epigenome. Indeed, there are several examples of compounds present in different sources of food shown to have a positive impact on DNA methylation (Chen and Xu, 2010) (summarised in **Figure 34A**). All these compounds mainly contribute to the productions of SAM, the donor of methyl-groups used by all the DNMTs and KMTs (Feil and Fraga, 2012) (**Figure 34B**).

Among all these compounds, folate is one of the most studied in the field of DNA methylation. Low folate diet is associated with global DNA hypomethylation and with several diseases, including cardiovascular diseases, multiple cancers and neural tube defects. However, the underlying mechanisms are far from being clarified. The most accepted hypothesis is based on the decrease of intracellular SAM causing a decrease in the activity of DNMT enzymes, leading to gradual and global DNA hypomethylation (Crider et al., 2012).

A very famous study in mice showed that a high folate diet can correct defects in DNA methylation profiles. The viable yellow agouti (Avy) mouse model, the colour of the coat depends on the expression of the Agouti gene, which is regulated by DNA methylation. Dolinoy and colleagues showed that pregnant mice exposed to bisphenol-A (BPA), an environmental endocrine disruptor found in plastics, have reduced DNA methylation at the Agouti gene promoter that shifts the coat colour of the offspring towards yellow. Maternal dietary supplementation with folic acid counteracts the hypomethylation caused by BPA exposure, restoring the normal coat colour of the offspring (Dolinoy et al., 2007). In two separate studies on healthy postmenopausal women, folate depletion resulted in hypomethylation in lymphocyte DNA, which can be reversed following folate repletion (Jacob et al., 1998; Rampersaud et al., 2000), suggesting that DNA methylation can be restored by folate diet also in humans.

Folate supplementation is also used in cancer therapy. As I mentioned before, cancer cells are characterized by a global hypomethylation, mainly affecting DNA repeats (Robertson, 2005). Folate is used to treat gliomas and causes the induction of SP1/SP3 proteins, which favour the expression of the DNMT3A and DNMT3B. This increase in the expression of the *de novo* DNMTs is associated with an increase in DNA methylation at repetitive elements, in particular at SINEs. This “re”-methylation correlates with a decrease in cancer cell proliferation and an increased sensibility to anti-tumour treatments (Hervouet et al., 2009).

All these findings support the famous slogan “5 fruits and vegetables per day”. However, it was proposed that excess in folate intake may also be associates with an increased risk to develop cancer, more precisely colorectal cancer (Kim, 2006). Other studies support the correlation between high plasmatic levels of folate and cancer, including prostate cancer and breast cancer (Smith et al., 2008). Although there is no precision of an optimal dose of folic acid or of its blood concentration, these findings suggest that a high folate-diet is far from being the solution for cancer treatment and that the precise role of folate in cancer prevention needs to be carefully studied and evaluated.

1.2.3. Chemotherapy: 5-azacytidine, what about normal cells?

Chemotherapy differs from the cases described above as it is rarely accidental. A genotoxic stress is "programmed" to inhibit cancer progression and promote the death of cancer cells, or to promote a higher sensitivity to other anti-tumour agents. However, in most of the cases, normal cells do not escape these increased doses of stress.

Drugs that specifically inhibit DNA methylation were discovered and synthesised almost 50 years ago (Christman, 2002; Issa, 2007). Cytosine analogues were developed as cytotoxic anti-cancer agents in the 1960's and tested in the clinics in the 1970's (Issa and Kantarjian, 2009). 5-azacytidine (AZA) is an analogue of cytosine, which can be incorporated into DNA, where it traps DNA methyltransferases and targets these enzymes for degradation (Jones and Taylor, 1980; Yoo and Jones, 2006). The absence of these enzymes then results in hypomethylation of the genome of the daughter cells and, eventually, to the reactivation of silenced gene expression, like that of tumour-suppressors, which are known to be silenced by hypermethylation of their promoter in cancer cells ((Ehrlich, 2002); see section DNA methylation and cancer). Early clinical trials in the 1970's and 1980's were based on the presumed cytotoxic effects of this drug. Tested mostly at very high doses for relatively short exposure times, the drug was found to be sporadically effective, more in hematologic malignancies than in solid tumours, but also quite toxic (Santini et al., 2001). After several pre-clinical trials, AZA was approved by the Food and Drug Administration (FDA) in 2004 and the drug called Vidaza. The effects of AZA, assessed by *in vitro* experiments, are quite non-specific on the genome. Moreover, its impact on cellular process may vary from the induction of differentiation, like muscle, adipocytes, and chondrocytes from AZA-treated primary MEFs (Taylor, 1993), of safeguard mechanism like apoptosis or senescence but also processes deleterious for tissue homeostasis like self-renewal and even invasion (Issa and Kantarjian, 2009).

In fact, the action of AZA varies with cellular proliferation. Indeed, DNA synthesis is needed to mediate incorporation of AZA into the double helix of DNA (Ren et al., 2011). Although, it is known that cancer cells proliferate faster compared to normal cells (Yang et al., 2011), and because of its non-targeted intravenous administration, there are pools of normal proliferating cells, in particular multipotent progenitors in the bone marrow (Friedenstein et al., 1974), which could also be affected (Ren et al., 2011). Although AZA-treated primary cells do not seem to show a general increase in gene transcription (Schmelz et al., 2004), a global hypomethylation can be induced at non-coding regions, like tandem or interspersed DNA repeats. For example, AZA-induced hypomethylation of tandem repeats is associated with an increase in mitotic recombination (this mechanism will be better explained in the section 2.1.3 of this chapter), as well as with accumulation of centromeric and pericentromeric repeats throughout the cell cycle (Bouzinba-Segard et al., 2006). This aberrant accumulation correlates with defects in chromosome

segregation and mislocalization of centromere-associated proteins that are essential for centromere function (Bouzinba-Segard et al., 2006) (this mechanism will be better explained in 2.1.3). Cytotoxic effects were, indeed, described mostly at high doses, including myelosuppression with neutropenic fever (Issa and Byrd, 2005). We cannot exclude that the treatment of normal cells with AZA may lead to global hypomethylation and promote new oncogenic events.

Although this compound is very promising for cancer treatment, in particular for myelodysplastic syndromes (Raj and Mufti, 2006), the associated cytotoxic effects highlight the necessity for cautious interpretation of laboratory and clinical data. Further exploration may provide an opportunity to improve the efficacy and reduce the side effects of this drug in a clinical setting.

2. DNA methylation in pathology

Given the pivotal role played by DNA methylation during development to regulate genomic output, and the impact that environment can have on DNA methylation profiles, it comes as no surprise that aberrant patterns of methylation are associated with many human diseases.

In the recent years, the interest in the relationship between DNA methylation and human diseases has increased. Although our knowledge significantly increased in the last years, the causal relationship between DNA methylation and diseases still needs clarification. Human diseases represent the best field to study the impact of alterations in the expression of genes coding for factors of the DNA methylation machinery on development and on the correct homeostasis and functioning of tissues and organs. The different diseases that will be described below can be divided in two categories: (i) diseases presenting somatic alterations, deregulated expression or alteration of splicing of their RNA precursors, and (ii) genetic diseases presenting germline mutations in the DNA methylation machinery genes.

In the first case, sporadic mutations/alterations may arise in a particular tissue or cell population contributing to drive disease or to favour its progression. However, whether these alterations and the associated anomalies in DNA methylation patterns are the cause or the consequence of the disease state is still a debated question. Autoimmune diseases and cancer are the main classes of disorders belonging to this first group. In the second case, mutations occur in the germ line and are transmitted to the progeny. These diseases represent unique occasions to study the role of one particular player in human development and in the onset of aberrant phenotypes. In this group, a particular attention will be paid to a rare recessive disease called ICF (Immunodeficiency Centromeric instability and Facial anomalies) syndrome.

2.1. Somatic alterations

2.1.1. Autoimmune diseases

Autoimmunity is characterized by the failure of an organism to recognize what is proper to the organism (self), raising an immune response against its own cells and tissues, primarily through the action of T- and B-lymphocytes (Hirst and Marra, 2009). Robertson group, in the mid-1980s, observed that the addition of AZA to human T-lymphocytes caused their temporary activation in the absence of antigen (Robertson, 2005). This was the first example of a possible impact of alterations to DNA methylation profiles in autoimmune diseases.

The link between DNA methylation and autoimmunity is best described in the context of the systemic lupus erythematosus (SLE) (Rahman and Isenberg, 2008). SLE is a chronic autoimmune condition that can virtually affect any organ. It is characterized by the production of auto-antibodies targeted to a variety of nuclear antigens and by a global hypomethylation profile (Jeffries and Sawalha, 2015).

Javierre and colleagues showed that DNMT1 and DNMT3B are down-regulated in SLE samples. This can provide an explication for the global hypomethylation observed in SLE. Moreover, this result suggests that the defects in DNA methylation may be associated with impaired *de novo* and maintenance activities (Javierre et al., 2010). Although the mechanisms leading to this decrease in DNMT3B expression still remains to be unravelled, Sunahori and colleagues proposed a mechanism to explain DNMT1 down-regulation. They observed that the phosphatase PP2Ac is highly expressed in SLE lymphocytes. One of the targets of this phosphatase is ERK kinase, which is degraded upon dephosphorylation. The loss of ERK causes in turn the down-regulation of DNMT1, contributing to the global loss of DNA methylation (Sunahori et al., 2013).

Different groups performed genome-wide analysis in naïve CD4+ T-cells of SLE patients. In all these analyses, the majority of the DMR were hypomethylated (Coit et al., 2013; Javierre et al., 2010). Surprisingly, hypomethylation of repetitive elements was not observed, which differs from results obtained in other diseases with altered expression of DNMT1 or DNMT3B, like cancer (Javierre et al., 2010). DNA methylation can also contribute indirectly to the clinical manifestations of patients. Indeed, Coit and colleagues observed hypomethylation of genes coding for interferons, like IFIT1 and 3, and their target genes, which are implicated in the proliferation and activation of T-lymphocytes. These results contributed to understand the hyper-responsiveness to type-I interferon described in SLE patients, which constantly activate an anti-viral reaction (Coit et al., 2013; Crow, 2014; Rahman and Isenberg, 2008). The rare hypermethylation events included two genes directly involved in antigen recognition and lymphocytes activation, CD247 and IL21R,

which were previously reported to be down-regulated in SLE patients lymphocytes (Coit et al., 2013). However, the causal relationship between anomalies in DNA methylation patterns, caused by the down-regulation of DNMT3B and DNMT1, and the onset of the disease is far to be clear.

2.1.2. Cancer

One of the first reports to establish a clear connection between DNA methylation and cancer came from Feinberg and Vogelstein in 1983, who found extensive hypomethylation in primary tumour samples in comparison with the corresponding normal tissue (Feinberg and Vogelstein, 1983). Such reports supported the idea that the cancer genome is globally hypomethylated. Hypomethylation of highly repeated DNA sequences, most notably tandem centromeric Satα, juxtacentromeric (pericentromeric) SatII, interspersed Alu, and LINE1 repeats, largely accounts for this global DNA hypomethylation (Ehrlich, 2009). Interestingly, a bisulfite sequencing study in colorectal tumours reported that this DNA hypomethylation in human cell lines is non-randomly distributed, and is mainly concentrated at sites of DNA attachment to the nuclear lamina (Liu et al., 2012) known to be late replicating and transcriptionally repressed in somatic cells. Examples of promoter-specific hypomethylation also exist, most notably hypomethylation of germ line gene promoters in various cancer cell lines (De Smet et al., 1999) that directly leads to their de-repression. These genes have also been referred to as "Cancer Testis" (CT) genes (Scanlan et al., 2002). The expression of germ line genes in cancer is thought to confer phenotypic traits, including immortality, invasiveness, and immune evasion that are essential for the survival and function of gametes but are hijacked in somatic cells to drive malignancy (Wang et al., 2011). In line with these ideas, their ectopic expression has been shown to drive malignant brain tumour growth in *Drosophila melanogaster* (Janic et al., 2010).

In contrast to global hypomethylation, hypermethylation in cancer cells occurs locally (Robertson, 2005). In 1989, Greger and colleagues reported hypermethylation of the Retinoblastoma gene (RB) in cells from patients with an eye cancer that begins in the retina, called retinoblastoma, suggesting a direct link between DNA methylation and silencing of this tumour suppressor (Greger et al., 1989). Hypermethylation occurs mainly at CGI, most of which are unmethylated in normal cells. Since the majority of CGI overlaps with gene promoters, their hypermethylation favours the silencing of the linked gene (Ehrlich, 2002; Issa, 2004). These genes are mainly involved in cell cycle control, DNA repair and apoptosis, and include p53, the aforementioned RB, BRCA1 (Breast Cancer susceptibility 1) and PTEN (Grønbaek et al., 2007). These events provide a growth advantage to tumour cells. Indeed, these proteins are known to control cell-cycle progression, to avoid uncontrolled proliferation (Hartwell and Kastan, 1994). Unchecked cell divisions can lead to increased genetic instability and can favour the formation of

DNMT enzymes	Alteration	Tumour type	Model studied
DNMT1	**Mutations**	Colon	Patients
	Overexpression	Acute Myeloid Leukemia, Chronic Myeiloid Leukemia, Breast, Liver, Pancreas, Esophagus	Patients
	Deletion	Lymphoma, Breast	Mouse models
DNMT3A	**Mutations**	Acute Myeloid Leukemia, Myelodysplastic Syndromes, Chronic Myelomonocytic Leukemia, Acute Lymphoblastic Leukemia	Patients, Mouse models
	Overexpression	Acute Myeloid Leukemia, Chronic Myeiloid Leukemia, Liver, Melanoma	Patients, Mouse models
	Deletion	Acute Myeloid Leukemia, Lung, Lymphoma	Mouse models
DNMT3B	**Overexpression**	Acute Myeloid Leukemia, Chronic Myeiloid Leukemia, Breast, Colon, Prostate	Patients, Mouse models
	Deletion	Acute Myeloid Leukemia, Lymphoma,	Mouse models
	Splice variants	Hepatocellular Carcinomas, Chronic Myeiloid Leukemia, Colon	Patients, Mouse models
TET enzymes			
TET1	**Mutations**	Endometroid Carcinoma, Acute Myeloid Leukemia, Chronic Myeiloid Leukemia, Chronic Lymphocytic Leukemia, Colorectal Carcinoma, Lung, Bladder, Lymphoma, Breast, Central Nervous System*, Kidney, Large Intestine, Liver*, Lung, Ovary*, Prostate, Skin, Stomach, Urinary tract*	Patients
	Deletions	Lymphoma	Mouse models
TET2	**Mutations**	Endometroid Carcinoma, Colorectal Carcinoma, Melanoma, Lung, Myeloproliferative Disorder	Patients
	Deletions	Lymphoma, Chronic myelomonocytic leukemia-like	Mouse models
TET3	**Mutations**	Endometroid Carcinoma, Chronic Lymphocytic Leukemia, Peripheral T-Cell Lymphoma, Acute Lymphocytic Leukemia, Breast*, Centroal Nervus System*, Kidney*, Large intestine, Liver, Lung, Ovary, Pancreas*, Prostate, Skin Stomach, Urinary Tract	
5mC and 5hmC readers			
MECP2	**Overexpression**	Breast	Patients
MBD1	**Mutations**	Lung, Endometrial Carcinoma, Colorectal Carcinoma, Breat, Melanoma	Patients
	Overexpression	Pancreas, Prostate	Patients, Mouse models
MBD2	**Mutations**	Colorectal Carcinoma, Breast	Patients
	Overepxression	Gastrointestinal Carcinoma, Breast, Glioblastoma	Patients, Mouse models
	Down-regulation	Prostate, Gastrointestinal Carcinoma	Patients
MBD3	**Overexpression**	Glioblastoma	Patients
	Down-regulation	Gastrointestinal Carcinoma	Patients, Mouse models
MBD4	**Down-regulation**	Lung, Endometrial Carcinoma, Colorectal Carcinoma, Breast, Melanoma	Patients, Mouse models
MBD5	**Mutations**	Lung, Bladder, Skin, Stomach, Colorectal Carcinoma, Uterine	Patients
MBD6	**Mutations**	Nerve, Cervical, Bladder, Stomach, Colorectal Carcinoma, Pancreatic, Uterine	Patients
UHRF1	**Mutations**	Lung, Endometrial Carcinoma, Colorectal Carcinoma	Patients

Table 5. Alterations of DNA methylation machinery factors in cancer. DNA methylation machinery factors are divided in functional groups: DNMT enzyme, TET enzymes and 5-methylcytosine and 5-hydroxymethylation readers. For each factor the alteration is indicated as well as the cancer in which the alteration was observed and the model system. Asterisks correspond to data from COSMIC database (http://cancer.sanger.ac.uk/cosmic). Some of the listed mutations have not been confirmed as somatic mutations. *Adapted from: Scourzic et al., 2015; Du et al., 2015; Rasmussen and Helin 2016; Zhang and Xu 2017.*

metastasis (Robertson, 2005). In tumours with a well-defined progression, such as colon cancer, aberrant hypermethylation is detectable in the earliest precursor lesions, indicating that it directly contributes to transformation and is not a late event that arises from genetic alterations, at least in this precise kind of tumour (Chan et al., 2002).

In agreement with aberrant changes to DNA methylation, DNMTs are frequently found to be deregulated in cancers (Jin et al., 2005; Park et al., 2006; Robertson et al., 1999; Schmidt et al., 2007). However, recent studies showed that alterations were not only detectable at the level of DNMTs, but at the level of the whole DNA methylation machinery (**Table 5**). A wide range of alterations are possible, including overexpression, down-regulation, mutations, deletions and expression of splice variants (Du et al., 2015; Rasmussen and Helin, 2016; Scourzic et al., 2015).

The general trend appears to be towards over-expression of DNMTs, which could explain the hypermethylation of tumour suppressor genes (Esteller, 2008). Moreover, knock-down of DNMT1 in oesophageal squamous cell carcinoma has a negative effect on the proliferation of tumour cells, suggesting that DNMT1 is required for the progression of cancer cells (Zhao et al., 2011). Somatic mutations in DNMT1 and DNMT3A coding genes are also a prominent feature of many tumours, contributing to the malignant transformation (Baylin and Jones, 2011). Interestingly, up to now, no mutation in the DNMT3B gene has been reported in tumours. Mutations in the catalytic domain or causing the deletion of this domain of DNMT1 and DNMT3A were shown to decrease the activity of the enzymes and to contribute to the global hypomethylation observed in cancer cells. Mutations in DNMT1 were described in colorectal cancer and in DNMT3A in leukaemia (Kanai et al., 2003; Yamashita et al., 2010). Mutations in DNMT1 consist mostly in one-base deletions upstream the catalytic domain, leading to the formation of a premature stop-codon and deletion of the entire catalytic domain (Kanai et al., 2003). The most common mutation in DNMT3A occurs in the catalytic domain of the protein ($DNMT3A^{R882H}$), reducing its activity down to 80% *in vitro*. A recent study showed that, in this case, the hypomethylation caused by the mutation precedes initiation of leukaemia, whereas hypermethylation events are mostly the consequence of cancer formation, suggesting that the cause/consequence link between mutation, changes in DNA methylation profiles and cancer initiation/progression has no easy answer (Spencer et al., 2017).

Cancer formation caused by the deletion of DNMT genes was addressed in mouse models. Interestingly, the lack of one of these enzymes in adult mice is mostly associated with lymphomas, suggesting that DNMT enzymes may act as tumour-suppressors in lymphocytes (Zhang and Xu, 2017). Moreover, the dosage of DNMT3B was observed to alter tumorigenesis. Vasanthakumar and colleagues bred $DNMT3B^{+/-}$ with a mouse model susceptible to B-cell lymphomas and

observed a dramatic acceleration of lymphomagenesis. This finding indicates that DNMT3B can act as a haploinsufficient tumor suppressor gene (Vasanthakumar et al., 2013).

Other alterations of DNMTs expression or integrity in cancer were found at the level of splicing. Aberrant splicing can induce the increase of variants expressed in normal cells, as well as driving the expression of aberrant splice variants encoding truncated proteins. This alteration seems to be a peculiarity of DNMT3B. Indeed, Godley and colleagues documented over 20 DNMT3B isoforms from different cancer cell lines and primary acute leukaemia cells that were aberrantly spliced at the 5' end of the gene, encoding truncated proteins lacking the C-terminal catalytic domain (Ostler et al., 2007). Ectopic expression of DNMT3B7, the most commonly expressed aberrant transcript in tumours, resulted in changes in the expression of fifty-one genes in 293T cells, including the overexpression of the CT gene MAGEA3 that was accompanied by the hypomethylation of its promoter. Two studies have also correlated the ectopic expression of truncated, catalytically inactive forms of DNMT3B, with the hypomethylation of juxtacentromeric SatII repeats, suggesting that these splice variants may impair normal enzyme function in a dominant negative fashion (Gopalakrishnan et al., 2009; Saito et al., 2002). More recently, the catalytically inactive splice variants DNMT3B3 and DNMT3B4, which are highly expressed in different cancer cell lines and in hepatocarcinoma cells respectively, were found to interact with catalytically active DNMT3A and DNMT3B enzymes as well as with DNMT3L. DNMT3B4 is a dominant-inhibitor of the catalytically active DNMT3A and DNMT3B enzymes whereas, in contrast, DNMT3B3 functions more like a weak-activator. These findings suggest that hypomethylation or hypermethylation depending on the DNMT3B splice variants that are expressed. However, the mechanisms leading to the formation of this splice variants and their role in tumorigenesis are still under investigation (Gordon et al., 2013).

Inactivation of TET enzymes, either by genomic deletions in mouse models or by somatic mutations in patients, were reported to be associated mostly with myeloid and lymphoid malignancies. Mutations inactivating TET3 seem to be less frequent in these cancer types, probably because of the low expression level of this enzyme in haematopoiesis compared to TET1 and TET2. Mutations in TET1 are mostly missense mutations but the functional consequences still need to be elucidated (Scourzic et al., 2015). The most frequently mutated gene of the TET family is TET2. Mutations targeting the catalytic domain, localized in the C-terminal part of the protein, were predicted to impair its function. Mutations can occur also at the N-terminal part of the protein, proposed to affect the function of the protein, although the underlying mechanism is still unknown. In myeloid cancers, heterozygous mutations in TET2 were observed, suggesting that partial loss of the protein may contribute to cellular transformation (Delhommeau et al., 2009). Moreover, in these cancers, mutations in TET2 are associated with decrease in 5hmC and increase

in 5mC levels, therefore promoting hypermethylation events (Ko et al., 2010). These events correlate with defects in myeloid differentiation, promoting myelodysplastic features and an expansion of the monocytic compartment. All together, these results suggest that TET2 may play a role in the pathogenesis of myeloid malignancies through the disturbance of myeloid differentiation (Pronier et al., 2011). Further studies will be required to understand whether mutations in TET2, or other TET enzymes, can drive tumorigenesis and which are the underlying mechanisms. Interestingly, mutations in TET2 coding gene are closely associated with mutations in DNMT3A coding gene in hematopoietic malignancies (Couronné et al., 2012).

Cancer is not only characterized by alterations in writing and erasing DNA methylation, but it is also associated with alterations in the reading of his mark. Among the genes coding for methyl-binding proteins, MBDs are the most frequently mutated in cancer. Colorectal and pancreatic cancers showed the highest total frequency of mutations in MDB genes, together with bladder and lung cancers, showing a mutation frequency over 10% (Du et al., 2015). Non-synonymous mutations, insertions, deletions, frame-shift mutations and mismatch mutations, may potentially lead to disrupted or aberrant binding of the MBDs, leading to mislocalization of their partners on the genome. Surprisingly, few mutations occur in the methyl-binding domain, while the majority falls outside this domain, suggesting an important role of these regions to maintain MBD functions. For example, mutations in the Intrinsically Disordered Region (IDR) of MBD2 gene were found to decrease its affinity for methylated DNA as well as the recruitment of the NuRD complex at methylated regions, therefore disrupting MBD2 functions (Desai et al., 2015). The deregulated expression of MBD proteins was also associated with tumorigenesis. Indeed, overexpression of MBD1 was associated with invasion and metastasis of pancreatic cancer (Xu et al., 2013). Conversely, the down-regulation of MBD2 and MBD3 was reported in gastric neoplastic tissues, but the underlying mechanism is still under investigation (Pontes et al., 2014).

Despite the wide “collection” of mutations in genes coding for writers, erasers and readers of DNA methylation, the hallmark of all kinds of cancer cells is the genome-wide hypomethylation, that most prominently affects repeated sequences (Ehrlich, 2009). Ting and colleagues observed an association between such global hypomethylation that affect cancer cells and a striking increase in the levels of transcripts of DNA repeats, mainly centromeric and pericentromeric tandem repeats, in a variety of human and mouse cancers, with the highest levels positively correlating with tumour grade (Ting et al., 2011). Interestingly, other non-coding RNAs emerge as candidate players in cancer development and progression (Cheetham et al., 2013). However, the causality relationship between hypomethylation, transcription of centromeric and pericentromeric repeats and aberrant phenotypes has not been clearly established. How these

sequences are transcribed and whether they act as driving force in cancer formation are important questions to answer.

2.1.3. Stress: a model to study the causal link between DNA methylation, transcription of pericentromeric and centromeric repeats and perturbed phenotypes

Transcripts originating from centromeric and pericentromeric regions have been described in mouse and humans, as well as in a broad range of organisms (Vourc'h and Biamonti, 2011). The accumulation of these transcripts in physiological contexts is cell-cycle regulated and the presence of these transcripts is crucial for the organization of centromeric chromatin and pericentromeric heterochromatin (Ferri et al., 2009; Lu and Gilbert, 2007) (*Cf* with **Chapter II**, 3.3). Interestingly, the accumulation of these transcripts throughout the cell-cycle is not only a feature of cancer cells, but can be induced by a large variety of stresses. Cellular stress can be induced by exposing cells to any environment outside of their optimal growth conditions, including high temperatures (heat shock), heavy metal, hazardous chemicals, ultra-violet radiation, hyperosmotic stress, and oxidative stress that may lead to DNA damage. Although originally described in response to heat shock (Jolly et al., 2004; Rizzi et al., 2004), the accumulation of human SatIII has been described for all of these conditions (Valgardsdottir et al., 2008). Although transcription primarily takes place at the pericentromeric heterochromatin blocks of the 9q12 locus, enriched in SatIII sequences, secondary regions enriched in SatIII sequences such as 1, 16, Y and acrocentric chromosomes are also transcribed. The role of non-coding SatIII transcripts is still unknown. Once synthesized, SatIII transcripts remain associated with the pericentromeric loci from which they originate. In human cells, activation of pericentromeric sequences at the 9q12 locus correlates with the formation of nuclear stress bodies enriched in specific transcription factors, including Heat Shock Factors (Vourc'h and Biamonti, 2011) and acetylated histones. As I mentioned in the previous parts of this manuscript, pericentromeric regions are characterized by high levels of DNA methylation (**Chapter II**). Interestingly, accumulation of pericentromeric transcripts upon heat shock was observed to occur independently of DNA demethylation (Eymery et al., 2009). This result suggests that accumulation of pericentromeric repeats may not be consequence of DNA hypomethylation, but, inversely, it may constitute a cause. Indeed, we cannot exclude that the transcription of pericentromeric satellite repeats may induce the loss of pericentromeric heterochromatin marks, including DNA methylation.

In the mouse, cellular stressed induced either upon treatment with the demethylating agent AZA as well as upon genotoxic stress, which can induce the formation of DNA double-strand breaks, was observed to induce an accumulation of centromeric transcripts (Bouzinba-Segard et

Gene	Cytogenetic location	Function	Disease	OMIM[1]	Identification
DNMT3A	2p23.3	DNMT	Overgrowth syndrome with intellectual disability (Tatton-Brown–Rahman syndrome)	615879	Tatton-Brown et al., 2014
DNMT1	19p13.2	DNMT	Hereditary sensory and autosomic neuropathy type 1 (HSAN1)	614116	Klein et al., 2011
MECP2	Xq28	MBD	Angelman syndrome	105830	Watson et al., 2001
MECP2	Xq28	MBD	Rett syndrome	312750	Amir et al., 1999
DNMT3B	20q11.21	DNMT	Immunodeficiency Centromeric instability and Facial anomalies 1 (ICF1)	242860	Xu et al., 1999

Table 6. List of genetic disorders caused by mutations in DNA methylation machinery genes. Identification in Online Mendelian Inheritance in Man (http://www.ncbi.nlm.nih.gov/gene). *Adapted from Berdasco and Esteller 2013.*

al., 2006; Hédouin et al., 2017). The forced accumulation of these repeats in mouse cell lines was associated with deleterious events. Indeed, the accumulation of centromeric transcripts throughout the cell cycle correlates with delocalization of CENP-A away from the centromere (Hédouin et al., 2017), and with impairment of mitosis and mislocalization of pericentromeric heterochromatin proteins (Bouzinba-Segard et al., 2006). All these results suggest that accumulation of centromeric repeats is associated with chromosomal instability, aneuploidy and centromere disorganization, which are all features of cancer cells. Moreover, we observed that accumulation of centromeric repeats and delocalisation of CENP-A were associated with senescence entry of mouse primary cells in a p53-dependent manner. In absence of p53, cells do not arrest, keeping high levels of centromeric transcripts and showing growing chromosomal instability (Hédouin et al., 2017). These results suggest that accumulation of centromeric transcripts may constitute an oncogenic hint to induce cellular senescence, with a mechanism resembling to the oncogene-induced senescence, described in the previous sections (*Cf* with 1.1 of this chapter). All together, these findings support the hypothesis that the accumulation of centromeric transcripts may constitute a cause of cancer formation, more than a consequence.

As a whole, studies on cellular stress, in human and mouse cells, suggest a role of pericentromeric and centromeric transcripts as putative driver of cancer formation. Moreover, cellular stress is allowing to better characterize the causal relationship between DNA methylation of pericentromeric and centromeric repeats, accumulation of transcripts originated from these regions and perturbed phenotype, which are all hallmarks of cancer cells.

2.2. Germline mutations

Germline mutations are known to cause genetic diseases. In contrast to somatic mutations, which arise in one particular tissue or cell population, germline mutations characterize all the cells of patients. Up to now, germline mutations in genes coding for DNA methylation writers and the reader MECP2 have been described to cause particular diseases (**Table 6**). Interestingly, all these diseases are associated with intellectual disability, confirming the crucial role of DNMTs and DNA methylation in neuronal differentiation and neuronal homeostasis (Feng et al., 2010; Watanabe et al., 2006). A particular attention will be paid to the ICF syndrome, which has been the model of study of my thesis after a description of the other genetic diseases.

2.2.1. Mutations in DNMT3A cause an "Overgrowth syndrome with intellectual disability"

The identification of genetic diseases caused by mutations in DNMT3A gene is the most recent, compared to all the other diseases described in this chapter. Exome-sequencing allowed the identification of *de novo* mutations in DNMT3A in eleven patients of a cohort of patients characterized by overgrowth (Tatton-Brown et al., 2014), and six additional cases were subsequently described to have inherited mutations in the same gene (Xin et al., 2017). All patients present dysmorphic facial features, increased height, intellectual disability, and variable additional features. This is an autosomal dominant disease. Moreover, only two of these mutations were also described in acute myeloid leukaemia, whereas all the other mutations described were not reported in cancer (Tatton-Brown et al., 2014). Interestingly, mutations occur in the PWWP and ADD domains that may reduce the binding capacity of the protein to chromatin or lead to decreased activity of the protein, respectively (Tatton-Brown et al., 2014; Xin et al., 2017). Interestingly, the mutation of the Arginine 767, which is located in the interaction domain between DNMT3A and DNMT3L, may impair their interaction, leading to decreased targeting and activity of DNMT3A (Tatton-Brown et al., 2014). Unfortunately, methylome analysis of these patients and functional tests on the activity of the mutant protein have not been performed yet.

2.2.2. Mutations in DNMT1 cause an "Hereditary sensory neuropathy with dementia and hearing loss"

Hereditary sensory neuropathy has an autosomal dominant inheritance with neuro-degeneration that clinically manifests with loss of sensation leading to painless extremity injuries, infections and resultant amputations (Klein et al., 2011). Although the genetic causes of this disease have been described for few of the patients, the cause of the diseases was unknown for the majority of the patients. Between 2011 and 2012, mutations in DNMT1 were described as causative in twenty-three patients (Klein et al., 2011; Winkelmann et al., 2012). Mutations occur in exons 21 and 22 of the coding gene (Klein et al., 2011; Winkelmann et al., 2012). Clinical manifestations consist of sensory impairment, sudomotor dysfunction (loss of sweating), dementia and sensorineural hearing loss. This disease is inherited in an autosomal dominant manner. Indeed, mutations are present within the targeting-sequence domain, through which DNMT1 is recruited to chromatin during S-phase and maintains this localization during G2/M phases. Moreover, mutations affect the proper conformation of the enzyme, reducing its activity (Fatemi et al., 2001; Song et al., 2012). Global levels of DNA methylation were assessed by mass spectrometry, showing an 8% decrease in DNA methylation. DNA methylation at SatII and Satα tandem repeats

and at LINE1 and Alu interspersed repeats was analysed by pyrosequencing, showing loss of 5-6% of DNA methylation at SatII. In contrast, methylation levels at centromeric Satα and LINE1 elements were not affected compared to healthy controls. The authors performed also DNA methylation analysis at promoter regions, performing the DNA methylation array HumanMethylation27 BeadChip. The reduction of DNA methylation was moderate. Moreover, the most significant variations in DNA methylation were related to ten promoters which were hypermethylated (Klein et al., 2011). The authors suggested that this local hypermethylation could be reminiscent to hypermethylation events observed in cancer. However, the absence of cancer history in this disease, the absence of genome-wide hypomethylation and the few hypermethylated promoters observed make the authors' observation quiet speculative.

2.2.3. Mutations in MECP2: Rett and Angelman syndromes

Somatic mutations and/or alterations in the expression of MECP2 coding gene are associated with a wide variety of neurological diseases, including autism and Prader-Willi syndrome (Samaco, 2004). However, at the moment, there are two syndromes that seem to be caused by germ line mutations in this gene. In 1999, the Rett syndrome was the first to be identified (Amir et al., 1999). This syndrome is inherited in a dominant X-linked manner and affects predominantly females, occurring in 1:10,000 births, although male patients were also described (Moretti and Zoghbi, 2006). The Rett syndrome is a progressive neurodevelopmental disorder characterized by arrested development between 6 and 18 months of age, regression of acquired skills, loss of speech, unusual stereotyped movements and intellectual disability (Zachariah and Rastegar, 2012). Mutations in MECP2 cause phenotypical alterations in the nervous system only in postnatal stages, consistent with the fact that MECP2 is essential once neuronal differentiation begins (Zachariah and Rastegar, 2012). This observation is reinforced by studies in mouse models showing that abrogation of MECP2 in adult mice results in severe neurological symptoms commonly observed in Rett patients, such as global shrinkage of the brain, increased neuronal cell density, reaction of dendritic arbors, reduction of synaptic proteins and altered astrocytic development (Nguyen et al., 2012). Mutations in MECP2 do not affect DNA methylation but lead to global hyperacetylation of histone H3 in brain tissues, which may alter chromatin structure and the correct expression programs (Shahbazian and Zoghbi, 2001). Unfortunately, a more precise analysis to identify genomic regions characterized by the presence of this aberrant hyperacetylation was not performed.

Angelman syndrome (AS) is a neurodevelopmental disorder characterised by severe mental retardation, inability to speak, ataxia, dysmorphic facial features, and a seizure disorder appearance (Scheffer et al., 1990). A variety of different genetic mechanisms involving the

chromosomic regions 15q11-13, known to be imprinted, have been described in association with AS (Jiang et al., 1999). However, a subset of AS patients has no detectable anomalies in this chromosomic region. Given the similarity between clinical manifestations of Rett syndrome and AS (Scheffer et al., 1990), Watson and colleagues decided to investigate whether mutations in MECP2 gene were detectable in AS patients negative for anomalies in 15q11-13 region. They observed that five patients on the almost fifty analysed, four females and one male, presented mutations in the MECP2 gene. These findings reinforced the crucial role of MECP2 in neuronal homeostasis and suggest a role for MECP2 in controlling the expression of genes hosted in the 15q11-13 region (Watson et al., 2001).

2.2.4. The Immunodeficiency, Centromeric instability and Facial anomalies syndrome

2.2.4.1. Historical overview

The immunodeficiency, Centromeric instability and Facial anomalies (ICF) syndrome is the founding member of a growing class of "chromatin diseases". In 1979, Tiepolo and colleagues were the first to describe a curious case of an infant presenting a severe combined immunodeficiency that was associated with peculiar cytogenetic abnormalities in blood cell cultures of the patient, namely multi-branching of chromosomes 1, 9, and 16 (Tiepolo et al., 1979). The term "ICF" was coined nine years later by Maraschio and colleagues (Maraschio et al., 1988) who identified a patient with clinical and cytogenetic features reminiscent of the case reported by Tiepolo, and subsequently proposed that this, along with several similarly described cases (Fryns et al., 1981; Howard et al., 1985; Valkova et al., 1987) represented a new syndrome, transmitted in an autosomal recessive manner.

Patients with similar clinical features and chromosomal abnormalities have since been identified, and the picture of the complex phenotype and underlying aetiology of this disease is still being drawn. ICF is an extremely rare disease, with around 70 patients reported worldwide. These patients mostly come from Europe, although cases in individuals of African, American, and Japanese descent have also been reported (Hagleitner et al., 2007). Although most cases are non-familial, some consanguinity has been recorded (Wijmenga et al., 2000). A recent increase in the number of diagnosed non-familial cases in Japan and Europe (Ehrlich et al., 2006) led to the assumption that the disease may be under-diagnosed.

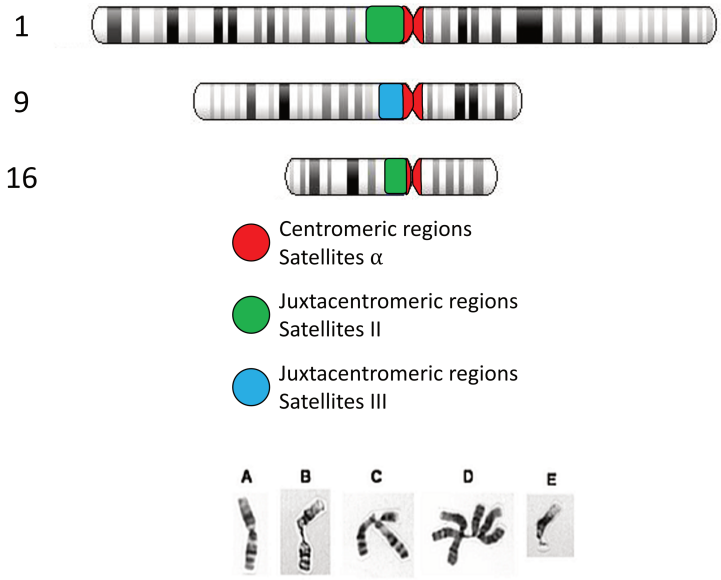


Figure 35. Chromosomal anomalies in the ICF syndrome. Top: Schematic showing chromosome 1, 9 and 16 Large heterochromatin regions surround the centromere, characterized by the presence of centromeric Satα. Juxtacentromeric regions are characterized by the presence of SatII for chromosome 1 and 16, while chromosome 9 present SatIII. Bottom: Examples of chromosomal anomalies in metaphases from ICF LCLs (A) normal Chr1, (B) Chr1 with decondensation in the qh region, (C) tri-radial (1;16)(q,q,p), (D) hexa-radial (1) (p,p,q,q,q,q), (E) deletion(1)(qh). *Adapted from Ehrlich 2003.*

2.2.4.2. Diagnosis, clinical and molecular features

ICF is a clinically heterogeneous disease, although it is almost invariantly characterised by a severe impairment of humoral immunity, characterized by lower levels of globin in the blood (hypoglobulinemia) or very low levels of the protective immune system proteins called immunoglobulins (agammaglobulinemia). The immunodeficiency is generally of the combined type, meaning that there is the absence or significant reduction in at least two immunoglobulin isotypes and in different combinations. However, B and T cell numbers are typically normal, indicating that the early stages of lymphocyte differentiation are not impaired. Membrane-bound IgM and IgD can be found at the surface of ICF B cell lines, indicating that V(D)J recombination is normal (Ehrlich et al., 2001). However, despite the presence of membrane-bound IgM, many patients showed low levels of IgM serum, likely indicating a defect in lymphocyte maturation or activation. In agreement with this, Blanco-Betancourt and colleagues noted accumulation of immature B cells in peripheral blood of ICF patients and the complete absence of memory B cells (Blanco-Betancourt et al., 2004). The result of such immunodeficiency is severe and leads to recurrent respiratory and gastrointestinal infections which often cause death at a young age. Other clinical features of ICF are present at varying degrees of penetrance and have been extensively reviewed by Hagleitner and colleagues (Hagleitner et al., 2007). Facial dimorphisms are usually mild and typically involve a flat nasal bridge, hypertelorism (widely spaced eyes), epicanthic folds, macroglossia (enlarged tongue), micrognathia (small jaw), and low set ears. Intellectual impairment and neurological defects may also be present, and include slow cognitive and motor development and psychomotor impairment (Hagleitner et al., 2007).

The characteristic chromosomal instabilities first documented by Tiepolo and colleagues represent an invariant molecular hallmark of the ICF syndrome that serves for diagnostic purposes (Ehrlich et al., 2006). Indeed, ICF was diagnosed by karyotyping of mitogen-stimulated peripheral blood cultures of paediatric patients presenting with recurrent unexplained infections in the presence of normal numbers of B cells (Ehrlich et al., 2006). Affected subjects show decondensation of juxtacentromeric heterochromatin of chromosomes 1, 9, and 16, leading to a wide variety of associated chromosomal anomalies, including chromosome breaks and rearrangements, whole-arm deletions, and often multibranched chromosomes containing three or more arms of chromosome 1 and 16 joined in the vicinity of the centromere (Carpenter et al., 1988; Fryns et al., 1981; Maraschio et al., 1988; Tiepolo et al., 1979; Tuck-Muller et al., 2000) (**Figure 35**).

The underlying molecular basis of these anomalies was thought to result from DNA hypomethylation of associated juxtacentromeric sequences. This idea was proposed by Maraschio and colleagues who wrote: “in an attempt to clarify the basic defect causing the chromosomal

abnormalities found in these (ICF) patients, we must consider that the centromeric regions of chromosomes 1, 9, 16, (15 and Y), represent a special class of heterochromatin characterised by positive staining with DA-DAPI, and by highly methylated DNA, and by under-condensation and association in metaphase after treatment with 5-azacytidine" (Maraschio et al., 1988). Viegas-Pequignot and colleagues subsequently confirmed that SatII DNA repeats on chromosomes 1 and 16, and SatIII DNA on chromosome 9, were hypomethylated in mitogen-stimulated leukocytes of four ICF patients, thus mimicking an embryonic-like methylation pattern (Jeanpierre et al., 1993). Treatment of pro-B cells with the DNA demethylating agent 5-azacytidine (but not with other non-demethylating genotoxins) induced pericentromeric chromosomal anomalies characteristic of ICF cells, suggesting a causal link between hypomethylation of SatII and SatIII regions and the formation of multibranched structures (Ji et al., 1997). Interestingly, although DNA hypomethylation is present in all cell types analysed, chromosomal anomalies arise in ICF lymphoblasts, bone marrow cells and fibroblast (Jeanpierre et al., 1993; Ehrlich et al., 2006).

DNA hypomethylation characterizing ICF syndrome extends beyond SatII and SatIII repeats of chromosomes 1, 9, and 16. Indeed, Kondo and colleagues reported hypomethylation of the non-satellite repeats NBL2 and D4Z4 in ICF patients, which was associated with an increased transcription of NBL2 sequences (Kondo et al., 2000).

2.2.4.3. Mutations in DNMT3B cause ICF1 syndrome

A genomic locus associated with ICF syndrome was first identified on a 9-centimorgan region of chromosome 20 by homozygosity mapping (Wijmenga et al., 1998). In the same year, the *de novo* methyltransferases, both human and mouse, were cloned (Okano et al., 1998). These two studies allowed to localize, by homozygosity mapping, DNMT3B gene in the 9-centimorgan region and led three groups, in 1999, to independently describe mutations in DNMT3B in a subset of ICF patients (Hansen et al., 1999; Okano et al., 1999; Xu et al., 1999). All the affected individuals studied were homozygotes for a mutation or compound heterozygotes. These mutations consisted of one nonsense mutation, a variety of missense mutations, and a three amino-acids insertion, which mostly localised to the C-terminal domain of the protein that contains the catalytic domain (Gowher and Jeltsch, 2002). These findings suggested that ICF1 results from an impairment in DNA methyltransferase activity. Indeed, Xu and colleagues directly demonstrated a loss of DNA methyltransferase activity associated with an ICF1 mutation. They introduce an ICF1 mutation in the DNMT3B cDNA and a methylable plasmid were co-transfected in human cells. They observed that mutated DNMT3B led to reduction in the methylation of cotransfected episomal plasmid DNA. However, some residual enzymatic activity is undoubtedly indispensable for survival, since nonsense mutations are always found as compound heterozygotes and complete

loss of DNMT3B function is likely to be lethal during embryonic development in humans as it is in mice (Okano et al., 1999). To date, more than 25 mutations associated with ICF1 syndrome have been identified in DNMT3B (Hagleitner et al., 2007; Jiang et al., 2005; Kutluğ et al., 2016; Rechavi et al., 2016; Sterlin et al., 2016; van den Boogaard et al., 2016; Weemaes et al., 2013), which explain 50 to 60% of ICF cases (Walton et al., 2014). ICF1 mutations cause a broad spectrum of biochemical defects in DNMT3B that lead to reduced enzymatic activity, including defects in homo-oligomerisation, S-adenosyl methionine (SAM) binding, SAM utilisation, and DNA binding (Moarefi and Chédin, 2011). ICF syndrome was the first disease to be linked to mutations in a DNMT gene.

In addition to SatII and SatIII, sub-telomeric repeats were reported to be hypomethylated (Yehezkel et al., 2008) with Lieberman and co-workers showing hypomethylation of 5/5 subtelomeric regions tested by MeDIP (Deng et al., 2010). This may explain why ICF1 cells express abnormally high levels of telomeric repeat containing RNA (TERRA) and are subject to telomere shortening (Yehezkel et al., 2008), leading to senescence at early passages (Yehezkel et al., 2013). Reintroduction of functional DNMT3B protein did not rescue the normal DNA methylation at sub-telomeric regions, while ectopic expression of DNMT3B in concert with DNMT3L partially restored DNA methylation (Yehezkel et al., 2013). More recently, Selig's lab showed that the increase in TERRA levels in ICF1 cells leads to the accumulation of DNA:RNA hybrids at sub-telomeric regions, which are associated with high levels of DNA damage at chromosome ends. These events may contribute to the chromosomal instability that characterizes the ICF syndrome (Sagie et al., 2017).

In addition to constitutive heterochromatin, mutations in DNMT3B also affect facultative heterochromatin. Two different studies show that inactive X is hypomethylated in female ICF1 patients (Hansen et al., 2000; Miniou et al., 1994), but this hypomethylation does not prevent the normal inactivation process (Bourc'his et al., 1999). In contrast to highly region-specific analyses of DNA hypomethylation, a report using whole-genome bisulfite sequencing of an ICF1 patient reported a global loss of 41% of methylation on all chromosomes, although the X chromosome was the most affected, with a loss of 63% of DNA methylation (Holger Heyn et al., 2012). Of note, hypomethylation of juxtacentromeric repeats was not restricted to chromosome 1, 9 and 16, but instead was found to a significant degree on all other chromosomes. Nevertheless, these results were obtained on transformed cell lines, suggesting a cautious interpretation.

Two mouse models harbouring ICF1 patient mutations have been created. These mice recapitulate some of the phenotypes observed in ICF1 syndrome and highlight the essential role of DNMT3B in development (Ueda et al., 2006; Velasco et al., 2010).

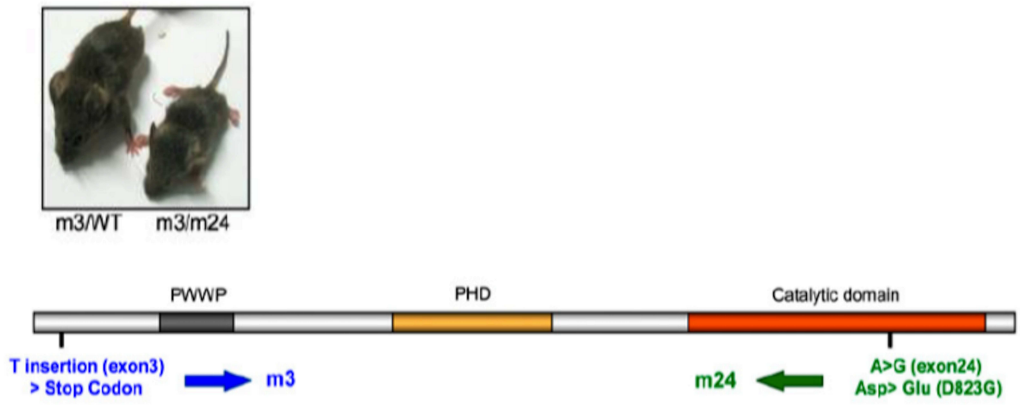


Figure 36. Dnmt3b hypomorphic mutant mouse. Top: Gross morphology of mEx3/WT versus mEx3/mEx24 mice. Note that very few mice survive after birth. Bottom: Schematic DNMT3B showing mutations used for the generation of ICF1 mice relative to protein domains. *Adapted from Velasco et al., 2010.*

Li and colleagues cloned six ICF1 mutations into mouse DNMT3B cDNA and studied the behaviour of each of these mutants with respect to their altered enzymatic activity, nuclear localisation of the coded protein, and interaction with DNMT3A (Ueda et al., 2006). Two mutations, A609T and D823G, that showed the greatest perturbation to DNMT3B function in biochemical assays, were subsequently selected for the generation of two ICF1 mouse models (Ueda et al., 2006). Although normal Mendelian frequencies both ICF1 mice were found at birth, most mutant mice died within 24 hours around birth, and only 4/62 mice survived to 3 weeks. Gross anatomy of these mice was normal but body weight was greatly reduced, corresponding to less than 2/3 of that of wild-type littermates. Facial anomalies similar to those observed in ICF patients were observed, including a shorter nose and wider nasal bridge. No significant differences between precursor and mature B cell populations were observed between ICF and littermate controls, suggesting that these mice do not represent a good model to study B cells defects observed in ICF1 patients. In contrast, T cell maturation appeared greatly perturbed in ICF mice and apoptotic analysis revealed that the majority of ICF thymocytes were dead or dying. At the molecular level, DNA hypomethylation was observed at major satellite repeats, mouse endogenous C type-retroviral DNA, and minor satellite repeats, although it was more pronounced at the latter. Methylation of minor satellite repeats was tested in a panel of tissues and all tissues tested were shown this hypomethylation was observed (Ueda et al., 2006).

A second ICF model was derived, which is a compound heterozygote (mEx3/mEx24), carrying mutations described in patients (Xu et al., 1999). Indeed, these mice bear the insertion of one base pair creating a frameshift and premature stop codon in the N terminal domain, and a missense mutation in the catalytic domain (**Figure 36**) (Velasco et al., 2010). These mice are viable but smaller compared to control littermates (**Figure 36**). DNA hypomethylation was observed at minor satellite repeats in ICF mice. In contrast to the previous report of Ueda and colleagues, the methylation status of the major satellite repeats was unaffected by mutations in DNMT3B, consistent with the previously described specificity of DNMT3B for minor satellite repeats (Okano et al., 1999; Velasco et al., 2010). Analysis of developmental defects not only revealed severe defects of the skull but also defects in the axial skeleton, with posterior homeotic transformations that transformed a thoracic vertebra into a lumbar one and a lumbar vertebra into a sacral one. Such alterations often result from the deregulated expression of homeotic (HOX) genes, which are key players in the establishment of positional identity along the antero-posterior axis. In line with these results, HOXA11 and HOXA13 expression were deregulated in mouse embryonic fibroblasts (MEF) of ICF1 mice. Transcriptome microarray analysis revealed the deregulated expression of 25 genes in the thymus of ICF1 mice, 21 of which are normally expressed in the testes. Among these germ line-specific genes MAELSTROM, SYCE1, SLC25A31, TEX11, and

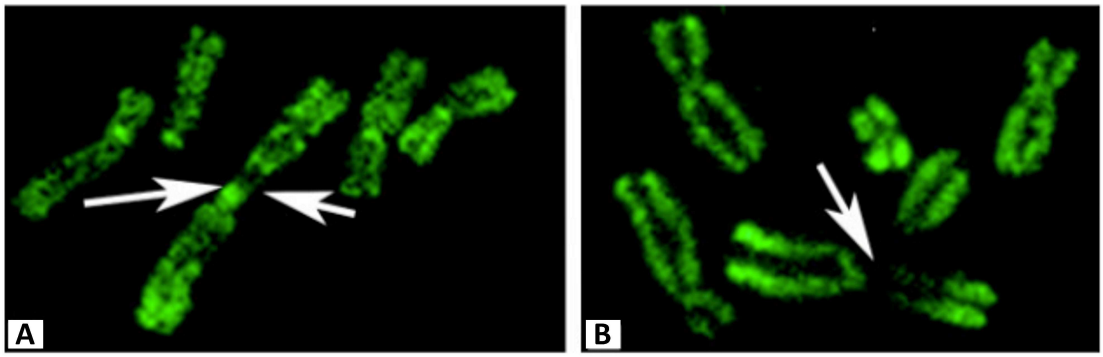


Figure 37. Variable methylation profiles at satellite repeats in ICF patients. Immunofluorescence with anti-5-mC antibody. A) ICF1 patients present the classical hypomethylation of SatII of chromosome 1 (small arrow), while Satα are normally methylated (large arrow). B) In an ICF patients without mutations in DNMT3B, the classical SatII and the Satα of chromosome 1 are hypomethylated (arrow). Satα of the other chromosomes of this patient are also hypomethylated. *Adapted from Jiang et al., 2005.*

DDX4 were expressed in ICF1 MEF, thymus, and the majority of ICF1 mouse somatic tissues tested. This illegitimate expression was associated with promoter hypomethylation. DNMT3B could still be detected at these regulatory regions, suggesting mutant DNMT3B can still be recruited at these loci but lack DNMT activity, resulting in promoter hypomethylation and aberrant gene expression. Homeotic transformation of ICF1 mouse model was reminiscent of transformations described after invalidation of one of its members, the transcriptional repressor E2F6 (Storre et al., 2002). More importantly, E2F6 was predicted to bind to most of the germline genes that were deregulated in ICF mouse somatic cells (Velasco et al., 2010). Chromatin-immunoprecipitation of DNMT3B in E2F6$^{-/-}$ reveals that DNMT3B recruitment at the promoter of this class of germline genes was E2F6-dependent. Co-immunoprecipitation experiments further confirmed the direct interaction between DNMT3B and E2F6. These findings highlight previously unknown mode of targeting DNA methylation at gene promoters (Velasco et al., 2010).

More recently, my host team confirmed the hypomethylation of the promoter of germline genes identified in the mouse model in ICF1 patients blood samples and primary fibroblasts, which was associated with aberrant transcription (Velasco et al., 2014). This germline signature represents a powerful biomarker, simple to assess, for the diagnosis of potential ICF1 patients. Nowadays, the diagnosis includes also the analysis of the methylation of SatII by Southern blot and exome-sequencing.

2.2.4.4. Mutations in the zinc finger ZBTB24 genes cause ICF2 syndrome

Genetic heterogeneity in the ICF syndrome could be speculated from the analysis of Miniou and colleagues who, in 1994, observed an additional hypomethylation of Satα in one ICF patient (Miniou et al., 1994). Subsequent studies provided evidence of other ICF patients showing hypomethylation of Satα, reproducing a embryonic DNA methylation profile of these regions (Miniou et al., 1997). The fact that the methylation status of satellite repeats distinguished two different classes of ICF was then formalized by Jiang and colleagues, who showed that an additional hypomethylation of Satα characterized all patients without mutations in DNMT3B gene (Jiang et al., 2005) (**Figure 37**).

Follow-up studies by Wijmenga and colleagues highlight genetic heterogeneity in the ICF syndrome when a screen for mutations in DNMT3B identified exonic mutations in only nine out of a total of fourteen ICF patients (Wijmenga et al., 2000). It was plausible that the remaining patients could harbour mutations within the DNMT3B promoter or intronic regions that affect transcriptional regulation or 3'UTR and regions that can be recognized by micro-RNA (miRNA). However, in this study, two ICF patients from consanguineous families for which mutations in DNMT3B were not detected, did not show homozygosity for the DNMT3B locus, strongly

suggesting the implication of another genetic locus. In one ICF patient, for which no mutation in DNMT3B coding sequences could be detected, Kubota and colleagues found no mutation in two putative DNMT3B promoter regions and suggested that a mutation may be present in a gene coding either for another DNMT or for a DNMT-associated protein (Kubota et al., 2004).

The existence of distinct subtypes of ICF with different genetic basis was confirmed with the finding that eleven of patients without mutations in DNMT3B gene possess mutations in the zinc-finger and BTB domain-containing 24 (ZBTB24) gene, and referred to as ICF2 patients (de Greef et al., 2011). A growing number of different mutations associated with ICF2 patients were then identified in the ZBTB24 gene (Cerbone et al., 2012; Chouery et al., 2012; Nitta et al., 2013; van den Boogaard et al., 2017; von Bernuth et al., 2014), which explain around 30% of ICF cases.

However, immunodeficiency is generally more pronounced in ICF1 patients compared to ICF2, who present intellectual disabilities in 100% of the cases (Weemaes et al., 2013; Sterlin et al., 2016). In contrast with ICF1 patients, van Boogaard and colleagues observed a gender bias in ICF2 patients, with 79% of male patients, that may suggest a more severe effect of ZBTB24 mutations in female patients (van Boogaard et al., 2017)

Relatively little is known about ZBTB24 itself. It belongs to same family of BCL6, which is essential for B cells development (Siggs and Beutler, 2012). When de Greef and colleagues identified mutations in ICF2 patients, they analysed the mRNA levels in different tissues and cell lines, showing that ZBTB24 is constitutively expressed, with the highest expression in naïve B cells compared to differentiated cells. These findings led the authors to propose a role for ZBTB24 in B cell development and/or B and T cell interaction (de Greef et al., 2011; Lee and Maeda, 2012). A very recent study, indeed, showed that one ICF2 patient presented low levels of T-cells supporting the hypothesis of a role of ZBTB24 in immune cells (Kelsen et al., 2017).

ZBTB24 belongs to the Zinc-finger BTB family and was described as a “protein of obscure physiological function” (Siggs and Beutler, 2012). ZBTB24 presents a BTB domain which is thought to mediate protein oligomerization and/or dimerization, an AT-hook domain through which it can bind to AT-rich DNA sequences and eight Zinc-finger domains in its C-terminal part, supporting that ZBTB24 is a transcription factor that binds directly to DNA (Siggs and Beutler, 2012). Ectopic expression of ZBTB24 in mouse ES cells showed that ZBTB24 localizes to chromocenters. Since this localization is not lost in TKO ES cells, it suggests that ZBTB24 binds DNA independently of DNA methylation (Nitta et al., 2013). Although the function of ZBTB24 remains unknown, its implication in the ICF syndrome strongly suggest that ZBTB24 is required for DNA methylation establishment and/or maintenance at centromeric and pericentromeric repeats. However, the underlying mechanism, as well as the target genomic loci of ZBTB24

remain largely unknown. The simplest hypothesis may be that ZBTB24 binds directly to centromeric and pericentromeric repeats, mediating the recruitment of *de novo* and/or maintenance DNMTs. Moreover, a yeast two-hybrid study identified the histone methyltransferase SUV39H1 and the H3K4 and H3K9 histone demethylase LSD1 as partners of ZBTB24, suggesting that ZBTB24 may affect DNA methylation indirectly, through modulation of histone modifications (Weimann et al., 2013).

Additional cases of ICF remaining with no described mutation in either DNMT3B or ZBTB24, have been referred to as ICFX (Weemaes et al., 2013). Most of the ICFX patients show molecular and clinical features more similar to ICF2 patients than ICF1, presenting hypomethylation of both SatII and Satα and intellectual disability (Jiang et al., 2005; Weemaes et al., 2013). The presence of ICFX patients strongly suggests that they could carry mutations in new genes implicated in the syndrome and, more in general, in DNA methylation pathways.

OBJECTIVES

The ICF syndrome is a very rare disease, with around 70 cases described world-wide. It is characterized by perturbations to DNA methylation landscapes, chromosomal instability, perturbed nuclear organization and perturbed gene expression programs, which are all hallmarks of several human diseases, including cancer, autoimmune disease and neurological disorders. Mutations in DNMT3B gene were the first to be identified in in 1999 in a subset of ICF patients, hereafter referred to as ICF1 whereas the remainder was named ICFX. Studies on ICF1 syndrome allowed my host team, together with others, to better characterize DNMT3B functions, identify its genomic targets as well as its mechanisms of recruitment at specific places on the genome (Velasco et al., 2010; Walton et al., 2011). In 2011, the advent of exome sequencing allowed the identification of mutations in ZBTB24 and defined the ICF2 subtype (de Greef et al., 2011). This finding was surprising since ZBTB24 is a factor without catalytic activity, calling for further characterization of its functions in DNA methylation and development. In addition, with ICF patients left with unknown mutations, it stressed the need to continue investigations into the causes of this rare but important disease.

Hence, the ICF syndrome, besides the medical aspect, it represents a unique opportunity to identify new candidate players in DNA methylation, in particular at (peri)centromeric repeats, in maintenance of genome integrity and normal development.

My thesis work capitalized on both mouse models and cohort of patients established in the host team. A mouse model with ICF mutations in the ZBTB24 gene inserted by knock-in, and primary cells from a unique cohort of ICF patients with all four genotypes represented the best tools to answer to the following specific questions:

- What is the role of ZBTB24 on establishment of DNA methylation patterns during early development?
- What is the role of ZBTB24 on maintenance of methylation patterns during differentiation in somatic cells?
- What is the impact of ZBTB24 mutations on mouse development?
- Is ZBTB24 an adapter of DNMT3B function, or of other DNMTs?
- What are the genomic regions targeted by ZBTB24?
- What are the other genes mutated in ICFX patients?
- What is the impact of mutations in ICF genes on the methylome of patients primary cells?

RESULTS

ARTICLE 1

Mutations in CDCA7 and HELLS cause immunodeficiency–centromeric instability–facial anomalies syndrome

Peter E. Thijssen, Yuya Ito, **Giacomo Grillo**, Jun Wang, Guillaume Velasco, Hirohisa Nitta, Motoko Unoki, Minako Yoshihara, Mikita Suyama, Yu Sun, Richard J.L F. Lemmers, Jessica C. de Greef, Andrew Gennery, Paolo Picco, Barbara Kloeckener-Gruissem, Tayfun Güngör, Ismail Reisli, Capucine Picard, Kamila Kebaili, Bertrand Roquelaure, Tsuyako Iwai, Ikuko Kondo, Takeo Kubota, Monique M. van Ostaijen-Ten Dam, Maarten J.D. van Tol, Corry Weemaes, Claire Francastel, Silvére M. van der Maarel & Hiroyuki Sasaki

Nature Communications, 2015

The ICF syndrome represents an important human model system for the identification of new players in DNA methylation pathways. The study of this very rare disease led to the identification of mutations in DNMT3B as the cause of the disease for 50% of patients (ICF1) (Xu et al., 1999). In 2011, mutations in ZBTB24 have been identified to cause the ICF syndrome in 30% of patients (ICF2) (deGreef et al., 2011). Before the implication in the ICF syndrome, the role of ZBTB24 had never been suspected. The presence of patients with unknown mutations, classified as ICFX, strongly suggests the presence of other genes implicated in the ICF syndrome, and more broadly, in DNA methylation pathways.

In collaboration with several paediatric immunology departments and Dutch and Japanese teams, we performed exome-sequencing on ICFX patients samples to identify new genes whose mutations are causative of the ICF syndrome. This study allowed the identification of mutations in CDCA7 and HELLS, characterizing ICF3 and ICF4 patients respectively. Although the implication of HELLS in DNA methylation was already well documented from studies in the mouse (Dennis et al. 2001; Huang et al., 2004; Dunican et al. 2013; Xi et al. 2007; Termanis et al., 2016), our study identifies CDCA7 as a new player in DNA methylation pathways. Moreover, we showed that knock-down of the newly identified factors in mouse primary fibroblast, where DNA methylation is already established, caused hypomethylation of centromeric minor satellite repeats, suggesting a role of these factors in DNA methylation maintenance, at least at these regions.

I contributed to this study performing the functional tests on the newly identified genes. I performed knock-down of DNMT3B, ZBTB24, CDCA7 and HELLS in primary MEF, followed by Southern blot to check the DNA methylation status of centromeric minor satellite repeats. Moreover, I performed Southern blots to verify the hypomethylation of Satα and SatII repeats of one of the ICF4 patients (patient of family G; Supplementary figure 1). This analysis has been performed on genomic DNA extracted from peripheral blood sample of the patient.

This study has been published in July 2015, in *Nature Communications*.

nature COMMUNICATIONS

ARTICLE

Received 2 Mar 2015 | Accepted 19 Jun 2015 | Published 28 Jul 2015

DOI: 10.1038/ncomms8870 OPEN

Mutations in *CDCA7* and *HELLS* cause immunodeficiency–centromeric instability–facial anomalies syndrome

Peter E. Thijssen[1], Yuya Ito[2], Giacomo Grillo[3], Jun Wang[1,†], Guillaume Velasco[3], Hirohisa Nitta[2], Motoko Unoki[2], Minako Yoshihara[4], Mikita Suyama[4], Yu Sun[1,†], Richard J.L F. Lemmers[1], Jessica C. de Greef[1], Andrew Gennery[5,6], Paolo Picco[7], Barbara Kloeckener-Gruissem[8,9], Tayfun Güngör[10], Ismail Reisli[11], Capucine Picard[12,13,14,15,16,17], Kamila Kebaili[18], Bertrand Roquelaure[19], Tsuyako Iwai[20], Ikuko Kondo[21], Takeo Kubota[22], Monique M. van Ostaijen-Ten Dam[23], Maarten J.D. van Tol[23], Corry Weemaes[24], Claire Francastel[3], Silvère M. van der Maarel[1] & Hiroyuki Sasaki[2]

The life-threatening Immunodeficiency, Centromeric Instability and Facial Anomalies (ICF) syndrome is a genetically heterogeneous autosomal recessive disorder. Twenty percent of patients cannot be explained by mutations in the known ICF genes DNA methyltransferase 3B or zinc-finger and BTB domain containing 24. Here we report mutations in the cell division cycle associated 7 and the helicase, lymphoid-specific genes in 10 unexplained ICF cases. Our data highlight the genetic heterogeneity of ICF syndrome; however, they provide evidence that all genes act in common or converging pathways leading to the ICF phenotype.

[1] Department of Human Genetics, Leiden University Medical Center, Leiden 2333ZA, The Netherlands. [2] Division of Epigenomics and Development, Department of Molecular Genetics, Medical Institute of Bioregulation, Kyushu University, Fukuoka 812-8582, Japan. [3] CNRS UMR7216, Epigenetics and Cell Fate, Université Paris Diderot, Sorbonne Paris Cité, 75205 Paris, France. [4] Division of Bioinformatics, Department of Multi-scale Research Center for Medical Science, Medical Institute of Bioregulation, Kyushu University, Fukuoka 812-8582, Japan. [5] Department of Paediatric Immunology, Newcastle Upon Tyne Hospital, NHS Foundation Trust, Newcastle Upon Tyne, UK. [6] Institute of Cellular Medicine, Newcastle University, Newcastle Upon Tyne NE1 4LP, UK. [7] Division of Pediatrics and Pediatric Rheumatology, G. Gaslini Scientific Institute, Genova 16147, Italy. [8] Institute of Medical Molecular Genetics, University of Zurich, Schlieren 8952, Switzerland. [9] Department of Biology, ETH Zurich, Zurich 8093, Switzerland. [10] Department of Oncology, University Children's Hospital, Zurich 8032, Switzerland. [11] Department of Pediatric Immunology and Allergy, Necmettin Erbakan University, Meram Medical Faculty, Konya 42080, Turkey. [12] Centre de Référence Déficits Immunitaires Héréditaires, AP-HP, 75743 Paris, France. [13] Centre d'Etude des Déficits Immunitaires, Hôpital Universitaire Necker-Enfants Malades, AP-HP, 75743 Paris, France. [14] Laboratoire de Génétique Humaine des Maladies Infectieuses, Inserm, 75743, Paris, France. [15] Université Paris Descartes, Institut Imagine, Sorbonne Paris, 75743 Paris, France. [16] Unité d'immunologie et d'hématologie pédiatrique, Hôpital Necker-Enfants Malades, 75743, Paris, France. [17] Inserm UMR 768, 75015 Paris, France. [18] Centre de Référence Déficits Immunitaires Héréditaires, Institut d'Hématologie et d'Oncologie Pédiatrique, CHU de Lyon, 69008 Lyon, France. [19] Service d'hépato-gastro-entérologie et nutrition, endocrinologie et néphrologie pédiatriques, Hôpital de la Timone, CHU Marseille, 13385 Marseille, France. [20] Department of Pediatric Hematology and Oncology, Shikoku Medical Center for Children and adults, Kagawa 765-8507, Japan. [21] Department of Pediatrics, Ooida Hospital, Kochi 788-0001, Japan. [22] Department of Epigenetic Medicine, Faculty of Medicine, Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, Yamanashi 409-3898, Japan. [23] Department of Pediatrics, Laboratory for Immunology, Leiden University Medical Center, Leiden 2333ZA, The Netherlands. [24] Department of Pediatric Infectious Diseases and Immunology, Radboud University Nijmegen Medical Center, Nijmegen 6500HC, The Netherlands. † Present address: Institutes of Biology and Medical Sciences, Soochow University, Suzhou 215123, China (J.W.); Shanghai Institute for Pediatric Research, Xin Hua Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China (Y.S.). Correspondence and requests for materials should be addressed to C.F. (email: claire.francastel@univ-paris-diderot.fr) or to S.M. (email: Maarel@lumc.nl) or to H.S. (email: hsasaki@bioreg.kyushu-u.ac.jp).

© 2015 Macmillan Publishers Limited. All rights reserved.

Immunodeficiency, Centromeric Instability and Facial Anomalies (ICF) syndrome is characterized by recurrent and often fatal respiratory and gastrointestinal infections as a consequence of hypo- or a-gammaglobulinemia in the presence of B cells[1,2]. Nearly all patients also present with distinct facial anomalies, including hypertelorism, flat nasal bridge and epicanthus[1,2]. Centromeric instability is the cytogenetic hallmark of ICF syndrome. It involves the juxtacentromeric heterochromatin repeats on chromosomes 1, 9 and 16 and is comparable to what is observed after treatment of cells with demethylating agents[3,4]. Therefore, CpG hypomethylation of juxtacentromeric satellite types II and III is a diagnostic for ICF syndrome, with additional hypomethylation of centromeric α-satellite repeats in *DNMT3B* mutation-negative patients[5,6].

Mutations in the DNA methyltransferase 3B (*DNMT3B*) gene (OMIM 602900; ICF1) account for ~50% of ICF cases, while ~30% of cases have mutations in the zinc-finger and BTB domain containing 24 (*ZBTB24*) gene (OMIM 614064; ICF2)[2,5]. *DNMT3B* is a *de novo* DNA methyltransferase, primarily acting during early development with a preference for CpG dense regions[7]. ICF1 mutations in the catalytic domain of *DNMT3B* result in largely reduced methyltransferase activity[8,9]. The function of *ZBTB24* is unknown, but it belongs to a family of ZBTB proteins of which many have regulatory roles in haematopoietic differentiation[10,11]. Despite the successful identification of ICF genes, the pathophysiological mechanism underlying the syndrome remains largely unresolved.

Previous studies indicated further genetic heterogeneity in ICF syndrome[2]. To identify the genetic cause in genetically unexplained cases (ICFX), we combined homozygosity mapping with whole-exome sequencing. By using an autosomal recessive inheritance model and prioritizing homozygous variants in consanguineous families, we now identify four different homozygous and potentially damaging variants in the cell division cycle associated 7 (*CDCA7*) gene in five ICFX patients (now referred to as ICF3). In an additional five ICFX patients, we identify compound heterozygous and homozygous variants in the helicase, lymphoid-specific (*HELLS*) gene (ICF4). We show that knockdown of both new ICF genes leads to hypomethylation of juxtacentromeric heterochromatin repeats in a murine cell model. Our results emphasize the genetic heterogeneity of ICF syndrome, nonetheless providing evidence that all four ICF genes are involved in at least one common pathway.

Results

Missense mutations in *CDCA7* cause ICF syndrome type 3. We selected 13 ICFX patients from 11 families, negative for mutations in *DNMT3B* or *ZBTB24*, of whom the clinicopathological characteristics are listed in Supplementary Table 1. Hypomethylation of pericentromeric satellite type II (Sat II), common to all ICF patients, and centromeric α-satellite DNA repeats, shown to be affected only in ICF2 and ICFX, was shown before for a subset of patients[2,5,6,12]. For an additional set of ICFX patients, we

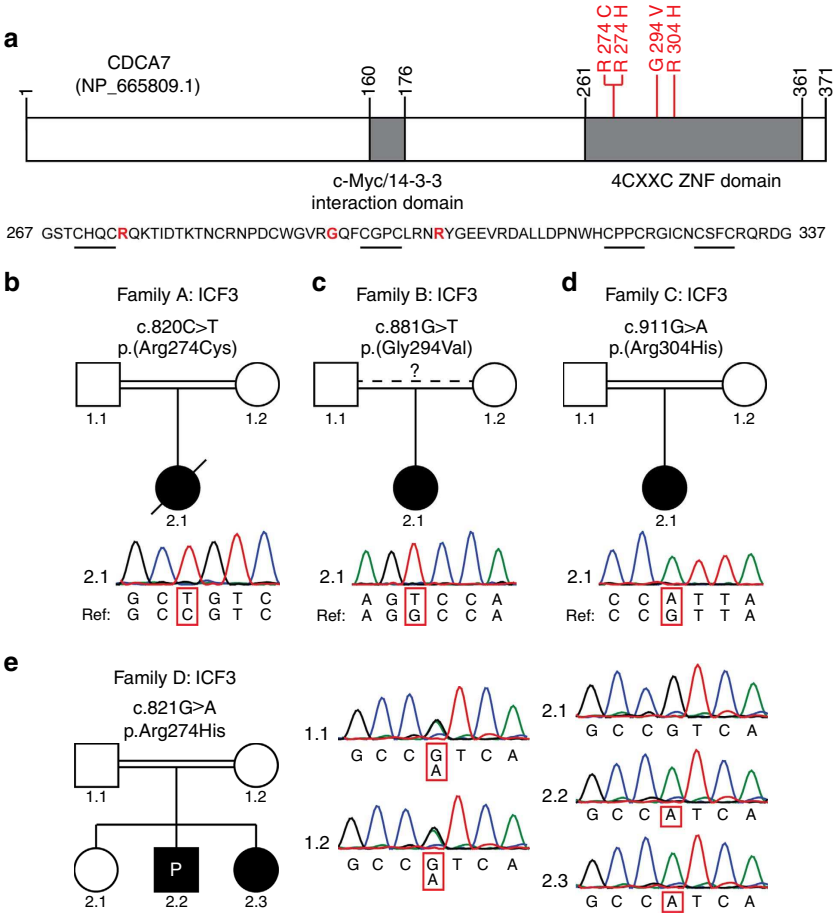


Figure 1 | Homozygous missense mutations in *CDCA7* in five ICF3 patients. (**a**) Schematic representation of CDCA7, with the identified homozygous missense mutations in red. Sequence outtake: 4-CXXC-type zinc-finger domain, CXXC motifs are underlined, mutated residues in red. (**b**–**d**) Sanger sequencing confirmation of missense mutations in CDCA7 in families A–C. All variants were homozygous, the reference sequence is displayed for comparison. (**e**) Sanger sequencing confirmation of the homozygous missense CDCA7 mutation in patients 2.2 (proband) and 2.3 of family D. Both parents are heterozygous for the variant, sibling 2.1 is unaffected and homozygous for the wild-type (wt) allele.

© 2015 Macmillan Publishers Limited. All rights reserved.

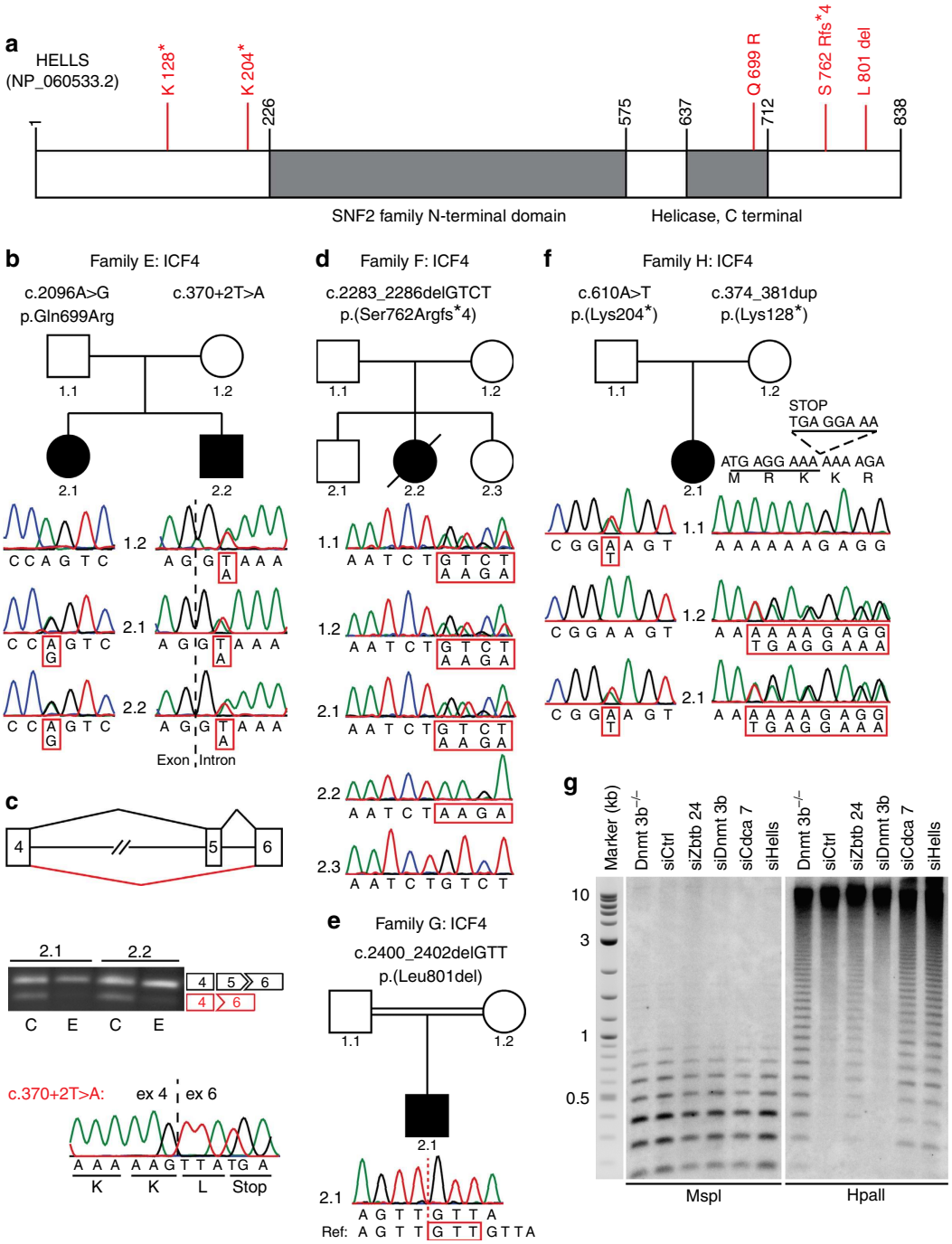


Figure 2 | Mutations in *HELLS* in five ICF4 patients. (**a**) Schematic representation of HELLS, with the identified mutations in red. (**b**) Sanger sequencing confirmation of HELLS mutations in family E. Only c.370 + 2T > A was identified in maternal DNA, indicating different allelic origins of both mutations, or *de novo* occurrence of the second mutation. (**c**) RT-PCR analysis of HELLS mRNA on treatment of patient-derived fibroblasts with cycloheximide (C) revealed that c.370 + 2T > A leads to complete skipping of exon 5 and disruption of the open reading frame. Ethanol-treated samples (E) served as controls, alternative splicing was confirmed using Sanger's sequencing in two independent experiments for both samples. (**d**) Sanger sequencing confirmation of a homozygous out-of-frame deletion in *HELLS* in family F. Both parents as well as unaffected sibling 2.1 are heterozygous for the deletion allele; unaffected sibling 2.3 is homozygous for the wt allele. (**e**) Sanger sequencing confirmation of a homozygous in-frame deletion in *HELLS* in family G. Both parents are heterozygous for the deletion allele. (**f**) Sanger sequencing confirmation of nonsense mutations in *HELLS* in family H. Different allelic origin was confirmed in parental DNA. (**g**) Southern blot analysis of minor satellite DNA methylation in *Dnmt3b*$^{-/-}$ and siRNA-treated wt MEFs after digesting DNA with MspI or its methylation-sensitive isoschizomer HpaII revealed CpG hypomethylation on knockdown of Zbtb24, Cdca7 and Hells. Molecular weights of the 2-Log DNA size marker are in kilobasepairs.

confirmed Sat II and α-satellite hypomethylation using Southern blot analysis (Supplementary Fig. 1). In five ICFX patients from four families we identified and confirmed homozygous missense mutations in *CDCA7*, all near the first two zinc-finger motifs in the conserved carboxyterminal 4-CXXC-type zinc-finger domain (Fig. 1a–e). Segregation with disease was confirmed in family D and all mutations were predicted to be pathogenic and have an allele frequency supporting pathogenicity (Fig. 1e, Supplementary

© 2015 Macmillan Publishers Limited. All rights reserved.

Table 1). *CDCA7* is involved in neoplastic transformation, *MYC*-dependent apoptosis and haematopoietic stem cell emergence; however, its molecular function is unknown[13,14]. All four zinc-finger motifs are completely conserved in the highly homologous 4-CXXC zinc-finger domain of the transcriptional repressor *CDCA7-Like* (*CDCA7L*; Supplementary Fig. 2). The repressive activity of CDCA7L is dependent on its 4-CXXC domain, and by homology *CDCA7* mutations in ICF3 may disrupt a similar function[15].

Mutations in HELLS cause ICF syndrome type 4. In an additional five ICFX patients from four families, we identified mutations in *HELLS* that were predicted to be pathogenic and with allele frequencies supporting pathogenicity (Fig. 2a, Supplementary Table 1, ICF4). In affected members of family E, we identified a missense mutation in the conserved helicase domain (c.2096A > G;p.[Gln699Arg]) and an intronic mutation leading to destruction of the splice donor site in intron 5 (c.370 + 2T > A; Fig. 2b). Different allelic origin is supported by analysis of maternal DNA, which carried only the splice site mutation (Fig. 2c). To analyse the effect of the lost splice donor site on mRNA processing, fibroblasts of both patients were treated with cycloheximide to inhibit nonsense-mediated decay. Upon treatment, reverse transcriptase–PCR (RT–PCR) analysis showed increased levels of a splice variant with complete skipping of exon 5, leading to a frameshift followed by a premature stop codon in exon 6 (Fig. 2c).

The patient in family F carries a homozygous out-of-frame deletion (c.2283_2286delGTCT;p.[Ser762Argfs*4]), resulting in a frameshift and introduction of a premature stop codon in exon 20. The unaffected siblings 2.1 and 2.3 were found to be heterozygous for the deletion or homozygous for the wild-type allele, respectively (Fig. 2d). In family G we found a deleterious homozygous in-frame deletion in the C-terminal domain of *HELLS*, leading to the deletion of Leucine 801 (c.2400_2402delGTT;p.[Leu801del]; Fig. 2e). The proband of family H carries a nonsense mutation (c.610A > T;p.[Lys204*]) and a duplication causing the insertion of a stop codon (c.374_381dup;p.[Lys128*]), suggesting that absence of HELLS is compatible with human life, whereas *Hells*$^{-/-}$ mice die perinatally[16]. Different allelic origin of the mutations was confirmed in parental DNA (Fig. 2f).

ICF genes converge at centromeric DNA methylation regulation. In mouse, HELLS is required for T-cell proliferation and mediates *de novo* DNA methylation, through its interaction with DNMT3B, dependent on its ATPase domain[16–19]. Genome-wide loss of CpG methylation, including centromeric repeats, has been observed in *Hells*$^{-/-}$ mice, reminiscent of what has been described in *Dnmt3b* knockout mice and in mouse models for ICF1 (refs 20,21). We show that transient depletion of HELLS, CDCA7 and ZBTB24, but not DNMT3B, resulted in decreased CpG methylation at centromeric repeats in wild-type mouse embryonic fibroblasts (MEFs, Fig. 2g, Supplementary Fig 3a,b). This confirms that DNMT3B acts during establishment of centromeric CpG methylation[22]. Moreover, it supports a role for ZBTB24 and CDCA7 in maintenance of CpG methylation at centromeric repeats, and, combined with previously published work, suggests that HELLS may be involved in both processes[19].

By identifying two new ICF syndrome genes this study highlights the genetic heterogeneity of the syndrome. The identification of at least one additional disease gene is expected with still a few cases remaining genetically unresolved. The complex, but highly overlapping, pathophysiology suggests that all ICF genes act in common or converging pathways involved in immunity, chromatin regulation and development. Convergence is supported by hypomethylation of pericentromeric repeats, which is common to all ICF subgroups, however the result of different defective pathways in the establishment and/or maintenance of CpG methylation.

Methods

Patients. All samples were obtained in an anonymized manner, and all families gave consent for genetic analyses. The study was approved by the Medical Ethical Committee of the Leiden University Medical Center, the local ethics committee of Necker-Enfants Malades Hospital, Paris, France and the Kyushu University Institutional Review Board for Human Genome/Gene Research. The patient from family A (referred to as patient 2 in ref. 23), patients from families B and C (referred to as pC and pS, respectively, in ref. 12), patient 2.2 from family D and the patients in family E (referred to as P4 and P8/P9 in ref. 5), as well as the three remaining ICFX cases (referred to as pN and P1 in ref. 12 and P4 in ref. 24) all show typical clinical features of ICF syndrome including hypo- or agammaglobulinemia in the presence of B cells, combined with the classical cytogenetic abnormalities involving chromosomes 1, 9 and 16, and hypomethylation of α-satellite DNA in addition to Sat II hypomethylation. The critical features of all patients are summarized in Supplementary Table 1; further details can be found in previous descriptions of these patients[2,5,6,12,23,24].

Gene identification by homozygosity mapping and sequencing. Homozygosity mapping was performed using the Sentrix HumanHap-300 Genotyping BeadChips (Illumina). To this end, 750-ng genomic DNA was processed with the Infinium II Whole-Genome Genotyping Assay (Illumina). After DNA amplification, fragmentation, precipitation and resuspension, DNA was applied to the BeadChip and incubated overnight, followed by enzymatic base extension, fluorescent staining of the beads and detection of fluorescent intensities using the BeadArray Reader (Illumina). To identify regions of homozygosity, B allele frequencies were assessed for all single-nucleotide polymorphisms using BeadStudio version 3.2 (Illumina). For a subset of patients, whole-exome sequencing was performed in the Neuromics project by deCODE Genetics (Reykjavik—Iceland) and analysed using deCODE Clinical Sequence Miner. Recessive analysis for multiple cases and controls and gene variant effect count (with variant effect predictor consequences moderate to high) were used to identify possible recessive mutations. For an additional set of patients, DNA libraries for whole-exome sequencing were constructed using the SureSelect Human All Exon V5 kit (Agilent Technologies) according to the manufacturer's instructions. Sequencing was performed on the Illumina HiSeq2500 platform to generate 100 bp paired-end reads. Reads were mapped to the reference human genome (UCSC hg19) with the Burrows-Wheeler Alignment tool (BWA v0.7.4)[25]. Duplicate reads were removed by Picard (v1.87). We called SNVs and indels using the Genome Analysis Toolkit (GATK v2.5-2)[26]. Annotations of variants were made using Annotate Variation (ANNOVAR)[27]. For confirmation, relevant exons and flanking sequences of *CDCA7* and *HELLS* were amplified using standard PCR; products were purified and analysed using Sanger sequencing. Sequence tracks were analysed and visualized using ContigExpress (Vector NTI, Invitrogen-Life Technologies). PCR primers are listed in Supplementary Table 2.

Cycloheximide treatment and RT–PCR analysis. Early-passage ($P < 6$) patient-derived primary fibroblasts were maintained in DMEM F12 (31331) supplemented with 20% heat-inactivated fetal calf serum, 1% penicillin–streptomycin, 1% sodium pyruvate and 1% HEPES (all Invitrogen-Life Technologies). Cells were treated with 250 μg ml^{-1} cycloheximide (dissolved in ethanol, 01810, Sigma-Aldrich) for 4 h using equal volumes of ethanol as control. After treatment, cells were harvested in Qiazol lysis reagent and RNA was isolated using the miRNeasy mini kit (both Qiagen) all according to the manufacturer's instructions. Total RNA (2 μg) was used for random primed cDNA synthesis using the RevertAid first-strand cDNA synthesis kit (Thermo Scientific). Transcripts were amplified using standard PCR with primers listed in Supplementary Table 2, separated by standard gel electrophoresis and sequenced by Sanger's sequencing.

Knockdown of gene expression in MEFs. The use of animal work has been reviewed by the Animal Experimentation Ethical Committee Buffon (CEEA-40), Paris, France, and was approved under the number CEB-06-2012. Female C57BL/6N pregnant mice, aged 3–6 months, were killed at 12.5 days post conception, and individual MEF clones were isolated from each embryo of the litter. Primary MEFs were cultivated for no more than two to three passages in complete media (DMEM glutamax supplemented with 10% fetal bovine serum, 100 U ml^{-1} penicillin, 100 μg ml^{-1} streptomycin, all from Life Technologies). Two rounds of transfection with synthetic short interfering RNAs (siRNAs) at a final concentration of 20 nM were performed using Interferin transfection reagent (Polyplus transfection) following the manufacturer's instructions: a first round on MEFs in suspension and the second when cells were allowed to adhere to the plate. Cells were collected after 48 h for genomic DNA extraction. Sequences of siRNAs used in the study are listed in Supplementary Table 2.

© 2015 Macmillan Publishers Limited. All rights reserved.

Satellite DNA methylation analysis using Southern blot analysis. Satellite II and α-satellite methylation in whole-blood DNA was analysed using Southern blot analysis of 2 μg genomic DNA digested with methylation-sensitive restriction enzymes HhaI for analysis of Sat-α or BstBI for analysis of Satellite II repeats. Both enzymes were purchased from Fermentas. After overnight digestion, DNA was separated by electrophoresis using 0.8% agarose gels. DNA was transferred overnight to Hybond-N+ membranes (GE Healthcare) and hybridized with a radioactive probe recognizing the α-satellite and satellite II repeats, respectively. Signals were detected by phosphoimaging or by exposure to X-ray films. For analysis of murine satellite repeats, genomic DNA was extracted and purified from MEFs using the NucleoSpin Tissue kit (Macherey-Nagel) according to the manufacturer's instructions. The DNA pellet was eluted in Tris EDTA buffer containing 20 μg ml^{-1} RNAse A. Genomic DNA from MEFs (500 ng) was digested with 20 units of MspI or HpaII (New England Biolabs) for 16 h to analyse the DNA methylation patterns of centromeric minor satellite repeats. The digested DNA fragments were separated by electrophoresis using 1% agarose gels and transferred overnight to Hybond-N+ membranes (GE Healthcare) in 20× SSC. After ultraviolet crosslink, the membranes were pre-hybridized in 6× SSC, 5× Denhardt and 0.1% SDS and then hybridized with 32P-labelled minor satellite oligonucleotide probe: (5′-ACATTCGTTGGAAACGGGATTTGTAGAACAGTG TATATCAATGAGTTACAATGAGAAACAT-3′).

Pre-hybridization and hybridization were carried out at 42 °C for 1 h. The membranes were washed three times in 6× SSC and 0.1% SDS at 37 °C and signals were detected by phosphorimaging using FLA 7000 phosphorimager (Fuji). Uncropped scans of the Southern blots are presented in Supplementary Figs 4 and 5.

Quantification of knockdown efficiency using qRT-PCR. Total RNA from MEFs was isolated using TRIzol Reagent (Life Technologies) according to the manufacturer's instructions. Contaminant genomic DNA was eliminated with TURBO DNA-free kit (Ambion). Reverse transcription was carried out using 1 μg DNA-free RNA and 50 μM random hexamers, 20 U of RNase Out and 100 U of RevertAid reverse transcriptase (Life Technologies). Complementary DNA reactions were used as templates for PCR reactions. Real-time PCR was performed using the light cycler-DNA MasterPLUS SYBR Green I mix (Thermo Scientific) supplemented with 0.5 μM of specific primer pairs (listed in Supplementary Table 2). Real-time qPCRs were run on a light cycler rapid thermal system (Light Cycler 480 2.0 Real time PCR system, Roche) with 20 s of denaturation at 95 °C, 20 s of annealing at 60 °C and 20 s of extension at 72 °C for all primers, and analysed by the comparative CT ($\triangle C_t$) method.

References

1. Hagleitner, M. M. *et al.* Clinical spectrum of immunodeficiency, centromeric instability and facial dysmorphism (ICF syndrome). *J. Med. Genet.* **45,** 93–99 (2008).
2. Weemaes, C. M. *et al.* Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects. *Eur. J. Hum. Genet.* **21,** 1219–1225 (2013).
3. Jeanpierre, M. *et al.* An embryonic-like methylation pattern of classical satellite DNA is observed in ICF syndrome. *Hum. Mol. Genet* **2,** 731–735 (1993).
4. Tiepolo, L. *et al.* Multibranched chromosomes 1, 9, and 16 in a patient with combined IgA and IgE deficiency. *Hum. Genet.* **51,** 127–137 (1979).
5. de Greef, J. C. *et al.* Mutations in ZBTB24 are associated with immunodeficiency, centromeric instability, and facial anomalies syndrome type 2. *Am. J. Hum. Genet.* **88,** 796–804 (2011).
6. Jiang, Y. L. *et al.* DNMT3B mutations and DNA methylation defect define two types of ICF syndrome. *Hum. Mutat.* **25,** 56–63 (2005).
7. Auclair, G., Guibert, S., Bender, A. & Weber, M. Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. *Genome Biol.* **15,** 545 (2014).
8. Gowher, H. & Jeltsch, A. Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases. *J. Biol. Chem.* **277,** 20409–20414 (2002).
9. Moarefi, A. H. & Chedin, F. ICF syndrome mutations cause a broad spectrum of biochemical defects in DNMT3B-mediated *de novo* DNA methylation. *J. Mol. Biol.* **409,** 758–772 (2011).
10. Nitta, H. *et al.* Three novel ZBTB24 mutations identified in Japanese and Cape Verdean type 2 ICF syndrome patients. *J. Hum. Genet.* **58,** 455–460 (2013).
11. Lee, S. U. & Maeda, T. POK/ZBTB proteins: an emerging family of proteins that regulate lymphoid development and function. *Immunol. Rev.* **247,** 107–119 (2012).
12. Velasco, G. *et al.* Germline genes hypomethylation and expression define a molecular signature in peripheral blood of ICF patients: implications for diagnosis and etiology. *Orphanet. J. Rare. Dis.* **9,** 56 (2014).
13. Gill, R. M., Gabor, T. V., Couzens, A. L. & Scheid, M. P. The MYC-associated protein CDCA7 is phosphorylated by AKT to regulate MYC-dependent apoptosis and transformation. *Mol. Cell Biol.* **33,** 498–513 (2013).
14. Guiu, J. *et al.* Identification of Cdca7 as a novel Notch transcriptional target involved in hematopoietic stem cell emergence. *J. Exp. Med.* **211,** 2411–2423 (2014).
15. Chen, K., Ou, X. M., Wu, J. B. & Shih, J. C. Transcription factor E2F-associated phosphoprotein (EAPP), RAM2/CDCA7L/JPO2 (R1), and simian virus 40 promoter factor 1 (Sp1) cooperatively regulate glucocorticoid activation of monoamine oxidase B. *Mol. Pharmacol.* **79,** 308–317 (2011).
16. Geiman, T. M. & Muegge, K. Lsh, an SNF2/helicase family member, is required for proliferation of mature T lymphocytes. *Proc. Natl Acad. Sci. USA* **97,** 4772–4777 (2000).
17. Ren, J. *et al.* The ATP binding site of the chromatin remodeling homolog Lsh is required for nucleosome density and de novo DNA methylation at repeat sequences. *Nucleic Acids Res.* **43,** 1444–1455 (2015).
18. Yu, W. *et al.* Genome-wide DNA methylation patterns in LSH mutant reveals de-repression of repeat elements and redundant epigenetic silencing pathways. *Genome Res.* **24,** 1613–1623 (2014).
19. Zhu, H. *et al.* Lsh is involved in *de novo* methylation of DNA. *EMBO J.* **25,** 335–345 (2006).
20. Ueda, Y. *et al.* Roles for Dnmt3b in mammalian development: a mouse model for the ICF syndrome. *Development* **133,** 1183–1192 (2006).
21. Velasco, G. *et al.* Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proc. Natl Acad. Sci. USA* **107,** 9281–9286 (2010).
22. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell* **99,** 247–257 (1999).
23. Kloeckener-Gruissem, B., Betts, D. R., Zankl, A., Berger, W. & Gungor, T. A new and a reclassified ICF patient without mutations in DNMT3B and its interacting proteins SUMO-1 and UBC9. *Am. J. Med. Genet. A* **136,** 31–37 (2005).
24. Kubota, T. *et al.* ICF syndrome in a girl with DNA hypomethylation but without detectable DNMT3B mutation. *Am. J. Med. Genet. A* **129A,** 290–293 (2004).
25. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26,** 589–595 (2010).
26. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).
27. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38,** e164 (2010).

Acknowledgements

We thank all patients and their families for their consent to contribute to our study. We thank K. Ichiyanagi and L. Daxinger for helpful advice, staff of the Kyushu University International Legal Office for help in legal and ethical affairs, and T.P. Pham, S.L. Höcker, M. Miyake and H. Furuumi for technical assistance. This work was supported by KAKENHI (22134006, 23249019 and 26253020), the National Institutes of Health/National Institute of Allergy and Infectious Diseases (R21 AI090135), European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 2012-305121 'Integrated European—omics research project for diagnosis and therapy in rare neuromuscular and neurodegenerative diseases (NEUROMICS)' and French National Research Agency—Program on physiopathology of rare diseases (ANR-09-GENO-035).

Author contributions

P.E.T., Y.I., G.V., J.W., J.C.d.G., C.W., C.F., S.M.v.d.M. and H.S. conceived and designed the experiments; P.E.T., Y.I., G.G., J.W., G.V., H.N., M.Y., M.S., J.C.d.G., M.M.v.O.-T.D. and M.J.D.v.T. performed the experiments; P.T., Y.I., G.G., J.W., G.V., H.N., M.Y., M.S., Y.S., R.J.L.F.L., J.C.d.G., M.M.v.O.-T.D. and M.J.D.v.T. analysed the data; M.U., A.G., P.P., B.K.-G., T.G., I.R., C.P., K.K., B.R., T.I., T.K., M.M.v.O.-T.D., M.J.D.v.T. and C.W. contributed samples; P.T., C.F., S.M.v.d.M. and H.S. wrote the paper; C.F., S.M.v.d.M. and H.S. jointly supervised the research.

Additional information

Accession Codes: Sequence data have been deposited to the Leiden Open Variation Database (LOVD): http://databases.lovd.nl/shared/individuals/CDCA7 (IDs CDCA7_000001, CDCA7_000002, CDCA7_000003, CDCA7_000004), http://databases.lovd.nl/shared/individuals/HELLS (IDs HELLS_000001, HELLS_000002, HELLS_000003, HELLS_000004, HELLS_000005, HELLS_000006).

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

© 2015 Macmillan Publishers Limited. All rights reserved.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Thijssen, P. E. *et al.* Mutations in *CDCA7* and *HELLS* cause immunodeficiency–centromeric instability–facial anomalies syndrome. *Nat. Commun.* 6:7870 doi: 10.1038/ncomms8870 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© 2015 **Macmillan Publishers Limited. All rights reserved.**

nature COMMUNICATIONS

DOI: 10.1038/ncomms12003 OPEN

Corrigendum: Mutations in *CDCA7* and *HELLS* cause immunodeficiency–centromeric instability–facial anomalies syndrome

Peter E. Thijssen, Yuya Ito, Giacomo Grillo, Jun Wang, Guillaume Velasco, Hirohisa Nitta, Motoko Unoki, Minako Yoshihara, Mikita Suyama, Yu Sun, Richard J.L.F. Lemmers, Jessica C. de Greef, Andrew Gennery, Paolo Picco, Barbara Kloeckener-Gruissem, Tayfun Güngör, Ismail Reisli, Capucine Picard, Kamila Kebaili, Bertrand Roquelaure, Tsuyako Iwai, Ikuko Kondo, Takeo Kubota, Monique M. van Ostaijen-Ten Dam, Maarten J.D. van Tol, Corry Weemaes, Claire Francastel, Silvère M. van der Maarel & Hiroyuki Sasaki

Nature Communications 6:7870 doi: 10.1038/ncomms8870 (2015); Published 28 Jul 2015; Updated 22 Jun 2016

In Fig. 1b of this Article, the sex of patient 2.1 in family A is incorrect, and should be depicted as male. The correct version of this figure appears below.

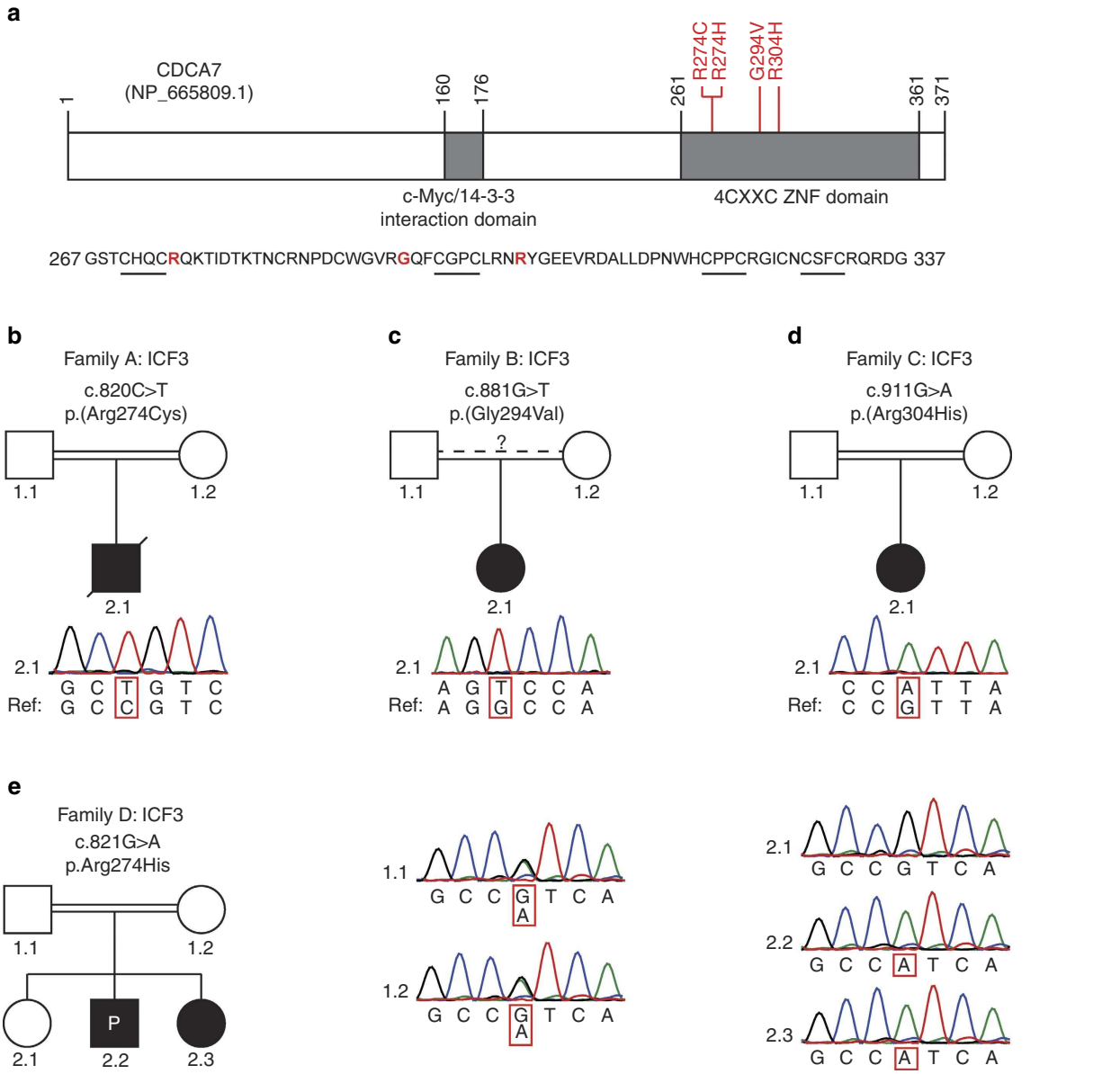
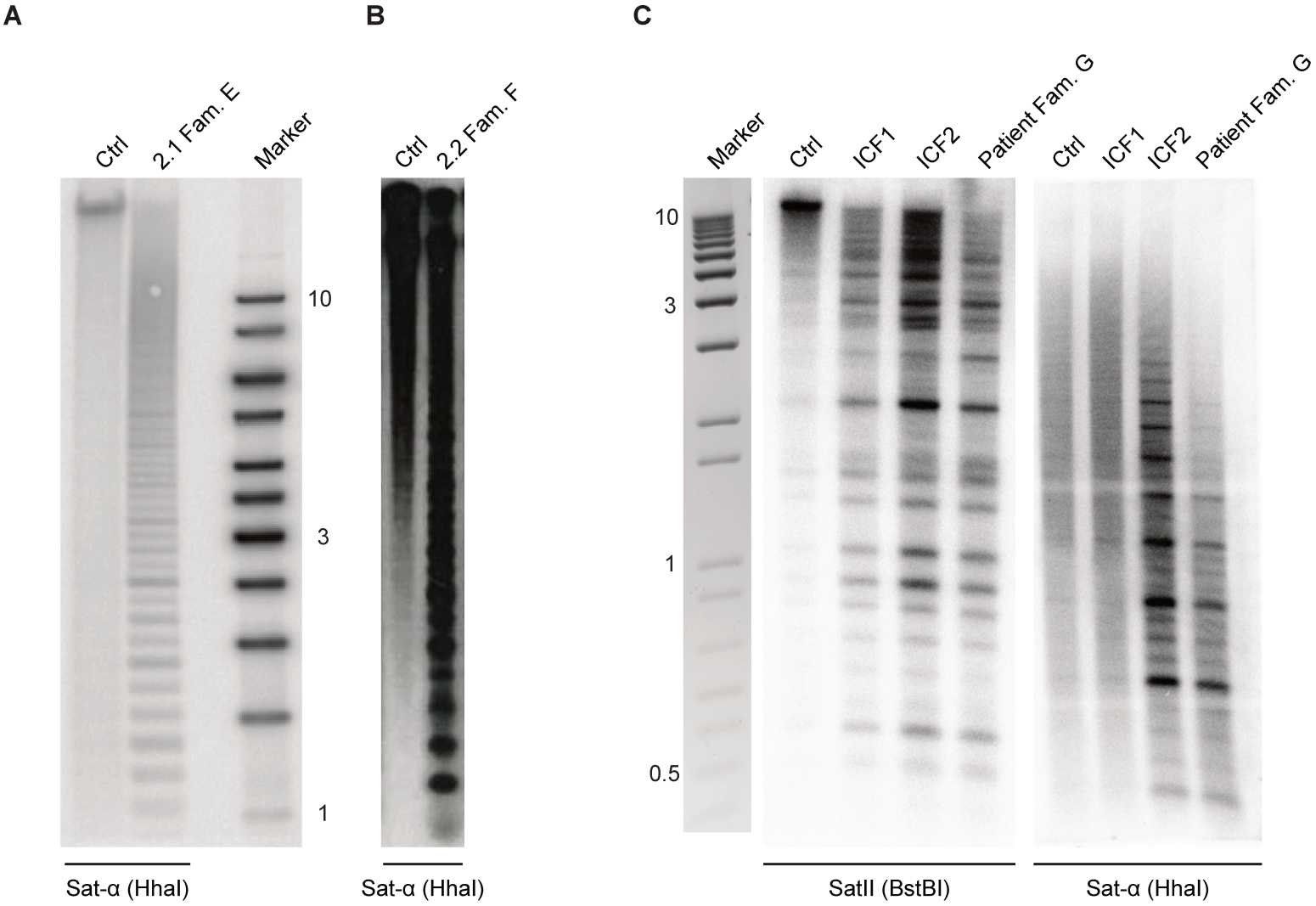
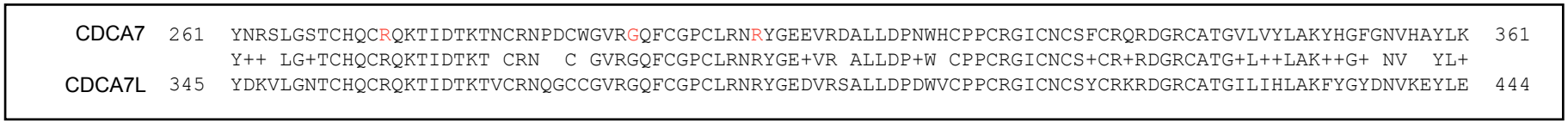


Figure 1

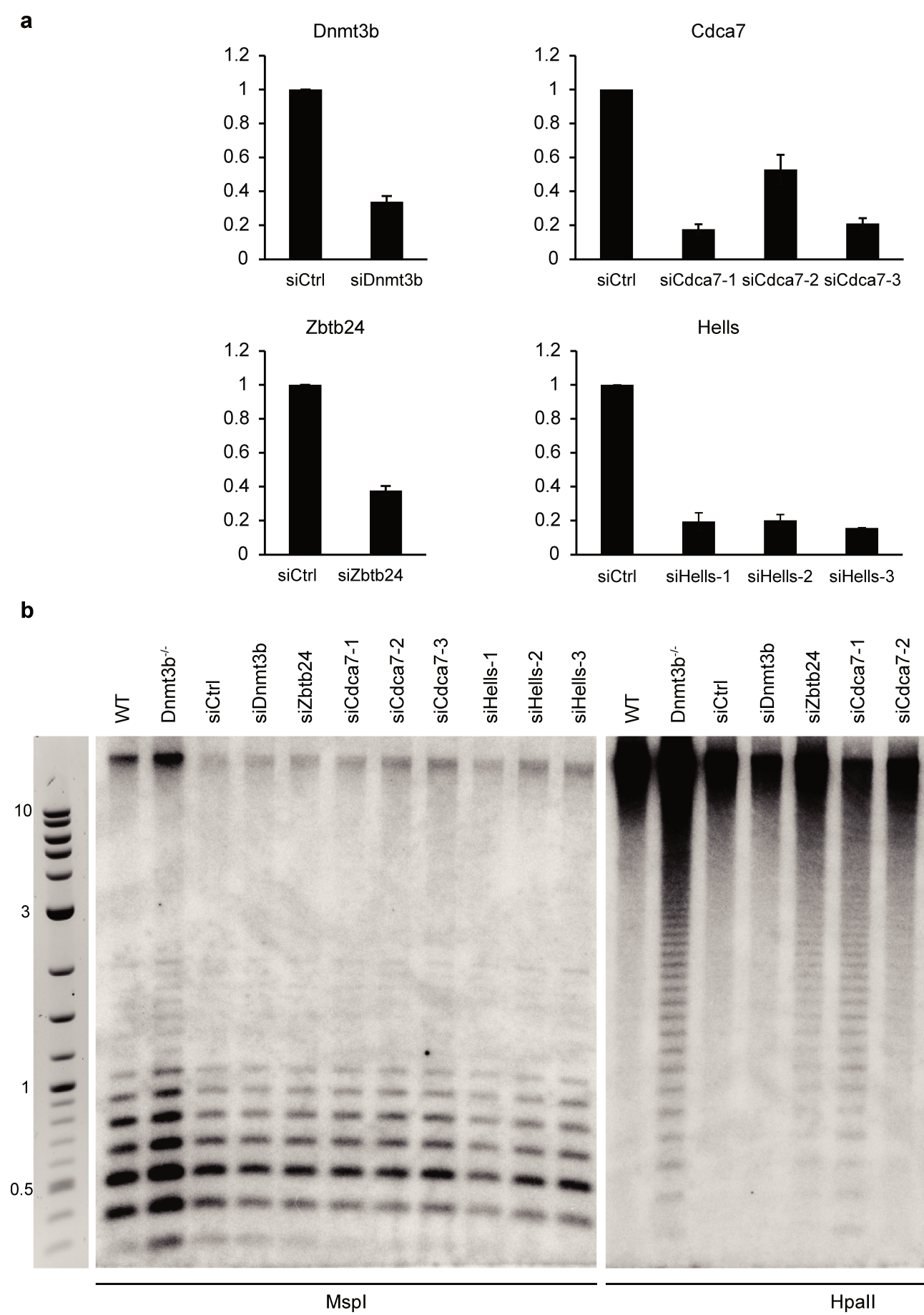
This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/



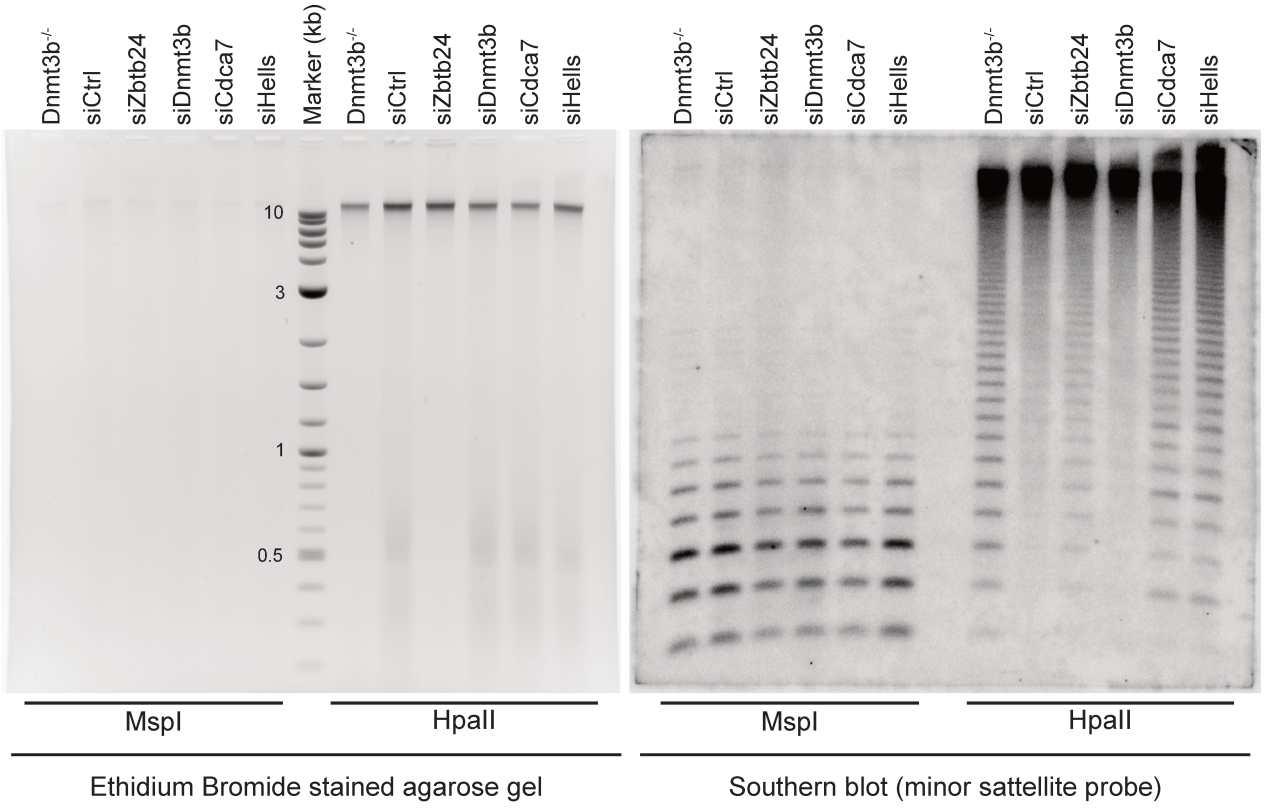
Supplementary Figure 1: Repeat hypomethylation in ICF patients from families E, F and G
(a, b) Sat-α methylation analysis using HhaI digestion followed by Southern blot analysis of healthy controls and patients 2.1 from family E and 2.2 from family F showed hypomethylation in the patient derived samples. **(c)** SatII (BstBI) and Sat-α (HhaI) methylation analysis by Southern blot analysis in DNA derived from a healthy control, an ICF1 patient, an ICF2 patient and the patient from family G confirmed SatII hypomethylation in all ICF samples and Sat-α hypomethylation in ICF2 and ICFX samples. Molecular weights of the DNA size markers are in kilo-basepairs.



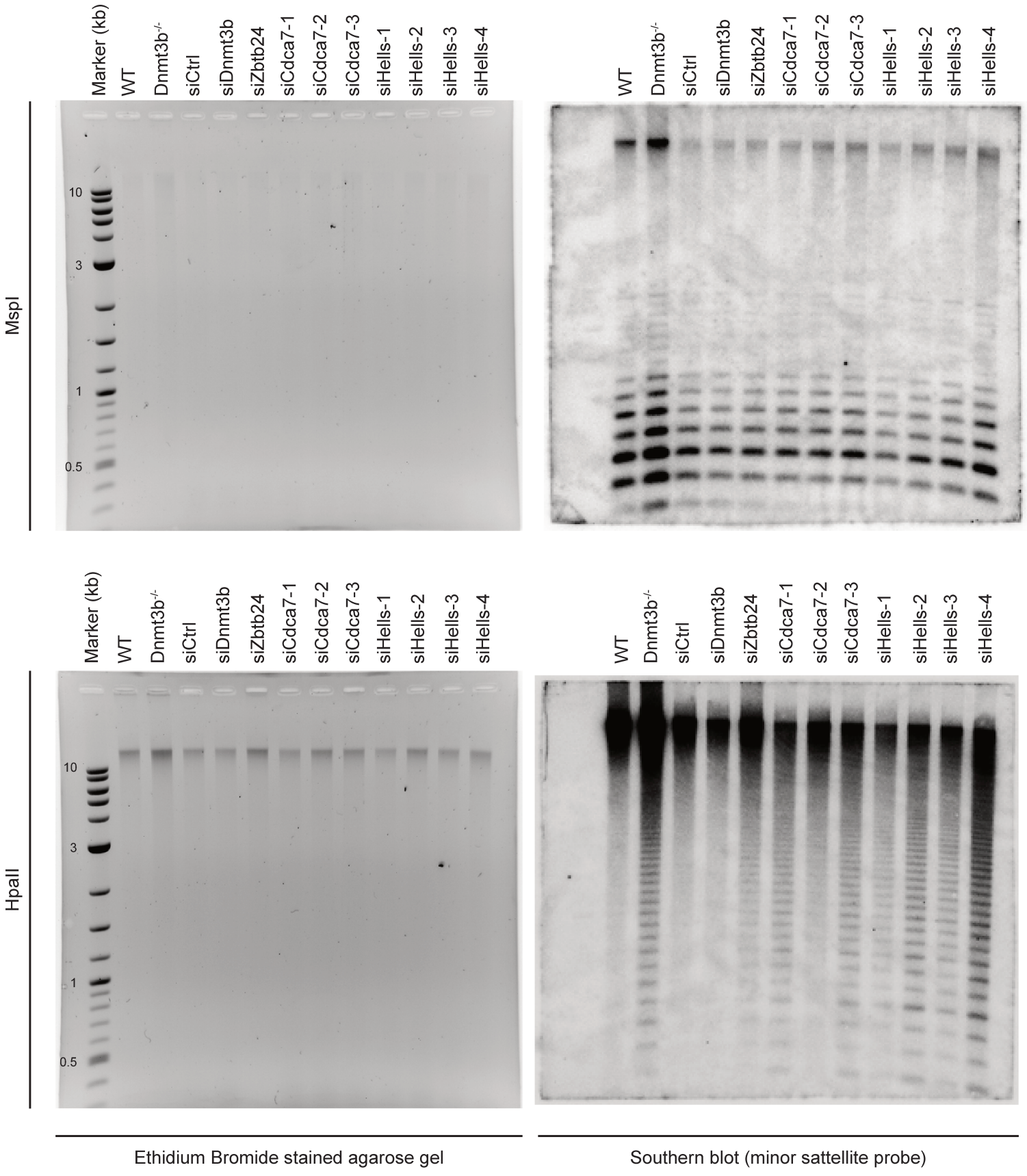
Supplementary figure 2: Aligment of the 4-CXXC domains of CDCA7 and CDCA7L
Alignment using blastp (NCBI) showed 76% sequence identity between the 4-CXXC domain of CDCA7 with that of CDCA7L. Identified mutations are highlighted in red.



Supplementary figure 3: confirmation of minor satellite hypomethylation through siRNA mediated knockdown in wildtype MEFs (a) qRT-PCR quantification of mRNA expression of the indicated transcript upon knockdown using the indicated siRNAs, siCtrl levels are set to 1, error bars indicate SEM of average knockdown levels of at least two independent experiments. **(b)** Replication and extension of the data in figure 2g showing minor satellite DNA hypomethylation in wild type MEFs upon knockout of Dnmt3b and knockdown of Zbtb24, Cdca7 and Hells with different siRNA's. Molecular weights of the 2-Log DNA size marker are in kilo-basepairs.



Supplementary Figure 4: Uncropped images of agarose gels and Southern blots
Uncropped images corresponding to Figure 2g. Left panel: ethidium bromide stained agarose gel, right panel: Southern blot probed for minor satellites. Molecular weights of the 2-Log DNA size marker are in kilo-basepairs.



Supplementary Figure 5: Uncropped images of agarose gels and Southern blots
Uncropped images corresponding to Supplementary Figure 3b. Left panels: ethidium bromide stained agarose gels, right panels: Southern blots probed for minor satellites. Upper panels: MspI digestion, lower panels: HpaII digestion. Molecular weights of the 2-Log DNA size marker are in kilo-basepairs.

Supplementary Table 1: demographic and clinical characteristics of patients

ICF type	ICF3	ICF3	ICF3	ICF3	ICF3	ICF4	ICF4	ICF4	ICF4	ICF4	ICF1	ICF2
Family (individual)	A	B	C	D (2.2)	D (2.3)	E (2.1)	E (2.2)	F	G	H	*n*= 23	*n*= 13
Gene defect	*CDCA7*	*CDCA7*	*CDCA7*	*CDCA7*	*CDCA7*	*HELLS*	*HELLS*	*HELLS*	*HELLS*	*HELLS*	*DNMT3B*	*ZBTB24*
Mutation Paternal allele	c.820C>T	c.881G>T	c.911G>A	c.821G>A	c.821G>A	c.2096A>G	c.2096A>G	c.2283_2286delGTCT	c.2400_2402delGTT	c.610A>T		
	p.(Arg274Cys)	p.(Gly294Val)	p.(Arg304His)	p.Arg274His	p.Arg274His	p.Gln699Arg	p.Gln699Arg	p.(Ser762Argfs*4)	p.(Leu801del)	p.(Lys204*)		
ExAC allele frequency	0	0	8.24E-06	8.24E-06	8.24E-06	0	0	0	8.26E-06	0		
Provean score	-7.306	-8.219	-4.566	-4.566	-4.566	-3.825	-3.825	N/A	-9.303	N/A		
PolyPhen-2 score	prob. damag.	prob. damag.	prob. damag.	prob. damag.	prob. damag.	prob. damag.	prob. damag.	N/A	N/A	N/A		
Mutation Maternal allele	c.820C>T	c.881G>T	c.911G>A	c.821G>A	c.821G>A	c.370+2T>A	c.370+2T>A	c.2283_2286delGTCT	c.2400_2402delGTT	c.374_381dup		
	p.(Arg274Cys)	p.(Gly294Val)	p.(Arg304His)	p.Arg274His	p.Arg274His	skip ex5	skip ex5	p.(Ser762Argfs*4)	p.(Leu801del)	(p.Lys128*)		
ExAC allele frequency	0	0	8.24E-06	8.24E-06	8.24E-06	0	0	0	8.26E-06	0		
Provean score #	-7.306	-8.219	-4.566	-4.566	-4.566	N/A	N/A	N/A	-9.303	N/A		
PolyPhen-2 score *	prob. damag.	prob. damag.	prob. damag.	prob. damag.	prob. damag.	N/A	N/A	N/A	N/A	N/A		
Sex	M	F	F	M	F	F	M	F	M	F		
Birth year	1982	1998	1978	2005	2012	2000	2007	2002	1999	2010		
Deceased (age in yrs)	yes (26)							yes (1.2)				
Consanguineous (reported)	yes	suspected	yes	yes	yes	no	no	unknown	yes	unknown		
Origin	Turkish	French	French	Turkish	Turkish	English	English	Japanese	French	Italian		
Birth weight (grams)	1960	730	1820	2800	2250	2400	3100	1770	3830	900	<p3 (3/20) ‡	<p3 (1/7) ‡
Motor developmental delay	yes	yes		no	no	no	yes			yes	9/16	7/8
Mental retardation	yes	yes	no	no	no	yes (mild)	no		yes	yes	9/20	13/13
Facial anomalies	yes	yes	no	yes	yes	yes	yes	yes	yes	yes	21/22	13/13
Gastrointestinal problems	yes	yes	yes	no	yes	yes	no	no	no	no	7/14	2/6
Infections	otitis pneumonia	otitis pneumonia myocarditis peritonitis	otitis pneumonia mastoiditis meningitis	otitis	pneumonia Candida	pneumonia Pneumocystis	no	pneumonia Pneumocystis	otitis pneumonia	bronchitis	8/13 16/16 Pneumocystis (n=2) Candida (n=4)	2/6 5/7 Pneumocystis (n=2) Candida (n=2)
Cytogenetic abnormalities Chr 1, 9 and 16	Stretching Multiradial configuration	Stretching Associations Rosettes Branching Micronuclei Decondensation	Stretching	Stretching Breaking Decondensation		Isochromosomes Multiradial configuration		Stretching Associations Rosettes Branching Micronuclei Decondensation	Translocations Deletions	Rosettes		
α-satellite hypomethylation	N/A	Yes (ref. 13)	Yes (ref. 13)	Yes (ref. 7)	N/A	Yes (this study)	N/A	Yes (this study)	Yes (this study)	N/A	No	Yes
Treatment						stem cell transplantation	stem cell transplantation					
IgG (g/l)	2.8	0.31	2	2.32	1.11	< 0.33	8.05 (maternal)	0.23	0.91	IVIG		
IgA (g/l)	< 0.04	0.05	0.27	< 0.07	0.06	< 0.07	< 0.3	0.06	< 0.04	< 0.04		
IgM (g/l)	0.23	< 0.06	0.25	0.03	0.04	0.10	< 0.22	0.05	0.04	< 0.04		

#: Protein Variation Effect Analyzer (provean.jcvi.org/)
*: Polymorphism Phenotyping v2 (genetics.bwh.harvard.edu/pph2/)
‡: Birthweight less than the third percentile for gestational age

Supplementary table 2: List of primers and siRNAs

List of primers:

human exons:	**Fw primer:**	**Rev primer:**	**Mutation:**
CDCA7 exon 8	GGACATTCTGTAAATATGCCATTTC	TCCACCCCTTCACCAATG	c.820C>T; c.821G>A
CDCA7 exon 8	GCAAGCCATTGTGATTGGTT	CATGTGGCTGAATGAATTGC	c.881G>T; c.911G>A
HELLS exon 5-6	GGCTCAATTGCAGGAAACAT	TTAAAACCTACATTTTACTCAACATCA	c.370+2T>A
			c.374_381dup
HELLS exon 8	GCTTCCCAAAGTGCTGTGAT	GCCATAAAAGTGCTGGAAGC	c.610A>T
HELLS exon 20	CCTGGCTAGATCATTCATGTTG	ACTCGGCCCAGGCTGTAATA	c.2096A>G
HELLS exon 21-22	CCGTCCTTCTTAATGCAAGC	CATGTGGCTGAATGAATTGC	c.2283_2286delGTCT
			c.2400_2402delGTT

Human transcripts:	**Fw primer:**	**Rev primer:**
HELLS ex3-5	TGACGAAAATGGAACAGCAA	CTCTTTTGACATGACCTCTGAAAT

Murine transcripts:	**Fw primer:**	**Rev primer:**
Dnmt3b	CGCCACCATGTGCAGGAGTAC	GACGCTCTTAGGTGTCACTTCTTCC
Zbtb24	GCTGGAGCACATGAGCCTG	TTGGCACTGGAGTCAGAGAAG
Cdca7	GACAGCTACATGAACGACGATG	GATGACATGTAGAACCCAGAGAG
Hells	ACACGAGCTGGTGGTTTAGG	GCCTGGAGATCAGACTGAGG

List of siRNA's

Target gene:	**siRNA sequence (5' -> 3')**
Dnmt3b	GGCUGUUCAGCCAGCACUUUAAU
Zbtb24	GAAGCCAUUUACUUGUGAA
Cdca7-1	AUGACAGUUGUGACAGCUUU
Cdca7-2	CCUCUUACCAGGUCAAGGU
Cdca7-3	AGGAGGAAGAGGAUAAGUA
Hells-1	GAGAAGUUGAUAAUCCAUA
Hells-2	GUCUACACUUCCUAACUGG
Hells-3	GUCUGAUGUUGCUCUUGAA

ARTICLE 2

Comparative methylome analysis of ICF patients: ZBTB24, CDCA7 and HELLS are required for DNA methylation at heterochromatic loci

Giacomo Grillo, Nizar Touleimat, Laure Ferry, Ivana Ivkovic, Florence Ribierre, Capucine Picard, Sophie Chantalat, Jean-François Deleuze, Guillaume Velasco* and Claire Francastel*

Under review, PLOS Genetics

RAW DATA

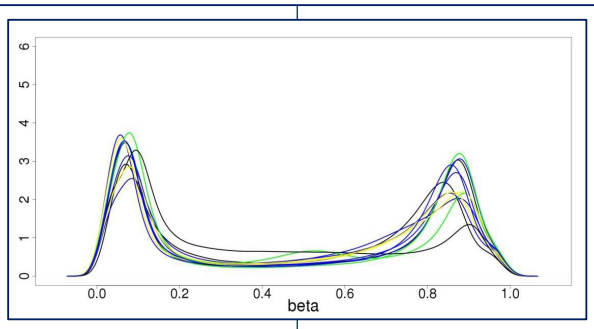
Sample and probe Quality Control

Color bias correction

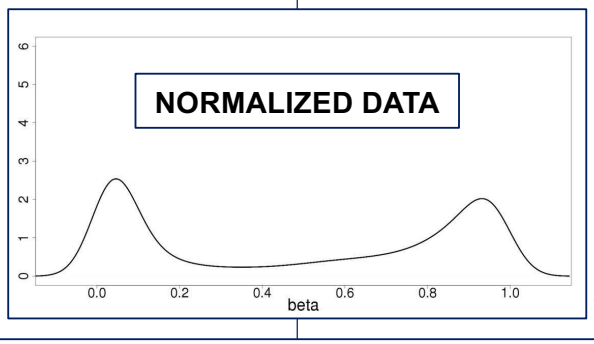
Background subtraction

Infinium I / Infinium II bias correction

Calculation of beta-values: beta = max(M,0)/[max(M,0)+max(U,0)+100]



Sample quantile NORMALIZATION



Touleimat & Tost, *2012*

Generation of Volcano plots and hierarchical clustering
Comparison of ICF subgroups relative to NA (median of probe beta-values)
Identification of DMRs (Δbeta-values ≥ 20% p-value<0,01)

With X and Y
ICF males
compared to NA males
(bed files)

Without Y
ICF females
compared to NA females
(bed files)

Without X and Y
All ICF
compared to all NA
(bed files)

Visualisation (IGV)
Intersect (Galaxy)
- Genomic tracks (UCSC Browser, Repeat Masker)
- Replication data (UW-Repli-Seq
- Chromain State Segmentation (ChromHMM)

The ICF syndrome is characterized by loss of DNA methylation at (peri)centromeric DNA repeats and chromosome instability (Ehrlich et al., 2006). ICF1 patients are characterized by mutations in DNMT3B (Xu et al., 1999). The discovery of mutations in factors devoid of the DNA methyltransferase activity, ZBTB24, CDCA7 and HELLS, as the cause of the ICF syndrome (deGreef et al., 2011; Thijssen et al., 2015), raised the question of the impact of mutations in these factors on the global levels of DNA methylation and their functional interplay, together and with DNMT3B. This discovery called for a comparative analysis of the perturbed DNA methylation landscapes in all the ICF subtypes, to evaluate the similarities and differences depending on the genotype and to identify their genomic and epigenomic contexts in normal cells. I provided a flow chart, shown on the left, to facilitate the understanding of the analysis.

We thus generated methylome maps, performing HumanMethylation450 BeadChip (HM450K), for ICF patients representing all four genotypes, using their uncultured blood cells. Strikingly, our methylome analysis clearly separated ICF1 patients from ICF2, ICF3 and ICF4 patients. While mutations in DNMT3B affected DNA methylation at CGI, mostly located on the inactive X chromosome and in the promoter region of germline genes, those in ZBTB24, CDCA7 and HELLS caused hypomethylation of CpG-poor regions, intergenic genomic locations and interspersed DNA repeats, and more generally, at genomic locations characterized by heterochromatin epigenomic features in normal cells.

Beyond the extensive list of DNA methylation biomarkers of dysfunction of any of the four factors in human diseases, our data strongly suggest a functional link between ZBTB24, CDCA7 and HELLS and heterochromatin domains of the human genome.

I contributed to this study analysing the normalised data, performing statistical analysis and participating in the preparation of the figures. I contributed to the validation of hypomethylated regions by MSRE-qPCR and BS. I performed the experiments elucidating the functional interplay between ZBTB24, CDCA7 and HELLS, including RT-qPCR in primary fibroblasts and LCLs derived from ICF patients, and transfection of tagged form of ZBTB24, CDCA7 and HELLS proteins in 293 cells, in order to elucidate their interaction and mutual stabilisation. Finally, I participate also to the draft of the manuscript.

This study is currently under review in *PLOS Genetics*.

Comparative methylome analysis of ICF patients: ZBTB24, CDCA7 and HELLS are required for DNA methylation at heterochromatic loci

Giacomo Grillo[1], Nizar Touleimat[2], Laure Ferry[1], Ivana Ivkovic[1], Florence Ribierre[2], Capucine Picard[3-6], Sophie Chantalat[2], Jean-François Deleuze[2], Guillaume Velasco[1,*] and Claire Francastel[1,*]

[1] Université Paris Diderot-Paris7, CNRS UMR7216, Epigénétique et Destin Cellulaire, Paris, France

[2] Centre National de Génotypage, Institut de Génomique, CEA, 91057 Evry, France

[3] Study Center for Primary Immunodeficiencies, Necker-Enfants Malades Hospital, Assistance Publique Hôpitaux de Paris (APHP), University Paris Descartes, 149 rue de Sèvres, 75015, Paris, France.

[4] Pediatric Immuno-Hematology Unit, Necker Children's Hospital, APHP, Université Paris Descartes, Paris, France.

[5] Laboratory of the Human Genetics of Infectious Diseases, INSERM UMR1163, Imagine Institute, Necker Medical School, Sorbonne Paris Cité, University Paris Descartes, Paris, France.

[6] French National Reference Center for Primary Immune Deficiencies (CEREDIH), Necker - Enfants Malades Hospital, Assistance Publique Hôpitaux de Paris, Paris, France.

[*] Corresponding authors

UMR7216 - Epigénétique et Destin Cellulaire

Université Paris 7 Diderot - Bâtiment Lamarck - 4ème étage

Case Courrier 7042

35, rue Hélène Brion

75205 Paris Cedex 13

Tél: +33 1 57 27 89 18

Fax: +33 1 57 27 89 10

e-mail: claire.francastel@univ-paris-diderot.fr; guillaume.velasco@univ-paris-diderot.fr

Short tittle: Methylome analysis in ICF patients with four different genotypes

Abstract

Perturbations of DNA methylation and its machinery are hallmarks of many human diseases. A special case is the ICF syndrome, the first inherited disease shown to originate from mutations in a DNA methyltransferase, DNMT3B. ICF is a rare autosomal recessive disorder diagnosed by the loss of DNA methylation at (peri)centromeric repeats and its associated chromosomal instability, and primary immunodeficiency. However, only a subset of patients carry mutations in DNMT3B (ICF1). In the remainder, the puzzling identification of mutations in factors devoid of methyltransferase activity, i.e. ZBTB24 (ICF2), CDCA7 (ICF3) and HELLS (ICF4), suggested that DNA methylation alterations downstream of the mutations in these genes would identify common target genomic regions and pathways. Here, we established the first comparative methylome maps from uncultured cells of ICF patients with all four genotypes, together with annotations of the perturbed methylation landscapes relative to CpG, genomic and epigenomic contexts in normal cells. Altogether, we provide evidence that, despite unifying hypomethylation of pericentromeric repeats that serves to establish diagnosis and a few common loci relevant for the etiology of the disease, methylation profiling clearly distinguished ICF1 from ICF2, 3 and 4 patients. In contrast to DNMT3B mutations, those in ZBTB24, CDCA7 and HELLS affect methylation at CpG-poor regions in intergenic genomic locations and interspersed DNA repeats, and more generally, at genomic locations characterized by heterochromatin epigenomic features in normal cells. More specifically, their integrity is required for the methylated status of a number of clusters of coding and non-coding genes, some of which are expressed in a monoallelic manner. Beyond providing an extensive list of biomarkers of dysfunction of the four factors relevant for human diseases, our study strongly supports that ZBTB24, CDCA7 and HELLS are required for the methylation status of heterochromatin domains of the human genome.

Author summary

Inherited monogenic disorders that disrupt components of the epigenetic machinery offer a unique opportunity to learn about (epi)genome maintenance. ICF syndrome is an immunological/neurological disorder with perturbed DNA methylation leading to chromosomal instability. It is caused by mutations in *de novo* DNA methyltransferase DNMT3B in some patients, but also in factors devoid of enzymatic activity (ZBTB24, CDCA7 or HELLS) in the remainder. Here, comparative methylation profiling of ICF patients with all four genotypes discriminated patients with mutations in these factors from those with mutations in DNMT3B. Combined to genomic and epigenomic annotations, it revealed domains with heterochromatin marks whose methylation status relies on the integrity of ZBTB24, CDCA7 and HELLS, among which are interspersed DNA repeats and monoallelically expressed clusters of coding and non-coding genes. Hence, in addition to biomarkers relevant for human diseases, our study unveiled new players required for DNA methylation of heterochromatin domains of the human genome.

Introduction

Cytosine DNA methylation is among the best-studied epigenetic modification in vertebrates and is essential for normal embryonic development [1,2]. After global erasure of DNA methylation patterns during early preimplantation development and germline formation, re-establishment of new patterns of DNA methylation requires the catalytic activity of DNMT3A and DNMT3B to add methyl-groups to unmodified DNA *de novo*, whereas DNMT1 is responsible for maintaining methylation patterns through cell divisions [3]. Hence, genome-wide distribution of DNA methylation is highly dynamic during early development. In contrast, once established, methylation patterns are very stable in somatic cells. Importantly, DNA methylation is not randomly distributed throughout the genome, and is preferentially found at repetitive sequences and transposable elements, imprinted loci, and inactive X-chromosome in female cells, where it is thought to control illicit transcription, recombination or transposition in somatic cells [4]. Against all odds, CpG dense regions known as CpG islands (CGI) and often located at gene promoters are largely devoid of DNA methylation, with the exception of CGIs found at promoters of genes that are developmentally regulated, associated with the germinal programs or linked to the inactive X chromosome in females [5]. Promoter methylation has been traditionally associated with gene repression although the link might not be that straightforward [6].

Given the pivotal role of DNA methylation in the control of gene expression and key biological processes, it is not surprising that perturbed DNA methylation patterns are hallmarks of many human diseases [7,8]. In this context, the DNA methylation machinery that classically comprises writers (DNMTs), erasers (the recently discovered TET enzymes involved in DNA demethylation [9], and readers (Methyl DNA Binding proteins) MBDs [10] is frequently perturbed. The existence of monogenic diseases caused by mutations in these factors has recently fueled the field with candidate players in DNA methylation pathways. Indeed, the search for causes of Immunodeficiency with Centromeric instability and Facial anomalies syndrome (ICF; OMIM no.

242860) first identified mutations in DNMT3B, and more recently in factors whose link with DNA methylation was not obvious, as the cause of perturbed DNA methylation patterns and the large clinical spectrum that affects patients.

ICF syndrome is a very rare autosomal recessive disorder, with less than 70 cases reported worldwide, which is most notably characterized by primary immunodeficiency [11]. Recurrent infections are the presenting symptom, usually in early childhood. More variable features include mild facial anomalies, intellectual disability, congenital malformations and developmental delay [12]. The ICF syndrome was referred to as a heterochromatin disease since hypomethylation markedly affects facultative heterochromatin on the inactive X chromosome in females and constitutive heterochromatin found at pericentromeric satellite repeats: Satellites type II (SatII) of chromosomes 1, and 16, and to a lesser extent Satellites type III (SatIII) of chromosome 9. Hypomethylation of satellite DNA is an invariant hallmark of the syndrome that leads to heterochromatin decondensation and chromosomal anomalies, which is used to establish diagnosis [13]. Mutations in the DNMT3B gene were the first identified genetic cause of the syndrome [14]. However, they accounted for only 55% of the cases, classified as ICF sub-type 1 (ICF1), suggesting a genetic heterogeneity consistent with a notable molecular heterogeneity whereby ICF patients without DNMT3B mutations exhibited additional hypomethylation at centromeric alpha-satellite repeats [15,16]. The clinical picture also highlighted distinct genotype-phenotype relationship with patients negative for DNMT3B mutations showing greater incidence of intellectual disability and patients with mutations in DNMT3B exhibiting a more severe immunodeficiency [17,18].

Search for the causes of the ICF syndrome in these patients identified three additional mutated genes: ZBTB24 (ICF2) [19], and more recently CDCA7 (ICF3) and HELLS (ICF4) [20], and contributed to the identification of new factors devoid of enzymatic activity but required for DNA methylation at DNA repeats and maintenance of genome stability. This discovery called for a comparative analysis of the perturbed methylation landscapes in ICF subtypes, to evaluate their

similarities and differences depending on genotype, and to document the genomic and epigenomic contexts affected by ICF mutations.

Results

Comparative analysis of DNA methylation landscapes distinguishes ICF1 from ICF2, 3 and 4 patients

We compared genome-wide DNA methylation profiles in uncultured cells from 15 ICF patients and 10 age- and gender-matched healthy controls. All four ICF genotypes were represented in our study, and patient clinical records are listed in S1 Table. We used the Illumina Infinium HumanMethylation450 BeadChip (HM450K) that assays >485,000 CpGs throughout the human genome covering almost 99% of RefSeq genes. After quality control and normalization (see Supporting Methods section) 479,259 probes were retained for further analysis. Probes were considered as differentially methylated between two genotypes when showing a difference in β-value of at least 0.2 (S1A Fig), a threshold recommended to detect differential methylation with 99% confidence from HM450K [21]. Differentially methylated probes (DMPs) between ICF and non-affected (NA) subjects represented 2 to 3% of total probes, of which between 68% and 82% where less methylated in ICF samples (S1B Fig). Comparison of DNA methylation values between ICF patients taken as a whole and NA subjects identified 44,555 DMPs on autosomes, of which a large fraction was hypomethylated (HypoMPs: 33,687 probes; 76% of total DMPs) compared to hypermethylated probes (HyperMPs: 10,868 probes; 24% of total DMPs) (S2A Fig), in agreement with the widespread DNA hypomethylation characteristic of the disease [13]. Unsurprisingly, unsupervised hierarchical clustering based on DNA methylation data clearly discriminated ICF from NA subjects (Fig 1A).

However, clustering also differentiated patients with mutations in DNMT3B (ICF1) from those with mutations in ZBTB24, CDCA7 or HELLS (ICF2, 3 and 4 respectively) (Fig 1A). Heat map representation of DNA methylation β-values for all autosomal DMPs between ICF and NA subjects further illustrated that NA subjects, ICF1 patients, or ICF2, 3, 4 patients form distinct

categories (Fig 1A). Pair-wise comparison of DNA methylation β-values on the 44,555 autosomal DMPs revealed a high correlation between ICF2, 3 and 4 patients, the strongest correlation being between ICF2 and ICF3 (Pearson correlation coefficient r=0.85) (Fig 1B), with almost 65% of shared DMPs (S2A Fig). High correlation was also found between ICF2 and ICF4 as well as ICF3 and ICF4 ($r > 0.7$; > 50% shared DMPs), in contrast to lower correlation between ICF1 and the other ICF subtypes ($r < 0.5$; <20% shared DMPs; Fig 1B). Relatively high correlation values were also observed between ICF1 and NA subtypes (r=0.74) and, to a slightly lesser extent, between ICF2, 3, and 4 and NAs (r=0.63-0.68; S2B Fig).

Altogether, analysis of DNA methylation patterns in ICF patients clearly segregated two groups with patients carrying mutations in DNMT3B from those who do not. In support of the high correlation between methylation landscapes in patients with mutations in ZBTB24, CDCA7 or HELLS, RT-qPCR analysis showed that CDCA7 expression is dependent on the integrity of ZBTB24 ([22]; S3A and S3B Figs). In addition, co-immunoprecipitation experiments suggested that protein interactions between ZBTB24, CDCA7 and HELLS might favor the stabilization of ZBTB24 protein (S3A and S3D Figs).

Distinct genomic context of DMPs in ICF1 compared to ICF2, 3 and 4 patients

Since the genomic regions with perturbed methylation profiles might be distinct among ICF subtypes, we analyzed the genomic and CpG contexts of DMPs depending on genotype. CpGs on the HM450K array have been annotated to identify CpG islands (CGI), 'shores' defined as the 2-kb regions near CGI, 'shelves' as the additional flanking 2 kb from CGI, and the 'open sea' as isolated CpGs. Shelves were almost unaffected in all ICF subtypes (Fig 1C and S4 Fig). Mutations in DNMT3B led to a pronounced decrease in CpG methylation at CGIs and shores, as illustrated by the shift from high towards low β-values (S4 Fig) and by the statistically significant enrichment in the percentage of HypoMPs in both categories relative to the HM450K array distribution (Fig

1C and Table 1). This was in clear contrast with mutations in ZBTB24, CDCA7 or HELLS that most prominently led to a decrease in CpG methylation within open sea (Fig 1C and S4 Fig; Table 1). Increased DNA methylation, although not the main feature of ICF syndrome (Fig S1B), mostly affected open sea in ICF1, 2 and 3 whereas CGIs seemed to be the main target in ICF4 patient (Fig 1D and S4 Fig; Table 1).

Table 1. Distribution of autosomal DMPs in ICF patients relative to CpG context (Illumina annotation).

	Hypomethylated probes (HypoMPs)				**Hypermethylated probes (HyperMPs)**			
	ICF1	**ICF2**	**ICF3**	**ICF4**	**ICF1**	**ICF2**	**ICF3**	**ICF4**
Total number of HypoMPs	6,942	8,414	9,623	8,708	1,921	2,661	2,166	4,120
Island (nb probes)	36% (2523)	3% (232)	4% (348)	4% (351)	3% (67)	11% (294)	15% (328)	44% (1808)
Shore (nb probes)	28% (1917)	9% (798)	10% (978)	13% (1092)	18% (340)	29% (775)	29% (632)	24% (985)
Shelf (nb probes)	6% (429)	11% (896)	11% (1067)	12% (1075)	16% (300)	16% (424)	13% (282)	5% (225)
Open sea (nb probes)	30% (2070)	77% (6488)	75% (7230)	71% (6190)	63% (1214)	63% (1214)	43% (924)	27% (1102)

We then categorized DMPs into six gene feature groups using Illumina annotation: promoter [probes annotated with respect to their distance to a Transcription Start Site (TSS): TSS200, TSS1500 or both], gene start (probes with multiple annotations including TSS200, 5'UTR and first exon), gene body, 3′ UTR, ambiguous (multiple annotations that cannot be categorized) and

intergenic. CpG hypomethylation most prominently affected intergenic regions and, to a lesser extent, gene bodies in all ICF patient samples. HyperMPs, although less represented, were also equally distributed in gene bodies and intergenic regions in all ICF patients (S5A Fig).

Because CGI is not a feature restricted to promoters, or isolated CpGs a feature of intergenic regions, we intersected Illumina annotations for gene context with that of CpG context to break down CGIs and open sea into gene feature groups. The categorization of hypomethylated open sea into gene features was quite similar among ICF subtypes, showing a higher proportion in intergenic regions and gene bodies (S5B Fig). In contrast, the fraction of hypomethylated CGIs was higher in ICF1 patients as expected (Fig 1C and S5B), although with a remarkable enrichment at intergenic regions and gene bodies relative to HM450K array composition (S5B Fig). CGIs hypermethylation mainly affected gene bodies in ICF2, 3 and 4 patients (S5C Fig). This feature was more striking in patient with HELLS mutations who exhibited the highest percentage of CGI hypermethylation (44%; Fig 1D and Table 1) mainly mapping to gene promoters, bodies and start (ICF4; S5C Fig). Open sea hypermethylation was a rare event in all ICF patients (S5C Fig).

As a whole, the distinctive CpG and genomic contexts in which HypoMPs are located depending on ICF genotype demonstrates that ZBTB24, CDCA7 and HELLS are not just guides for DNMT3B targeting to methylation sites.

Unifying features in ICF patients

Since ICF patients share molecular and phenotypic features [17,18], we evaluated common perturbed methylation landscapes in our cohort of patients. Only a few CpGs (433 probes) were hypomethylated in all ICF patients regardless of their genotype (S6A Fig). The most prominently affected locus in all ICF subtypes was a large region spanning around 750kb on chromosome 5q31.3, containing 3 gene clusters encoding protocadherins alpha (PCDHA), β (PCDHB) and gamma (PCDHG), although their hypomethylation was more pronounced in ICF1 patients (S6B

and S6C Figs; S2 Table). We used COBRA (see Supporting Methods section; Validation of DNA methylation data) to confirm CpG hypomethylation at PCDHB13 and PCDHB18 promoters in blood cells from ICF patients (S6D Fig). Gene Ontology (GO) analysis of common HypoMPs using Enrichr showed enrichment for biological processes associated with these PCDH genes, and with Olfactory Receptors (OR) owing to the abundance of HypoMPs mapping to OR scattered throughout the genome (S3 Table). These common HypoMPs mainly mapped to intergenic regions and open sea (S7A Fig), suggesting that, at least at these loci, ZBTB24, CDCA7 and HELLS may serve as guides for DNMT3B recruitment.

Although CpG distribution on the HM450K is skewed towards CGIs and genes, a number of probes (63,738 probes; 13.4% of total probes) overlap with repeated sequences or with pericentromeric cytogenetic bands. Since hypomethylation of DNA repeats is a hallmark of all ICF patients [13], we analyzed the methylation status of these probes in the four ICF subtypes. We intersected datasets of HypoMPs with the Repeat Masker track (see Supporting Methods section) available from the UCSC genome browser (S7B Fig). Although less than 5% of probes designed in DNA repeats were hypomethylated in ICF patients, this represented around one-third of HypoMPs in ICF samples (Table 2; 22% in ICF1 to 37% in ICF2). The breakdown of repeat-associated probes showed that the distribution of HypoMPs into different families of DNA repeats was statistically different from that on the array, with Long Terminal Repeat (LTR) elements and Long Interspersed Nuclear Elements (LINE) being more represented in ICF2, 3 and 4 patients (54-56% or repeat-associated HypoMPs compared to 42.4% in ICF1; S7B Fig). From 12% (ICF2) up to 25% (ICF1) of probes designed in satellite repeat family were hypomethylated in patients, showing hypomethylation of centromeric repeats at 8q11.1 (GSAT repeats) and 12p11.1 (GSATII repeats) (Fig 2A and 2B), in addition to the characteristic hypomethylation of SatII and III repeats located in 1q12, 9q12 and 16q11.2. However, in contrast to previous studies showing CpG hypomethylation proximal to the hypomethylated blocks of heterochromatin located at 1q12, 9q12 and 16q11.2 in the context of transformed ICF cell lines [23] or cancer [24], hypomethylation

spreading into neighboring cytogenetic bands 1q21.1, 9q13 and 16q12.1 was not significant in ICF2, 3 and 4 patients and rather limited to the heterochromatin-proximal probes in ICF1 (S8A to S8C Figs).

Table 2. Distribution of hypoMPs within DNA repeats represented on the HM450K in ICF patient cells

Number of repeat-associated probes on HM450K chip		Number of HypoMPs in DNA repeats				% of HypoMPs (relative to total number of probes in each category)				% of HypoMPs mapping to repeats (related to S6 Fig)			
		ICF1	ICF2	ICF3	ICF4	ICF1	ICF2	ICF3	ICF4	ICF1	ICF2	ICF3	ICF4
63,738	**Total HypoMPs**	**6,942**	**8,414**	**9,623**	**8,708**								
		1,531	**3,111**	**3,494**	**3,127**	**22.1**	**37.0**	**36.3**	**35.9**	**2.4**	**4.9**	**4.9**	**2.8**
21,103	**SINE**	405	796	901	816	1.9	3.8	4.3	3.9	26.4	25.6	25.8	26.1
16,7	**LINE**	350	935	1,038	941	2.1	5.6	6.2	5.6	22.9	30.1	29.7	30.1
12,739	**LTR**	299	815	914	760	2.3	6.4	7.2	6.0	19.5	26.2	26.2	24.3
5,859	**DNA**	146	333	336	340	2.5	5.7	5.7	5.8	9.5	10.7	9.6	10.9
563	**Satelllite**	141	69	112	101	25.0	12.3	19.9	17.9	9.2	2.2	3.2	3.2
2,366	**Simple_repeat**	90	103	121	97	3.8	4.4	5.1	4.1	5.9	3.3	3.5	3.1
3,163	**Low_complexity**	85	43	54	54	2.7	1.4	1.7	1.7	5.6	1.4	1.5	1.7
1,245	**Others**	15	17	18	18	1.2	1.4	1.4	1.4	1.0	0.5	0.5	0.5

Preferential hypomethylation of CpG islands in ICF1 patients

In agreement with ICF1 patients forming a distinct group, a number of probes were specifically hypomethylated in ICF1 samples, among which several classes of normally heavily methylated CGIs from the inactive X chromosome, germline gene promoters and Differentially Methylated Regions (DMR) that control imprinted genes expression. In male ICF patients, hypomethylation on the active X chromosome was barely detectable (S9A Fig), suggesting that the hypomethylation

detected in female patients came from the inactive X chromosome. In ICF1 female patients, 30% of probes mapping to the X chromosome (3,377 out of 11,232 probes) were hypomethylated (Fig 3A and S9A), mostly at CGIs (58% of HypoMPs; Table 3). The X chromosome was also affected in ICF2 and ICF3 female patients (Fig 3A and S9A), although to a lesser extent and mainly in open sea (57% in ICF2 and 53% in ICF3; Table 3). The few X-linked HypoMPs common to all females ICF (181 probes; S9B Fig) corresponded to isolated CpGs located in promoters and intergenic regions (S9C and S9D Figs). On the Y chromosome, 7 to 10% of the probes were hypomethylated, mainly within CGIs in ICF1 or open sea in ICF2 and ICF4 male patients (S9E Fig). We validated that the loss of CpG methylation at MAGEB10 CGI promoter was more prominent in female ICF1, using COBRA (S10A Fig) and pyrosequencing (Fig 3C and S10B Fig). In contrast, MSRE-qPCR validated that hypomethylation of isolated CpG within MAGEB3 promoter is common to all ICF female patients (S10C Fig).

Table 3. Distribution of X-linked HypoMPs in female ICF female patients with respect to CpG context (Illumina annotations).

	% of HypoMPs in each category (Number of HypoMPs)		
	ICF1	**ICF2**	**ICF3**
Total number of HypoMPs	**3,377**	**750**	**1,071**
Island (nb probes)	**58% (1,965)**	4% (32)	7% (71)
Shore (nb probes)	23% (762)	21% (154)	20% (209)
Shelf (nb probes)	4% (139)	19% (140)	21% (222)
Open sea (nb probes)	15% (511)	**57% (424)**	**53% (569)**

GO analysis of hypomethylated promoters (HypoMPs within TSS200 and TSS1500; Illumina annotation) as a consequence of DNMT3B mutations in ICF1 cells, also revealed a highly significant enrichment for biological processes related to germ cells (S11A Fig). Hence, we compared our datasets of hypomethylated promoters to lists of genes which expression is restricted or enriched in germ cells (Protein Atlas and reference [25]). Since many germline genes are linked to the X chromosome, we used datasets of HypoMPs according to the sex of the patients. Out of the ≈2,000 genes with elevated expression levels in germ cells compared to other tissues, we identified 707 (1,677 probes) and 542 (1,298 probes) genes with abnormally hypomethylated promoters in female and male ICF patients, respectively, which could be reduced to 179 and 142 genes when considering the category "tissue enriched" (S4 and S5 Tables). As suspected from previous studies [26,27], hypomethylation of germline gene CGI promoters was a strong signature of ICF1 as illustrated by the heatmap representation of CpG methylation levels within the promoters of these genes (Fig 3B and Table 4). In contrast, the reduced CpG methylation at germline genes was less dramatic in ICF2, ICF3 and ICF4 patients, and mainly affected open sea (Table 4). We confirmed that the loss of CGI methylation at MAEL, SPO11 and TDRD6 promoters was specific to ICF1 patients, using COBRA (S11B Fig) and pyrosequencing (Fig 3C and S11C Fig). COBRA assay also confirmed that CpG hypomethylation in open sea of ADAM2, shore of SPO11 and within CGI of C15ORF60, was more prominent in ICF2, ICF3 and ICF4 (S11D Fig).

Table 4. Distribution of HypoMPs mapping to germline genes (database [25]) with respect to CpG context (Illumina annotation).

	Females / Hypomethylated probes			Males / Hypomethylated probes		
	ICF1	**ICF2**	**ICF3**	**ICF1**	**ICF2**	**ICF4**
Total number of HypoMPs	**748**	**392**	**385**	**454**	**272**	**275**
Island (nb probes)	**51% (378)**	3% (10)	5% (18)	**63% (284)**	5% (13)	7% (19)
Shore (nb probes)	24% (181)	16% (61)	15% (57)	24% (108)	11% (31)	19% (51)
Shelf (nb probes)	2% (15)	6% (22)	3% (12)	3% (14)	5% (13)	5% (13)
Open sea (nb probes)	23% (174)	**76% (299)**	**77% (298)**	11% (48)	**79% (215)**	**70% (192)**

Since DMRs often overlap with CGIs[28], we compared datasets of hypomethylated CGIs in ICF1 patients with DMRs annotated in a previous study[29,30]. DMRs were not affected with the exception of the DMR identified in the ZNF597 CGI-promoter, one of the three somatic DMRs described so far [30], which was hypomethylated in ICF1 patients. We used COBRA, pyrosequencing and bisulphite sequencing to confirm CpG hypomethylation of ZNF597 promoter in blood cells of ICF1 patients, suggestive of a loss of somatic imprinting at this DMR (S12A to S12C Figs), which remained undetectable using COBRA in ICF2, ICF3 or ICF4 patients (S12A Fig). COBRA assay further showed the conserved hypomethylated signature of the ZNF597 DMR in another tissue, ICF1-derived skin fibroblasts, suggesting that DNMT3B might be required for establishment of DNA methylation at somatic imprints that occurs post-implantation (S12D Fig).

Alteration of DNA methylation at gene clusters in ICF2, ICF3 and ICF4

GO analysis performed on HypoMPs common to ICF2, ICF3 and ICF4 first indicated that genes related to neuron biology and function seemed to be particularly affected in these patients (S13A Fig). We further confirmed their hypomethylated status in blood cells from ICF2, 3 and 4 patients, using COBRA for the promoters of Dopamine Receptor D3 (DRD) and Necdin (NDN) (S13B Fig), and MSRE-qPCR for the promoter of Neurexin 3 (NRXN3) gene (S13C Fig). As shown by GO analysis, we observed that the 484 HypoMPs common to ICF2, ICF3 and ICF4 subtypes (S13D Fig) mapped to several clusters of functionally related genes (S13E Fig) including, in addition to the Protocadherin family (PCDH) common to all ICF, Olfactory Receptors (OR) on several chromosomal locations, Late Cornified Envelope (LCE), small proline-rich protein (SPRR), NLR Family Pyrin Domain Containing (NLRP), Keratin Associated Protein (KRTAP), Zinc Finger proteins (ZNF), and Defensin β (DEFB) (S6 Table). In addition to these clusters of coding genes, our analysis also revealed CpG hypomethylation within several clusters of small regulatory non-coding RNAs (ncRNA) hosted in imprinted loci [31,32]. Among these, we identified HypoMPs on chromosome 15 and 19 respectively linked to large clusters of small nucleolar RNAs (snoRNAs) embedded within the Prader-Willi locus (PWS), and to the largest microRNA gene cluster of the human genome C19MC (Fig 4A and 4B).

We validated the significant drop in CpG methylation levels, almost exclusively in ICF2, ICF3 and ICF4 patients, at SNORD115-9 and SNORD115-5 promoters using MSRE-qPCR (Fig 4C and S14A Fig), at SNORD115-14 and SNORD116-7 promoters using bisulfite sequencing (Fig 4D) and COBRA (S14B Fig), and at the C19MC locus using MSRE-qPCR at the MIR521-2 and MIR518D promoters (Fig 4E and S14C Fig). Quantification of methylation levels at the MIR518D promoter also confirmed the loss of methylation to less than 20% in ICF2, 3 and 4 patients compared to the 50% found in NA and ICF1 subjects as expected for an imprinted locus (Fig 4F). Although less affected than clusters of ncRNA on chromosome 15 and 19, we also found

HypoMPs within other imprinted clusters of microRNAs (C14MC) and SNORD (SNORD114) hosted in the imprinted DLK1-DIO3 locus on chromosome 14 (S15A Fig). The greater hypomethylation in ICF2, 3 and 4 patients was confirmed using COBRA (S15B Fig), although quantification by pyrosequencing showed that the CpGs in the MIR376C promoter (C14MC cluster) were not equally affected (S15C Fig).

We then tested whether the HypoMPs identified as common to ICF2, ICF3 and ICF4 were tissue specific. COBRA assays showed that hypomethylation within promoters of NDN, MIR376C, SNORD115-14 was also observed in low-passages primary skin fibroblasts derived from ICF2, 3 and 4 patients (S15D Fig). We extended this validation to low-passages skin fibroblasts derived from an ICF2 patient (pP patient; S15D Fig) who was not initially included in the DNA methylation array analysis. Since defects in DNA methylation was observed in at least two cell types from patients, they may well originate from defects in establishment at early stages of development, especially in the case of most of the imprinted genes for which the methylation status is set at early stages. In favor of this hypothesis, we showed that knockdown of ZBTB24, CDCA7 or HELLS expression in somatic cells was not sufficient to perturb DNA methylation levels at unique genes, imprinted genes or at centromeric and pericentromeric DNA repeats (S16 Fig).

To test whether the genomic regions affected by mutations in ZBTB24, CDCA7 or HELLS share common epigenetic features in non-affected subjects, we annotated the differentially HypoMPs identified in ICF patients according to ENCODE data from healthy cells, publically available through the UCSC Genome Browser. We compared HypoMPs datasets to a recent categorization of chromatin states segmentation of the human genome into fifteen states grouped, further categorized in ten broad classes to predict functional elements (see Supporting Methods section), using the ChromHMM track from the B-lymphoblastoid cell line (LCL) GM12878. We also interrogated genome-wide DNA replication timing data using the UW Repli-seq track (see Supporting Methods section) to extract replication timing values available for GM12878, cells sorted into six cell cycle fractions (G1, 4 fractions in S phase, G2), for the genomic regions of

interest. These analyses showed that the majority of HypoMPs were statistically enriched in the 'Heterochromatin' and 'Weakly transcribed' chromatin states and rather late replicating categories (Fig 5A and 5B). However, there was again a striking difference between ICF1 and ICF2, 3, 4 patients. More than 90% of HypoMPs mapped to the two chromatin states aforementioned in ICF2, 3, 4 compared to less than 70% in ICF1. In contrast, predicted functional promoters and enhancers represented 25% of HypoMPs in ICF1 but less than half of that in ICF2, 3, 4 cells (Fig 5A). Along the same lines, more than 60% of HypoMPs in ICF2, 3, 4 mapped to late-replicating regions (categories S3, S4 and G2) compared to the same proportion mapping to early-replicating regions (G1, S1, S2) in ICF1 patients (Fig 5B). Consistent with the prominent location of HypoMPs in 'Heterochromatin' chromatin state, they also mapped to regions normally enriched for H3K9me3 and H3K27me3 histone marks according to Epigenomics Roadmap data from normal hematopoietic and lymphoid cells (S7 to S10 Tables).

Discussion

We report here methylome analysis in patients with a monogenic disease that affects DNA methylation landscapes, comparing for the first time the four genotypes reported so far in ICF syndrome. Besides the identification of a few hundreds of genomic loci affected by mutations in any of the four incriminated factors, clustering of methylation data and analyses of genomic and epigenomic context clearly discriminated patients with DNMT3B mutations from those with mutations in ZBTB24, CDCA7 or HELLS. The discovery of mutations in factors devoid of DNA methyltransferase activity and with a virtually unknown connection with DNA methylation, with the exception of HELLS in the mouse [33,34], raised the question of their link with DNA methylation. Our data lend support to a key role of ZBTB24, CDCA7 or HELLS in the methylation of CpG-poor loci in intergenic regions identified as quiescent chromatin domains, of several classes of repetitive elements, of gene families with random or imprinted monoallelic expression, and, in general, of loci with heterochromatin features in healthy cells.

The identification of two groups of ICF patients based on high throughput methylation data was consistent with the heterogeneity already described in ICF syndrome with respect to loss of methylation at DNA repeats [15,16]. ICF patients also exhibit both common and distinctive clinical manifestations despite unifying cytogenetic defects that serve to establish diagnosis [12]. Immunodeficiency is more severe in ICF1 patients whereas ICF2, 3 and 4 patients are more affected by intellectual disability [17,18]. The deregulated expression of a few genes involved in development, neurogenesis and immunity was proposed to account for clinical manifestations in ICF1 [13,35]. Here, we identified additional common and distinguishing loci with perturbed DNA methylation in the blood of patients, that could be used as biomarkers for the disease as we have already described [26]. Hypomethylation of genes with functions in the central nervous system (CNS) provides biomarkers that are particularly relevant for neurodevelopmental defects that affect ICF2, ICF3 and ICF4 patients, as exemplified by studies reporting that this type of blood-

based biomarkers can reflect DNA methylation changes in various tissues including in the CNS [36,37].

Furthermore, the existence of monogenic diseases affecting DNA methylation machinery or landscapes offers a unique opportunity to uncover actors and mechanisms and infer genomic regions controlled by these factors. Many studies have provided clues into DNMT3B function and preferential genomic and CpG contexts. DNMT3B was shown to actively methylate transcribed gene bodies [38,39], although we could not infer such preference from our methylome analysis owing to the design of the HM450K array on which CpGs in gene bodies are poorly represented. Nonetheless, our methylome analysis underscores the prominent role played by DNMT3B in the methylation of CGIs, in particular on the inactive X chromosome in females and at germ line gene promoters, in line with previous data obtained in the mouse [40–44] or inferred from studies in ICF1 patients [26,27,45–47]. Beyond that, we provide an extensive list of germ line genes at which DNMT3B is required for their methylated state in the soma. Since our data were not biased by cell culture or transformation effects, and experimentally validated in two somatic tissues, it represents a strong signature for DNMT3B dysfunction that can be applied to other pathological contexts like cancer in which DNMT3B catalytically inactive splice variants have been identified [48–50].

DNA methylation is not randomly distributed on the genome implying that *de novo* DNMTs must be guided to specific places. Several studies have reported that CpG content and chromatin context promote or prevent recruitment of *de novo* DNMTs [5,39,51–53]. In addition, transcription factors, by virtue of their binding to specific sequences, may also shape DNA methylation patterns by recruiting DNMTs [54]. Indeed, the transcriptional repressor E2F6 was shown to be required for DNA methylation [55] and DNMT3B recruitment at CGI promoters of germ line genes [40]. Likewise, the member of the structural maintenance of chromosomes (SMC) family SMCHD1, through its hinge domain, engages with methylated promoters at clustered PCDH genes to maintain DNA methylation patterns and transcriptional repression [56] and is required for DNMT3B recruitment at CGI promoters on the inactive X chromosome and inactive

X-linked promoters, respectively [40,55,57]. The recent discovery of ICF patients with mutations in factors that lack a DNMT activity but have DNA binding ability suggested that they could recruit DNMT3B activity at specific places on the genome. We indeed identified a few hundreds of hypomethylated CpGs, like in the PCDH gene clusters, whose DNA methylation is lost following mutations in any of the four factors suggesting that, at least for ZBTB24 and CDCA7 with DNA binding activities they could function as recruiting factors for DNMT3B activity at these loci. However, genomic loci affected by loss of DNMT3B functions seemed to be prominently distinct from that affected by mutations in ZBTB24, CDCA7 or HELLS as shown by clustering of DNA methylation data. Furthermore, and in contrast to CGI promoters of germ line and inactive X-linked genes that are mainly affected by DNMT3B loss of function, mutations of ZBTB24, CDCA7 and HELLS mainly lead to hypomethylation of CpG-poor regions whether of the X chromosome in female patients or at CpG-poor promoters of protein-coding and non-coding genes as well as intergenic regions. Although we could infer from these data that ZBTB24, CDCA7 and HELLS are involved in targeting the DNA methylation machinery at specific genomic loci, they also imply that these factors are not just guides for DNMT3B to methylation sites.

Before their implication in the ICF syndrome, the potential role of ZBTB24 and CDCA7 in DNA methylation had never been suspected. On the contrary, the role of HELLS, also known as LSH (Lymphoid-specific helicase), in DNA methylation has been well documented in the mouse, especially at tandem and interspersed DNA repeats [34,58,59], certain gene clusters like HOX, RHOX and PCDHB[60], and at intergenic and CpG poor promoters in mouse embryonic fibroblasts [61]. Methylome analysis of ICF4 patient suggests that the function of HELLS at CpG poor regions and LTRs is conserved between mouse and humans and that, besides the PCDHB gene cluster and satellite repeats, DNMT3B recruitment is mostly independent of HELLS.

We have extended the analysis of the impact of ZBTB24, CDCA7 and HELLS mutations to single-copy genes that are affected in ICF2, 3 and 4 but not in ICF1 patients. Among the genomic loci affected by their loss of integrity, we found a number of coding and non-coding genes organized

in clusters. These clusters, for most of them, are expressed in a monoallelic manner, resulting from either random or imprinted choice [62,63]. Among those, we found several clusters of small non-coding RNAs: SNORD115-116 on chromosome 15, the primate-specific microRNA gene cluster C19MC on chromosome 19, the SNORD114 and the miRNA cluster C14MC on chromosome 14. C19MC and C14MC are large microRNA imprinted clusters whose expression is restricted to the placenta and undifferentiated cells [32,64,65]. SNORD115 and 116 (Small Nucleolar RNA, C/D Box 115 and 116 Clusters) clusters of small ncRNAs are hosted in the imprinted locus on chromosome 15q11-q13 whose deletion is the genetic cause of the Prader-Willi syndrome (PWS, OMIM#176270), a neurobehavioral disorder [31]. A protein complex containing the zinc finger protein ZNF274 and the H3K9 lysine methyltransferase SETDB1 were recently involved in DNA methylation of the PWS imprinting center (PWS-IC) that controls the transcriptional repression of SNORD116 cluster [66]. Our data adds a new zinc finger protein, ZBTB24, whose loss of function specifically alters DNA methylation at SNORD clusters without affecting the PWS-IC. Interestingly, together with the suspected role of C19MC and C14MC in brain function [67], it is tempting to speculate that the DNA hypomethylation of these clusters of non-coding RNA that we have characterized in ICF2, ICF3 and ICF4 patients may contribute to perturbed neurological functions that characterize these ICF sub-types.

Compared to the obvious contribution of DNMTs to DNA methylation, virtually nothing was known about the contribution of newly identified factors causally involved in ICF syndrome that are devoid of methyltransferase activity but whose loss of integrity has dramatic impact on methylation patterns and development. The remarkable location of hypomethylated probes within genomic domains with heterochromatin-associated marks in normal cells might suggests that loss of DNA methylation in patients with mutations in ZBTB24, CDCA7 or HELLS may primarily result from the loss of repressive histone marks. This idea seems to be reinforced by a two-hybrid screen that identified SUV39H1, the major enzyme that generates H3K9me3, as a ZBTB24 partner [68]. Likewise, HELLS was shown to cooperate with the G9A/GLP complex of histone

methyltransferases at a number of gene promoters, although it was clearly implicated in *de novo* DNA methylation during the early steps of embryonic development [69]. Since many heterochromatin regions resist to the loss of ZBTB24, CDCA7 or HELLS, it argues against a general mechanism through which mutations in these factors would primarily affect heterochromatin chromatin marks. We previously demonstrated that ZBTB24, CDCA7 and HELLS are required for methylation maintenance in the very specific case of centromeric repeats in the mouse [20]. In contrast, our loss of function experiments performed in human primary fibroblasts shows that it is not the case at human (peri)centromeric repeats and imprinted loci, suggesting that ZBTB24, CDCA7 and HELLS are involved in establishment of DNA methylation profiles at early steps during human development.

Whether ZBTB24, CDCA7 and HELLS function as guides for the DNA methylation machinery or through more intricate mechanisms that involve histone modifications remains to be tested although it will clearly depend on the stage of development, the type of tissue and the genomic context. Nonetheless, such comparative genome-wide characterization of DNA methylation profiles in an extremely rare disease established ZBTB24, CDCA7 and HELLS as serious candidates in future studies on the mechanisms of DNA methylation at repetitive and mobile elements, and more specifically, in future studies on the mechanisms of DNA methylation at heterochromatin regions of the human genome.

Material and Methods

Full experimental procedures are provided in the Supporting Information section.

Healthy donors and patients

Blood samples were collected in EDTA-containing tubes to prevent coagulation. Peripheral blood from 10 healthy volunteers (aged 2, 4, 5, 11, 15, 24, 26, 27, 36 and 42 years) were collected and numbered from 1 to 10. Donors 1, 4, 5, 7 and 10 are women. Our cohort of patients included eight ICF1 with mutations in DNMT3B, four ICF2 patients with mutations in ZBTB24, two ICF3 patients with mutations in CDCA7 and one ICF4 patients with mutations in HELLS. These ICF patients have been previously described [26] and their clinical picture is summarized in S1 Table. Approval for this study was obtained from the institutional review board of Necker Hospital and informed consent was obtained from all patients or their families (for minors), in agreement with the Helsinki Declaration (CNIL authorization: no. 908,256, October 14th, 2008).

Primary cells and cell lines

Primary skin fibroblasts from non-affected subjects and ICF patients have been previously described [26]. Additional primary skin fibroblasts from ICF patient P10 were derived by expanding primary cultures of skin biopsies.

Lymphoblastoid cell lines (LCLs) have also already been described [16].

Sample preparation for Genome-Wide DNA methylation analysis

A sample of 1 μg of genomic DNA from cells lines or whole blood was converted by sodium bisulfite using the EpiTect® 96 Bisulfite Kit (Qiagen). Bisulfite converted samples were then

hybridized on the Illumina Infinium HumanMethylation450 BeadChip (Illumina) according to the manufacturer's protocol at the genotyping facility of the Centre National de Génotypage (Evry, France) [70].

Genome-Wide DNA methylation data processing

Illumina 450K microarray signals were extracted with GenomeStudio software (Illumina) and were processed and normalized by an updated version of the previously developed pipeline [70]. The procedure is described in the Supporting Information section.

RNA extraction and analysis of gene expression

Total RNA was isolated as previously described [26]. Primer sequences are provided in S11 Table.

Statistical analysis

For the annotation of probes relative to the various categories defined in terms of CpG density, genomic and epigenomic contexts, the proportion of probes assigned to each category was calculated, and category distributions were compared through a Chi-square test.

Student's t-tests were used to determine the statistical significance between values of RT-qPCR, MSRE-qPCR and bisulfite pyrosequencing experiments.

Webservers and web resources

Coriell Cell Repositories (USA) (http://ccr.coriell.org/)

MethPrimer (http://www.urogene.org/methprimer/)

GenomeStudio Illumina (https://www.illumina.com/techniques/microarrays/array-data-analysis-experimental-design/genomestudio.html)

Galaxy tools (http://main.g2.bx.psu.edu)

IGV visualization tool (http://software.broadinstitute.org/software/igv/)

UCSC genome Browser (http://genome.ucsc.edu)

Enrichr Bioinformatics Resources (http://amp.pharm.mssm.edu/Enrichr)

Roadmap Epigenome Project (http://www.roadmapepigenomics.org)

Human Protein Atlas project (http://www.proteinatlas.org/humanproteome/testis)

Data access

The data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE95040.

Acknowledgements

The authors would like to thank Pr Jonathan Weitzman for comments on this manuscript, Florent Hubé for his advices with bioinformatics tools, Drs Paola Ballerini (Hôpitaux Universitaires Est Parisien, Trousseau) for the gift of whole blood from control children, Dr Kamila Kebaili (Centre de Référence Déficits Immunitaires Héréditaires, CHU de Lyon) for the gift of whole blood from ICF4 patient, Dr Bertrand Roquelaure (Hôpital de la Timone, CHU Marseille) for the gift of whole blood from ICF3 pC patient, and all the patients and their family members for their participation in this study.

References

1. Bird A. The essentials of DNA methylation. Cell. 1992;70: 5–8.

2. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nat Rev Genet. 2013;14: 204–220. doi:10.1038/nrg3354

3. Goll MG, Bestor TH. Eukaryotic cytosine methyltransferases. Annu Rev Biochem. 2005;74: 481–514. doi:10.1146/annurev.biochem.74.010904.153721

4. Weber M, Schübeler D. Genomic patterns of DNA methylation: targets and function of an epigenetic mark. Curr Opin Cell Biol. 2007;19: 273–280. doi:10.1016/j.ceb.2007.04.011

5. Illingworth RS, Bird AP. CpG islands - "A rough guide." FEBS Lett. 2009;583: 1713–1720. doi:10.1016/j.febslet.2009.04.012

6. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012;13: 484–492. doi:10.1038/nrg3230

7. Scarano MI, Strazzullo M, Matarazzo MR, D'Esposito M. DNA methylation 40 years later: Its role in human health and disease. J Cell Physiol. 2005;204: 21–35. doi:10.1002/jcp.20280

8. Robertson KD. DNA methylation and human disease. Nat Rev Genet. 2005;6: 597–610. doi:10.1038/nrg1655

9. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. Science. 2009;324: 930–935. doi:10.1126/science.1170116

10. Klose RJ, Bird AP. Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci. 2006;31: 89–97. doi:10.1016/j.tibs.2005.12.008

11. Maraschio P, Zuffardi O, Dalla Fior T, Tiepolo L. Immunodeficiency, centromeric heterochromatin instability of chromosomes 1, 9, and 16, and facial anomalies: the ICF syndrome. J Med Genet. 1988;25: 173–180.

12. Hagleitner MM, Lankester A, Maraschio P, Hulten M, Fryns JP, Schuetz C, et al. Clinical spectrum of immunodeficiency, centromeric instability and facial dysmorphism (ICF syndrome). J Med Genet. 2007;45: 93–99. doi:10.1136/jmg.2007.053397

13. Ehrlich M, Jackson K, Weemaes C. Immunodeficiency, centromeric region instability, facial anomalies syndrome (ICF). Orphanet J Rare Dis. 2006;1: 1.

14. Xu GL, Bestor TH, Bourc'his D, Hsieh CL, Tommerup N, Bugge M, et al. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature. 1999;402: 187–191. doi:10.1038/46052

15. Wijmenga C, Hansen RS, Gimelli G, Björck EJ, Davies EG, Valentine D, et al. Genetic variation in ICF syndrome: evidence for genetic heterogeneity. Hum Mutat. 2000;16: 509–517.

16. Jiang YL, Rigolet M, Bourc'his D, Nigon F, Bokesoy I, Fryns JP, et al. DNMT3B mutations and DNA methylation defect define two types of ICF syndrome. Hum Mutat. 2005;25: 56–63. doi:10.1002/humu.20113

17. Weemaes CMR, van Tol MJD, Wang J, van Ostaijen-ten Dam MM, van Eggermond MCJA, Thijssen PE, et al. Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects. Eur J Hum Genet EJHG. 2013;21: 1219–1225. doi:10.1038/ejhg.2013.40

18. Sterlin D, Velasco G, Moshous D, Touzot F, Mahlaoui N, Fischer A, et al. Genetic, Cellular and Clinical Features of ICF Syndrome: a French National Survey. J Clin Immunol. 2016;36: 149–159. doi:10.1007/s10875-016-0240-2

19. de Greef JC, Wang J, Balog J, den Dunnen JT, Frants RR, Straasheijm KR, et al. Mutations in ZBTB24 Are Associated with Immunodeficiency, Centromeric Instability, and Facial Anomalies Syndrome Type 2. Am J Hum Genet. 2011;88: 796–804. doi:10.1016/j.ajhg.2011.04.018

20. Thijssen PE, Ito Y, Grillo G, Wang J, Velasco G, Nitta H, et al. Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome. Nat Commun. 2015;6: 7870. doi:10.1038/ncomms8870

21. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. Genomics. 2011;98: 288–295. doi:10.1016/j.ygeno.2011.07.007

22. Wu H, Thijssen PE, de Klerk E, Vonk KK, Wang J, den Hamer B, et al. Converging disease genes in ICF syndrome: ZBTB24 controls expression of CDCA7 in mammals. Hum Mol Genet. 2016; ddw243.

23. Brun M-E, Lana E, Rivals I, Lefranc G, Sarda P, Claustres M, et al. Heterochromatic Genes Undergo Epigenetic Changes and Escape Silencing in Immunodeficiency, Centromeric Instability, Facial Anomalies (ICF) Syndrome. Chadwick BP, editor. PLoS ONE. 2011;6: e19464. doi:10.1371/journal.pone.0019464

24. Fournier A, McLeer-Florin A, Lefebvre C, Duley S, Barki L, Ribeyron J, et al. 1q12 chromosome translocations form aberrant heterochromatic foci associated with changes in nuclear architecture and gene expression in B cell lymphoma: Aberrant heterochromatic foci and lymphomagenesis. EMBO Mol Med. 2010;2: 159–171. doi:10.1002/emmm.201000067

25. Djureinovic D, Fagerberg L, Hallstrom B, Danielsson A, Lindskog C, Uhlen M, et al. The human testis-specific proteome defined by transcriptomics and antibody-based profiling. Mol Hum Reprod. 2014;20: 476–488. doi:10.1093/molehr/gau018

26. Velasco G, Walton EL, Sterlin D, Hédouin S, Nitta H, Ito Y, et al. Germline genes hypomethylation and expression define a molecular signature in peripheral blood of ICF patients: implications for diagnosis and etiology. Orphanet J Rare Dis. 2014;9: 56. doi:10.1186/1750-1172-9-56

27. Simo-Riudalbas L, Diaz-Lagares A, Gatto S, Gagliardi M, Crujeiras AB, Matarazzo MR, et al. Genome-Wide DNA Methylation Analysis Identifies Novel Hypomethylated Non-Pericentromeric Genes with Potential Clinical Implications in ICF Syndrome. Dante R, editor. PLOS ONE. 2015;10: e0132517. doi:10.1371/journal.pone.0132517

28. Barlow DP, Bartolomei MS. Genomic Imprinting in Mammals. Cold Spring Harb Perspect Biol. 2014;6: a018382–a018382. doi:10.1101/cshperspect.a018382

29. Court F, Tayama C, Romanelli V, Martin-Trujillo A, Iglesias-Platas I, Okamura K, et al. Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment. Genome Res. 2014;24: 554–569. doi:10.1101/gr.164913.113

30. Hanna CW, Peñaherrera MS, Saadeh H, Andrews S, McFadden DE, Kelsey G, et al. Pervasive polymorphic imprinted methylation in the human placenta. Genome Res. 2016;26: 756–767. doi:10.1101/gr.196139.115

31. Cassidy SB, Schwartz S, Miller JL, Driscoll DJ. Prader-Willi syndrome. Genet Med. 2012;14: 10–26. doi:10.1038/gim.0b013e31822bead0

32. Noguer-Dance M, Abu-Amero S, Al-Khtib M, Lefevre A, Coullin P, Moore GE, et al. The primate-specific microRNA gene cluster (C19MC) is imprinted in the placenta. Hum Mol Genet. 2010;19: 3566–3582. doi:10.1093/hmg/ddq272

33. Myant K, Stancheva I. LSH Cooperates with DNA Methyltransferases To Repress Transcription. Mol Cell Biol. 2008;28: 215–226. doi:10.1128/MCB.01073-07

34. Dennis K, Fan T, Geiman T, Yan Q, Muegge K. Lsh, a member of the SNF2 family, is required for genome-wide methylation. Genes Dev. 2001;15: 2940–2944.

35. Jin B, Tao Q, Peng J, Soo HM, Wu W, Ying J, et al. DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function. Hum Mol Genet. 2007;17: 690–709. doi:10.1093/hmg/ddm341

36. Davies MN, Volta M, Pidsley R, Lunnon K, Dixit A, Lovestone S, et al. Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. Genome Biol. 2012;13: R43.

37. Walton E, Hass J, Liu J, Roffman JL, Bernardoni F, Roessner V, et al. Correspondence of DNA Methylation Between Blood and Brain Tissue and Its Application to Schizophrenia Research. Schizophr Bull. 2016;42: 406–414. doi:10.1093/schbul/sbv074

38. Yang X, Han H, De Carvalho DD, Lay FD, Jones PA, Liang G. Gene Body Methylation Can Alter Gene Expression and Is a Therapeutic Target in Cancer. Cancer Cell. 2014;26: 577–590. doi:10.1016/j.ccr.2014.07.028

39. Baubec T, Colombo DF, Wirbelauer C, Schmidt J, Burger L, Krebs AR, et al. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature. 2015;520: 243–247. doi:10.1038/nature14176

40. Velasco G, Hubé F, Rollin J, Neuillet D, Philippe C, Bouzinba-Segard H, et al. Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. Proc Natl Acad Sci U S A. 2010;107: 9281–9286. doi:10.1073/pnas.1000473107

41. Borgel J, Guibert S, Li Y, Chiba H, Schübeler D, Sasaki H, et al. Targets and dynamics of promoter DNA methylation during early mouse development. Nat Genet. 2010;42: 1093–1100. doi:10.1038/ng.708

42. Walton EL, Francastel C, Velasco G. Maintenance of DNA methylation: Dnmt3b joins the dance. Epigenetics. 2011;6: 1373–1377. doi:10.4161/epi.6.11.17978

43. Hackett JA, Reddington JP, Nestor CE, Dunican DS, Branco MR, Reichmann J, et al. Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. Development. 2012;139: 3623–3632. doi:10.1242/dev.081661

44. Auclair G, Guibert S, Bender A, Weber M. Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. Genome Biol. 2014;15. doi:10.1186/s13059-014-0545-5

45. Bourc'his D, Miniou P, Jeanpierre M, Molina Gomes D, Dupont J, De Saint-Basile G, et al. Abnormal methylation does not prevent X inactivation in ICF patients. Cytogenet Cell Genet. 1999;84: 245–252. doi:15269

46. Hansen RS, Stöger R, Wijmenga C, Stanek AM, Canfield TK, Luo P, et al. Escape from gene silencing in ICF syndrome: evidence for advanced replication time as a major determinant. Hum Mol Genet. 2000;9: 2575–2587.

47. Heyn H, Vidal E, Sayols S, Sanchez-Mut JV, Moran S, Medina I, et al. Whole-genome bisulfite DNA sequencing of a DNMT3B mutant patient. Epigenetics. 2012;7: 542–550. doi:10.4161/epi.20523

48. Gopalakrishnan S, Van Emburgh BO, Shan J, Su Z, Fields CR, Vieweg J, et al. A Novel DNMT3B Splice Variant Expressed in Tumor and Pluripotent Cells Modulates Genomic DNA Methylation Patterns and Displays Altered DNA Binding. Mol Cancer Res. 2009;7: 1622–1634. doi:10.1158/1541-7786.MCR-09-0018

49. Ostler KR, Davis EM, Payne SL, Gosalia BB, Exposito-Cespedes J, Le Beau MM, et al. Cancer cells express aberrant DNMT3B transcripts encoding truncated proteins. Oncogene. 2007;26: 5553–5563.

50. Saito Y, Kanai Y, Sakamoto M, Saito H, Ishii H, Hirohashi S. Overexpression of a splice variant of DNA methyltransferase 3b, DNMT3b4, associated with DNA hypomethylation on pericentromeric satellite regions during human hepatocarcinogenesis. Proc Natl Acad Sci. 2002;99: 10060–10065.

51. Rose NR, Klose RJ. Understanding the relationship between DNA methylation and histone lysine methylation. Biochim Biophys Acta BBA - Gene Regul Mech. 2014;1839: 1362–1372. doi:10.1016/j.bbagrm.2014.02.007

52. Du J, Johnson LM, Jacobsen SE, Patel DJ. DNA methylation pathways and their crosstalk with histone methylation. Nat Rev Mol Cell Biol. 2015;16: 519–532. doi:10.1038/nrm4043

53. Schübeler D. Function and information content of DNA methylation. Nature. 2015;517: 321–326. doi:10.1038/nature14192

54. Marchal C, Miotto B. Emerging Concept in DNA Methylation: Role of Transcription Factors in Shaping DNA Methylation Patterns. J Cell Physiol. 2015;230: 743–751. doi:10.1002/jcp.24836

55. Pohlers M, Truss M, Frede U, Scholz A, Strehle M, Kuban R-J, et al. A Role for E2F6 in the Restriction of Male-Germ-Cell-Specific Gene Expression. Curr Biol. 2005;15: 1051–1057. doi:10.1016/j.cub.2005.04.060

56. Chen K, Hu J, Moore DL, Liu R, Kessans SA, Breslin K, et al. Genome-wide binding and mechanistic analyses of Smchd1-mediated epigenetic regulation. Proc Natl Acad Sci. 2015;112: E3535–E3544.

57. Gendrel A-V, Apedaile A, Coker H, Termanis A, Zvetkova I, Godwin J, et al. Smchd1-Dependent and -Independent Pathways Determine Developmental Dynamics of CpG Island Methylation on the Inactive X Chromosome. Dev Cell. 2012;23: 265–279. doi:10.1016/j.devcel.2012.06.011

58. Huang J, Fan T, Yan Q, Zhu H, Fox S, Issaq HJ, et al. Lsh, an epigenetic guardian of repetitive elements. Nucleic Acids Res. 2004;32: 5019–5028. doi:10.1093/nar/gkh821

59. Dunican DS, Cruickshanks HA, Suzuki M, Semple CA, Davey T, Arceci RJ, et al. Lsh regulates LTR retrotransposon repression independently of Dnmt3b function. Genome Biol. 2013;14: R146. doi:10.1186/gb-2013-14-12-r146

60. Xi S, Zhu H, Xu H, Schmidtmann A, Geiman TM, Muegge K. Lsh controls Hox gene silencing during development. Proc Natl Acad Sci. 2007;104: 14366–14371.

61. Yu W, McIntosh C, Lister R, Zhu I, Han Y, Ren J, et al. Genome-wide DNA methylation patterns in LSH mutant reveals de-repression of repeat elements and redundant epigenetic silencing pathways. Genome Res. 2014;24: 1613–1623. doi:10.1101/gr.172015.114

62. Eckersley-Maslin MA, Spector DL. Random monoallelic expression: regulating gene expression one allele at a time. Trends Genet. 2014;30: 237–244. doi:10.1016/j.tig.2014.03.003

63. Chess A. Monoallelic Gene Expression in Mammals. Annu Rev Genet. 2016;50: 317–327. doi:10.1146/annurev-genet-120215-035120

64. Seitz H, Royo H, Bortolin M-L, Lin S-P, Ferguson-Smith AC, Cavaillé J. A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. Genome Res. 2004;14: 1741–1748.

65. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, et al. A Mammalian microRNA Expression Atlas Based on Small RNA Library Sequencing. Cell. 2007;129: 1401–1414. doi:10.1016/j.cell.2007.04.040

66. Cruvinel E, Budinetz T, Germain N, Chamberlain S, Lalande M, Martins-Taylor K. Reactivation of maternal SNORD116 cluster via SETDB1 knockdown in Prader-Willi syndrome iPSCs. Hum Mol Genet. 2014;23: 4674–4685. doi:10.1093/hmg/ddu187

67. Laufer BI, Singh SM. A macro role for imprinted clusters of MicroRNAs in the brain. MicroRNA. 2012;1: 59–64.

68. Weimann M, Grossmann A, Woodsmith J, Özkan Z, Birth P, Meierhofer D, et al. A Y2H-seq approach defines the human protein methyltransferase interactome. Nat Methods. 2013;10: 339–342. doi:10.1038/nmeth.2397

69. Myant K, Termanis A, Sundaram AYM, Boe T, Li C, Merusi C, et al. LSH and G9a/GLP complex are required for developmentally programmed DNA methylation. Genome Res. 2011;21: 83–94. doi:10.1101/gr.108498.110

70. Touleimat N, Tost J. Complete pipeline for Infinium ® Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. Epigenomics. 2012;4: 325–341. doi:10.2217/epi.12.21

Supporting information

Supporting Material and Methods

S1 Fig. DNA methylation defects in uncultured cells of ICF patients

S2 Fig. Common and distinguishing DNA methylation defects in ICF patients

S3 Fig. ZBTB24, CDCA7 and HELLS protein interactions

S4 Fig. CpG context of DMPs in ICF and non-affected subjects (NA)

S5 Fig. Genomic distribution of hypomethylated probes in ICF patients

S6 Fig. Genomic localization of HypoMPs common to all ICF subtypes

S7 Fig. Genomic features of HypoMPs common to all ICF subtypes

S8 Fig. Visualization of global hypomethylation next to hypomethylated pericentromeric repeats of chromosomes 1, 9 and 16

S9 Fig. Genomic features of HypoMPs on the sex chromosomes of ICF patients

S10 Fig. Validation of CpG hypomethylation within MAGE gene promoters in ICF patients

S11 Fig. Gene ontology and validation of CpG hypomethylation within germline gene promoters in ICF1 patients

S12 Fig. Validation of CpG hypomethylation of ZNF597 promoter in ICF1 patients

S13 Fig. Gene ontology and genomic features of HypoMPs common to ICF2, ICF3 and ICF4 patients

S14 Fig. Validation of CpG hypomethylation at clusters of small ncRNAs

S15 Fig. Validation of CpG hypomethylation at clusters of small ncRNAs

S16 Fig. DNA methylation analysis following knockdown of ZBTB24, CDCA7 or HELLS in human somatic cells

S1 Table: Genetic status of ICF patients enrolled in the genome-wide DNA methylation study

S2 Table: Hypomethylated gene clusters common to all ICF patients

S3 Table: **Gene ontology analysis of HypoMPs common to all ICF patients (GO Biological Process 2015, EnrichR)**

S4 Table: **Germline-enriched genes hypomethylated in their promoter region in ICF female patients**

S5 Table: **Germline-enriched genes hypomethylated in their promoter region in ICF male patients**

S6 Table: Hypomethylated gene clusters common to ICF2, ICF3 and ICF4 patients

S7 Table: **Histone marks enriched in genomic regions that show hypomethylation (whole HypoMPs) in ICF1 (Epigenomics Roadmap HM ChIP-seq; EnrichR)**

S8 Table: **Histone marks enriched in regions that show hypomethylation of genes promoters (TSS200; TSS1500; Illumina annotation) in ICF1 (Epigenomics Roadmap HM ChIP-seq; EnrichR)**

S9 Table: **Histone marks enriched in genomic regions that show hypomethylation (whole HypoMPs) in ICF2, ICF3 and ICF4 (Epigenomics Roadmap HM ChIP-seq; EnrichR)**

S10 Table: **Histone marks enriched in regions that show hypomethylation of genes promoters (TSS200; TSS1500; Illumina annotation) in ICF2, ICF3 and ICF4 (Epigenomics Roadmap HM ChIP-seq; EnrichR)**

S11 Table: List of primers and techniques in which they were used

Funding statement

This work was supported by ANR (http://www.agence-nationale-recherche.fr/en/; ANR-14-CE14-0011 METHYL-MEMORY), Fondation Jérôme Lejeune (https://www.fondationlejeune.org/en/), Ligue Nationale contre le Cancer (https://www.ligue-cancer.net/) and GEFLUC (http://www.gefluc-paris.fr/). GG was supported by the French Ministry of Research and Fondation ARC (http://www.fondation-arc.org/).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interest

The authors declare that there is no competing interest.

Figure captions

Fig 1. Genome wide DNA methylation profiling distinguishes two groups of ICF patients. (A) Unsupervised hierarchical clustering of ICF patients (ICF1, ICF2, ICF3 and ICF4) and non-affected (NA) subjects, based on β-values of the HM450K probes mapping to autosomes after normalization (479,259 probes retained for analysis). Below is the heat map of β-values of the 44,555 probes located on autosomes that showed a difference of DNA methylation of at least 20% (delta $\beta \geq 0.2$) between at least one ICF sub-type and NA subjects. High methylation levels are shown in red and low methylation levels in blue, according to the scale bar next to the Figure. Each column represents a patient and each row represents a probe. (B) Correlation plot comparing the β-values of the 44,555 probes between 2 ICF sub-types. The plot shows the median β-values of one ICF subtype vs. another one, and probes in red or blue represent Differentially Methylated probes (DMPs) between two genotypes. Probes shown in red are significantly more methylated in the ICF subtype indicated on the Y axis and probes in blue are significantly more methylated in the ICF subtype indicated on the X axis (differences >20%). β-values of probes that do not change between two genotypes are shown as grey dots. The black dashed line represents a linear regression based on β-values of all probes analyzed and black line represents a linear regression based on DMPs. The Pearson correlation coefficient (r) is indicated above each scatter plot. (C and D) Distribution of DMPs relative to CpG context. Bar plots show the distribution of (C) HypoMPs and (D) HyperMPs relative to CpGs islands, shores, shelves and open sea in the different ICF subtypes. The distribution of the HM450K probes relative to each CpG context is also indicated. P-values (Chi-square test) assessing significant enrichment in a given category relative to the HM450K array composition are indicated (*** $p < 0.001$).

Fig 2. Hypomethylation at centromeric and pericentromeric regions in ICF syndrome. Heat maps of β values of (A) 87 and (B) 188 probes within the cytogenetic bands 8q11.1 and 12p11.1,

respectively. ICF subtypes are indicated above the heat map. DNA methylation levels are indicated as in the legend to Fig 1. Genomic coordinates are indicated on the right of the heat map.

Fig 3. DNA methylation profiling of X chromosome and germline genes in ICF patients. (A) Heat map of β values of 10,316 DMPs mapping to the X chromosome in ICF female patients. ICF subtypes are indicated above and name of patients below the heat map. Histogram on the right of the heat map shows the percentage of common hypomethylated probes (HypoMPs) within each ICF subtype. (B) Heat maps of β values of 1,677 and 1,268 HypoMPs within the promoter of germline genes identified respectively in female (left) and male (right) ICF patients. DNA methylation levels are indicated as in the legend to Fig. 1. (C) DNA methylation levels of CpGs within CGIs of MAGEB10 (left) and MAEL (right) promoters using pyrosequencing. The methylation levels at each CpG were measured in NA male and female subjects (CTL, M or F; grey lines) and ICF1 patients (P5 and P3; blue lines). Dashed orange line represent CpG methylation levels at the X-linked MAGEB10 CGI promoter in ICF1 female patient. Error bars represent standard error. Significant p-values from Student's t-test are indicated by asterisks (* $p < 0.05$; ** $p < 0.01$).

Fig 4. CpG Hypomethylation within non-coding RNA clusters located on chromosomes 15 and 19. (A) Heat map of β values of 379 probes within the SNORD cluster embedded in the Prader-Willi Syndrome locus on chromosome 15 and (B) Heat map of β values of 21 probes within the C19MC cluster located on chromosome 19. Genomic coordinates are indicated on the right of the heat map. DNA methylation levels and category of subjects are indicated as in legend to Fig. 1. (C to E) Validation of HM450K DNA methylation data using various techniques depending on CpG context (see Supplementary Methods section). CpG methylation (C) at the SNORD115-9 promoter was determined by MSRE-qPCR, (D) at the SNORD115-14 promoter by bisulfite

sequencing and (E) at the MIR521-2 promoter by MSRE-qPCR. For MSRE-qPCR, the probe ID and the methylation-sensitive enzyme are indicated in parenthesis under the gene name. For bisulfite sequencing, white and black dots represent unmethylated and methylated CpGs, respectively, within the analyzed sequence (black line). (F) DNA methylation of CpGs within the promoter of MIR518D gene using pyrosequencing assay. The methylation levels of this CpG (probe ID: cg10583119) was measured in non-affected (CTL, M or F, age; in grey) and ICF subjects with different genotypes. Error bars represent standard error. Significant p-values from Student's t-test are indicated by asterisks (* $p < 0.05$; ** $p < 0.01$).

Fig 5. Epigenomic features of HypoMPs in ICF syndrome. (A and B) Distribution of HypoMPs relative to (A) 10 chromatin states and (B) replication timing patterns established in the GM12878 lymphoblastoid cell line. The repartition of HM450K probes within each chromatin and replication context is indicated. 450K: HM450K probes. HypoMPs: differentially hypomethylated probes between ICF and NA subjects. P-values (Chi-square test) assessing significant enrichment in a given category relative to the HM450K array composition (*** p< 0,001).

Captions for supporting figures

S1 Fig. DNA methylation defects in uncultured cells of ICF patients. (A) Volcano plots representing the methylation difference between ICF and NA individuals. The median differences in β-values are plotted on the X axis and the statistical significance [-log10 (p-value)], on the Y axis. (ICF1, n = 8; ICF2 n = 4; ICF3 n = 2 and non-affected NA, n = 10). Differentially methylated probes with a delta beta ≥0.2 and further considered for analysis, are shown in black. (B) Pie charts showing the number of differentially methylated probes (DMPs) identified on autosomes in each ICF subtype and the corresponding percentage relative to the total number of probes on the HM450K array (upper row). Their distribution in hyper- and hypo-methylated categories is shown in pie charts on the lower row.

S2 Fig. **Common and distinguishing DNA methylation defects in ICF patients.** (A) Venn diagram representing the intersection of datasets of DMPs identified in each ICF subtype. The pairwise comparison of DMPs between ICF subtypes is presented in the table next to the Venn diagram. The numbers of common DMPs are indicated with their percentage relative to the total number of DMPs identified in ICF. (B) Scatter plots comparing median β-values for individual CpGs between non-affected (NA) and each ICF sub-types as described in Figure 1C. The Pearson correlation coefficient (r) is indicated above each scatter plot.

S3 Fig. **ZBTB24, CDCA7 and HELLS protein interactions.** (A) Expression levels of DNMT3B, ZBTB24, CDCA7 and HELLS were assessed by RT-qPCR, normalized to U6 snRNA levels and presented as a fold change relative to control skin fibroblasts from healthy donors (CTL). ICF patients from subtypes 1 (pR), 2 (P10) and 3 (pS and pC) are indicated on the X-axis. (B) Expression levels of ZBTB24 and CDCA7 in lymphoblastoid cell lines (LCLs) were assessed by RT-qPCR, normalized to U6 snRNA levels and presented as a fold change relative to control LCL1 from healthy donor. LCLs were from healthy donor (LCL1 and LCL2), ICF1 (pH, pT and 8714),

ICF2 (pV and pD) and ICF4 patients (pU). Error bars represent standard error. Significant p-values from Student's t-test are indicated by asterisks (* $p < 0.05$; ** $p < 0.01$). (C) Western blot analysis of the detection of ZBTB24, CDCA7 and HELLS nuclear proteins in HEK-293 cells transiently transfected with the expression vectors encoding tagged versions of ZBTB24 and CDCA7, or HELLS proteins. Antibodies against the V5 or ZBTB24 epitopes were used to discriminate exogenous from endogenous proteins. Ponceau red coloration was used as a loading control. A molecular weight scale in indicated on the left of the panel. (D) Co-immunoprecipitation of V5-ZBTB24 with HELLS or CDCA7 from protein extracts of HEK-293 cells transiently co-transfected with V5-ZBTB24 and HELLS or CDCA7 expression vectors, using antibodies against the V5-tag or pre-immune IgG. Co-immunoprecipitation of CDCA7 and HELLS from protein extracts of HEK-293 cells transiently transfected with HELLS and CDCA7 expression vectors, using antibodies against HELLS protein or control antibodies. INP: input lane; IP: immunoprecipitation; IgG: control immunoprecipitation.

S4 Fig. **CpG context of DMPs in ICF and non-affected subjects (NA).** Cumulative frequency plots showing the distribution of DMPs β-values relative to CGIs, shores, shelves and open sea classification (Illumina annotation) depending on genotype. The median β-values are represented on the X-axis and the frequencies (counts) on the Y-axis.

S5 Fig. **Genomic distribution of hypomethylated probes in ICF patients.** (A) Histogram representing the distribution of hypo- (HypoMPs; left) and hyper- (HyperMPs; right) methylated probes according to the different gene features: promoter, gene start, gene body, 3'UTR, ambiguous and intergenic (Illumina annotation) depending on genotype. The distribution of probes represented on the HM450K array (450K) relative to CpG context is indicated. (B) and (C) Histograms representing the distribution of DMPs after the breakdown of CGI and open sea CpG annotations into gene features (Illumina annotation). (B) Distribution of HypoMPs. (C)

Distribution of HyperMPs. The distribution of probes represented on the HM450K array (450K) relative to each category is indicated.

S6 Fig. **Genomic localization of HypoMPs common to all ICF subtypes. (**A) Venn diagram representing the intersection of datasets of HypoMPs identified in ICF1, 2, 3 or 4 patients. The numbers of common HypoMPs are indicated for each intersection. The total number of HypoMPs for each subtype is indicated under each ICF category. (B) Histogram showing the chromosomal distribution of the 433 HypoMPs that are common to all ICF subtypes, showing a higher proportion on certain chromosomes including Chr5 (PCDH cluster) mentioned in the text. (C) Heat map representing β-values for DMPs on whole Chr 5 identified in ICF patients and healthy subjects. The localization of the PCDH cluster is indicated. High methylation levels are shown in red and low methylation levels in blue, according to the scale bar next to the figure. Each row represents a patient and each column represents a probe. (D) Validation of CpG hypomethylation within PCDHB13 and PCDHB18 promoters in ICF patients. Genomic DNA from peripheral blood of non-affected (CTL M36), ICF1 (P3), ICF2 (pV), ICF3 (pS) and ICF4 (pU) patients were subjected to COBRA analysis using the methylation-sensitive Taqα1 enzyme. In control cells, PCDHB13 was partially methylated whereas PCDHB18 was almost fully methylated. In contrast, both gene promoters were hypomethylated in ICF cells with a more prominent hypomethylation in ICF1 cells compared to ICF2, 3, and 4. The CpG analyzed (probe ID) and the absence (-) or presence (+) of the enzyme are indicated in parenthesis under the gene name. U and M refer to the position of unmethylated (undigested) and methylated (digested) fragments, respectively.

S7 Fig. **Genomic features of HypoMPs common to all ICF subtypes.** (A) Histograms representing the distribution of the 433 HypoMPs common to all ICF subtypes relative to gene features (left) and CpG context (right) (Illumina annotations). (B) Bar plots showing the

distribution of HypoMPs common to all ICF subtypes relative to DNA repeats (Intersect with Repeatmasker). The distribution of probes represented on the HM450K array (450K) within each family of repeat is indicated. P-values (Chi-square test) assessing significant enrichment in a given family of repeat relative to the HM450K array composition: ***p< 0.001; **p<0.01.

S8 Fig. **Visualization of global hypomethylation next to hypomethylated pericentromeric repeats of chromosomes 1, 9 and 16.** Heat maps representing β-values of HypoMPs identified in ICF patients and healthy subjects at pericentromeric regions and their adjacent cytogenetic bands of (A) Chr 1, (B) Chr9 and (C) Chr 16. The position of centromeric and pericentromeric cytogenetic bands is illustrated on the chromosome cartoon. Centromeric alpha-satellites (α-Sat) are represented in red, pericentromeric Sat II or Sat III repeats in green and the cytogenetic band corresponding to the heat map in yellow. No probe on the HM450K array map to the pericentromeric cytogenetic band 9q12. High methylation levels are shown in red and low methylation levels in blue, according to the scale bar next to the figure. Each column represents a patient and each row represents a probe.

S9 Fig. **Genomic features of HypoMPs on the sex chromosomes of ICF patients. (**A) Snapshot from Integrative Genomics Viewer (IGV) of the chromosomal distribution of HypoMPs in males and females ICF patients depending on genotype. The chromosome number is indicated above the tracks. The height of the peaks is proportional to the density of HypoMPs. The sex chromosomes are highlighted by a red rectangle. (B) Venn diagram representing the intersection of datasets of HypoMPs identified in female ICF patients. The total number of HypoMPs for each subtype is indicated under each ICF category. (C) and (D) Histograms representing the distribution of the 181 HypoMPs on the X chromosome common to female ICF subtypes relative to gene features (left) and CpG context (right). (E) Histogram representing the percentage of HypoMPs on the sex

chromosomes in male ICF subtypes. The table summarizes the distribution of HypoMPs relative to CpG context with the number of HypoMPs and the percentage relative to the number of probes on the Y chromosome represented on the HM450K array.

S10 Fig. **Validation of CpG hypomethylation within MAGE gene promoters in ICF patients.** (A) Validation of CpG hypomethylation within CGI promoter of MAGEB10 promoter using COBRA and pyrosequencing assay. COBRA was performed on genomic DNA from peripheral blood DNA of non-affected (CTL F42), ICF1 (P3), ICF2 (P10), ICF3 (pS and pC) female patients. The methylation-sensitive enzyme is indicated into parenthesis (HpyCH4IV; 3 sites analyzed) under the gene name. The absence (-) or presence (+) of the enzyme are indicated in parenthesis under the gene name. U and M referred to the position of unmethylated (undigested) and methylated (digested) fragments, respectively. (B) Validation of CpG hypomethylation within CGI promoter of MAGEB10 promoter using bisulfite pyrosequencing assay. The methylation levels at each CpG were measured on genomic DNA from blood of non-affected (CTL; grey), ICF2 (P10, blue), ICF3 (pS and pC, light blue) patients. Dashed lines represent methylation levels at the interrogated genomic sequence in female ICF patients. (B) Validation of CpG hypomethylation at MAGEB3 promoter by MSRE-qPCR. Genomic DNA from peripheral blood of non-affected (CTL M36 and F42), ICF1 (P5 and P3), ICF2 (P10), ICF3 (pS and pC) patients was used in this assay. Data showed a loss of DNA methylation in all females ICF tested, at levels comparable to that obtained from the HM450K array (patient P3: 33% loss by MSRE-qPCR compared to 26% loss on HM450K array; patient P10: 50% compared to 46%; pS/pC: 42% from both assays). The probe ID and the methylation-sensitive enzyme (HpaII) are indicated in parenthesis under the gene name. Error bars represent standard error. Significant p-values from Student's t-test are indicated by asterisks (* $p < 0.05$; ** $p < 0.01$).

S11 Fig. **Gene ontology and validation of CpG hypomethylation within germline gene promoters in ICF1 patients.** (A) Top Gene Ontology Biological Process terms of HypoMPs in promoters (Illumina annotation) in ICF1 versus NA subjects obtained from Enrichr. (B) Validation of DNA hypomethylation of CGI of MAEL, TDRD6 and SPO11 promoters. Genomic DNA from peripheral blood from the indicated subjects was analyzed by COBRA using the indicated methylation-sensitive enzymes (BstUI for MAEL and TDRD6; HpyCH4IV for SPO11). When applicable, the probe ID was mentioned. The absence (-) or presence (+) of the enzyme are indicated in parenthesis under the gene name. U and M referred to the position of unmethylated (undigested) and methylated (digested) fragments, respectively. (C) Validation of CpG hypomethylation within the CGI of SPO11 promoter using bisulfite pyrosequencing assay. The methylation levels at each CpG were measured on gemomic DNA from blood of non-affected (CTL; grey) and ICF1 patients (P3 and P5; orange). (D) Validation of CpG hypomethylation within promoters of ADAM2, SPO11 and C15ORF60 promoters was validated by COBRA as described in (B).

S12 Fig. **Validation of CpG hypomethylation of ZNF597 promoter in ICF1 patients.** (A) Validation of CpG methylation within the promoter of ZNF597 on genomic DNA from peripheral blood of non-affected subject (CTL, M36) and ICF patients (ICF1, P3; ICF2, pV; ICF3, pS; ICF4, pU) using COBRA. The probe ID and the methylation-sensitive enzyme (HpyCH4IV; 2 sites analyzed) are indicated in parenthesis under the gene name. The absence (-) or presence (+) of the enzyme are indicated in parenthesis under the gene name. U and M referred to the position of unmethylated (undigested) and methylated (digested) fragments, respectively. (B) Validation on genomic DNA from blood of non-affected subject (CTL) and ICF1 patients (P3 and P5) using bisulfite pyrosequencing assay. The methylation levels at each CpG were measured on genomic DNA from blood of non-affected (CTL; grey) and ICF1 patients (P3 and P5; orange). Significant p-values from Student's t-test are indicated by asterisks (* $p < 0.05$; ** $p < 0.01$). (C) Validation on

genomic DNA from primary skin fibroblasts of ICF1 P3 patient by bisulfite sequencing. For bisulfite sequencing, white and black dots represent unmethylated and methylated CpGs, respectively, within the analyzed sequence (black line). (D) Validation of CpG methylation within the promoter of ZNF597 on genomic DNA from primary skin fibroblasts derived from non-affected subject (CTL WT) and ICF patients (ICF1, pR; ICF2, P10; ICF3, pS) using COBRA as described in (A). Error bars represent standard error.

S13 Fig. **Gene ontology and genomic features of HypoMPs common to ICF2, ICF3 and ICF4 patients.** (A) Top Gene Ontology Biological Process terms of HypoMPs common to ICF2, 3 and ICF4 obtained from Enrichr. (B) Validation of CpG methylation within the promoter of DRD3 and NDN on genomic DNA from peripheral blood of non-affected subject (CTL, M36) and ICF patients (ICF1, P3; ICF2, pV; ICF3, pS; ICF4, pU) using COBRA. For NDN promoter, the methylation of two probes in the promoter of NDN were analyzed (NDN* and NDN**). The probe ID and the methylation-sensitive enzyme are indicated in parenthesis under the gene name. The absence (-) or presence (+) of the enzyme are indicated in parenthesis under the gene name. U and M referred to the position of unmethylated (undigested) and methylated (digested) fragments, respectively. (C) Validation of CpG hypomethylation within the NRXN3 promoter on genomic DNA from blood of the indicated patients and healthy subjects (CTL) using MSRE-qPCR. The percentage of relative DNA methylation is indicated on the Y-axis. The probe ID and the methylation-sensitive enzyme are indicated in parenthesis under the gene name. Error bars represent standard error. Significant p-values from Student's t-test are indicated by asterisks (* $p < 0.05$; ** $p < 0.01$). (D) Venn diagram representing the intersection of datasets of HypoMPs identified in ICF2, 3 and 4 patients. The total number of HypoMPs is indicated below each ICF category. (E) Histogram showing the chromosomal distribution of the 484 HypoMPs common to ICF2, 3 and ICF4. The position of certain gene clusters mentioned in the text is indicated.

S14 Fig. **Validation of CpG hypomethylation at clusters of small ncRNAs.** (A) Validation of hypomethylation at SNORD115-5 using MSRE-qPCR on genomic DNA from blood of the indicated patients and healthy subjects (CTL). The percentage of relative DNA methylation is indicated on the Y-axis. The probe ID and the methylation-sensitive enzyme are indicated in parenthesis under the gene name. (B) Validation of CpG hypomethylation of SNORD115-14, SNORD115-5 promoters on genomic DNA from peripheral blood of non-affected subject (CTL, M36) and ICF patients (ICF1, P3; ICF2, pV; ICF3, pS; ICF4, pU) using COBRA. The probe ID and the methylation-sensitive enzyme are indicated in parenthesis under the gene name. The absence (-) or presence (+) of the enzyme are indicated in parenthesis under the gene name. U and M referred to the position of unmethylated (undigested) and methylated (digested) fragments, respectively. (C) Validation of hypomethylation at MIR521-2 and MIR518D by MSRE-qPCR as in (A). Error bars represent standard error. Significant p-values from Student's t-test are indicated by asterisks (* $p< 0.05$; ** $p< 0.01$).

S15 Fig. **Validation of CpG hypomethylation at clusters of small ncRNAs.** (A) Heat map of β-values of DMPs localized in the C14MC cluster of miRNAs on chromosome 14q32.31. Genomic coordinates are indicated on the right of the heat map. High methylation levels are shown in red and low methylation levels in blue, according to the scale bar next to the figure. Each column represents a patient and each row represents a probe. (B) Validation of CpG hypomethylation of SNORD114-1 promoter on genomic DNA from peripheral blood of non-affected subject (CTL, M36) and ICF patients (ICF1, P3; ICF2, pV; ICF3, pS; ICF4, pU) using COBRA. The probe ID and the methylation-sensitive enzyme are indicated in parenthesis under the gene name. The absence (-) or presence (+) of the enzyme are indicated in parenthesis under the gene name. U and M referred to the position of unmethylated (undigested) and methylated (digested) fragments, respectively. (C) Validation of DNA hypomethylation at 5 CpGs within the MIR376C promoter using bisulfite pyrosequencing assay. The methylation levels at each CpG were measured on

gemomic DNA from blood of non-affected (CTL; grey), ICF2 (pV and P10, blue), ICF3 (pS and pC, light blue) and ICF4 (pU; green) patients. At least 40% of methylation was lost at CG2 and CG3 in the two ICF2 patients, one ICF3 patient (pS) and the ICF4 patient. ICF2 pV and ICF4 pU patients were the most affected and showed almost complete loss of DNA methylation at certain CpG analyzed (CG2 and CG3;) (D) Validation of hypomethylated CpG within NDN, MIR376C and SNORD115-14 promoters on primary skin fibroblasts derived from healthy controls (CTL) and ICF patients (ICF1, pR; ICF2, pP; ICF3, pS; ICF2, P10) using COBRA as in (B). Of note, ICF2 patient pP was not included in the HM450K study (Supplemental Table S1).

S16 Fig. DNA methylation analysis following knockdown of ZBTB24, CDCA7 or HELLS in human somatic cells. (A) Expression levels of ZBTB24, CDCA7 and HELLS were assessed by RT-qPCR after knockdown of ZBTB24, CDCA7 and HELLS using RNA interference in primary skin fibroblasts from a healthy subject, normalized to U6 snRNA levels and presented as a fold change relative to control siRNA. (B) CpG methylation levels at MIR521-2, MIR518D, SNORD115-5, SNORD115-9 and NRXN3 promoters were assessed by MSRE-qPCR as described previously. Genomic DNA from primary skin fibroblast of ICF2 patient P10 patient with germline mutations in ZBTB24 was used as a positive control where centromeric and pericentromeric satellite repeats are hypomethylated. Error bars represent standard error. **(**C) Southern blot analysis of DNA methylation at Satellite II (BstBI) and Satellite α (HhaI) after the knockdown of ZBTB24, CDCA7 or HELLS in primary skin fibroblasts from a healthy donor. Genomic DNA from primary skin fibroblasts of healthy (CTL) or ICF2 (P10) subjects were used as negative and positive controls, respectively. Molecular weights of the DNA size markers (Marker) are in kilobase pair.

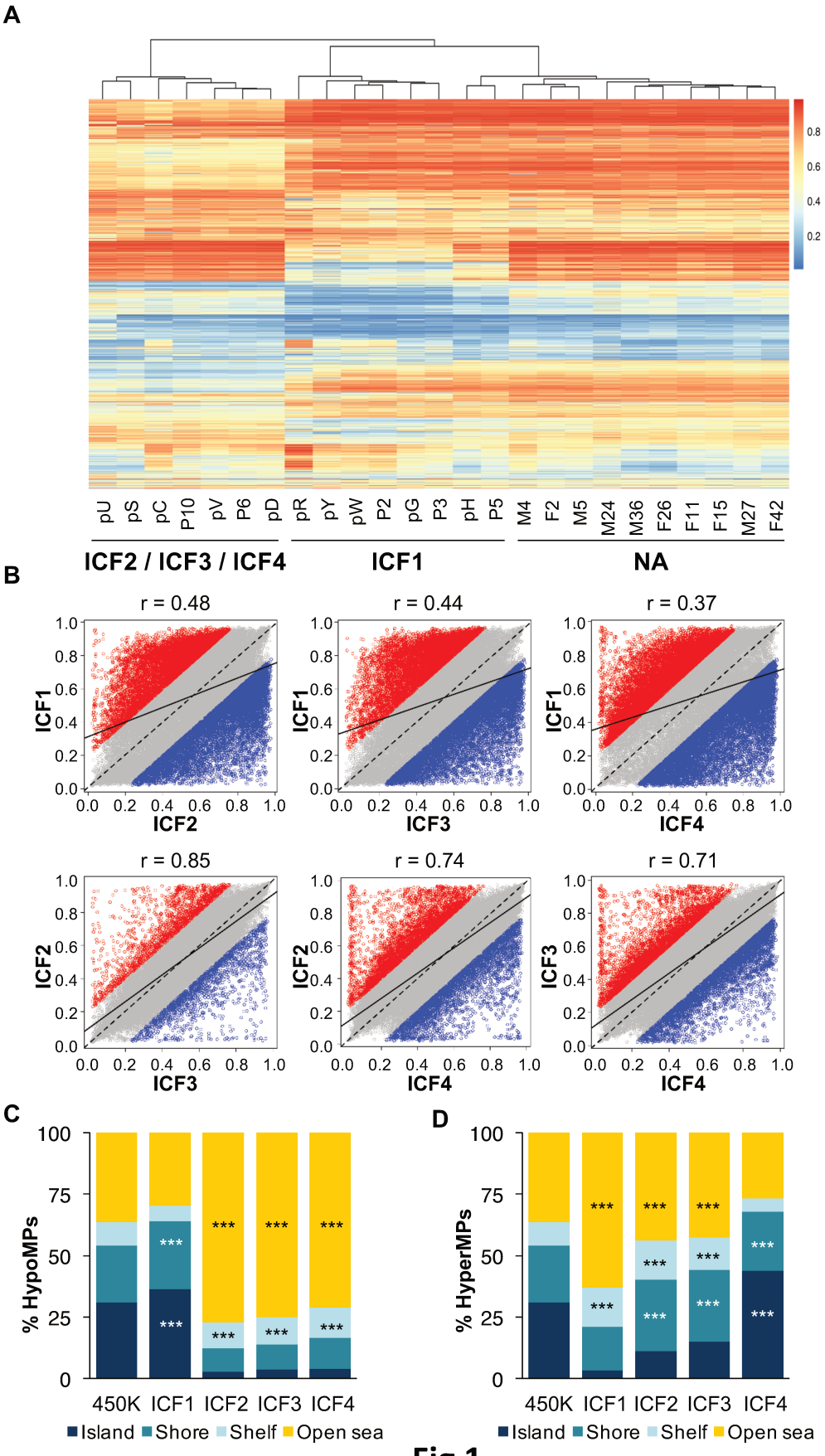

A
0.8
0.6
0.4
0.2
pU pS pC P10 pV P6 pD pR pY pW P2 pG P3 pH P5 M4 F2 M5 M24 M36 F26 F11 F15 M27 F42
ICF2 / ICF3 / ICF4
ICF1
NA
B
r = 0.48
r = 0.44
r = 0.37
r = 0.85
r = 0.74
r = 0.71
ICF1
ICF2
ICF3
ICF4
0.0 0.2 0.4 0.6 0.8 1.0
C
D
% HypoMPs
% HyperMPs
100
75
50
25
0
***
450K ICF1 ICF2 ICF3 ICF4
Island Shore Shelf Open sea

Fig 1

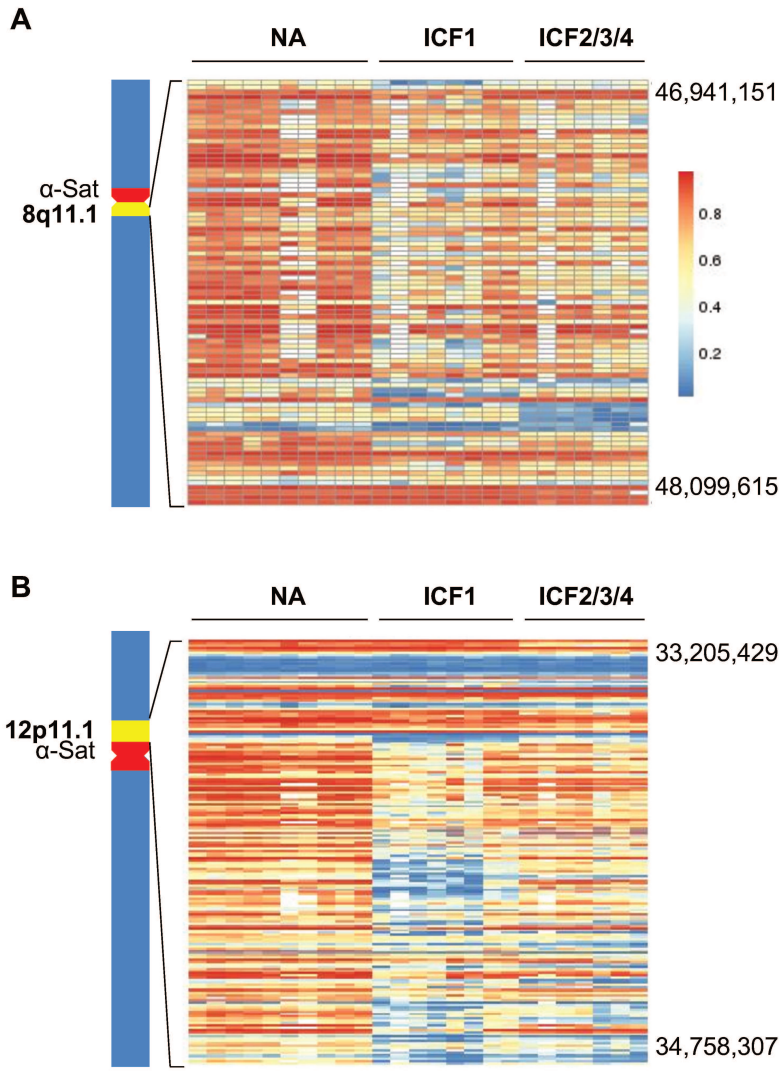

A
NA
ICF1
ICF2/3/4
46,941,151
α-Sat
8q11.1
0.8
0.6
0.4
0.2
48,099,615
B
NA
ICF1
ICF2/3/4
33,205,429
12p11.1
α-Sat
34,758,307

Fig 2

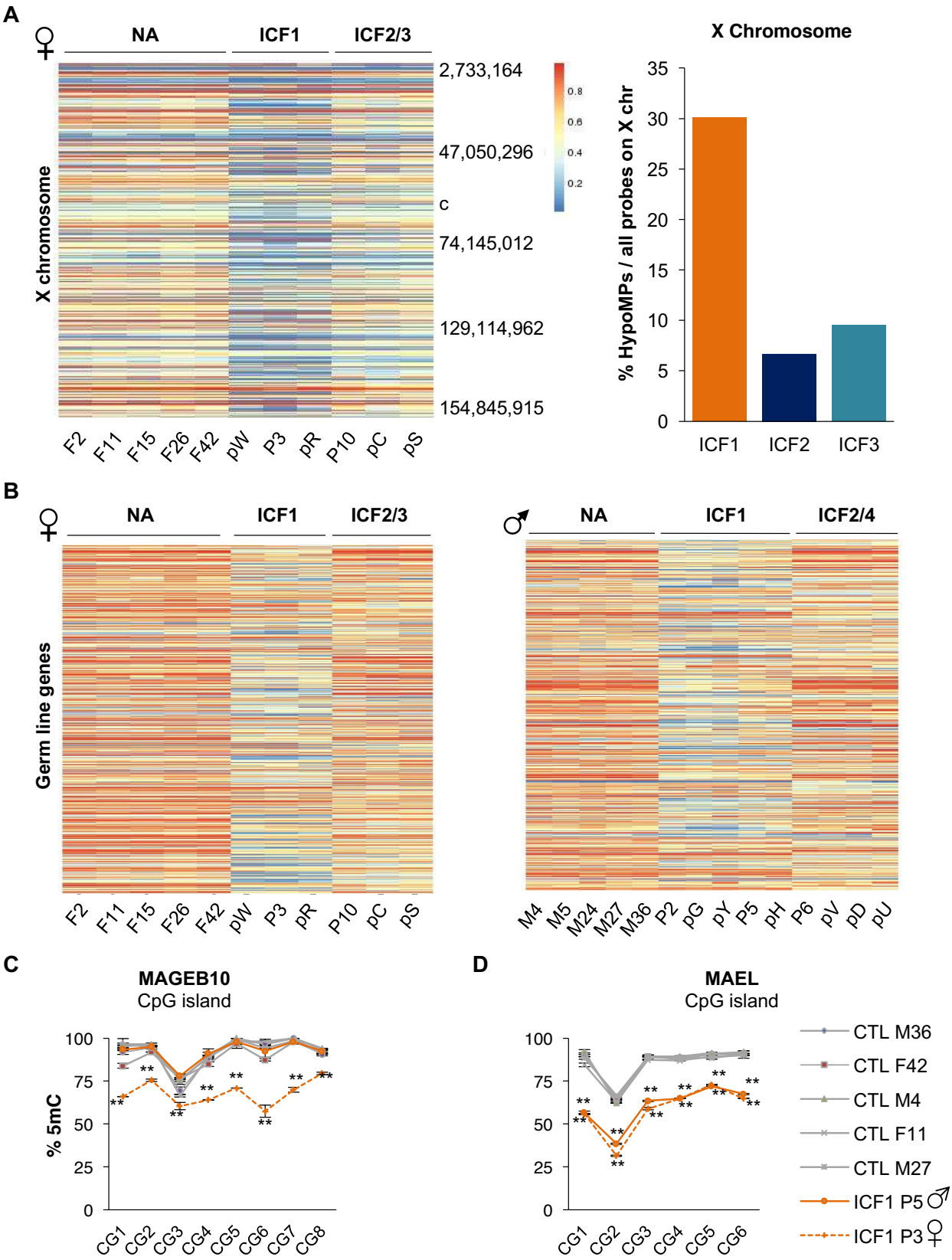

A
NA
ICF1
ICF2/3
X chromosome
2,733,164
47,050,296
c
74,145,012
129,114,962
154,845,915
F2 F11 F15 F26 F42 pW P3 pR P10 pC pS
X Chromosome
% HypoMPs / all probes on X chr
ICF1 ICF2 ICF3
B
NA
ICF1
ICF2/3
ICF2/4
Germ line genes
F2 F11 F15 F26 F42 pW P3 pR P10 pC pS
M4 M5 M24 M27 M36 P2 pG pY P5 pH P6 pV pD pU
C
MAGEB10
CpG island
% 5mC
CG1 CG2 CG3 CG4 CG5 CG6 CG7 CG8
D
MAEL
CpG island
CG1 CG2 CG3 CG4 CG5 CG6
CTL M36
CTL F42
CTL M4
CTL F11
CTL M27
ICF1 P5
ICF1 P3

Fig 3

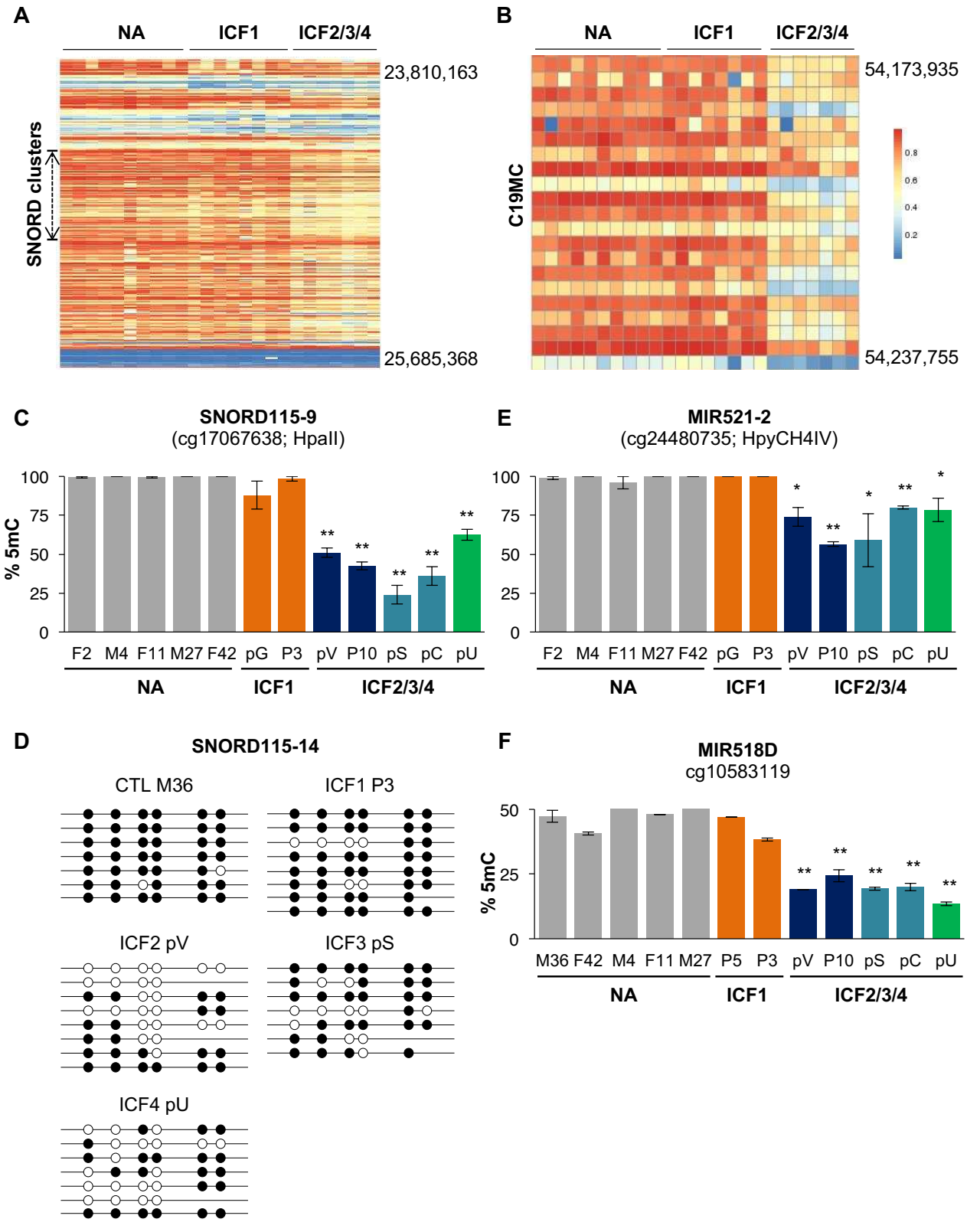

A
NA
ICF1
ICF2/3/4
23,810,163
SNORD clusters
25,685,368
B
NA
ICF1
ICF2/3/4
54,173,935
C19MC
0.8
0.6
0.4
0.2
54,237,755
C
SNORD115-9
(cg17067638; HpaII)
% 5mC
100
75
50
25
0
F2 M4 F11 M27 F42 pG P3 pV P10 pS pC pU
NA
ICF1
ICF2/3/4
E
MIR521-2
(cg24480735; HpyCH4IV)
% 5mC
100
75
50
25
0
F2 M4 F11 M27 F42 pG P3 pV P10 pS pC pU
NA
ICF1
ICF2/3/4
D
SNORD115-14
CTL M36
ICF1 P3
ICF2 pV
ICF3 pS
ICF4 pU
F
MIR518D
cg10583119
% 5mC
50
25
0
M36 F42 M4 F11 M27 P5 P3 pV P10 pS pC pU
NA
ICF1
ICF2/3/4

Fig 4

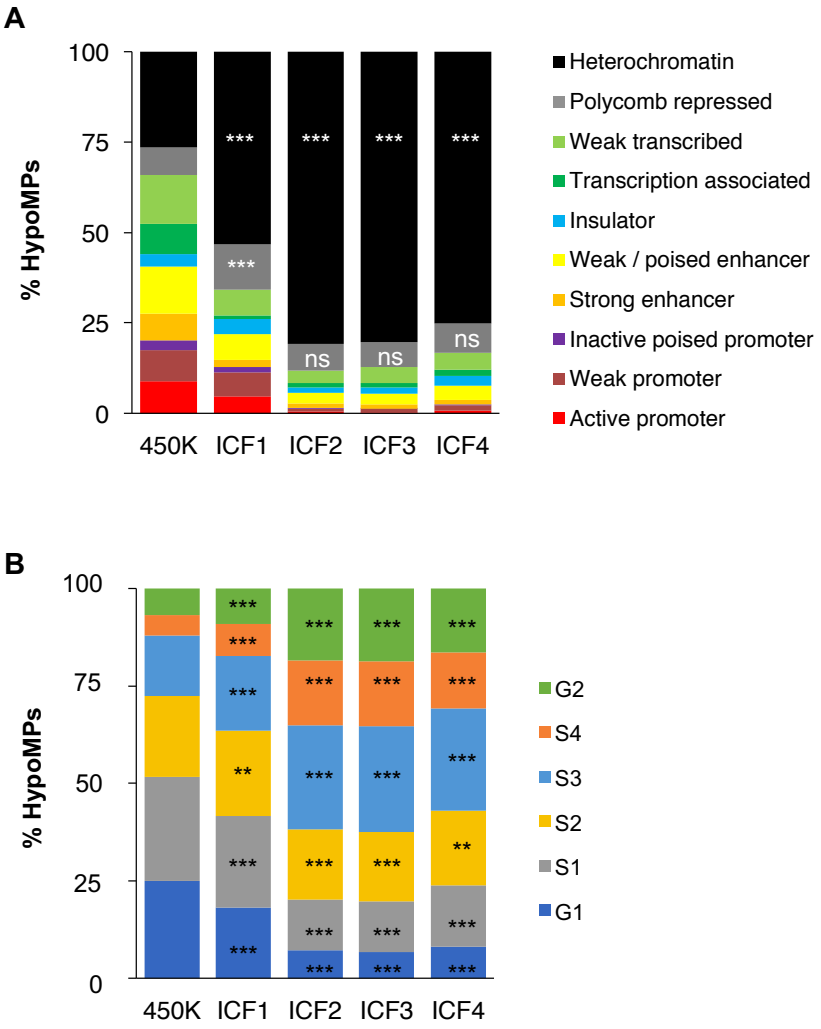

A
100
75
50
25
0
% HypoMPs
***
***
***
***
***
ns
ns
ns
450K
ICF1
ICF2
ICF3
ICF4
Heterochromatin
Polycomb repressed
Weak transcribed
Transcription associated
Insulator
Weak / poised enhancer
Strong enhancer
Inactive poised promoter
Weak promoter
Active promoter
B
100
75
50
25
0
% HypoMPs
450K
ICF1
ICF2
ICF3
ICF4
G2
S4
S3
S2
S1
G1

Fig 5

SUPPORTING MATERIAL AND METHODS

Primary cells and cell lines

Primary skin fibroblasts (passage 10 for non-affected subjects, passage 5 for pR; passage 14 for pC and pP; passage 9 for pS) have been previously described [1]. Primary skin fibroblasts from P10 ICF patient were derived by expanding primary cultures of skin biopsies and cultured in alpha-MEM supplemented with 20% Fetal Calf Serum, 1% Penicillin-Streptomycin, non-essential amino-acids and 1M HEPES (PAA, The Cell Culture Company).

Lymphoblastoid cell lines (LCLs) have also been previously described [2] and include LCLs from healthy donors (LCL1 and 2) and LCLs pH, pT and 8714 for ICF1 patients, pV and pD for ICF2; pC and pU for ICF3 and ICF4 patients. LCL 8714 were obtained from the Coriell Cell Repositories (USA).

Primary skin fibroblasts and LCLs were cultured as mentioned previously [1] in DMEM or RPMI 1640, respectively, supplemented with 15% Fetal Calf serum, glutamine and antibiotics (Invitrogen).

Expression vectors and cell transfection

ZBTB24, CDCA7 and HELLS cDNAs (BC050933, BC066169, BC100394; TransOMIC technologies) were cloned into the pTCP vector. Oligonucleotides encoding V5 or HA tags were inserted in frame with full-length ZBTB24 or CDCA7, at the 5' or 3' end, respectively, using Q5 site-directed mutagenesis kits (New England Biolabs) according to manufacturer's instructions, to generate pTCP-V5-ZBTB24 and pTCP-CDCA7-HA expression vectors. Sequences of primers are provided in the S11 Table.

Transient transfections were carried out in HEK-293 cells and primary skin fibroblasts using JetPEI (POLYplus transfection) following manufacturer’s instruction with 5 µg of V5-ZBTB24, CDCA7-HA or HELLS expressing vectors. GFP expressing vector (pEGFP-C1 vector; Clontech) was used as a control for transfection efficiency.

Knockdown of gene expression in primary skin fibroblast

Two rounds of transfection with synthetic short interfering RNAs (siRNAs) at a final concentration of 20nM were performed using Lipofectamine RNAiMax Transfection Reagent (ThermoFisher Scientific) following the manufacturer's instructions: a first round on fibroblasts in suspension and the second when cells were allowed to adhere to the plate. Cells were collected 48 h later for RNA and genomic DNA extraction. Sequences of siRNAs used in the study have been previously provided [3].

Preparation of genomic DNA

Genomic DNA was extracted and purified from cell lines or whole blood using the NucleoSpin®Tissue kit (Macherey-Nagel) according to manufacturer's instructions. The DNA pellet was eluted in TE containing 20 µg/mL RNAse A (Euromedex).

Genome-Wide DNA methylation data processing

Illumina 450K microarray signals were extracted with GenomeStudio software (Illumina) and processed and normalized by an updated version of the previously developed pipeline [4] that includes the following steps: background signal subtraction, filtering of probes that can generate spurious signals caused by hybridization to alternate genomic sequences highly homologous to the intended target [5] and Infinium I / Infinium II bias correction. For this last step, The Infinium I methylated (M) and unmethylated (U) signals were used as 'anchors' to normalize Infinium II M and U signal intensity at the level of probe categories. We computed beta-values for each probe using the classical formula proposed by Illumina: beta = max(M,0)/[max(M,0)+max(U,0)+100], where, for a given probe, M and U correspond to the methylated and the unmethylated signals intensities, respectively. We defined 16 functional categories by crossing the "gene association" and "CpG density" annotations proposed by [6]. We performed a multi-samples quantile

normalization in order to make the distribution of probe intensities comparable before differential analysis. To compare the methylation status of a given probe between two genotypes, we computed the median of probe beta-value for each genotype. Using the median rather than the mean makes the group average calculation less sensitive to extreme values and to outliers. For autosomal probes analysis, male and female samples were normalized together excluding probes located on sexual chromosomes. For the analysis of probes located on sexual chromosome X and Y chromosomes, male and female samples were normalized separately. Beta-values associated with a detection p-value greater than 0.01 were flagged and not taken into account for differential analysis. Due to the small number of samples representing each genotype (ICF and Non-affected), statistical test approaches were not suitable for the differential methylation analysis. We considered a probe as differentially methylated (DMP) between two genotypes when the beta-value difference was greater or equal to 0.2 (interpreted as 20% methylation difference) between the median beta-values. DMPs were associated with genes and grouped into functional gene components according to Illumina annotations (Promoter, TSS, gene body, 3'UTR). Thus, we aggregated differential methylation data at the gene level and considered that a gene's component had a differential methylation profile if showing at least one differential probe and no contradictory differential probe (probe with opposite methylation dynamics).

Student t-test, Volcano plots, hierarchical clustering representations, correlation calculations, methylation density plots, heatmaps and histograms have been generated with R statistical environment functions. Hierarchical clustering is based on Euclidean distance and Ward agglomerative method.

In-house R scripts were used for between-samples normalization, differential probes and genes identification and by gene components methylation signal summarization.

Webservers and databases

Datasets of differentially methylated probes were compared to publically available data downloaded from the UCSC genome Browser using Galaxy tools. We used the following tracks: Repeating Elements by RepeatMasker (Repeat masker [7]; Replication Timing by Repli-seq from ENCODE/University of Washington (UW Repli-seq [8]) and Chromatin State Segmentation by HMM from ENCODE/Broad (Broad ChromHMM [9]). We used Enrichr Bioinformatics Resources [10,11] to assess Gene Ontology, Biological process, and histone modifications extracted from the Roadmap Epigenome Project (White blood cells dataset). We used the Integrative Genomics Viewer (IGV) as a visualization tool for the exploration of DMPs datasets [12,13]. The list of genes whose expression is restricted or enriched in germ cells was obtained from the Human Protein Atlas project.

Validation of DNA methylation data

We used three different techniques, namely MSRE-qPCR (Methylation Sensitive Restriction Enzyme-based qPCR), COBRA (Combined Bisulfite Restriction Analysis) and bisulfite sequencing to validate DNA methylation data generated by HM450K array. Because this technology allows the interrogation of DNA methylation at single CpG site resolution on bisulfite converted DNA, we mostly performed validation using methods requiring bisulfite conversion, as it is the case for COBRA, bisulfite-cloning-sequencing or bisulfite pyrosequencing experiments. The last two methods were usually preferred to validate DNA methylation status of several CpGs in close proximity. When COBRA was not appropriate, the MSRE-qPCR method was used to analyze DNA methylation of single CpG within a methylation-sensitive restriction site. Primers used in the three validation techniques aforementioned are listed in S11 Table.

MSRE-qPCR

We digested 300 ng of genomic DNA with 10 U of HpaII or HpyCH4IV (New England BioLabs), followed by qPCR to measure the percentage of digestion relative to undigested control

DNA. Each data shown in qPCR analysis is the result of at least three independent experiments performed on at least three independent genomic extractions.

COBRA

Bisulfite conversion of genomic DNA was performed with the EpiTect bisulfite kit (Qiagen). We performed PCR amplification of converted DNA using platinum Taq DNA polymerase (Thermo Fischer Scientific). We used the online tool MethPrimer to design primers on bisulfite-treated DNA. PCR programs consisted of 50 cycles of PCR with annealing temperatures varying from 52 to 54°C. For COBRA, we digested the PCR product with 10 U Taqα1, BSTUI or HpyCH4IV for 2 h at 65°C, 60°C or 37°C, respectively, and used an equal amount of PCR product for the undigested control.

Bisulfite sequencing

Bisulfite conversion and PCR amplification were performed as described for COBRA assay. PCR fragments were then cloned in pGEM®-T Easy Vector System I (Promega). Individual clones (at least 8) were screened for the presence of inserts using EcoRI digestion and sequenced using the T7 universal primer (Eurofins genomics, Germany). For pyrosequencing method, PCR amplification was performed using biotinylated reverse primers. PCR products were then converted to single-stranded DNA templates with the Pyromark Vacuum Prep Workstation (Qiagen) and sequencing was performed with primers on a PyroMark Q24 System (Qiagen) according to manufacturer instructions. Data were extracted and analyzed using PyroMark Q24 2.0.6 software (Qiagen).

Analysis of DNA Methylation by Southern blotting

Genomic DNA (400 ng) from primary skin fibroblasts, before or after RNA interference, was digested with 20 units of BstBI or HhaI (New England Biolabs) for 16h to analyze DNA methylation patterns at Sat II and α-Sat repeats, respectively. The digested DNA fragments were separated by electrophoresis using 1% agarose gels and transferred overnight to Hybond-N+

membranes (GE Healthcare) in 20X SSC. After UV crosslinking, the membranes were pre-hybridized in 6X SSC, 5X Denhardt and 0.1% SDS and then hybridized with ^{32}P-labeled Sat II or α-Sat oligonucleotide probes (5'-TCGAGTCCATTCGATGAT-3' or 5'-ATGTGTGCATTCAACTCACAGAGTTGAAC-3'). Pre-hybridization and hybridization was carried out at 42°C for 1h. The membranes were washed 3 times in 6X SSC and 0.1%SDS at 37°C and then subjected to phosphor-imaging using FLA 7000 phosphor-imager (Fuji).

RNA extraction and analysis of gene expression

Total RNA was isolated from cell lines using TRI Reagent® (Sigma) according to manufacturer's instructions and as previously described [1]. Contaminant genomic DNA was eliminated with TURBO DNA-free kit (Ambion). Reverse transcription was carried out using 500ng DNA-free RNA in a reaction containing 50μM random hexamers (Roche), 20U of RNase inhibitor (New England Biolabs) and 20U of RevertAid Reverse Transcriptase (Fermentas). Complementary DNA reactions were used as templates for PCR reactions. Real-time PCR was performed using the SensiFAST SYBR NoROX Kit (Bioline) supplemented with 0.2μM specific primer pairs. Real-time quantitative PCR were run for 40 cycles on a light cycler rapid thermal system (LightCycler®480 2.0 Real time PCR system, Roche) with 20 sec of denaturation at 95°C, 20 sec of annealing at 60°C and 20 sec of extension at 72°C for all primers, and analyzed by the comparative CT method (ΔΔCT) using U6 RNA as an invariant RNA. Each data shown in RT-qPCR analysis is the result of at least three independent experiments performed on at least three independent RNA extractions. Primer sequences are provided in S11 Table.

Co-immunoprecipitations and western blot analysis

HEK-293 cells were transfected with V5-ZBTB24, CDCA7-HA and HELLS expression vectors using JetPEI (POLYplus transfection) following manufacturer's instructions. HEK-293

were harvested 36 hr post-transfection and nuclear proteins were extracted as previously described [14] using lysis buffer [50 mM Tris pH8, 150 mM NaCl, 1mM EDTA, 5% glycerol, 0,5% Triton X-100 and protease inhibitor mixture (Roche)]. Nuclear extracts were then sonicated (Bioruptor, Diagenode, 5 cycles of 15 s ON/30 s OFF). Cell debris were pelleted by centrifugation at 16,000g for 10 min and supernatants were collected. For immunoprecipitations, 2 µg of specific antibodies against V5 epitope tag (sc81594, Santa Cruz Biotechnology), HELLS (sc46665, Santa Cruz Biotechnology) or control IgG (Mouse IgG, M7023, Sigma-Aldrich) were added to 200µg of protein extract and incubated at 4 °C overnight on a wheel. Next, 30 µL of 0.1% BSA-saturated protein G Sepharose beads (GE Healthcare) were added to each sample, followed by incubation at room temperature for 2 hr on a wheel. Beads were collected by centrifugation for 1 min at 1,000 × g and washed five times in lysis buffer. Beads were then boiled in 30 µL of Laemmli 1X buffer (50 mM Tris pH7, 10% glycerol, 2% SDS, 0.01% bromophenol blue, 5% 2-Mercaptoethanol) for 5 min and immunoprecipitated proteins were collected by centrifugations for 1 min at 16,000 × g. Proteins were separated on 10% SDS/PAGE gels and blotted onto PVDF transfer membrane (ThermoScientific).

Analysis of ZBTB24, CDCA7 and HELLS protein levels following overexpression of their respective expression vectors was performed in HEK-293. After 48 hr, nuclei were lysed using RIPA buffer [50mM Tris pH7.4, 150mM NaCl, 1mM EDTA, 1% NP40, 0.5% Na-deoxycholate, 0.1% SDS and protease inhibitor mixture (Roche)]. 100 µg of nuclear extracts were separated on 10% SDS/PAGE gels and blotted onto PVDF transfer membrane (ThermoScientific). Membranes were incubated for 2H at room temperature with the antibodies mentioned above and the presence of each protein was revealed using ECL kit (Pierce).

For western blot experiments, anti-ZBTB24 (PM085, MBL Life Science), anti-CDCA7 (15249-1-AP, Proteintech), V5 epitope tag (sc81594, Santa Cruz Biotechnology) or anti-LSH/HELLS (sc46665, Santa Cruz Biotechnology) antibodies were used and proteins were visualized using ECL kit (Pierce).

SUPPLEMENTARY REFERENCES

1. Velasco G, Walton EL, Sterlin D, Hédouin S, Nitta H, Ito Y, et al. Germline genes hypomethylation and expression define a molecular signature in peripheral blood of ICF patients: implications for diagnosis and etiology. Orphanet J Rare Dis. 2014;9: 56. doi:10.1186/1750-1172-9-56

2. Jiang YL, Rigolet M, Bourc'his D, Nigon F, Bokesoy I, Fryns JP, et al. DNMT3B mutations and DNA methylation defect define two types of ICF syndrome. Hum Mutat. 2005;25: 56–63. doi:10.1002/humu.20113

3. Thijssen PE, Ito Y, Grillo G, Wang J, Velasco G, Nitta H, et al. Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome. Nat Commun. 2015;6: 7870. doi:10.1038/ncomms8870

4. Touleimat N, Tost J. Complete pipeline for Infinium ® Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. Epigenomics. 2012;4: 325–341. doi:10.2217/epi.12.21

5. Chen Y, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics. 2013;8: 203–209. doi:10.4161/epi.23470

6. Price EM, Cotton AM, Lam LL, Farré P, Emberly E, Brown CJ, et al. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. Epigenetics Chromatin. 2013;6: 1.

7. Jurka J. Repbase update: a database and an electronic journal of repetitive elements. Trends Genet. 2000;16: 418–420.

8. Hansen RS, Thomas S, Sandstrom R, Canfield TK, Thurman RE, Weaver M, et al. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proc Natl Acad Sci. 2010;107: 139–144. doi:10.1073/pnas.0912402107

9. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011;473: 43–49. doi:10.1038/nature09906

10. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013;14: 1.

11. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016;44: W90–W97. doi:10.1093/nar/gkw377

12. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29: 24–26. doi:10.1038/nbt.1754

13. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14: 178–192. doi:10.1093/bib/bbs017

14. Méndez J, Stillman B. Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. Mol Cell Biol. 2000;20: 8602–8612.

15. Sterlin D, Velasco G, Moshous D, Touzot F, Mahlaoui N, Fischer A, et al. Genetic, Cellular and Clinical Features of ICF Syndrome: a French National Survey. J Clin Immunol. 2016;36: 149–159. doi:10.1007/s10875-016-0240-2

16. Brun M-E, Lana E, Rivals I, Lefranc G, Sarda P, Claustres M, et al. Heterochromatic Genes Undergo Epigenetic Changes and Escape Silencing in Immunodeficiency, Centromeric Instability, Facial Anomalies (ICF) Syndrome. Chadwick BP, editor. PLoS ONE. 2011;6: e19464. doi:10.1371/journal.pone.0019464

17. Turleau C, Cabanis MO, Girault D, Ledeist F, Mettey R, Puissant H, et al. Multibranched chromosomes in the ICF syndrome: immunodeficiency, centromeric instability, and facial anomalies. Am J Med Genet. 1989;32: 420–424. doi:10.1002/ajmg.1320320331

18. Weemaes CMR, van Tol MJD, Wang J, van Ostaijen-ten Dam MM, van Eggermond MCJA, Thijssen PE, et al. Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects. Eur J Hum Genet EJHG. 2013;21: 1219–1225. doi:10.1038/ejhg.2013.40

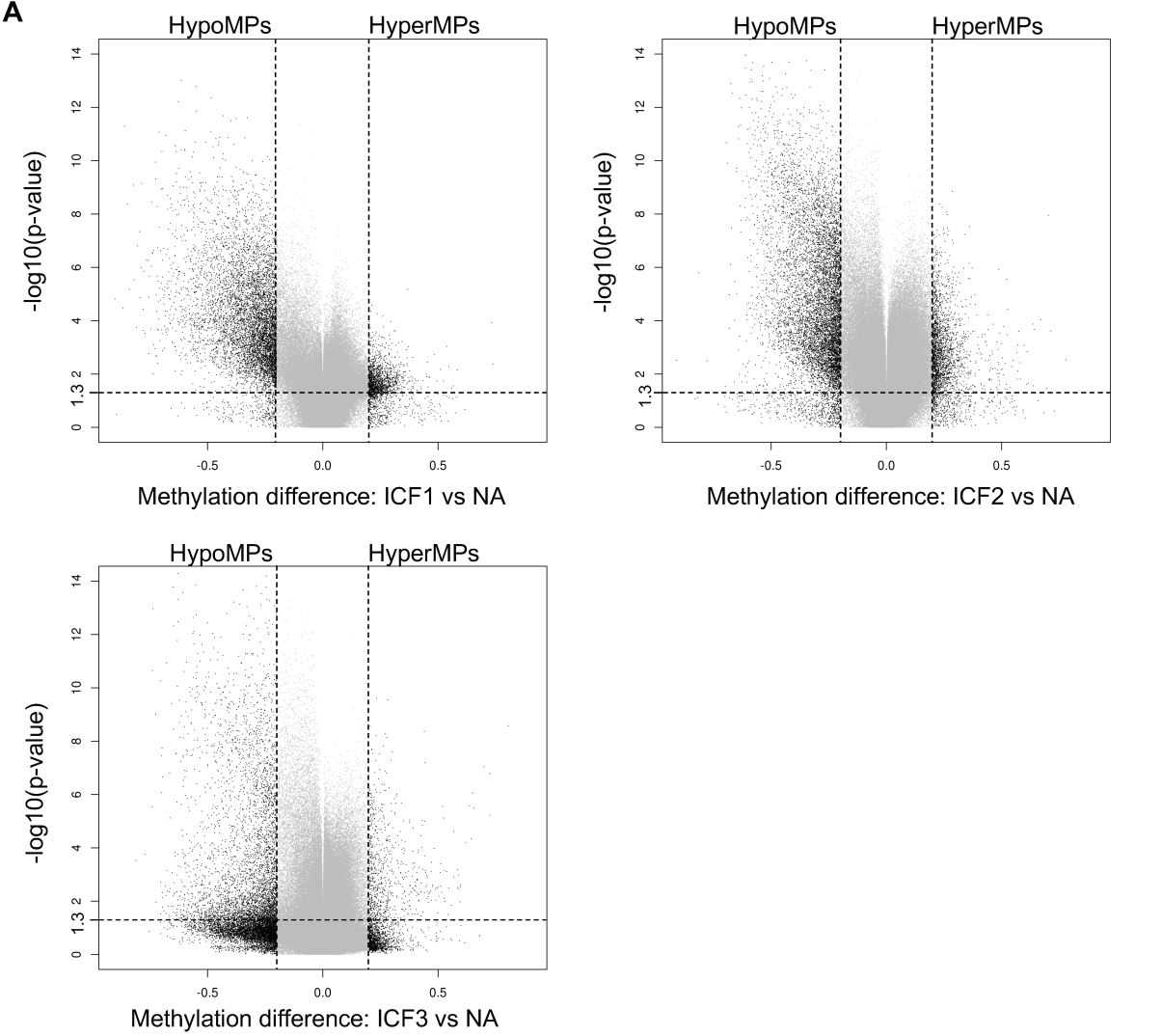
19. Chouery E, Abou-Ghoch J, Corbani S, El Ali N, Korban R, Salem N, et al. A novel deletion in *ZBTB24* in a Lebanese family with immunodeficiency, centromeric instability, and facial anomalies syndrome type 2. Clin Genet. 2012;82: 489–493. doi:10.1111/j.1399-0004.2011.01783.x

20. Nitta H, Unoki M, Ichiyanagi K, Kosho T, Shigemura T, Takahashi H, et al. Three novel ZBTB24 mutations identified in Japanese and Cape Verdean type 2 ICF syndrome patients. J Hum Genet. 2013;58: 455–460.

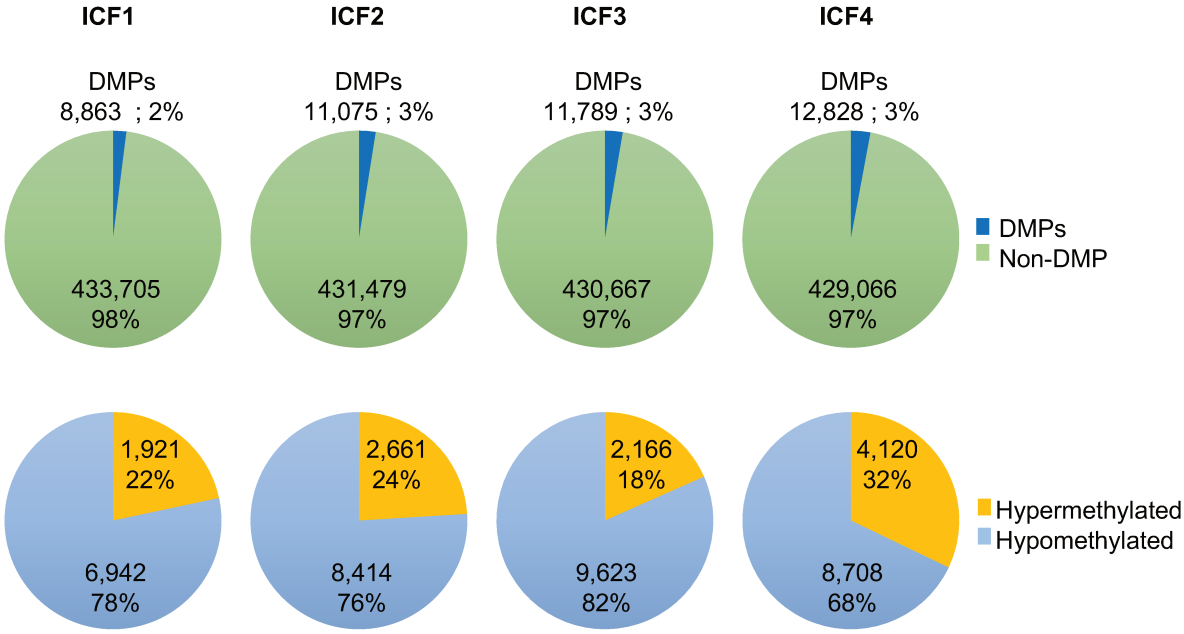
21. Maraschio P, Zuffardi O, Dalla Fior T, Tiepolo L. Immunodeficiency, centromeric heterochromatin instability of chromosomes 1, 9, and 16, and facial anomalies: the ICF syndrome. J Med Genet. 1988;25: 173–180.

S1 Fig

A

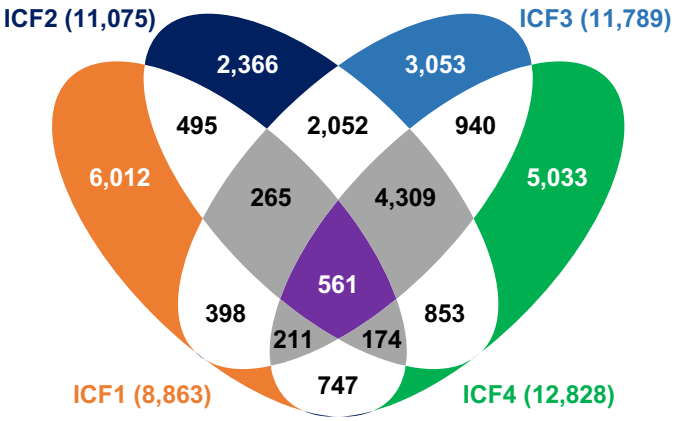

HypoMPs
HyperMPs
-log10(p-value)
Methylation difference: ICF1 vs NA
Methylation difference: ICF2 vs NA
Methylation difference: ICF3 vs NA

B


ICF1
ICF2
ICF3
ICF4
DMPs
8,863 ; 2%
11,075 ; 3%
11,789 ; 3%
12,828 ; 3%
433,705
98%
431,479
97%
430,667
97%
429,066
97%
DMPs
Non-DMP
1,921
22%
2,661
24%
2,166
18%
4,120
32%
6,942
78%
8,414
76%
9,623
82%
8,708
68%
Hypermethylated
Hypomethylated

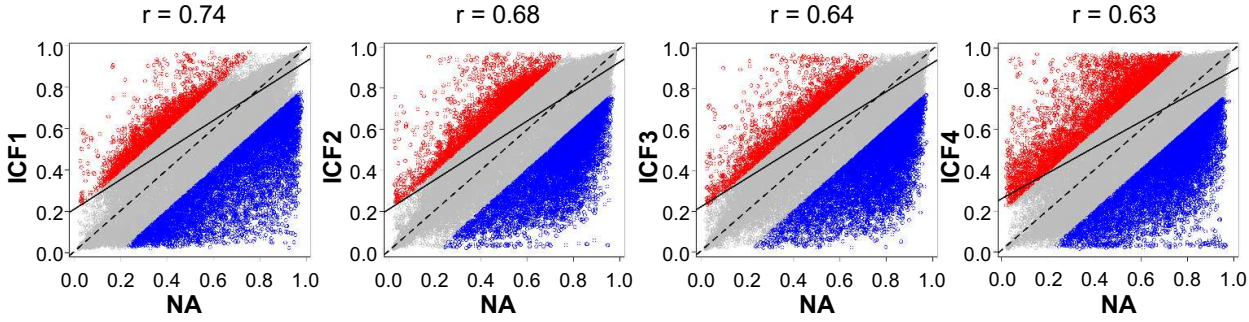
S2 Fig

A

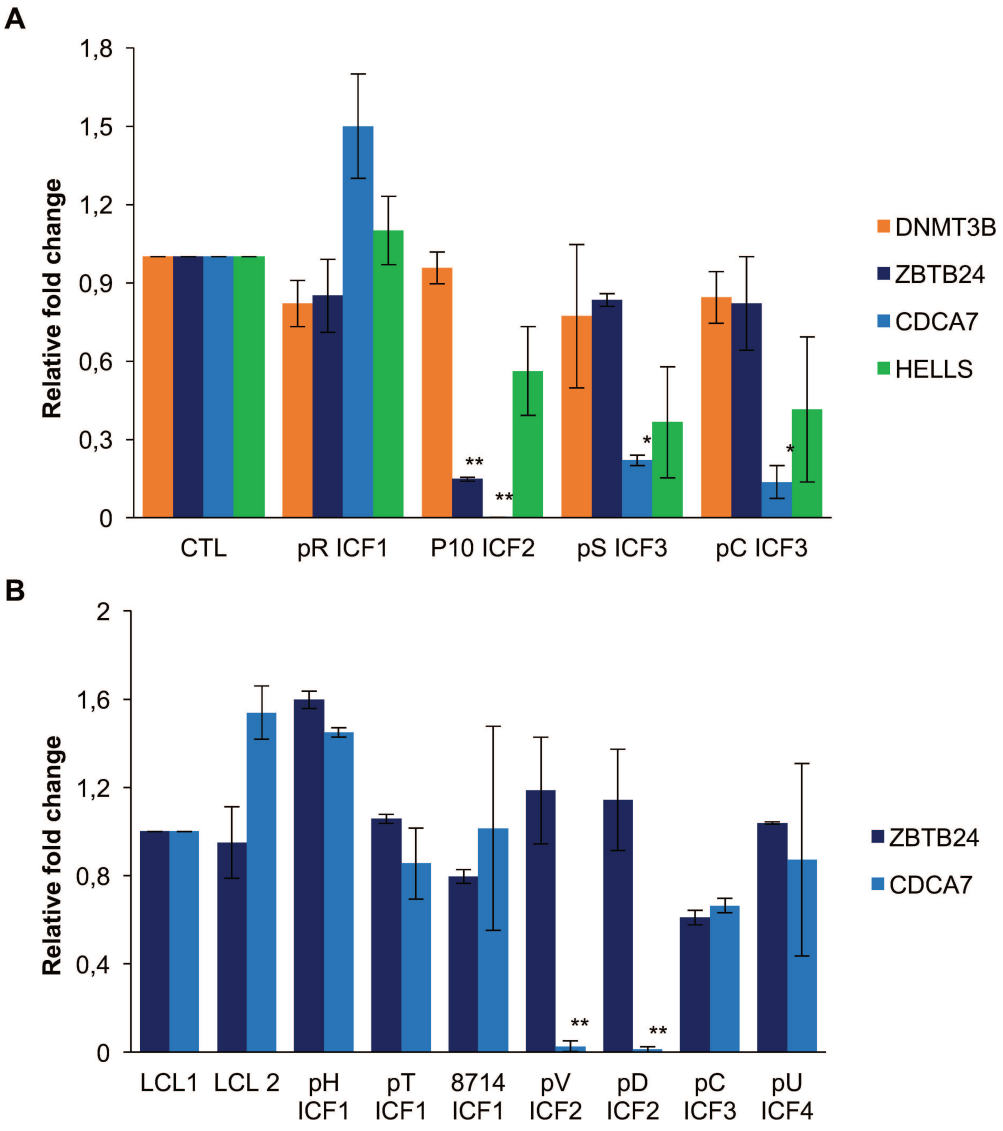


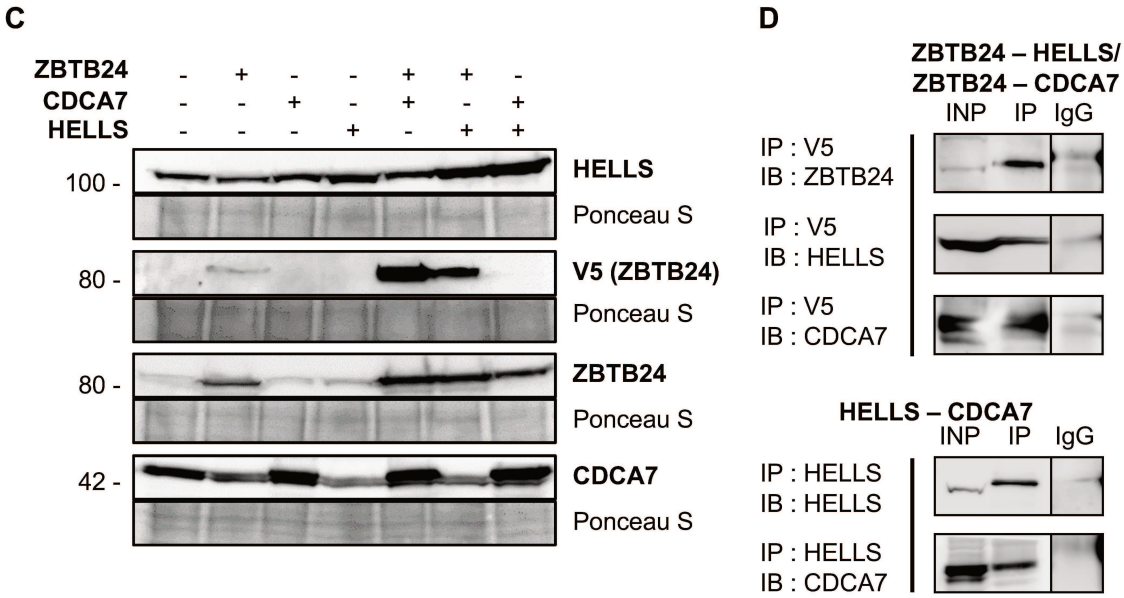
Autosomes	ICF1	ICF2	ICF3	ICF4
ICF1	8,863	1,495 (16.9%)	1,435 (16.2%)	1,693 (19.1%)
ICF2		11,075	**7,187 (64.9%)**	**5,897 (53.2%)**
ICF3			11,789	**6,021 (51.1%)**
ICF4				12,828
Total DMPs 44,555				
Total HypoMPs 33,687			**Total HyperMPs 10,868**	

B

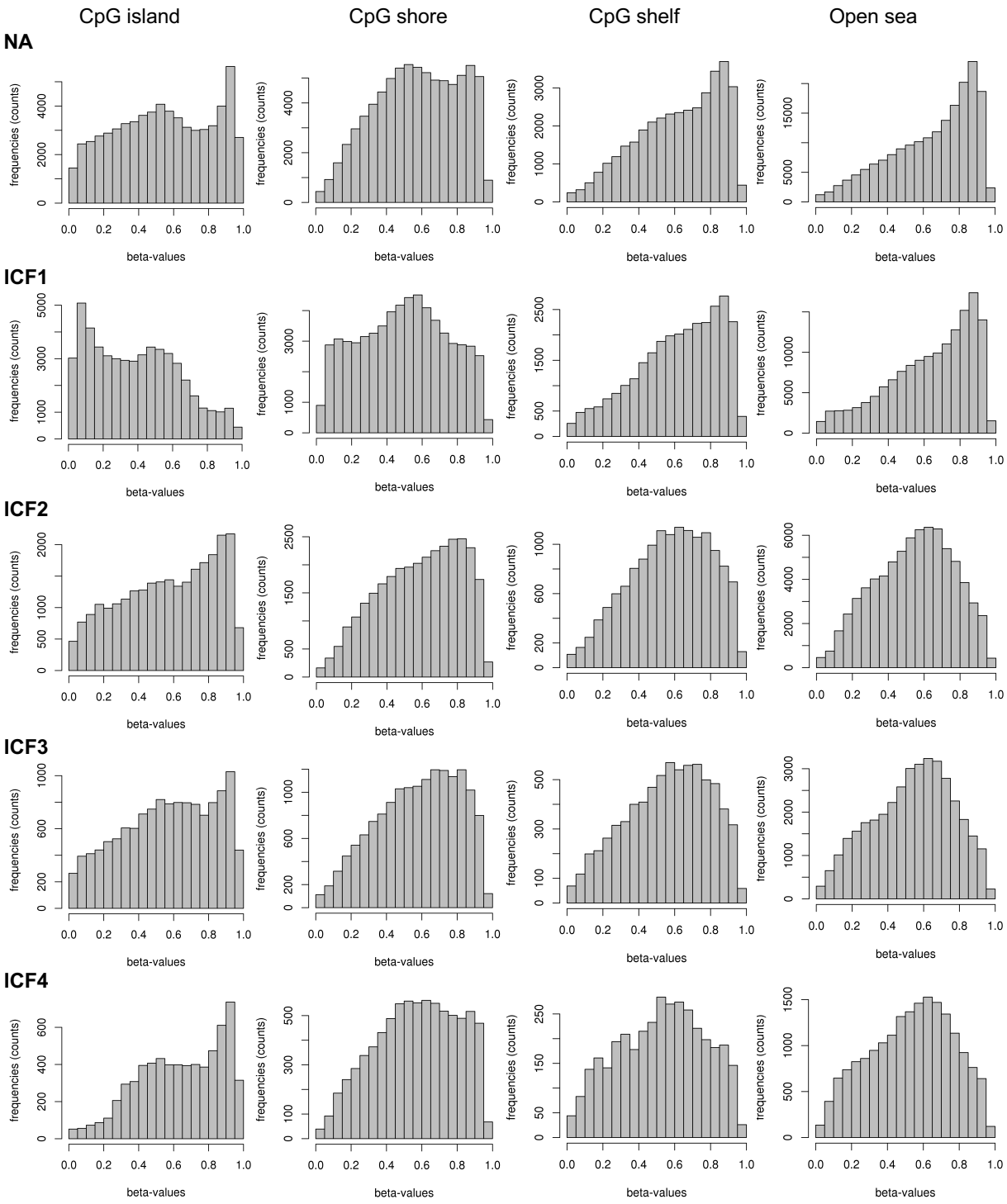


S3 Fig


A
Relative fold change
1,8
1,5
1,2
0,9
0,6
0,3
0
CTL
pR ICF1
P10 ICF2
pS ICF3
pC ICF3
DNMT3B
ZBTB24
CDCA7
HELLS
B
Relative fold change
2
1,6
1,2
0,8
0,4
0
LCL1
LCL 2
pH ICF1
pT ICF1
8714 ICF1
pV ICF2
pD ICF2
pC ICF3
pU ICF4
ZBTB24
CDCA7

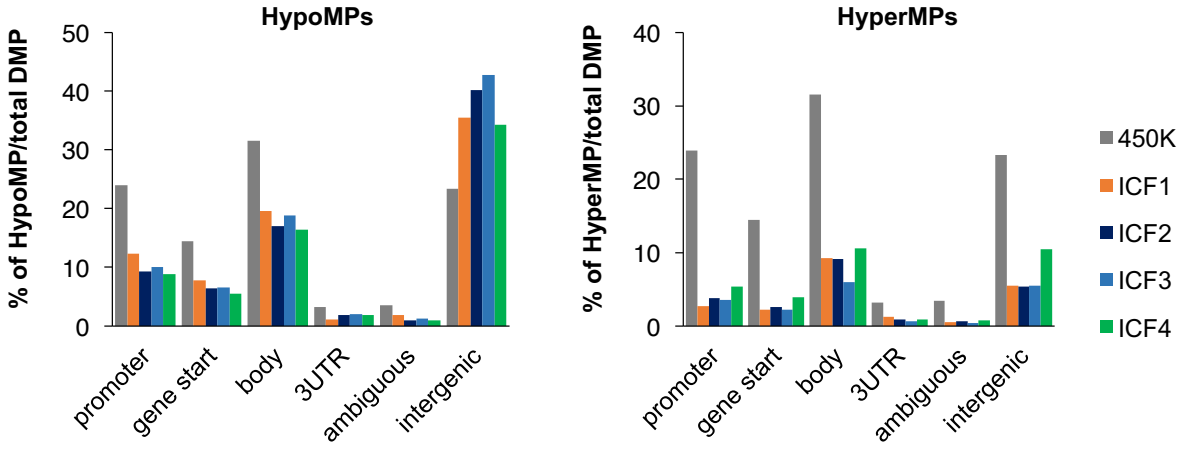

C
ZBTB24
CDCA7
HELLS
100 -
80 -
80 -
42 -
HELLS
Ponceau S
V5 (ZBTB24)
Ponceau S
ZBTB24
Ponceau S
CDCA7
Ponceau S
D
ZBTB24 – HELLS/
ZBTB24 – CDCA7
INP IP IgG
IP : V5
IB : ZBTB24
IP : V5
IB : HELLS
IP : V5
IB : CDCA7
HELLS – CDCA7
INP IP IgG
IP : HELLS
IB : HELLS
IP : HELLS
IB : CDCA7

S4 Fig

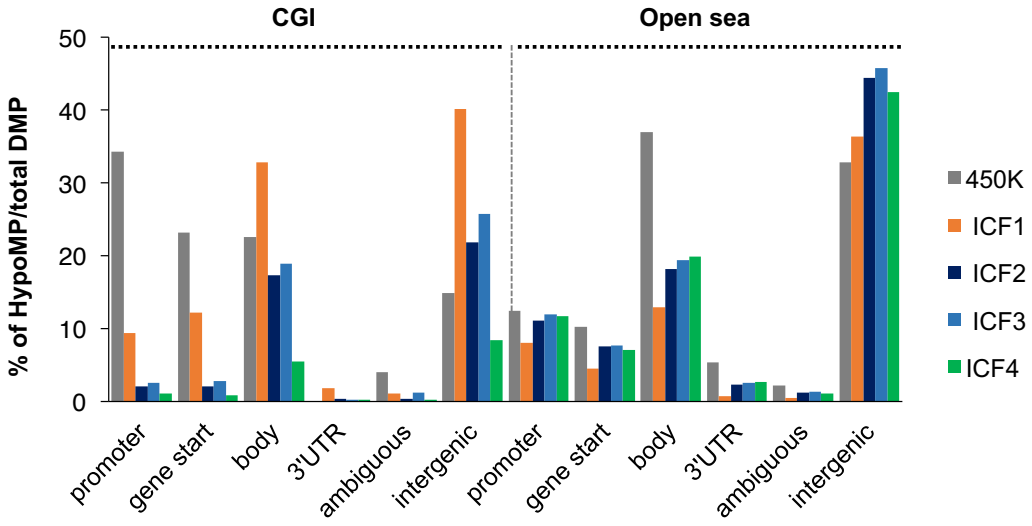

CpG island
CpG shore
CpG shelf
Open sea
NA
ICF1
ICF2
ICF3
ICF4
frequencies (counts)
beta-values

S5 Fig

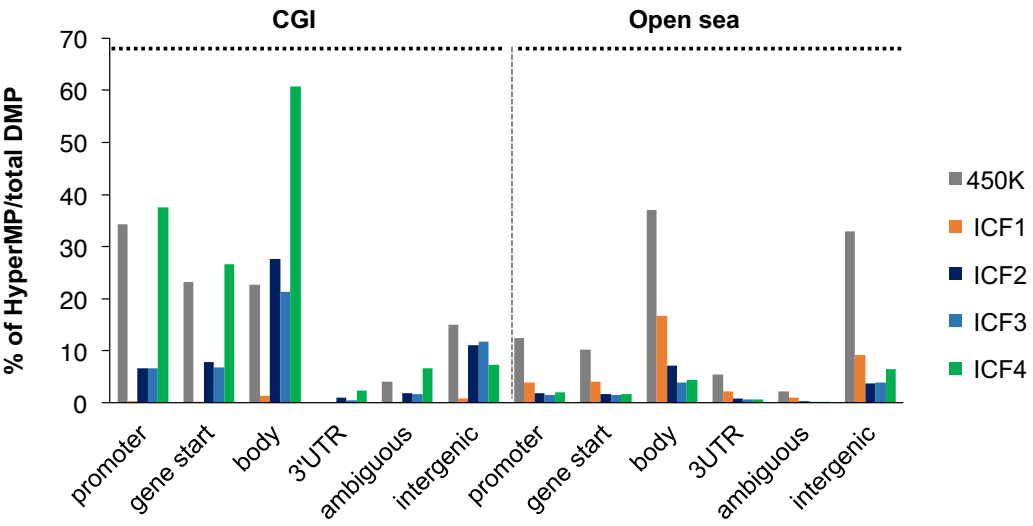
A


HypoMPs
% of HypoMP/total DMP
50
40
30
20
10
0
promoter
gene start
body
3UTR
ambiguous
intergenic
HyperMPs
% of HyperMP/total DMP
40
30
20
10
0
promoter
gene start
body
3UTR
ambiguous
intergenic
450K
ICF1
ICF2
ICF3
ICF4

B

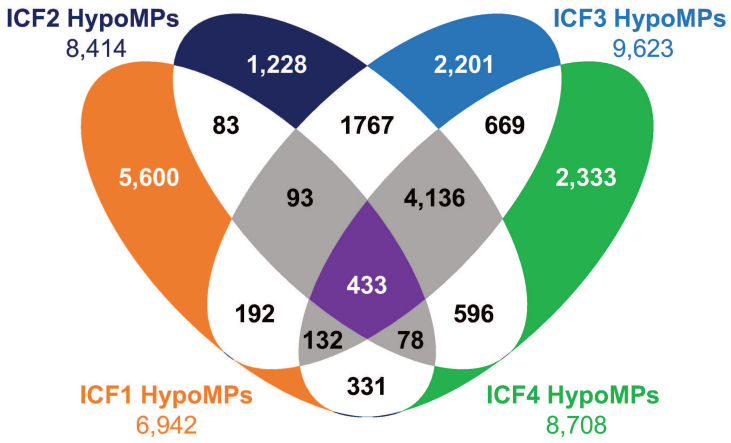

CGI
Open sea
% of HypoMP/total DMP
50
40
30
20
10
0
promoter
gene start
body
3'UTR
ambiguous
intergenic
promoter
gene start
body
3'UTR
ambiguous
intergenic
450K
ICF1
ICF2
ICF3
ICF4

C

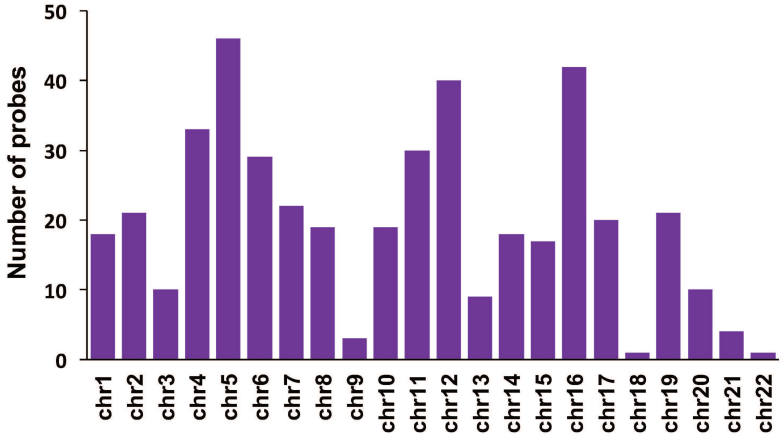

CGI
Open sea
% of HyperMP/total DMP
70
60
50
40
30
20
10
0
promoter
gene start
body
3'UTR
ambiguous
intergenic
promoter
gene start
body
3'UTR
ambiguous
intergenic
450K
ICF1
ICF2
ICF3
ICF4

S6 Fig

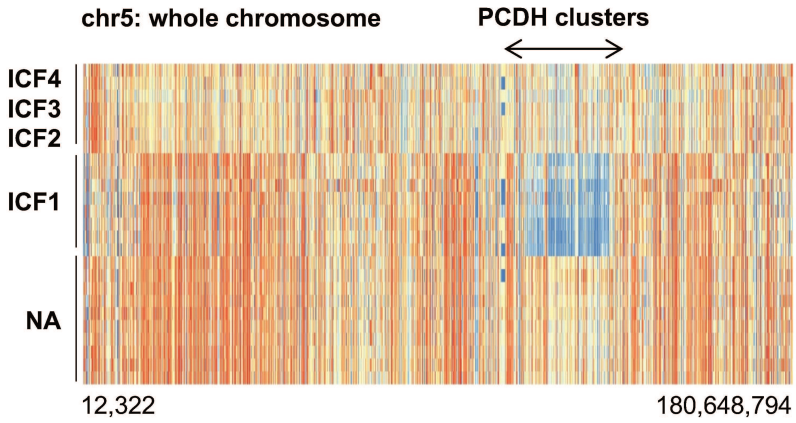
A


ICF2 HypoMPs
8,414
ICF3 HypoMPs
9,623
ICF1 HypoMPs
6,942
ICF4 HypoMPs
8,708
1,228
2,201
5,600
2,333
83
1767
669
93
4,136
433
192
596
132
78
331

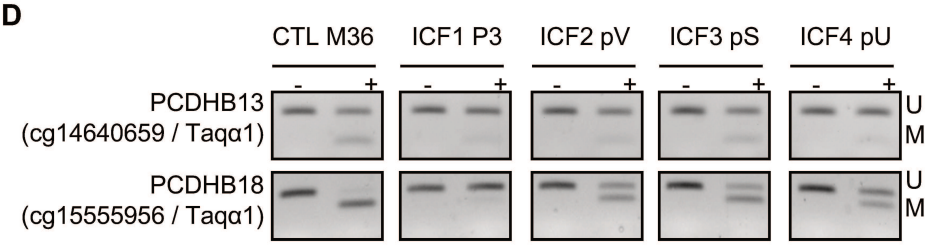
B


Number of probes
0
10
20
30
40
50
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22

C

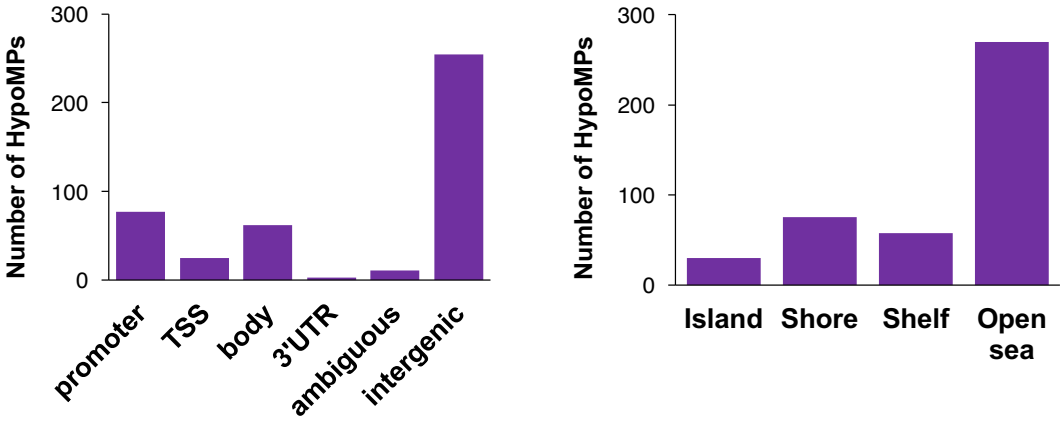

chr5: whole chromosome
PCDH clusters
ICF4
ICF3
ICF2
ICF1
NA
12,322
180,648,794

D

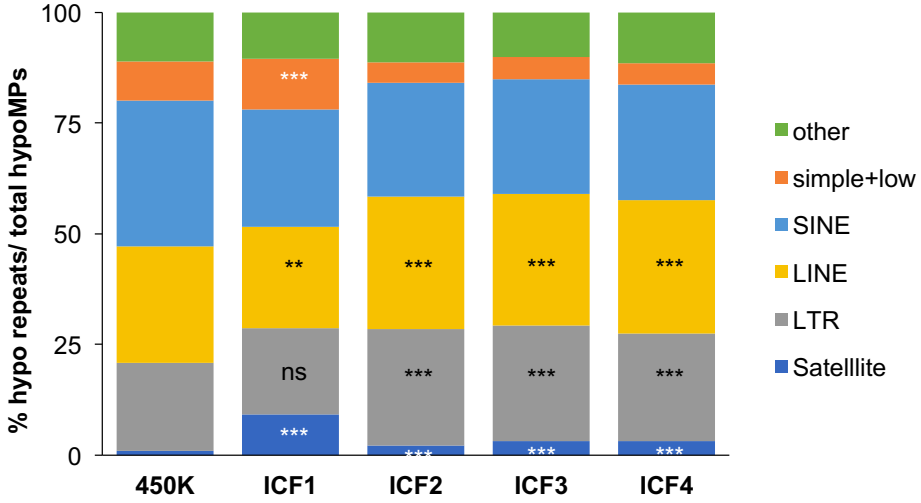

CTL M36
ICF1 P3
ICF2 pV
ICF3 pS
ICF4 pU
-
+
PCDHB13
(cg14640659 / Taqα1)
PCDHB18
(cg15555956 / Taqα1)
U
M

S7 Fig

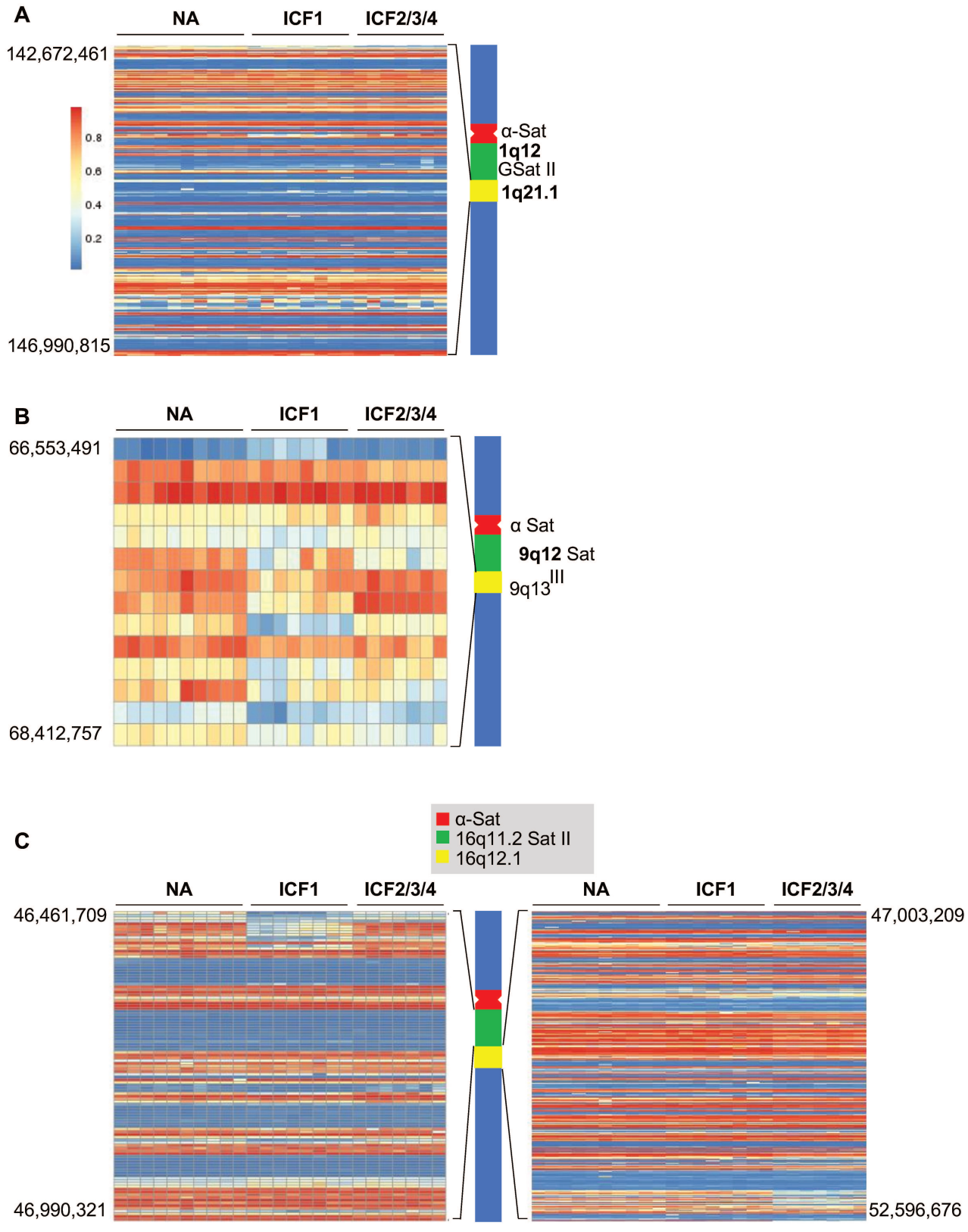
A


Number of HypoMPs
300
200
100
0
promoter
TSS
body
3'UTR
ambiguous
intergenic
Number of HypoMPs
300
200
100
0
Island
Shore
Shelf
Open sea

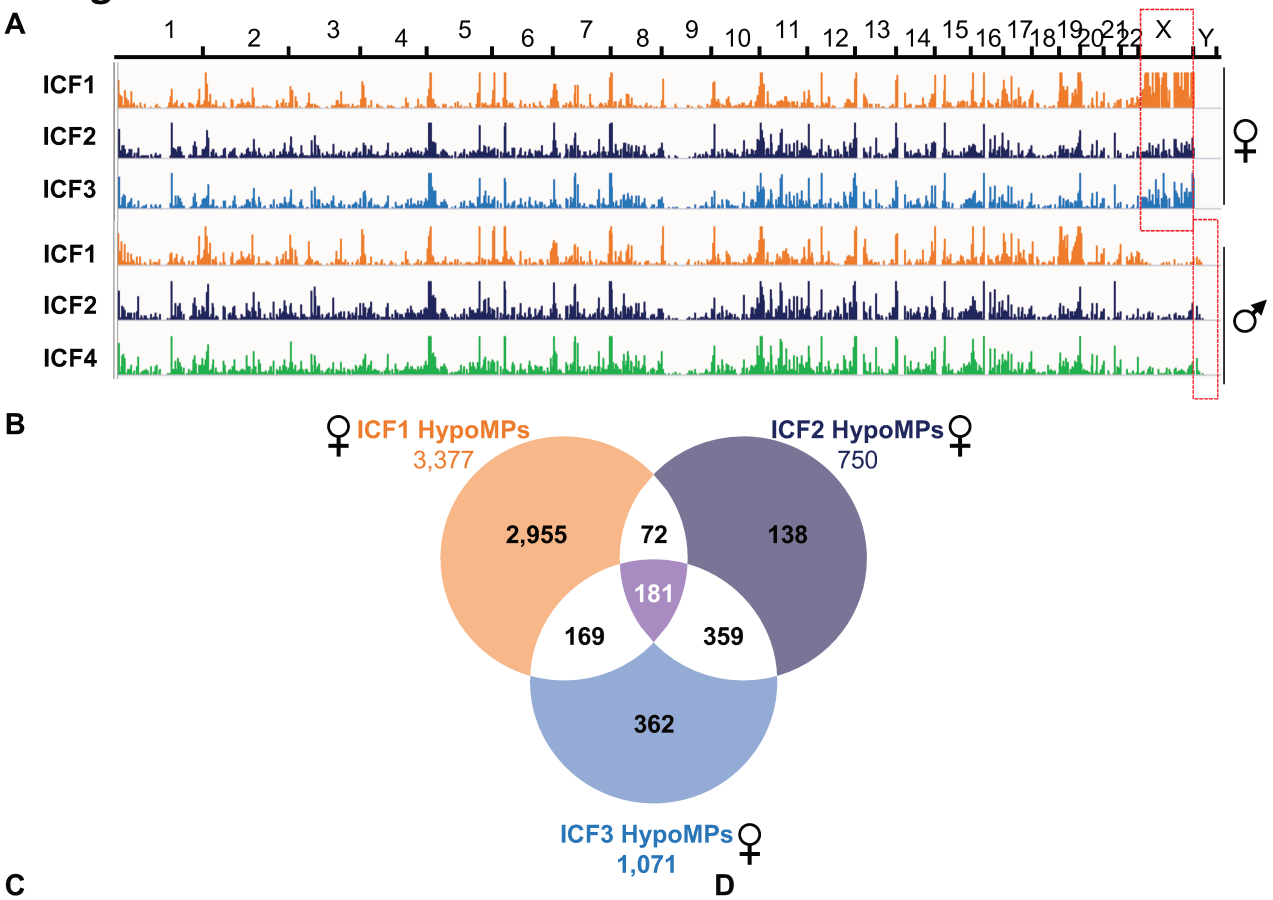
B


% hypo repeats/ total hypoMPs
100
75
50
25
0
450K
ICF1
ICF2
ICF3
ICF4
***
**
ns
***
other
simple+low
SINE
LINE
LTR
Satelllite

S8 Fig


A
NA
ICF1
ICF2/3/4
142,672,461
146,990,815
0.8
0.6
0.4
0.2
α-Sat
1q12
GSat II
1q21.1
B
NA
ICF1
ICF2/3/4
66,553,491
68,412,757
α Sat
9q12 Sat
9q13III
C
α-Sat
16q11.2 Sat II
16q12.1
NA
ICF1
ICF2/3/4
NA
ICF1
ICF2/3/4
46,461,709
46,990,321
47,003,209
52,596,676

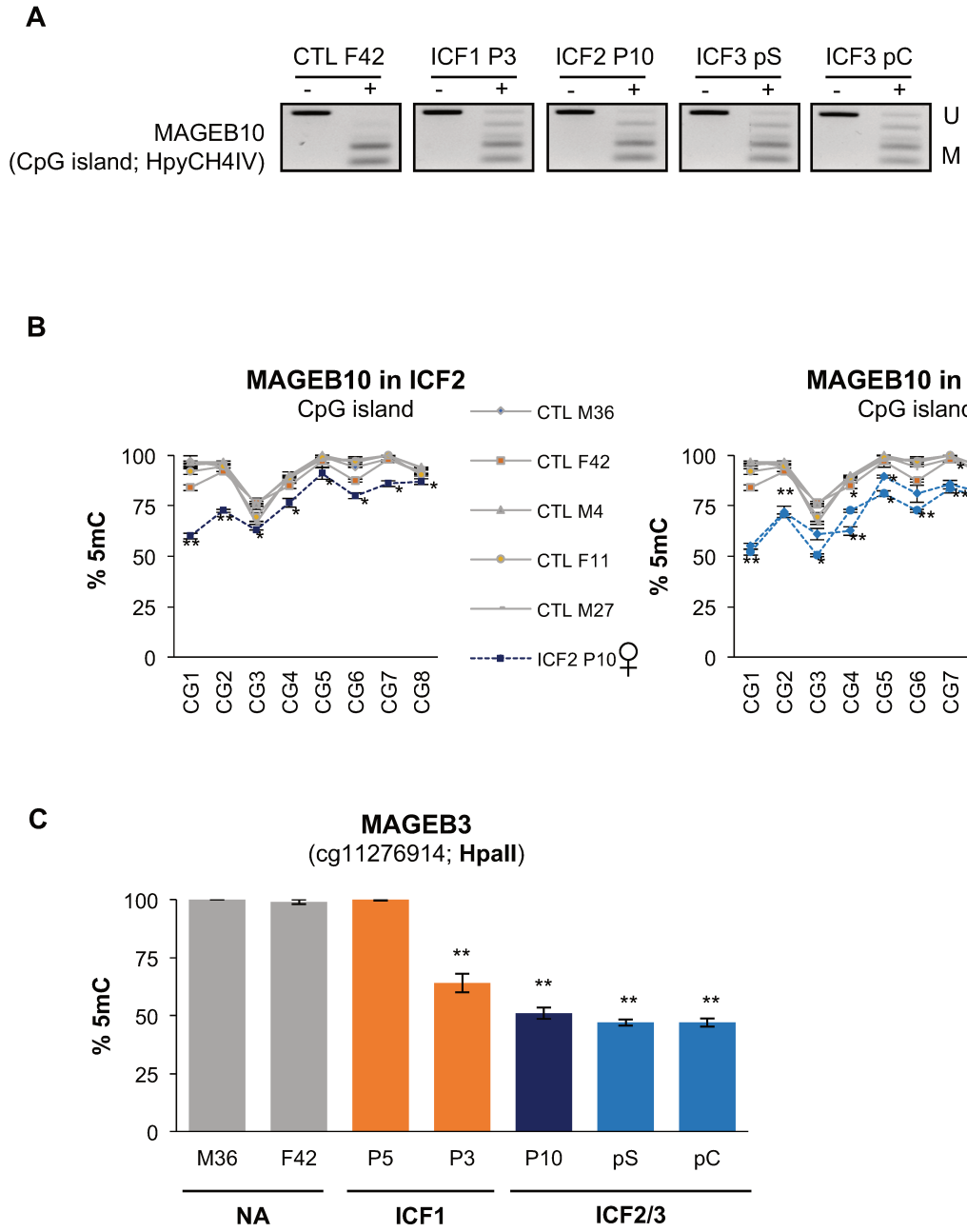
S9 Fig



Y chromosome (416 total probes)

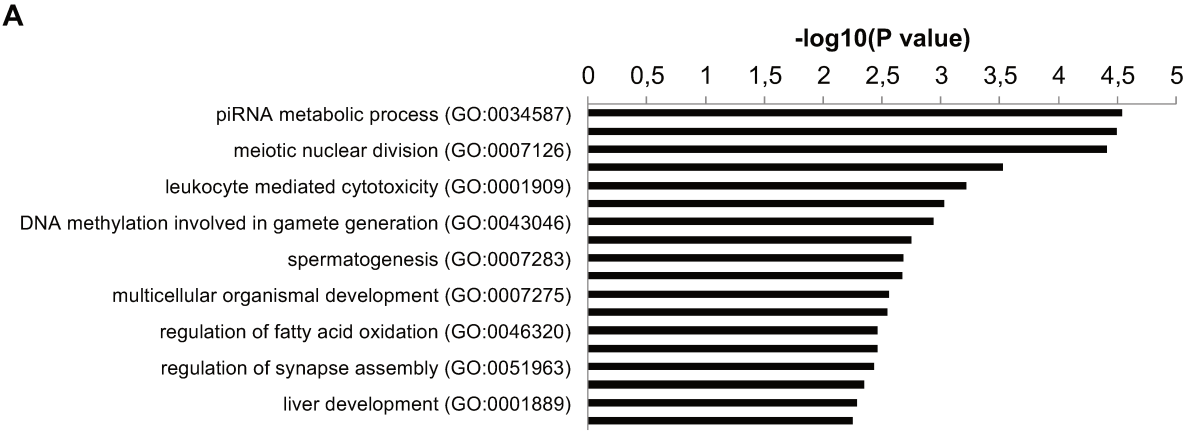
ICF	HypoMPs	CpG island	CpG Shore	CpG Shelf	CpG open sea
ICF1	28	13 (46%)	7 (25%)	6 (21%)	2 (7%)
ICF2	37	2 (5%)	10 (27%)	12 (32%)	13 (35%)
ICF4	35	4 (11%)	9 (26%)	6 (17%)	16 (46%)

S10 Fig

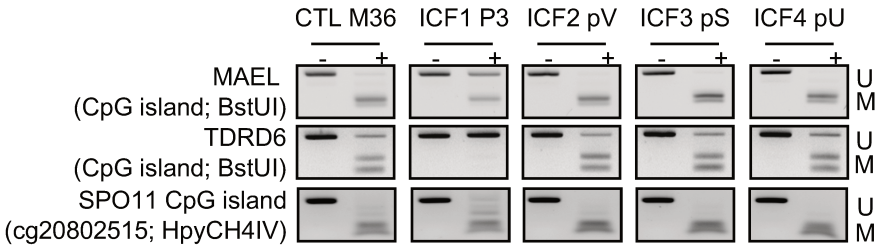

A
CTL F42
ICF1 P3
ICF2 P10
ICF3 pS
ICF3 pC
- + - + - + - + - +
MAGEB10
(CpG island; HpyCH4IV)
U
M
B
MAGEB10 in ICF2
CpG island
MAGEB10 in ICF3
CpG island
% 5mC
100
75
50
25
0
CG1 CG2 CG3 CG4 CG5 CG6 CG7 CG8
CTL M36
CTL F42
CTL M4
CTL F11
CTL M27
ICF2 P10♀
ICF3 pS♀
ICF3 pC♀
C
MAGEB3
(cg11276914; HpaII)
% 5mC
M36 F42 P5 P3 P10 pS pC
NA
ICF1
ICF2/3

S11 Fig

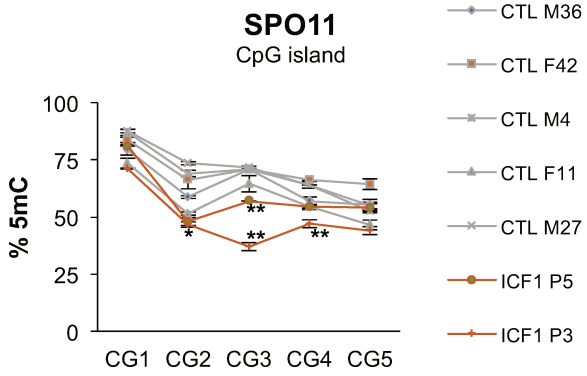
A


-log10(P value)
0 0,5 1 1,5 2 2,5 3 3,5 4 4,5 5
piRNA metabolic process (GO:0034587)
meiotic nuclear division (GO:0007126)
leukocyte mediated cytotoxicity (GO:0001909)
DNA methylation involved in gamete generation (GO:0043046)
spermatogenesis (GO:0007283)
multicellular organismal development (GO:0007275)
regulation of fatty acid oxidation (GO:0046320)
regulation of synapse assembly (GO:0051963)
liver development (GO:0001889)

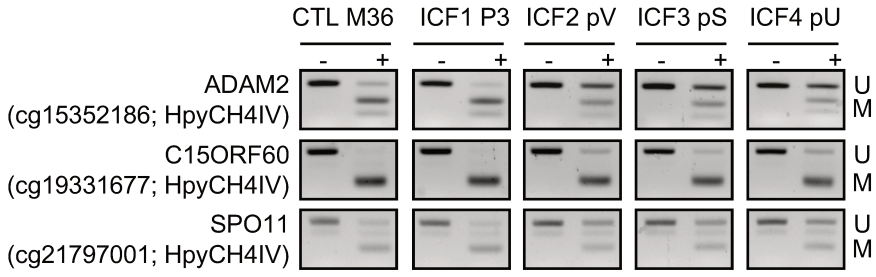
B


CTL M36
ICF1 P3
ICF2 pV
ICF3 pS
ICF4 pU
MAEL
(CpG island; BstUI)
TDRD6
(CpG island; BstUI)
SPO11 CpG island
(cg20802515; HpyCH4IV)
U
M

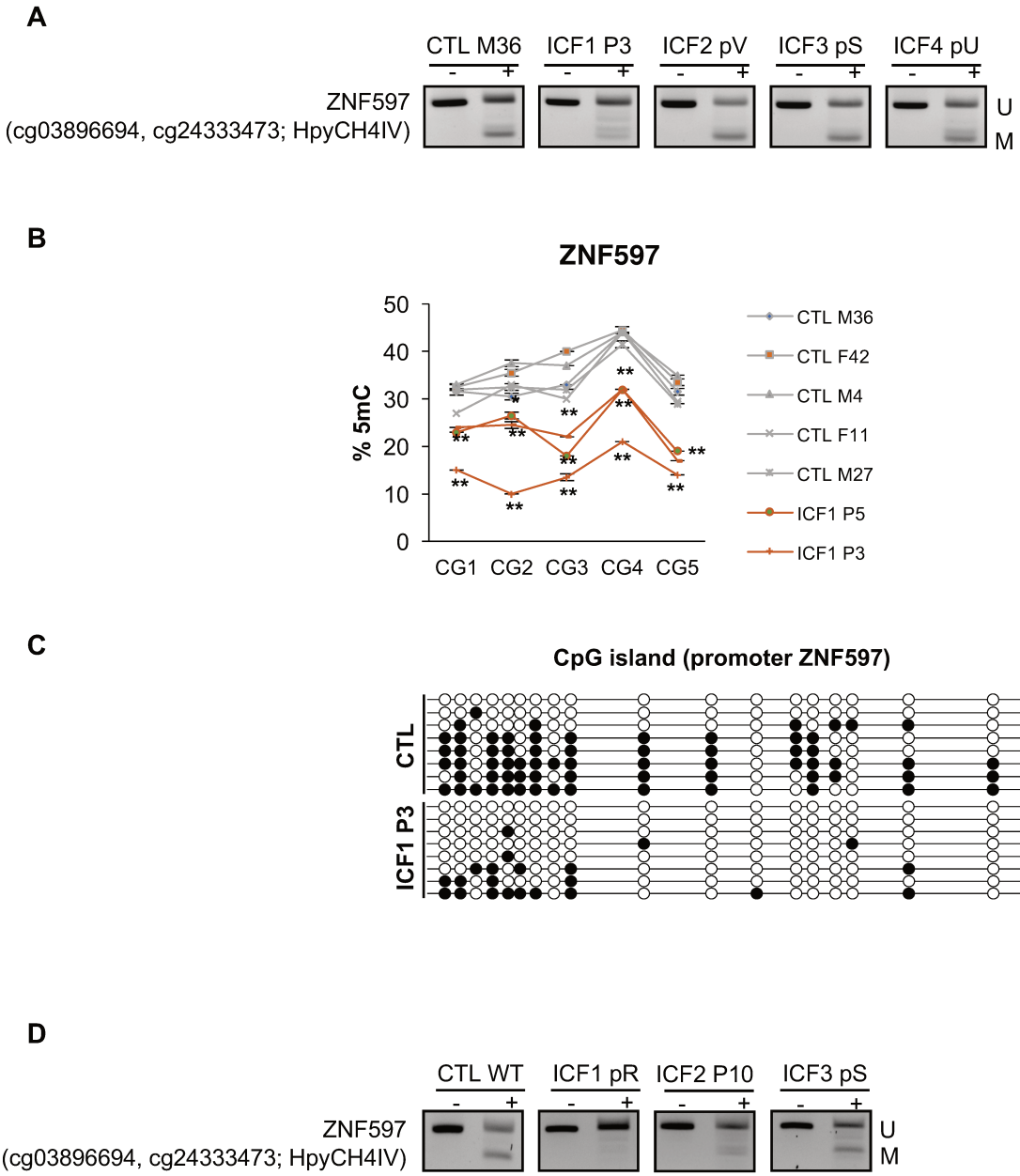
C


SPO11
CpG island
% 5mC
100
75
50
25
0
CG1 CG2 CG3 CG4 CG5
CTL M36
CTL F42
CTL M4
CTL F11
CTL M27
ICF1 P5
ICF1 P3

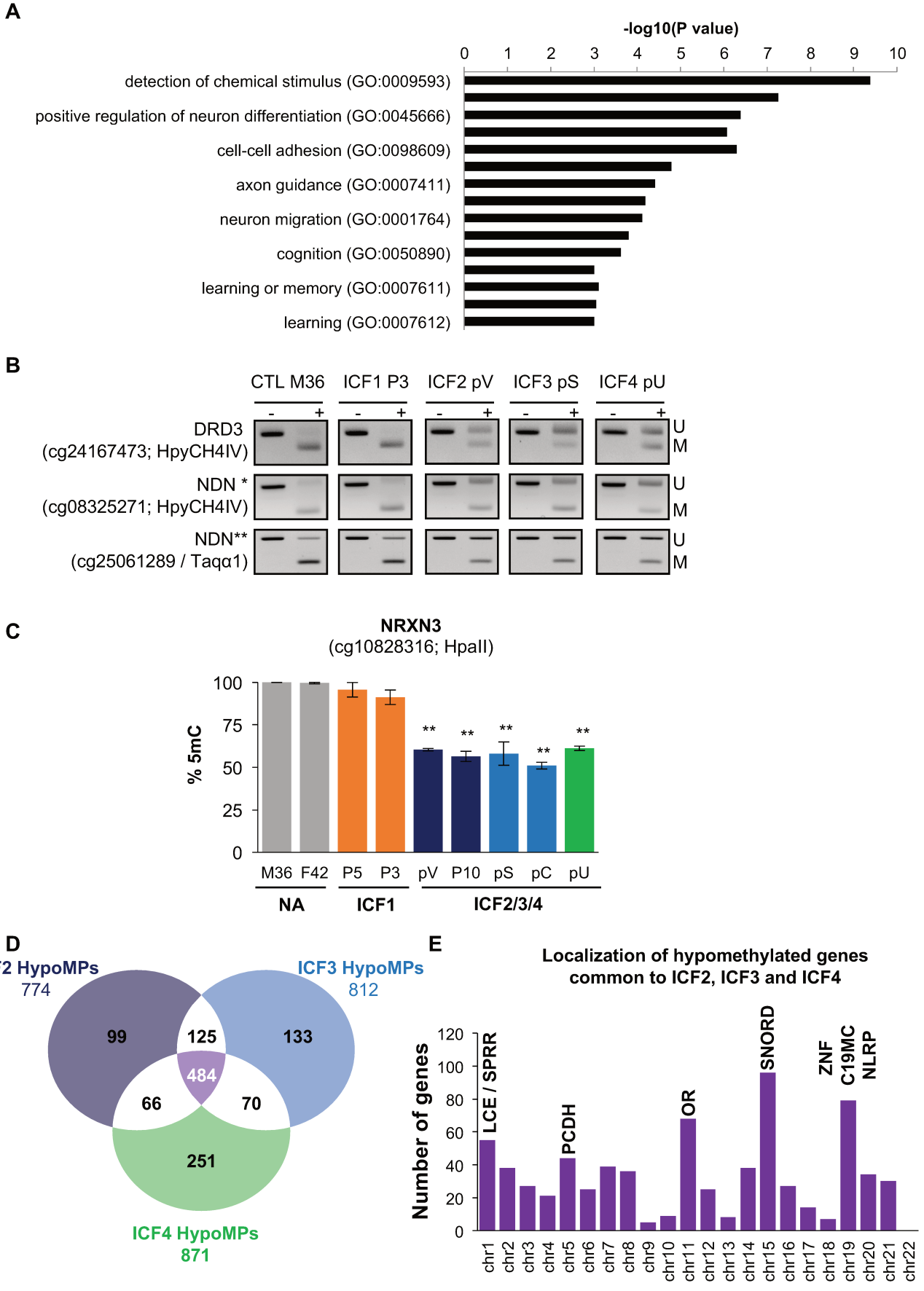
D


CTL M36
ICF1 P3
ICF2 pV
ICF3 pS
ICF4 pU
ADAM2
(cg15352186; HpyCH4IV)
C15ORF60
(cg19331677; HpyCH4IV)
SPO11
(cg21797001; HpyCH4IV)
U
M

S12 Fig

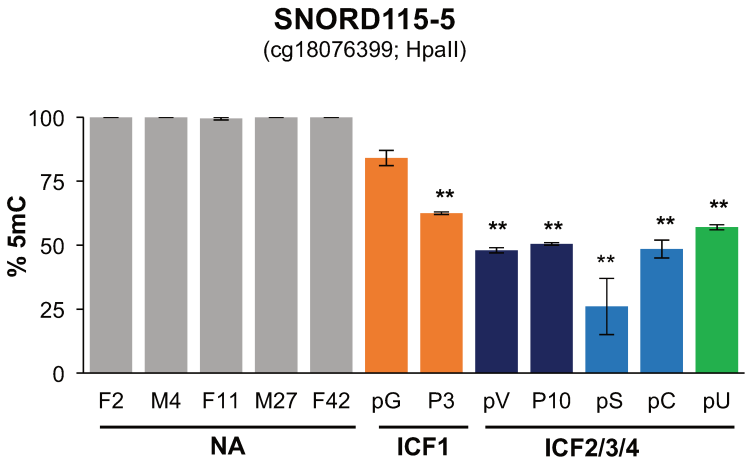

A
CTL M36
ICF1 P3
ICF2 pV
ICF3 pS
ICF4 pU
- + - + - + - + - +
ZNF597
(cg03896694, cg24333473; HpyCH4IV)
U
M
B
ZNF597
% 5mC
50
40
30
20
10
0
CG1 CG2 CG3 CG4 CG5
CTL M36
CTL F42
CTL M4
CTL F11
CTL M27
ICF1 P5
ICF1 P3
C
CpG island (promoter ZNF597)
CTL
ICF1 P3
D
CTL WT
ICF1 pR
ICF2 P10
ICF3 pS
- + - + - + - +
ZNF597
(cg03896694, cg24333473; HpyCH4IV)
U
M

S13 Fig

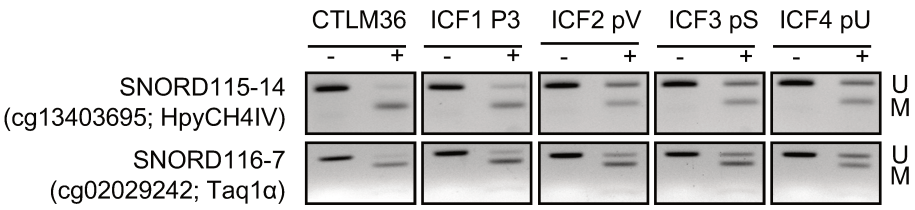

A
-log10(P value)
0 1 2 3 4 5 6 7 8 9 10
detection of chemical stimulus (GO:0009593)
positive regulation of neuron differentiation (GO:0045666)
cell-cell adhesion (GO:0098609)
axon guidance (GO:0007411)
neuron migration (GO:0001764)
cognition (GO:0050890)
learning or memory (GO:0007611)
learning (GO:0007612)
B
CTL M36
ICF1 P3
ICF2 pV
ICF3 pS
ICF4 pU
DRD3
(cg24167473; HpyCH4IV)
NDN *
(cg08325271; HpyCH4IV)
NDN**
(cg25061289 / Taqα1)
U
M
C
NRXN3
(cg10828316; HpaII)
% 5mC
100
75
50
25
0
M36 F42 P5 P3 pV P10 pS pC pU
NA
ICF1
ICF2/3/4
D
ICF2 HypoMPs
774
ICF3 HypoMPs
812
ICF4 HypoMPs
871
99
125
133
484
66
70
251
E
Localization of hypomethylated genes
common to ICF2, ICF3 and ICF4
Number of genes
120
100
80
60
40
20
0
LCE / SPRR
PCDH
OR
SNORD
ZNF
C19MC
NLRP
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22

S14 Fig

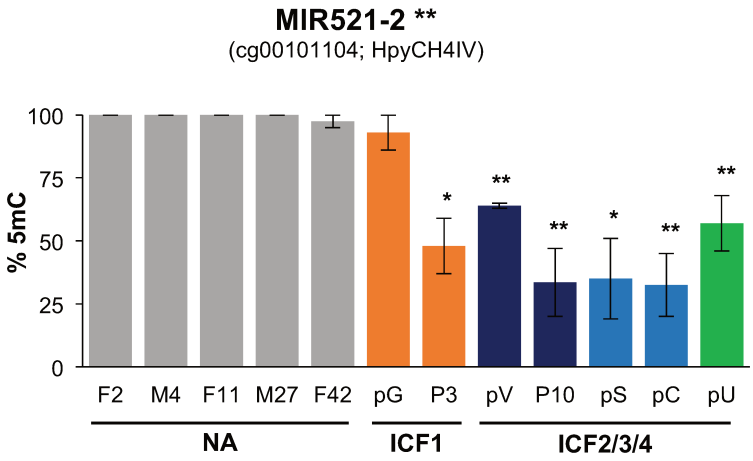
A

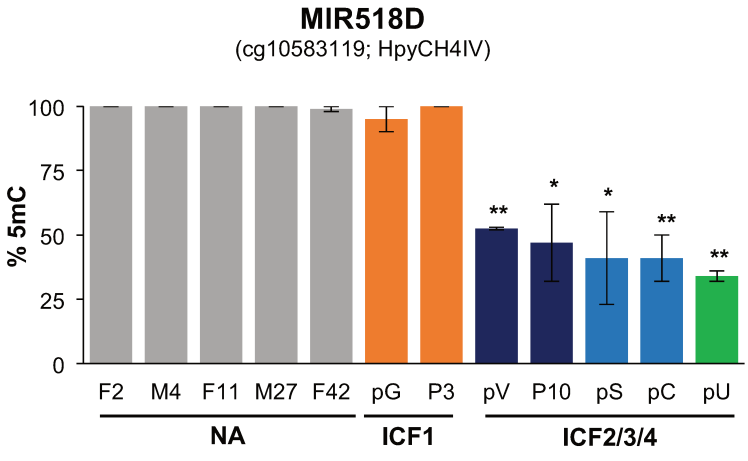

SNORD115-5
(cg18076399; HpaII)
% 5mC
100
75
50
25
0
F2 M4 F11 M27 F42 pG P3 pV P10 pS pC pU
NA
ICF1
ICF2/3/4

B

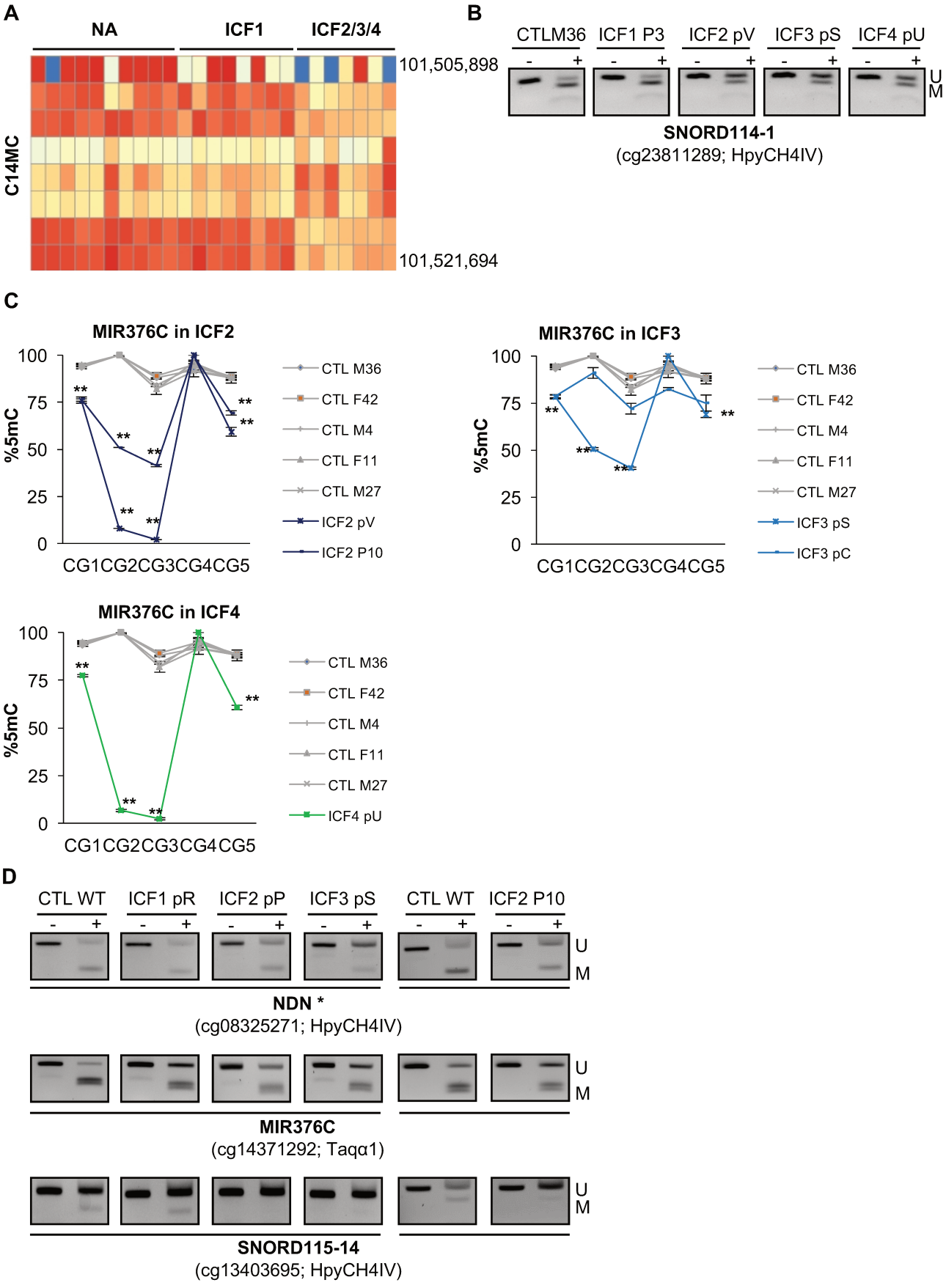

CTLM36
ICF1 P3
ICF2 pV
ICF3 pS
ICF4 pU
- + - + - + - + - +
SNORD115-14
(cg13403695; HpyCH4IV)
SNORD116-7
(cg02029242; Taq1α)
U
M
U
M

C


MIR521-2 **
(cg00101104; HpyCH4IV)
% 5mC
100
75
50
25
0
F2 M4 F11 M27 F42 pG P3 pV P10 pS pC pU
NA
ICF1
ICF2/3/4

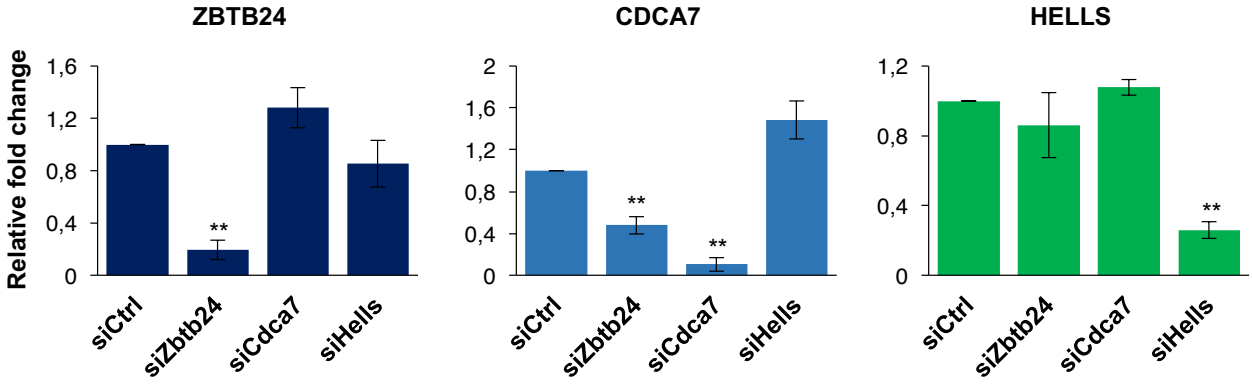

MIR518D
(cg10583119; HpyCH4IV)
% 5mC
100
75
50
25
0
F2 M4 F11 M27 F42 pG P3 pV P10 pS pC pU
NA
ICF1
ICF2/3/4

S15 Fig

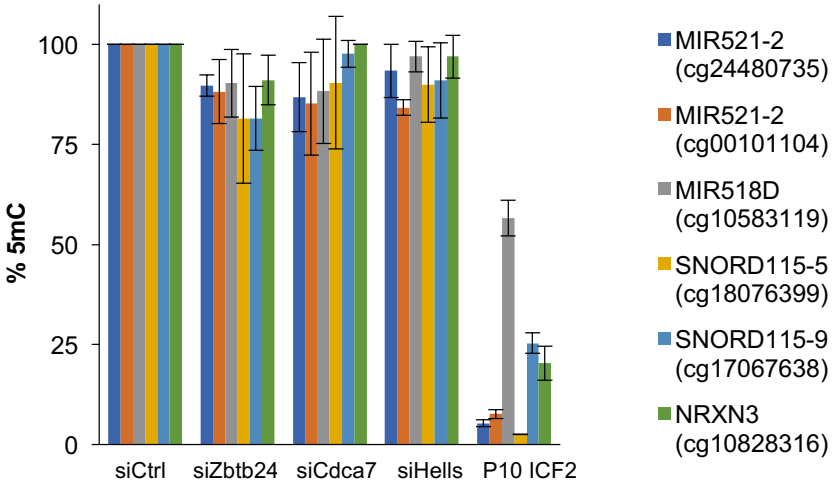

A
NA
ICF1
ICF2/3/4
101,505,898
C14MC
101,521,694
B
CTLM36
ICF1 P3
ICF2 pV
ICF3 pS
ICF4 pU
U
M
SNORD114-1
(cg23811289; HpyCH4IV)
C
MIR376C in ICF2
%5mC
CG1CG2CG3CG4CG5
CTL M36
CTL F42
CTL M4
CTL F11
CTL M27
ICF2 pV
ICF2 P10
MIR376C in ICF3
ICF3 pS
ICF3 pC
MIR376C in ICF4
ICF4 pU
D
CTL WT
ICF1 pR
ICF2 pP
ICF3 pS
CTL WT
ICF2 P10
NDN *
(cg08325271; HpyCH4IV)
MIR376C
(cg14371292; Taqα1)
SNORD115-14
(cg13403695; HpyCH4IV)

S16 Fig

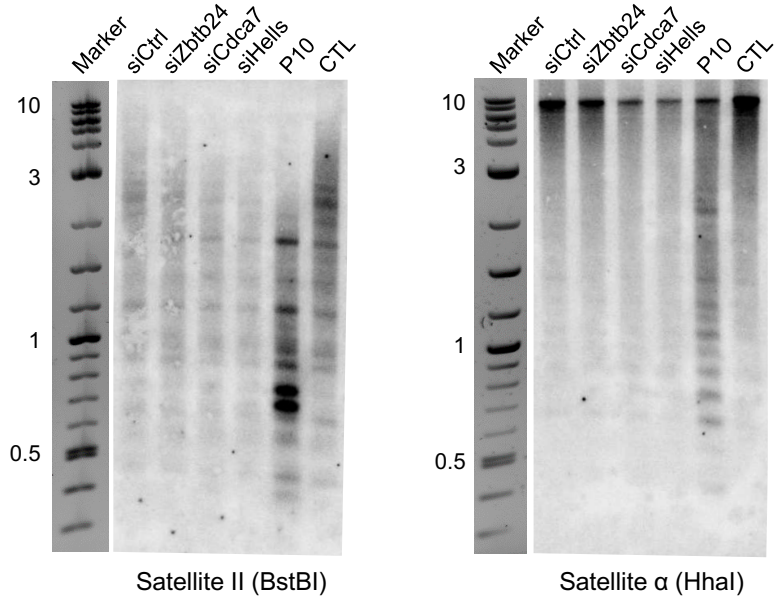
A


ZBTB24
CDCA7
HELLS
Relative fold change
1,6
1,2
0,8
0,4
0
2
1,6
1,2
0,8
0,4
0
1,2
0,8
0,4
0
**
**
**
siCtrl
siZbtb24
siCdca7
siHells

B


% 5mC
100
75
50
25
0
siCtrl
siZbtb24
siCdca7
siHells
P10 ICF2
MIR521-2 (cg24480735)
MIR521-2 (cg00101104)
MIR518D (cg10583119)
SNORD115-5 (cg18076399)
SNORD115-9 (cg17067638)
NRXN3 (cg10828316)

C


Marker
siCtrl
siZbtb24
siCdca7
siHells
P10
CTL
10
3
1
0.5
Satellite II (BstBI)
Satellite α (HhaI)

S1 Table: Genetic status of ICF patients enrolled in the genome-wide DNA methylation study

Patients (Gender)	Mutations and amino acid changes in DNMT3B	Mutations and amino acid changes in ZBTB24	Mutations and amino acid changes in CDCA7	Mutations and amino acid changes in HELLS	Birth year	Follow-up Status [reference]
ICF1	**NM_006892.3 (protein NP_008823.1)**					
p2 (M)	g.310C >T / g.2162 T >C p. Arg104* / p.Ile721Thr				2008	Necker Hospital (Paris, France) [1,2]
p3 (F)	g.1747G >A / g.1747G >A p. Gly583Ser / p. Gly583Ser				2009	Necker Hospital (Paris, France) [1,2]
p5 (M)	g.2629A>G g.2629A>G p. Lys770Glu / p. Lys770Glu				1991	(Hotel-Dieu de France Hospital (Beirut, Lebanon) [3]
pG (M)	g.158dupT / g.158dupT p. Ile41fs42* / p.Ser780Leu				1985	Deceased age 16 [2,4]
pH (M)	g.2428G>A / g.2428G>A p.Val818Met / p.Val818Met				1990	University Hospital Center (Nancy, France) [5]
pR (F)	g.2450A>G / g.2450A>G p. Asp817Gly / p. Asp817Gly				1991	Deceased age 22, 5y after HSCT [2,5]
pW (F)	g.1838T>C g.2476C>T p.Val613Ala / p.Arg826Cys				2004	NA [6]
pY (M)	g.2645C>T g.2645C>T p.Thr775Ile/ p.Thr775Ile				1996	Necker Hospital (Paris, France) [1,2]
ICF2		**NM_014797.2 (protein NP_055612.2)**				
p6 (M)		g.396_397del / g.396_397del p. His132fs19* / p. His132fs19*			2003	Hotel-Dieu de France Hospital (Beirut, Lebanon) [7]
p10 (F)		g.1369C>T / g.1369C>T p.Lys263* / p.Cys327fs54*			1996	Necker Hospital (Paris, France) [2,8]
pD (M)		c.396_397del / c.396_397del p. His132fs19* / p. His132fs19*			1997	Hotel-Dieu de France Hospital (Beirut, Lebanon) [7]
pP (F)		g.1369C>T / g.1369C>T p.Arg457* / p.Arg457*			1981	NA [9]
pV (M)		g.396_397del / g.396_397del p. His132fs19* / p. His132fs19*			1998	Hotel-Dieu de France Hospital (Beirut, Lebanon) [7]
ICF3			**NM_031942.2 (protein NP_665809.1)**			
pC (F)			g.881G>T / g.881G>T p.Gly294Val / p.Gly294Val		1998	La Timpone Hospital (Marseille, France) [5,10]
pS (F)			g.911G>A / g.911G>A p.Arg304His / p.Arg304His		1978	NA [2,5]
ICF4				**NM_018063.4 (protein NP_060533.2)**		
pU (M)				g.2400_2402delGTT / g.2400_2402delGTT p.Leu801del / p.Leu801del	1999	Pediatric Hematology/Oncology Institute (Lyon, France) [10]

M = male; F = female ; NA = not available; HSCT = Hematopoietic Stem Cell Transplantation

S2 Table: Hypomethylated gene clusters common to all ICF patients

Cluster name	Chr. band	Genes	GO	ChromHMM	Replication timing	Known regulators
Protocadherins alpha (PCDHA)	5q31.3	PCDHA1, PCDHA2, PCDHA3, PCDHA4, PCDHA5, PCDHA6, PCDHA7, PCDHA8, PCDHA9, PCDHA10, PCDHA11	homophilic cell adhesion via plasma membrane adhesion molecules (GO:0007156); cell-cell adhesion (GO:0098609); cell-cell adhesion via plasma-membrane adhesion molecules (GO:0098742); nervous system development (GO:0007399); calcium-dependent cell-cell adhesion via plasma membrane cell adhesion molecules (GO:0016339); synapse assembly (GO:0007416)	Heterochromatin, Repetitive	S2	DNMT3B, SMCHD1
Protocadherins beta (PCDHB)		PCDHB4, PCDHB6, PCDHB7			S2, S3	
Protocadherins gamma (PCDHGA and PCDHGB)		PCDHGA1, PCDHGA2, PCDHGA3, PCDHGA4, PCDHGA5, PCDHGA6, PCDHGA7, PCDHGA8, PCDHGA9, PCDHGA10, PCDHGA11, PCDHGA12, PCDHGB1, PCDHGB2, PCDHGB3, PCDHGB4, PCDHGB5, PCDHGB6, PCDHGB7			S2, S3	

S3 Table: **Gene ontology analysis of HypoMPs common to all ICF patients (GO Biological Process 2015, EnrichR)**

Term	P-value	Adjusted P-value	Z-score	Combined Score	Genes
homophilic cell adhesion via plasma membrane adhesion molecules (GO:0007156)	3,31E-20	5,68E-17	-2,21E+00	8,25E+01	CLSTN2;CDH9;PCDH18;CDH8;CDH6;PCDHA7;PCDHA6;PCDHGA7;PCDHGA2;PCDH7;PCDHB14;PCDHB13;PCDHB12;PCDHA12;PCDHGA12;PCDHB2;PCDHB1;PCDHB18;CDH11;FAT1;PCDHB6;PCDHB5;FAT3;PCDHB4;CDH18;PCDHB7
cell-cell adhesion (GO:0098609)	2,28E-18	1,31E-15	-2,21E+00	7,57E+01	CLSTN2;CDH9;PCDH18;CDH8;CDH6;CLDN22;PCDHA7;PCDHA6;PCDHGA7;PCDHGA2;PCDH7;PCDHB14;PCDHB13;PCDHB12;PCDHA12;PCDHGA12;PCDHB2;PCDHB1;PCDHB18;CDH11;FAT1;PCDHB6;PCDHB5;FAT3;PCDHB4;PCDH18;PCDHB7
cell-cell adhesion via plasma-membrane adhesion molecules (GO:0098742)	2,02E-18	1,31E-15	-2,21E+00	7,57E+01	CLSTN2;CDH9;PCDH18;CDH8;CDH6;CLDN22;PCDHA7;PCDHA6;PCDHGA7;PCDHGA2;PCDH7;PCDHB14;PCDHB13;PCDHB12;PCDHA12;PCDHGA12;PCDHB2;PCDHB1;PCDHB18;CDH11;FAT1;PCDHB6;PCDHB5;FAT3;PCDHB4;PCDH18;PCDHB7
detection of chemical stimulus involved in sensory perception of smell (GO:0050911)	4,58E-08	1,97E-05	-2,29E+00	2,48E+01	OR2T27;OR2C1;OR4K5;OR10H3;OR1E2;OR14J1;OR10H2;OR5I1;OR4K1;OR4M2;OR4K17;OR2W1;OR2Y1;OR4A47;OR52B2;OR4C6;OR5D16;OR9G4;OR4C15;OR4A15;OR4F6;OR4C46;OR4A16

Term	P-value	Adjusted P-value	Z-score	Combined Score	Genes
detection of chemical stimulus involved in sensory perception (GO:0050907)	6,61E-08	2,27E-05	-2,29E+00	2,45E+01	OR2T27;OR2C1;OR4K5;OR1E2;OR14J1;OR10H3;OR5I1;OR10H2;OR4K1;OR4M2;OR4K17;OR2W1;OR2Y1;OR4A47;TAS2R1;OR52B2;OR4C6;OR5D16;OR9G4;OR4C15;OR4A15;OR4F6;OR4C46;OR4A16
synapse assembly (GO:0007416)	8,77E-08	2,51E-05	-2,17E+00	2,30E+01	GPM6A;PCDHB2;DNER;NRXN3;PCDHB6;PCDHB5;PCDHB14;PCDHB4;PCDHB13
detection of chemical stimulus (GO:0009593)	2,30E-07	5,64E-05	-2,29E+00	2,24E+01	OR2T27;OR2C1;OR4K5;OR10H3;OR1E2;OR14J1;OR10H2;OR5I1;OR4K1;OR4M2;OR4K17;OR2W1;OR2Y1;OR4A47;TAS2R1;OR52B2;OR4C6;OR5D16;OR9G4;OR4C15;OR4A15;OR4F6;OR4C46;OR4A16
calcium-dependent cell-cell adhesion via plasma membrane cell adhesion molecules (GO:0016339)	6,14E-06	1,25E-03	-2,37E+00	1,58E+01	PCDHB2;PCDHB6;PCDHB5;PCDHB14;PCDHB4;PCDHB13
synapse organization (GO:0050808)	6,52E-06	1,25E-03	-2,20E+00	1,47E+01	GPM6A;PCDHB2;DNER;NRXN3;PCDHB6;PCDHB5;PCDHB14;PCDHB4;PCDHB13;PDZRN3
synaptic transmission (GO:0007268)	5,00E-04	7,16E-02	-2,32E+00	6,13E+00	GABRB3;SLC12A5;GABRA6;PCDHB14;HTR1B;PCDHB13;GRIK2;CARTPT;GABRG3;CDH8;GLRA1;GRIN2A;PCDHB2;PCDHB6;PCDHB5;PCDHB4
regulation of neurological system process (GO:0031644)	7,21E-04	8,85E-02	-2,42E+00	5,87E+00	GLRA1;GRIN2A;NPY2R;CARTPT;ADRA2C
neuron migration (GO:0001764)	1,56E-03	1,67E-01	-2,20E+00	3,93E+00	GPM6A;FOXG1;DNER;NDN;FAT3;PITX2;ATOH1

S4 Table: **Germline-enriched genes hypomethylated in their promoter region in ICF female patients** (1035 tissue-enriched genes, database http://www.proteinatlas.org/humanproteome/testis)

ICF subtype (number of genes)	**Hypomethylated germline gene promoters in female patients**
ICF1 specific (69)	ADAD1;AURKC;BIRC8;BOLL;C10orf53;C12orf40;C12orf54;C17orf104;CCDC105;CCDC144NL;CCDC155;CCDC79;CXorf58;CT45A1;DMRTC2;DPEP3;DUSP21;ESX1;FAM46D;FAM71B;FIGLA;FOXR1;FOXR2;FTHL17;GRID2;HIST3H3;HSPB9;IL13RA2;INSL6;LDHAL6A;LYPD4;MAEL;MARCH11;MOV10L1;ODF3L1;OOEP;PAGE1;PASD1;PIWIL1;PIWIL3;PNLDC1;POLR2D;POTED;PRR23A;PRR23B;PRSS21;PRSS50;RBM44;RBMXL2;RFPL3S
ICF2 specific (20)	C20orf173;DMRTB1;FATE1;GUCA1A;H1FNT;IRGC;KCNV2;MORC1;MYCBPAP;MYO1H;OPN5;PDHA2;PTCHD3;PSORS1C1;RDM1;SMCP;TNP1;TTLL2;TTLL8;UBL4B
ICF3 specific (17)	SEPT14;ADAM29;AKAP4;C10orf120;C6orf163;CPXCR1;CT47A1;FAM71E2;GLT6D1;MAGEA4;MAGEB16;OR3A2;SPACA3;RHOXF1;TKTL2;ZNF679;ZNRF4
Common to ICF1 and ICF2 (6)	C2orf78;CRISP2;CSNK1A1L;GLUD2;RBM46;SLC25A31
Common to ICF1 and ICF3 (9)	ASZ1;C1orf94;FMR1NB;LUZP4;MAGEA12;MAGEB2;PAGE2;PAGE2B;PRR23C
Common to ICF2 and ICF3 (30)	ADAM18;ADAM2;ATXN3L;C2orf16;CATSPER1;CNBD1;DRD3;FAM170A;KCNU1;KIF2B;MAGEB1;MAGEB18;MAGEC1;MAGEB6;MAGEC2;NANOS2;OR4M1;OR4N4;PDILT;POM121L12;PPEF1;SEL1L2;SPAM1;SPANXA1;SPANXA2;SPATA8;TMEM225;TRIML1;TUBA3C;ZNF479
Common to ICF1, ICF2 and ICF3 (28)	BEND2;DCAF8L1;DCAF8L2;DDX53;FAM24A;KIF4B;LDHC;LELP1;LRIT1;LYZL4;MAGEA1;MAGEB10;OTUD6A;MBD3L1;PAGE3;PAGE5;PKD2L2;SAGE1;SPO11;TAF7L;TAS2R1;TMCO5A;TPTE;TRIM42;USP29;XAGE5;ZCCHC13;ZNF716

S5 Table: **Germline-enriched genes hypomethylated in their promoter region in ICF male patients** (1035 tissue enriched genes, database http://www.proteinatlas.org/humanproteome/testis)

ICF subtype (number of genes)	**Hypomethylated germline gene promoters in male patients**
ICF1 specific (50)	ACTL9;ASZ1;AURKC;C10orf53;C12orf40;C1orf146;CCDC144NL;CCDC155;CCDC79;CRISP2;CT45A1;DMRTC2;FIGLA;DPEP3;H3F3C;HIST3H3;HSF2BP;HSPB9;INSL3;LYPD4;MAEL;MARCH11;MOV10L1;OOEP;PDCL2;PIWIL1;PIWIL3;PNLDC1;POLR2D;POTED;PRR23A;PRR23B;PRSS21;PRSS50;PTCHD3;RBM44;RBM46;RGPD8;SLC25A2;SPACA1;SYCE1;TBL2;TDRD12;TMPRSS12;TSSK6;UBE2U;YBX2;ZNF560;ZNF57;ZPBP2
ICF2 specific (12)	ADAM29;CCDC54;FAM71E2;IQCF5;LELP1;ODF3L1;OR3A2;OR4M1;OTUD6A;SPACA3;SPAM1;SPATA8
ICF4 specific (25)	ADAD2;C10orf120;C11orf42;C14orf177;C2orf16;C4orf51;CCDC36;CPXCR1;CSNK1A1L;DCAF4L2;FAM71F1;H1FNT;MAGEB1;MAGEA1;MAGEC2;NMS;PAGE5;PRM2;PRNT;SPATA19;SPATA21;TCEB3B;TPTE2;XAGE5;ZNRF4
Common to ICF1 and ICF2 (4)	CTAGE1;NANOS2;RBMXL2;SPO11
Common to ICF1 and ICF4 (11)	ADAD1;FOXR1;FTMT;GRID2;INSL6;LDHAL6A;PKD2L2;PRR23C;PSMA8;TRIM36;WDR88
Common to ICF2 and ICF4 (29)	ADAM18;ADAM2;CATSPER1;DCAF8L1;DRD3;FAM170A;FATE1;GLT6D1;KCNU1;KIF2B;MAGEB18;MAGEB6;MBD3L1;MAGEC1;MYCBPAP;OPN5;OR4N4;PAGE3;PDILT;POM121L12;SEL1L2;SPANXA1;SPANXA2;TAS2R1;TRIM42;TRIML1;TUBA3C;ZNF479;ZNF716
Common to ICF1, ICF2 and ICF4 (11)	C2orf78;FAM24A;FAM46D;GAPDHS;KIF4B;LRIT1;LYZL4;SLC25A31;TMCO5A;TPTE;USP29

S6 Table: Hypomethylated gene clusters common to ICF2, ICF3 and ICF4 patients

Cluster name	Chr. band	Genes	GO	ChromHMM	Replication timing	Remarks
Late Cornified Envelope (LCE)	1q21.3	LCE1C, LCE2A, LCE2D, LCE3C, LCE3E, LCE4A, LCE6A	keratinization (GO:0031424)	Heterochromatin, Repetitive	S3	
Small proline-rich proteins (SPRR)	1q21.3	SPRR2D, SPRR2F, SPRR3	keratinization (GO:0031424); epidermis development (GO:0008544); epithelium development (GO:0060429); keratinocyte differentiation (GO:0030216); epidermal cell differentiation (GO:0009913)	Heterochromatin, Repetitive	S2, S3	
Olphactory Receptors (OR)	11p15.4	OR51A7, OR51L1, OR51Q1, OR56A5, OR52L1, OR10A6, HBG2 HBE1	detection of chemical stimulus involved in sensory perception of smell (GO:0050911)	Heterochromatin, Repetitive	S3, S4	Random monoallelic expression
Olphactory Receptors (OR)	11p11.2	OR4B1, OR4X2, OR4X2, OR4C3, OR4C12, TRIM49B		Heterochromatin, Repetitive	S2, S3	
Olfactory Receptors (OR)	11q11	OR4A16, OR4A15, OR4C15, OR4C16, OR4C16, OR4C11, OR4S2, OR5D13, OR5D13, OR5D14, OR5L1, OR5L1, OR5D18, OR5D16, OR5W2, OR5AS1, OR8K5, OR5J2, OR5T2, OR8K3, OR8K3, OR8K1, OR5AR1, OR5AR1, OR9G1, OR9G4, OR5G3, OR5AK2, OR9Q1, OR6Q1, OR6Q1, OR9Q1, OR9Q2, OR10W1, LRRC55, TRIM51		Heterochromatin, Repetitive	S4, G2	
Olfactory Receptors (OR)	12q13.2	OR6C1, OR6C4, OR6C74, OR6C74		Heterochromatin, Repetitive	S3	

Olfactory Receptors (OR)	14q11.2	OR4K2, OR4K1, OR4K5, OR4K17, OR4M1, OR11H6, OR11G2		Heterochromatin, Repetitive	S1,S2,S3	
NLR Family Pyrin Domain Containing (NLRP)	19q13.42-q13.43	NLRP9, RFPL4A, NLRP11, NLRP4, NLRP13, NLRP8, NLRP5,	neuron death (GO:0070997); embryo implantation (GO:0007566); regulation of type I interferon production (GO:0032479)	Heterochromatin, Repetitive	S1, S2	
Keratin Associated Protein (KRTAP)	21q22.11	KRTAP24-1, KRTAP26-1, KRTAP13-2, KRTAP19-3, KRTAP19-4, KRTAP19-7, KRTAP8-1, KRTAP7-1, KRTAP11-1	Cellular Component: Cellular component: intermediate filament (GO:0005882)	Heterochromatin, Repetitive	S4	
Zinc Finger proteins (ZNF)	19p12	ZNF132, ZNF208, ZNF208, ZNF30-AS1, ZNF490, ZNF599, ZNF676, ZNF714, ZNF714, ZNF714, ZNF714, ZNF728	transcription, DNA-templated(GO:0006351); regulation of transcription (GO:0006355)	Heterochromatin, Repetitive	S4, G2	Pericentromeric - region chr19 late replication, satellite repeats, heterochromatin
snoRNA cluster 109-116-115 (SNORD)	15q11.2	SNHG14, SNORD116-9, SNORD116-15, SNORD116-17, SNORD116-18, SNORD116-24, PWAR1, SNORD115-15, SNORD115-18, SNORD115-19, SNORD115-22, SNORD115-28, SNORD115-31, SNORD115-36, SNORD115-37, SNORD115-38, SNORD115-39, SNORD115-40, SNORD115-45	RNA editing and/or alternative splicing	Heterochromatin, Repetitive SNORD 116- Weak transcription/transcri ption elongation (SNORD109 and 115)	S4, G2	Parent-of-origin imprint - Prader-Willi/Angelman syndrome (PWS) region
miRNAs cluster C19MC	19q13.42		Post-transcriptional regulation of gene expression	Heterochromatin, Repetitive	S3	Maternal-specific methylation imprint
Defensin (DEFB)	20q11.21	DEFB118, DEFB119, DEFB121, DEFB122	response to bacterium (GO:0009617)	Heterochromatin	S1, S2	Pericentromeric-DEFB115 et 116 are commmon to all ICF patients

S7 Table: **Histone marks enriched in genomic regions that show hypomethylation (whole HypoMPs) in ICF1 (Epigenomics Roadmap HM ChIP-seq; EnrichR)**

Term	P-value	Adjusted P-value	Z-score	Combined Score
H3K27me3 Mobilized CD34 Primary Cells	4,15E-34	1,59E-31	-1,74E+00	1,23E+02
H3K9me3 CD34 Cultured Cells	8,27E-10	3,96E-08	-1,25E+00	2,14E+01
H3K27me3 CD3 Primary Cells	6,02E-05	1,36E-03	-1,33E+00	8,81E+00
H3K27me3 CD8 Naive Primary Cells	4,55E-04	7,26E-03	-1,34E+00	6,61E+00
H3K27me3 CD34 Primary Cells	1,03E-03	1,46E-02	-1,25E+00	5,27E+00
H3K9me3 Peripheral Blood Mononuclear Primary Cells	8,66E-04	1,28E-02	-1,05E+00	4,58E+00
H3K27me3 CD4+ CD25int CD127+ Tmem Primary Cells	1,48E-03	1,95E-02	-1,15E+00	4,51E+00

S8 Table: **Histone marks enriched in regions that show hypomethylation of genes promoters (TSS200; TSS1500; Illumina annotation) in ICF1 (Epigenomics Roadmap HM ChIP-seq; EnrichR)**

Term	P-value	Adjusted P-value	Z-score	Combined Score
H3K27me3 Mobilized CD34 Primary Cells	1,24E-07	4,60E-05	-1,74E+00	1,73E+01
H3K9me3 CD34 Cultured Cells	4,37E-06	3,24E-04	-1,29E+00	1,03E+01
H3K9me3 Mobilized CD34 Primary Cells	3,05E-05	1,25E-03	-1,33E+00	8,89E+00
H3K9me3 Peripheral Blood Mononuclear Primary Cells	2,20E-05	1,11E-03	-1,25E+00	8,49E+00
H3K27me3 CD34 Primary Cells	2,45E-03	3,46E-02	-1,26E+00	4,23E+00
H3K27me3 CD8 Memory Primary Cells	2,52E-03	3,46E-02	-1,21E+00	4,06E+00
H3K27me3 Peripheral Blood Mononuclear Primary Cells	5,52E-03	5,83E-02	-1,21E+00	3,43E+00

S9 Table: **Histone marks enriched in genomic regions that show hypomethylation (whole HypoMPs) in ICF2, ICF3 and ICF4 (Epigenomics Roadmap HM ChIP-seq; EnrichR)**

Term	P-value	Adjusted P-value	Z-score	Combined Score
H3K27me3 Mobilized CD34 Primary Cells	4,15E-34	1,59E-31	-1,74E+00	1,23E+02
H3K9me3 CD34 Cultured Cells	8,27E-10	3,96E-08	-1,25E+00	2,14E+01
H3K27me3 CD3 Primary Cells	6,02E-05	1,36E-03	-1,33E+00	8,81E+00
H3K27me3 CD8 Naive Primary Cells	4,55E-04	7,26E-03	-1,34E+00	6,61E+00
H3K27me3 CD34 Primary Cells	1,03E-03	1,46E-02	-1,25E+00	5,27E+00
H3K9me3 Peripheral Blood Mononuclear Primary Cells	8,66E-04	1,28E-02	-1,05E+00	4,58E+00
H3K27me3 CD4+ CD25int CD127+ Tmem Primary Cells	1,48E-03	1,95E-02	-1,15E+00	4,51E+00

S10 Table: **Histone marks enriched in regions that show hypomethylation of genes promoters (TSS200; TSS1500; Illumina annotation) in ICF2, ICF3 and ICF4 (Epigenomics Roadmap HM ChIP-seq; EnrichR)**

Term	P-value	Adjusted P-value	Z-score	Combined Score
H3K9me3 Peripheral Blood Mononuclear Primary Cells	5,17E-09	9,87E-07	-1,31E+00	1,81E+01
H3K9me3 Mobilized CD34 Primary Cells	2,43E-05	8,79E-04	-1,32E+00	9,28E+00
H3K9me3 CD4+ CD25- Th Primary Cells	4,72E-06	2,27E-04	-7,47E-01	6,27E+00
H3K27me3 Mobilized CD34 Primary Cells	5,50E-03	6,63E-02	-1,43E+00	3,87E+00
H3K9me3 CD4 Memory Primary Cells	3,70E-04	7,88E-03	-5,58E-01	2,70E+00
H3K9me3 CD8 Naive Primary Cells	2,13E-03	3,22E-02	-5,33E-01	1,83E+00
H3K9me3 Treg Primary Cells	1,69E-03	2,92E-02	-5,05E-01	1,79E+00

S11 Table: List of primers and techniques in which they were used

Name	Sequence (5' to 3')	Method
PCDHB13 F	ATTGGATATAATTTTTTTATTTGAGATAGT	*COBRA*
PCDHB13 R	TTAATA ACATAAATCCATCATAAATCTAAC	*COBRA*
SNORD116-7 F	GTATTTTATTGATTGAAGGTTTTGG	*COBRA*
SNORD116-7R	TATATCTCCCCCACACAAATATTTC	*COBRA*
SNORD115-14F	GTTTGTTGTTGGGGGTTATTTT	*COBRA*
SNORD115-14R	CACATAATACCCTCTCCTTTCCTTA	*COBRA*
SNORD114-1F	AAAGGTAGTGTTAATAAATGGTAGATAA	*COBRA*
SNORD114-1R	TCCCAACAAAATAATTCAAAAAACT	*COBRA*
MIR376C F	AGAGTGATGGAAGGTGAATTTTAGA	*COBRA*
MIR376C R	TCCACCTTTTAAATA CCAAAAAAAA	*COBRA*
TDRD6 F	AGGAGATGTAGGTTGTGTTTAAAATTT	*COBRA*
TDRD6 R	AATTCAAAAAACCCAATAAAAAAAA	*COBRA*
SPO11 isl F	GTTTTAGGGTTTTATAGGTGGGAT	*COBRA*
SPO11 isl R	CTCCAAAAACCCCTATCCTAAAC	*COBRA*
MIR521-2 F	GATTGTGTGTATTTGTTTTGGAATTTA	*COBRA*
MIR521-2 R	AAAAAACCATCTCCAATTCCTTACT	*COBRA*
DRD3 F	TTTGGATGTAAGATTTTTGAAGTGT	*COBRA*

Name	Sequence (5' to 3')	Method
DRD3 R	TACAAAACCCCTCTCCTTAACTTAC	*COBRA*
NDN F1	TTGTAAAGTGGATGTTTGAGGTT	*COBRA*
NDN R1	CCCAAAAAAAAATATAAAACCCAAAT	*COBRA*
NDN2 F2	TTGTTTTAGTATTTTTGGTTTGGGT	*COBRA*
NDN R2	AAAAAACCTTTCAACATCAACCTA	*COBRA*
ZNF597 F	GTTTTTTGATAGGAGTTGTAGAAAG	*COBRA*
ZNF597 R	ATTCAATTAATCAACACAAAAAAAA	*COBRA*
SNRPN F	ATTATGGTTGTTTGGTTTTTGGA	*COBRA*
SNRPN R	CCACACAAAAAACTCTAAACTCAACTAC	*COBRA*
PCDHB18F	AGTTTAAAATGTTTAATTTTTTTAAAGTAA	*COBRA*
PCDHB18R	TTAACTCTCTCCACAACTACTCTTC	*COBRA*
ADAM2F	GAGTTTTTGAGTTTTGTTTTTTGTAGG	*COBRA*
ADAM2R	ATTTATAAATCTTCCTATTCCTCAACTATT	*COBRA*
MAGEB10F	AGGAGTTAAAATATGGATTTATTTT	*COBRA*
MAGEB10R	CCTCACTACAAAACCTCACCTC	*COBRA*
C15ORF60F	GATATGGGTTGTTTGAGGGAAG	*COBRA*
C15ORF60R	TCAACAAAAACCAATAAACAAACTC	*COBRA*
SNORD115-9F	TTTTAGTGTTGTTTGGGGTTATGA	*COBRA*
SNORD115-9R	TAAAAAACCCTTTACCCTCCC	*COBRA*

Name	**Sequence (5' to 3')**	**Method**
MIR518D F	TAAATTGTGGTTTATTTTGGAGGAT	*COBRA*
MIR518D R	CCAAATATTCACTTAAATTCCAACAC	*COBRA*
SPO11 F	TGGGATTTTAATTTAGGTTTGTTTG	*COBRA*
SPO11 R	AAATACCTCCAAAACCCTTTACACT	*COBRA*
Name	**Sequence (5' to 3')**	**Method**
SNORD115-14F	GTT TGTTGTTGGGGGTTATTTT	*Bisulfite sequencing*
SNORD115-14R	CACATAATACCCTCTCCTTTCCTTA	*Bisulfite sequencing*
ZNF597 F	GTTTTTTGATAGGAGTTGTAGAAAG	*Bisulfite sequencing*
ZNF597 R	ATTCAATTAATCAACACAAAAAAAA	*Bisulfite sequencing*
Name	**Sequence (5' to 3')**	**Method**
MAGEB 10 F	AGGAGTTAAAATATGGATTTATTTT	*Pyrosequencing*
MAGEB 10 R	Biot-CCTCACTACAAAACCTCACCTC	*Pyrosequencing*
MAGEB 10 S1	TTATTTTATAATAAAGAG	*Pyrosequencing*
MAGEB 10 S2	GTTTTTGTTGGTAGTTTT	*Pyrosequencing*
SPO11 F	GTTTTAGGGTTTTATAGGTGGGAT	*Pyrosequencing*
SPO11 R	Biot-CTCCAAAAACCCCTATCCTAAAC	*Pyrosequencing*
SPO11 S	TTTGAGAGTTTTGCGTTT	*Pyrosequencing*
MIR518D F	TAAATTGTGGTTTATTTTGGAGGAT	*Pyrosequencing*
MIR518D R	Biot-CCAAATATTCACTTAAATTCCAACAC	*Pyrosequencing*
MIR518D S	GTGATAATTTATATTTTT	*Pyrosequencing*
ZNF597 F	GTTTTT TGATAGGAGTTGTAGAAAG	*Pyrosequencing*

ZNF597 R	Biot-ATTCAATTAATCAACACAAAAAAAA	*Pyrosequencing*
ZNF597 S	TGATAGGAGTTGTAGAAA	*Pyrosequencing*
MIR376C F	AGAGTGATGGAAGGTGAATTTTAGA	*Pyrosequencing*
MIR376 R	Biot-TCCACCTTTTAAATACCAAAAAAAA	*Pyrosequencing*
MIR376C S	AAGGTTTGTTTTGGGATA	*Pyrosequencing*
Name	**Sequence (5' to 3')**	**Method**
NRXN3 F	ACAAATGATGCTCTCTCCCTGT	*MSRE-qPCR*
NRXN3 R	ATTGAACCTGTCATTTCTTTGC	*MSRE-qPCR*
MAGEB3 F	CACAGTGCACAAACTAGGTGCT	*MSRE-qPCR*
MAGEB3 R	AGACTGTGCTAGGATAGGGACA	*MSRE-qPCR*
SNORD115-5 F	TGTGAGCTCTTTGGTGTCCGT	*MSRE-qPCR*
SNORD115-5 R	GCCACTGAACCATGGAAGGAC	*MSRE-qPCR*
SNORD115-9 F	AGTGCTGCTTGGGGTCATG	*MSRE-qPCR*
SNORD115-9 R	ACACTGATGTGTCTCAGACGTGG	*MSRE-qPCR*
MIR521-2 F1	CTCAGGTGATCCTCCCGTC	*MSRE-qPCR*
MIR521-2 R1	ATTCTCCAGGATAGACCATAGGC	*MSRE-qPCR*
MIR521-2 F2	GAACTCAAGTGCACCCTTG	*MSRE-qPCR*
MIR521-2b R2	TTCCTGAGTTTCGTATCTGC	*MSRE-qPCR*
MIR518D F	TGTGGTCTATCCTGGAGGATG	*MSRE-qPCR*
MIR518D R	AACAGCAATGGATGTACAGCC	*MSRE-qPCR*
Name	**Sequence (5' to 3')**	**Method**
U6 F	CTCGCTTCGGCAGCACA	*RT-qPCR*

U6 R	AACGCTTCACGAATTTGCGT	*RT-qPCR*
DNMT3B F	AGAGCAGTCTGAGCCTGAGAC	*RT-qPCR*
DNMT3B R	CGAGCTTGATCTTCGGCCTCTGAT	*RT-qPCR*
ZBTB24 F	GCTGGAGCACATGAGCCTG	*RT-qPCR*
ZBTB24 R	TTGGCACTGGAGTCAGAGAAG	*RT-qPCR*
CDCA7 F	TCCGACTCACAATCAAGGAGAC	*RT-qPCR*
CDCA7 R	GGTCATCTTCATTCATGTAGCC	*RT-qPCR*
LSH F	ACA CTG CTG TGA TTA CCC CG	*RT-qPCR*
LSH R	AGA CAT GCG AGC CTT TTC CA	*RT-qPCR*
Name	**Sequence (5' to 3')**	**Method**
V5ZBTB24 F	CTGCTGGGCCTGGATAGCACCGCAGACACAACCCCGGAG	*Plasmid constructs*
V5ZBTB24 R	GGGTTCGGAATCGGTTTGCCCATTTTCTCCAGAAGACTCTGAAGAG	*Plasmid constructs*
CDCA7HA F	GCCGGATTATGCGTAGAGCCTGAAGATTGTC	*Plasmid constructs*
CDCA7HA R	ACATCATACGGATACGCTTGCATTTCAAATTCC	*Plasmid constructs*

UNPUBLISHED RESULTS

The ICF syndrome represents an important human disease to identify new players and new mechanisms of DNA methylation. Work in the host team previously characterized a mouse model for the ICF1 syndrome, carrying mutations described in patients. This study allowed the identification of a subfamily of germline genes whose promoter is methylated by DNMT3B. Moreover, the E2F6 repressor was described to recruit DNMT3B at these promoters to ensure their methylated status and transcriptional silencing (Velasco et al., 2010). Compared to the obvious contribution of DNMTs to DNA methylation, virtually nothing is known about the contribution of newly identified factors causally involved in ICF syndrome that are devoid of methyltransferase activity but which loss of integrity has dramatic impact on methylation patterns and development.

To better characterize the role of ZBTB24 in development and in DNA methylation machinery, we generated a mouse model carrying mutations in ZBTB24 that reproduce those already described in three ICF2 brothers. Although ZBTB24 seemed to be essential for development in the mouse, as previously described (Wu et al., 2016), it seemed to be dispensable for *in vitro* differentiation of mouse ES cells. We studied the role of ZBTB24 during differentiation, taking advantage of a dynamic system of study and we observed that ZBTB24 is required to establish DNA methylation *via* direct recruitment of DNMT3B at genomic repeats, tandem minor satellite repeats and LINE1T interspersed DNA repeats. Moreover, we have evidence that ZBTB24 control indirectly the establishment of DNA methylation, modulating the establishment of H3K9me3 at some genomic loci. This study supports a dual role of ZBTB24 in the establishment of DNA methylation in the first steps of cell differentiation.

I contributed to this study performing dissection experiments at different stages of embryonic development. I performed Southern blot analysis on genomic DNA extracted directly from embryos at 11.5 dpc. I participated to the derivation of mouse ES cells at 3.5 dpc. I adapted WT and $Z24^{*/*}$ cells, as well as all the other mouse ES cell lines used in this study, to grow in a feeders-free condition in 2i growing medium. I optimized and performed the embryoid body differentiation by “hanging drop” as well as the neuron specific differentiation of mouse ES cells, followed by either analysis of differentiation markers by RT-qPCR or methylation analysis of major and minor satellite repeats by Southern blot. I prepared the genomic DNA to perform our RRBS analysis (performed in collaboration with Michael Weber’s lab) and I participated to the first analysis of the normalised data. I then performed all the ChIP-qPCR analyses in different cells lines, taking advantage of a dynamic system of study.

This study still needs further validations and analysis before being submitted.

1. ZBTB24 is essential for embryonic development but is dispensable for ES cell differentiation

The discovery of ZBTB24 mutations as a genetic cause of the ICF syndrome (ICF2) raised the question of its role in DNA methylation and embryonic development. ZBTB24 being highly conserved between mouse and humans, we designed mutations in ZBTB24 that reproduce those already described in three ICF2 brothers to generate an ICF2 mouse model (Chouery et al., 2012) (**Fig. 1A; see Methods section**). This mutation is a deletion of one base in the exon 2 which encodes the BTB and the AT-hook domains of ZBTB24 protein, leading to a frame-shift of 19 amino-acids and the appearance of a premature stop codon (**Fig. 1B; see Methods section**). The truncated protein product of the mutant allele only contains the dimerization and oligomerization BTB domain, losing the AT-hook and zinc finger domains which are essential for the ability of ZBTB24 to bind to DNA (**Fig. 1B; see Methods section**). The mentioned ICF2 patients and the mouse model are homozygous for the mutation.

Heterozygous mice ($Z24^{+/*}$) have the same visible phenotype as their wild-type littermates and were born at the expected Mendelian ratio (**Fig. 2A**). In contrast, the breeding of heterozygous mice yielded no viable homozygous ZBTB24 mutant mice ($Z24^{*/*}$) from a total of 102 offspring (**Fig. 2A**). We were able to detect $Z24^{*/*}$ embryos, although not beyond 11.5 dpc (**Fig. 2A**), indicating that implantation was not affected by the mutations. The characterization of $Z24^{*/*}$ embryos at post-implantation stages showed that $Z24^{*/*}$ embryos were much smaller in size compared to WT and $Z24^{+/*}$, suggesting an impairment or a delay in embryonic development (**Fig. 2B**). To confirm this observation, we analysed expression of BRACHYURY, a mesoderm marker highly expressed in the primitive streak cells that migrate to the posterior epiblast of the embryo between 6.5 and 7.5 dpc (Tam and Behringer, 1997). We performed whole-mount *in situ* hybridisation on embryos at 7.5 dpc and observed that $Z24^{*/*}$ embryos did not exhibit a complete migration of the primitive streak cells to the posterior of the embryo compared to WT and $Z24^{+/*}$ embryos (**Fig. 2B**). These results further support that loss of ZBTB24 functions affects the normal development of the embryos. To determine whether ZBTB24 mutations lead to developmental anomalies of the embryos, we looked for morphological defects of the $Z24^{*/*}$ at 9.5 and 11.5 dpc. Whereas examination at 9.5 dpc only allowed to highlight an important developmental delay, we observed the occurrence of developmental morphological defects of $Z24^{*/*}$ embryos at 11.5 dpc. These embryos clearly lack head structures and may also suffer from limb buds anomalies (**Fig. 3**). Altogether, our results support an essential role of ZBTB24 in embryonic development as already reported in a different mouse model (Wu et al., 2016). Interestingly, our study also

highlights a role of ZBTB24 in development pathways that lead to head and brain formation during embryogenesis.

We then derived mouse embryonic stem cells (mES cells) to study the role of ZBTB24 in cellular differentiation. We isolated blastocysts at 3.5 dpc and did not notice any significant difference between $Z24^{*/*}$ and WT or $Z24^{+/*}$ embryos (**Fig. 4A**). Following the protocol described by Czechanski and colleagues (Czechanski et al., 2014), we were able to derive 2 male WT, 5 male $Z24^{+/*}$, 2 female $Z24^{+/*}$ and 2 male $Z24^{*/*}$ mES cells clones (**Fig. 4B**). The mES cells were isolated and maintained in 2i media (Ying et al., 2008) which contains the two small-molecule inhibitors PD0325901 (inhibitor of the MEK signalling pathway) and CHIR99021 (inhibitor of the GSK3 signalling pathway), in combination with LIF (Leukemia Inhibition Factor), in the absence of serum and feeder cells (Tosolini and Jouneau, 2016). We confirmed the naïve pluripotent state of our derived cells by checking the described homogeneous expression of the pluripotency factors NANOG and OCT3/4 detected by immunofluorescence (**Fig. 4C**).

To evaluate the differentiation potential of our ES, we performed two types of kinetics of differentiation that promote i) the formation of both endoderm and mesoderm cells (embryoid bodies differentiation) or ii) the formation of cortical neurons, originated from the ectoderm (neuron-specific differentiation) (Kurosawa, 2007). These analyses were first performed on the WT clone #1 and the $Z24^{*/*}$ clone #1. Embryoid bodies differentiation was performed using the "hanging drop" technique (**Fig. 5A; see Methods section**), in presence of serum and absence of LIF. To test the differentiation potential of $Z24^{*/*}$ compared to WT mES cells, we followed the kinetics of expression of pluripotency markers by RT-qPCR (**see Methods section**). During differentiation, both WT and $Z24^{*/*}$ mES cells showed a drastic silencing of the pluripotency factors KLF4 and REX1 starting at day 1 and day 3 of differentiation, respectively, suggesting that the pluripotency program is normally silenced in the absence of a functional ZBTB24 (**Fig. 5B**). We then evaluated the expression of mesoderm markers BRACHYURY and WNT3 (Lako et al., 2001; Tam and Behringer, 1997), and endoderm markers GATA4 and GATA6, which are specifically expressed in the primitive endoderm *in vivo* and *in vitro* (Cai et al., 2008). We observed that $Z24^{*/*}$ ES cells showed altered kinetics of mesoderm as well as endoderm markers expression during differentiation. At day 5, the appearance of differentiation markers was comparable between WT and $Z24^{*/*}$ cells. In contrast, it seems that defects of BRACHYURY and WNT3 expression, as well as that of GATA4 and GATA6, occur during late stages of differentiation, mainly at day 12. (**Fig. 5C and 5D**). As a whole, although the expression of differentiation markers seems to be altered in late stages of the differentiation, their kinetics of appearance is comparable between WT and $Z24^{*/*}$ cells, suggesting that ZBTB24 functions is dispensable for *in vitro* differentiation of mouse ES cells into mesoderm and endoderm cells.

We then checked the ability of mES cells to form ectoderm lineages using the neuron-specific differentiation protocol described by Wichterle et al., (2009). In this differentiation protocol we added retinoic acid (RA) during embryoid bodies formation, followed by the addition of corticosterone, which an agonist of the hedgehog pathway, to promote neuronal differentiation (Traiffort et al., 2010). Neuron progenitors were then plated on plates coated with poly-ornithine and laminin to promote neuron differentiation and dendrites formation (**Fig. 6A**). The analysis of expression of pluripotency factors by RT-qPCR during differentiation showed a drastic silencing of the pluripotency factors KLF4 and REX1 starting at day 1 and day 3 of differentiation for WT and Z24$^{*/*}$ cells, respectively (**Fig. 6B**). Moreover, expression of the neuron-specific markers, PAX6 and NEUN (Gusel'nikova and Korzhevskiy, 2015; Mastick and Andrews, 2001), were not significantly different between WT and Z24$^{*/*}$ cells (**Fig. 6C**). To test the ability of WT- and Z24$^{*/*}$-derived neurons to form networks and dendritic arborisations we analysed the expression of the neuron-specific marker **β**III-tubulin (Menezes and Luskin, 1994) by immunofluorescence. We used H3K9me3 as a nuclear marker, which is also a typical marker of heterochromatin (Schotta et al., 2004). **β**III-tubulin was expressed in both WT- and Z24$^{*/*}$-derived neurons, and we could not observe significant differences between WT and Z24$^{*/*}$ differentiated neurons with respect to the formation of networks and dendritic arborisations (**Fig. 6D**). In addition, we did not observe any difference in the intensity and the repartition of H3K9me3 signal (**Fig. 6D**). Our results suggested that ZBTB24 functions are dispensable for *in vitro* differentiation of mouse ES cells into cortical neurons.

As a whole, our results underline an essential role of ZBTB24 in post-implantation development *in vivo*, around gastrulation. However, our results do not highlight a major implication in the formation of the three germ layers lineage *in vitro*.

2. Loss of ZBTB24 functions impairs DNA methylation at tandem and interspersed DNA repeats

We first verified that the levels of DNMT proteins were not affected in Z24$^{*/*}$ cells (**Fig. 7**). However, mutations in ZBTB24 coding gene were associated with hypomethylation of centromeric and pericentromeric repeats in ICF patients (de Greef et al., 2011). To test the impact of mutations in ZBTB24 on DNA methylation levels at centromeric and pericentromeric repeats, called respectively minor and major satellite repeats, we performed Southern blot on genomic DNA extracted from three different Z24$^{*/*}$ embryos at 11.5 dpc, when DNA methylation is already established (Auclair et al., 2014), and we found that all of them showed hypomethylation of both minor and major satellite repeats (**Fig. 8**). These results are in agreement with our previous results

showing that ZBTB24 is required for the maintenance of DNA methylation at centromeric minor satellite repeats in the mouse (Thijssen et al., 2015). However, the hypomethylation of centromeric and pericentromeric repeats may originate from defects that occur during establishment of DNA methylation at early steps of development from blastocyst stage.

This wave of *de novo* DNA methylation can be reproduced in culture during ES cell differentiation. mES cells cultured in 2i are known to have low expression of *de novo* DNA methyltransferases DNMT3A and DNMT3B and low DNA methylation levels (Habibi et al., 2013; Leitch et al., 2013). Upon induction of mES cell differentiation, expression of DNMT3A and DNMT3B increases and the wave of *de novo* methylation occurs (Leitch et al., 2013). We analysed the levels of DNA methylation at minor satellite repeats during embryoid formation by Southern blot. In WT mES cells, minor satellite repeats became fully methylated at d3 of differentiation (**Fig. 9A**). Interestingly, Z24$^{*/*}$ mES cells already showed a drastic hypomethylation at minor satellite repeats in 2i conditions compared to WT mES cells. During differentiation, mutations in ZBTB24 led to a delay in the establishment of DNA methylation compared to WT cells, resulting in an incomplete DNA methylation establishment after 5 days of differentiation, suggested by the presence of fragments at low molecular weights (**Fig. 9A**). In contrast to minor satellite repeats, we did not observe any difference in DNA methylation levels at major satellite repeats between WT and Z24$^{*/*}$ mES cells during differentiation (**Fig. 9B**).

To identify genomic sequences at which functional ZBTB24 is required for their methylation, we generated single-base profiles of cytosine methylation by Reduced-Resolution Bisulfite Sequencing (RRBS) in WT and Z24$^{*/*}$ mES cells. This analysis was performed on two clones of WT and Z24$^{*/*}$ mES cells, in 2i media and at day 3 (d3) of differentiation. We quantified methylation levels for approximately 1,300,000 CpGs *per* sample at an average sequencing depth of 74x (**Table 1**). Unsupervised hierarchical clustering based on DNA methylation data clearly distinguished mES cells in 2i from differentiated mES cells (**Fig. 10A**). In addition, in 2i media, clustering differentiated Z24$^{*/*}$ clones from WT clones (**Fig. 10A**). At d3 of differentiation, the DNA methylation profile of Z24$^{*/*}$ clone #1 was more similar to WT clones compared to Z24$^{*/*}$ clone #2, suggesting that the methylation profile of Z24$^{*/*}$ clone #2 was more affected compared to Z24$^{*/*}$ clone #1 (**Fig. 10A**). The global levels of DNA methylation were very low in WT clones maintained in 2i, as expected from previous studies (Habibi et al., 2013), and increased at d3 of differentiation, as illustrated by the shift from low towards high DNA methylation values (**Fig. 10B**). Both Z24$^{*/*}$ clones in 2i conditions showed lower DNA methylation levels compared to WT clones, as illustrated by the shift from high towards low DNA methylation values in both Z24$^{*/*}$ clones compared to WT clones (**Fig. 10B**). At d3 of differentiation, DNA methylation levels increased in both Z24$^{*/*}$ clones, although to a lesser extent compared to WT clones (**Fig. 10B**).

Hence, our results showed that loss of ZBTB24 functions partially impaired the genome-wide establishment of DNA methylation *in vitro*.

We thus compared DNA methylation values between WT and Z24$^{*/*}$ samples in 2i conditions and at d3 of differentiation. Genomic regions were considered differentially methylated between the two genotypes when showing a difference of at least 20% of DNA methylation and a *P*-value<0.05. We identified 10,947 and 199 differentially methylated regions (DMRs) between WT and Z24$^{*/*}$ clones in 2i and at d3, respectively (**Fig. 10C**). We then focused our analysis on hypomethylated regions that comprise 99% of DMR (10,940 in 2i condition and 195 at d3) (**Fig. 10C**). Hypomethylated CpGs were mainly located within CpG poor regions with only a few within CpG islands (13 and 2 in 2i and at d3, respectively) (**Fig. 11A**). The genomic repartition of hypomethylated regions showed that most of these regions were found more than 1kb away of the TSS of annotated genes (**Fig. 11B**).

The chromosomal distribution of hypomethylated regions revealed that all the chromosomes were almost equally affected by hypomethylation in 2i conditions, whereas at d3 of differentiation, most of the hypomethylated regions (37 out of 190) were hosted on the chromosomes 7 and X (**Fig. 11C**). Since the mouse chromosome 7 is enriched in imprinted genes and clusters (Caspary et al., 1998), we further analysed the chromosomal coordinates of hypomethylated regions on chromosome 7. In agreement with our previous observations in human ICF2 patients (Article 2), the SNORD116 non-coding RNA cluster within the Prader-Willi locus was hypomethylated in Z24$^{*/*}$ ES cells, both in proliferation and at d3 of differentiation. We validated SNORD116 hypomethylation by MSRE-qPCR that showed a reduction of 70% of DNA methylation in Z24$^{*/*}$ clone (**Fig. 12A**; the results were obtained analysing WT and Z24$^{*/*}$ clone #1). Our analysis further revealed hypomethylation of the promoter region of the imprinted gene H19. We used pyrosequencing to validate H19 hypomethylation in Z24$^{*/*}$ cells and indicate a loss of 20% of DNA methylation at all the CpGs analysed (**Fig. 12B**; the experiment was performed once and on the WT and Z24$^{*/*}$ clone #1).

Since X chromosome is particularly enriched in LINE elements (Boyle et al., 1990), and in light of our methylomic data in ICF2 patients (Article 2), we tested the impact of mutations in ZBTB24 on DNA methylation of interspersed DNA repeats as LINE and LTR elements. RRBS analysis revealed that 52 and 59% of hypomethylated regions in Z24$^{*/*}$ cells contain DNA repeats, both in 2i and at d3 of differentiation respectively (**Fig. 13A**). The breakdown of repeat-associated hypomethylated regions into different families of DNA repeats showed that ERVs, in particular the class of IAP, and LINE1 elements were highly hypomethylated in Z24$^{*/*}$ cells, both in 2i and at d3 of differentiation (**Fig. 13A and 13B**). IAP elements represented 8 and 19% of the total hypomethylated repeats in 2i and at d3, respectively (**Fig. 13A and 13B**). LINE1 elements were

mostly represented among hypomethylated repeats, accounting for 69 and 50% of all the hypomethylated repeats, in 2i and d3 respectively (**Fig. 13A and 13B**). LINE1 elements are characterized by the presence of specific classes of promoters, including LINE1A and LINE1T. In addition, these two families of LINE elements are known to be the most retrotransposition-competent LINEs (Sookdeo et al., 2013). LINE1A and LINE1T promoters were highly hypomethylated in 2i and, to a lesser extent, at d3 of differentiation (**Fig. 13C**). In contrast, we did not detect hypomethylation of SINE elements (**Fig. 13**).

3. ZBTB24 function is important for the recruitment of DNMTs

Our results highlighted an essential role of ZBTB24 to establish DNA methylation at minor satellite repeats (this study). In light of these data, we investigated the functional link between ZBTB24 and DNMT3B, the DNMT that specifically methylates minor satellite repeats (Okano et al., 1999; Velasco et al., 2010). Using ChIP experiment, we demonstrated that ZBTB24 binds to minor satellite repeats in ES cells and that its binding was significantly increased at d3 of differentiation (**Fig. 14A**), although ZBTB24 expression slightly decreased from 2i conditions to d3 of differentiation (**Fig. 14B**). We then performed ChIP experiments using an anti-DNMT3B antibody and we observed that the recruitment of DNMT3B drastically increased at d3 of differentiation in WT cells (**Fig. 14C**). In contrast, loss of ZBTB24 impaired DNMT3B recruitment at minor satellite repeats at d3 of differentiation in $Z24^{*/*}$ cells (**Fig. 14C**). The enrichment of DNMT3B at minor satellite repeats correlated with its increased expression in WT cells. In contrast, although DNMT3B expression in $Z24^{*/*}$ cells was comparable to WT cells, DNMT3B was not recruited at minor satellite repeats at d3 of differentiation (**Fig. 7**). These results strongly suggested that the binding of ZBTB24 at minor satellite repeats is required for DNMT3B recruitment and further establishment of DNA methylation at theses repeats. In contrast, although ZBTB24 binds to major satellite repeats at d3 of differentiation (**Fig. 14D**), ChIP experiments using an anti-DNMT3A antibody showed that presence of ZBTB24 was dispensable for the recruitment of DNMT3A, the DNMT that specifically methylate major satellite repeats during ES differentiation (**Fig. 14E**). Of note, as previously mentioned, DNMT3A was expressed at the same levels in WT and $Z24^{*/*}$ cells during differentiation (**Fig. 7**).

All together, these findings suggested that ZBTB24 binds directly to minor and major satellite repeats. However, loss of ZBTB24 functions impairs the establishment of DNA methylation at minor satellite repeats but not at major satellite repeats. Our results support a role of ZBTB24 in mediating the recruitment of DNMT3B at centromeric repeats.

To assess a broader role of ZBTB24 in mediating the recruitment of DNMTs on the genome, we analysed the coverage between our datasets of hypomethylated regions in Z24$^{*/*}$ cells in 2i and d3 of differentiation and published RRBS data generated in knock-out DNMT3A or DNMT3B embryos at 8.5 dpc (Auclair et al., 2014). Although ES cells in 2i or at d3 of differentiation are not truly comparable to specific *in vivo* developmental stages, this preliminary analysis suggested a stronger functional link between ZBTB24 and DNMT3B than with DNMT3A (**Fig. 15A**). In particular, 100% of the hypomethylated LINE1T elements were commonly affected in Z24$^{*/*}$ cells and DNMT3B knock-out embryos, whereas only 55% were shared with Dnmt3a knock-out embryos (**Fig. 15A**). Using ChIP assays, we further showed that, in WT cells, both ZBTB24 and DNMT3B binds to LINE1T promoter at d3 of differentiation, whereas we could not detect DNMT3A at this region (**Fig. 15B and 15C**). Importantly, 50% of DNMT3B recruitment was lost in Z24$^{*/*}$ cells thus reinforcing the fact that ZBTB24 contributes to DNMT3B recruitment at LINE1T promoter to establish DNA methylation (**Fig. 15C**).

As a whole, our results revealed that ZBTB24 protein contributes to the control of DNA methylation at DNA repeats and at certain imprinted loci. In particular, ZBTB24 seems to be required for the recruitment of DNMT3B at LINE1T promoter to establish DNA methylation.

4. ZBTB24 controls DNA methylation at interspersed DNA repeats LINE1 and IAP by promoting H3K9me3 establishment

Since DNA methylation of a large proportion of interspersed repeats is dependent on pre-existing H3K9me3 (Rose and Klose, 2014), it raises the question of an indirect role of ZBTB24 in the establishment of DNA methylation, by modulating the establishment of H3K9me3 interspersed repeats. We then decided to focus on LINE elements because our ChIP experiments showed that ZBTB24 binds LINE1A promoter in 2i and to both LINE1A and LINE1T promoters at d3 of differentiation (**Fig. 16A**). In contrast, we did not detect ZBTB24 at IAP 5'UTR, 3'UTR and SNORD116 regions (**Fig. 16B**).

Using ChIP experiments we investigated the enrichment of H3K9me3 at DNA repeats in Z24$^{*/*}$ cells compared to WT cells in most of LINE1 elements, including ORF2 and 3'UTR, as well as specific promoter regions, like LINE1A and LINE1T promoters.

The H3K9me3 histone mark increased at LINE1 ORF2 and 3'UTR regions during the differentiation of WT but not of Z24$^{*/*}$ mES cells (**Fig. 17A**). Similarly, H3K9me3 levels increased at LINE1A and LINE1T promoters during differentiation of WT but not of Z24$^{*/*}$ mES cells (**Fig. 17B**). We extended the analysis of H3K9me3 levels to 5'UTR and 3'UTR of IAP elements and, although no significant difference was detected at 5'UTR region, H3K9me3 did not increase at

3'UTR region during of Z24$^{*/*}$ cells (**Fig. 17C**). Similarly to DNA repeats, loss of H3K9me3 levels at SNORD116 was observed between WT and Z24$^{*/*}$ cells at d3 of differentiation (**Fig. 17D**). In contrast, at ERVL repeats, which are not hypomethylated during differentiation of Z24$^{*/*}$ mES cells, we observed H3K9me3 levels comparable to WT mES cells (**Fig. 17D**). For the majority of genomic loci analysed, we observed that H3K9me3 levels increased in absence of DNA methylation (TKO cells) (**Fig. 17**).

All together, these findings suggested that ZBTB24 favours the deposition of H3K9me at repetitive elements. One possible scenario could be that ZBTB24 directly binds to these repeats, like LINE1A and LINE1T promoter and allows the recruitment of the H3K9 methylation enzymatic machinery, like SUV39H1, directly to their genomic targets. Although ZBTB24 does not directly bind to 5'UTR and 3'UTR regions of IAP elements, could be important for the recruitment of SETDB1 at their genomic targets. To test this model, we performed ChIP experiment that showed an enrichment of SUV39H1 at LINE1A and LINE1T promoters in 2i and at d3 of differentiation in WT cells (**Fig. 18A**). In contrast, the recruitment of SUV39H1 was significantly decreased in Z24$^{*/*}$ cells at d3 at both LINE1A and LINE1T promoters (**Fig. 18A**), although its expression was not affected in Z24$^{*/*}$ (**Fig. 18B**). ChIP using an anti-SETDB1 antibody showed that SETDB1 was detectable at IAP 5'UTR and IAP 3'UTR, and its recruitment did not increase during differentiation in WT cells (**Fig. 18C**). We also detected SETDB1 at SNORD116 region and its recruitment significantly increased at d3 in WT cells (**Fig. 18C**). In Z24$^{*/*}$ cells, although we did not observe any difference in SETDB1 recruitment at IAP 5'UTR, we observed that SETDB1 recruitment was completely lost in 2i at IAP 3'UTR. The loss of SETDB1 recruitment was also observed at SNORD116 non-coding RNAs cluster at d3 of differentiation in Z24$^{*/*}$ mES cells (**Fig. 18C**), although the expression of SETDB1 was not affected by the loss of ZBTB24 (**Fig. 18D**).

Importantly, our results indicate that ZBTB24 could favour the deposition of H3K9me3 at interspersed repeats, including LINE1 and IAP elements, and at the SNORD116 non-coding cluster and thus influence the establishment of DNA methylation at these regions. Altogether, our study strongly suggests a dual role of ZBTB24 in the establishment of DNA methylation at genomic loci by direct and indirect mechanisms: i) ZBTB24 directly mediates the recruitment of DNMT3B at centromeric minor satellite repeats and at LINE1T promoter. ii) ZBTB24 modulates the establishment of H3K9me3 at interspersed repeats, including LINE1A and LINE1T promoters and IAP 3'UTR region, and at SNORD116 non-coding RNA cluster, favouring the recruitment of SUV39H1 and SETDB1, indirectly promoting the establishment of DNA methylation.

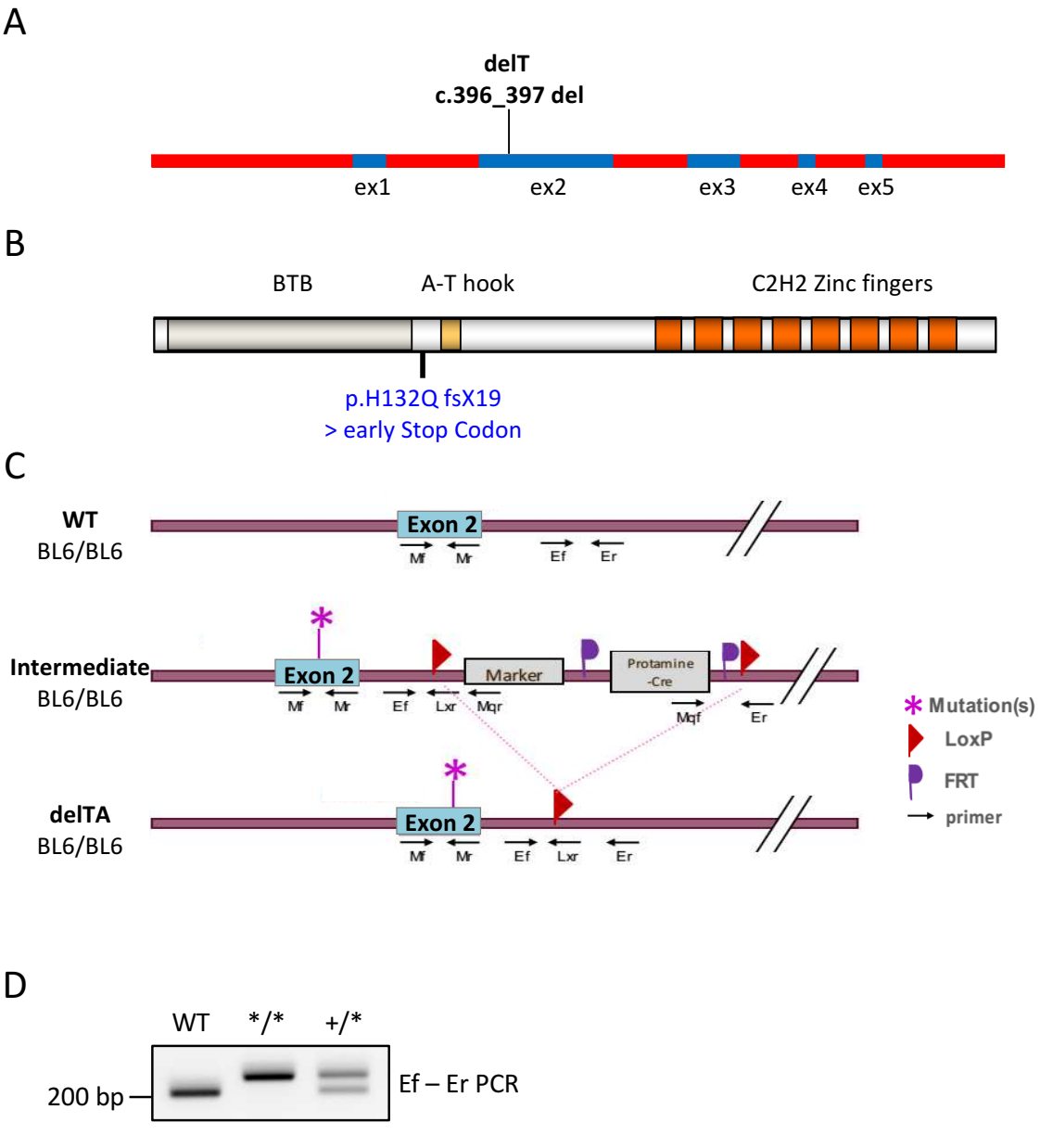


Figure 1. Generation of the ZBTB24 mutant mouse. *(A)* A Map of the ZBTB24 mutant allele, showing the position and the type of mutation. *(B)* Representation of ZBTB24 functional protein domains: BTB, AT-hook, and the C2H2 zinc finger domains. The position of the introduced mutation is indicated. *(C)* Schematic representation of the WT and mutated ZBTB24 allele. The genetic background, the position of the loxP sites, the mutation, and the primers used for the genotyping of embryos are indicated. *(D)* An example of agarose gel electrophoresis photo showing the different PCR products (Ef-Er primers) and corresponding genotypes. Molecular weight is shown next to the gel.

A

Stage (dpc)	WT	Z24$^{+/*}$	Z24$^{*/*}$	total
3.5	2 (16.5%)	8 (67%)	2 (16.5%)	12
6.5	1 (17%)	2 (33%)	3 (50%)	6
7.5	4 (20%)	11 (55%)	5 (25%)	20
8.5	10 (24%)	22 (52%)	10 (24%)	42
9.5	9 (45%)	10 (50%)	1 (5%)	20
11.5	1 (12,5%)	2 (25%)	5 (62,5%)	8
13	4 (40%)	6 (60%)	0 (0%)	10
18	4 (50%)	4 (50%)	0 (0%)	8
Neonate	35 (34 %)	67 (66 %)	0 (0%)	102

B

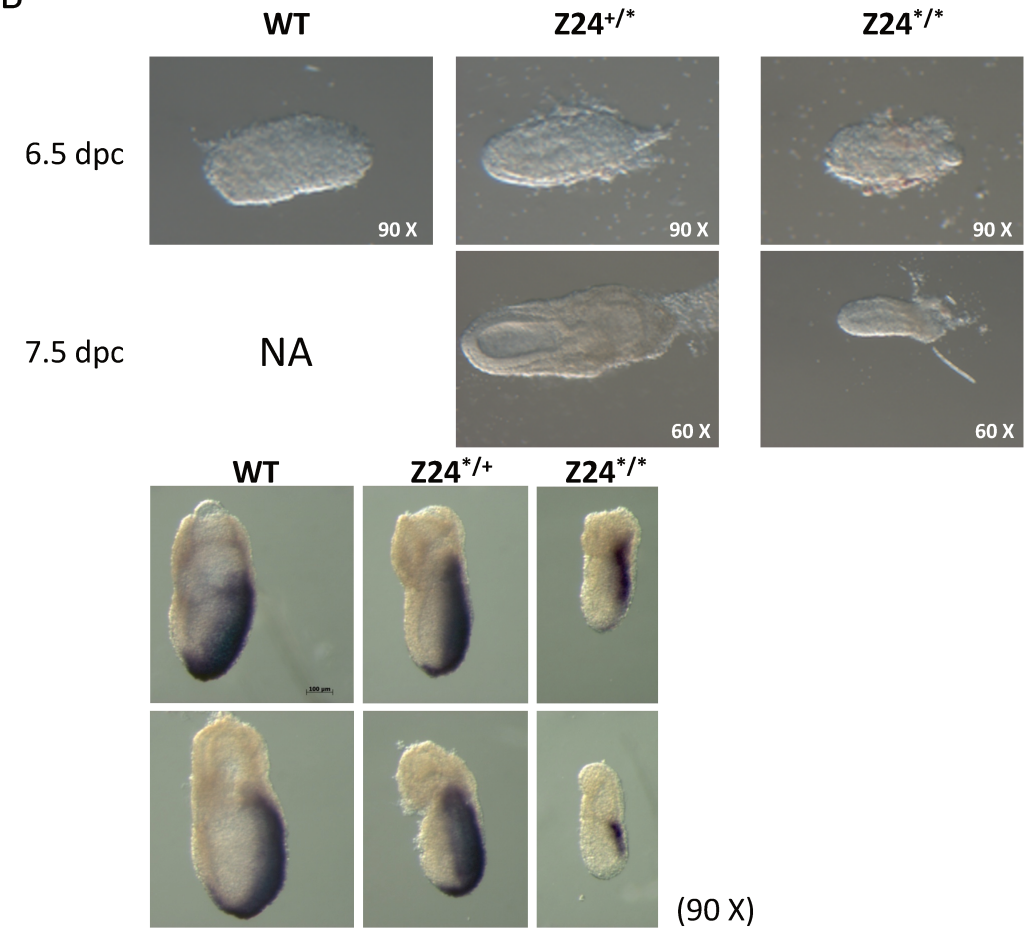


Figure 2. Loss of ZBTB24 functions lead to developmental delay from 6.5 dpc and embryonic lethality. *(A)* Table summarising the number of WT, Z24+/* and Z24*/* mice observed at the different embryonic day The total of observed embryos is also indicated. *(B)* Pictures of WT, Z24+/* and Z24*/* embryos at post-implantation stages 6.5 and 7.5 dpc (Top panels). Expression of BRACHYURY (blue staining) in WT, Z24+/* and Z24*/* embryos at 7.5 dpc, analysed by whole-mount *in situ* hybridization (Bottom panels) . WT embryos at 7.5 dpc were not available (NA).

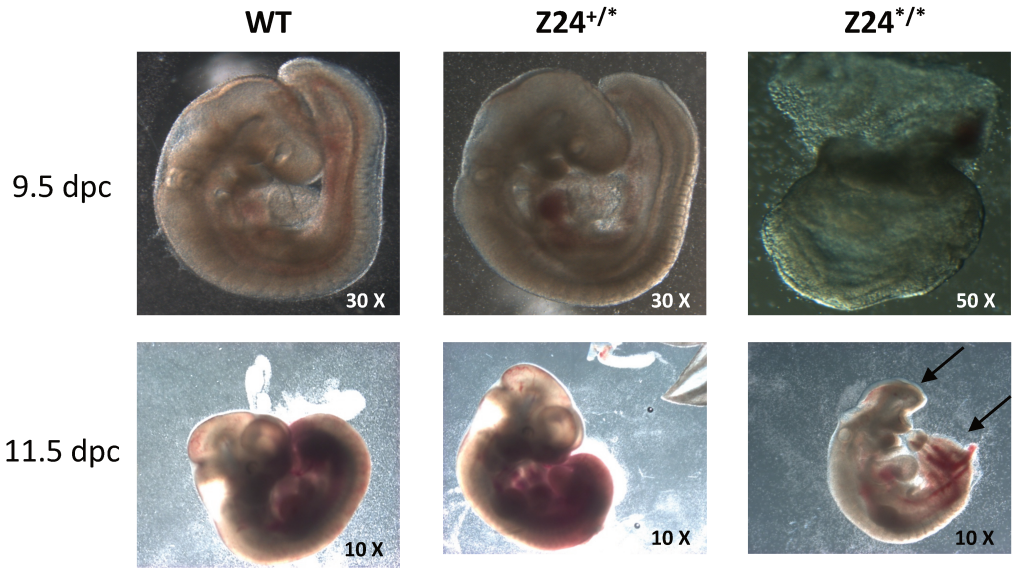
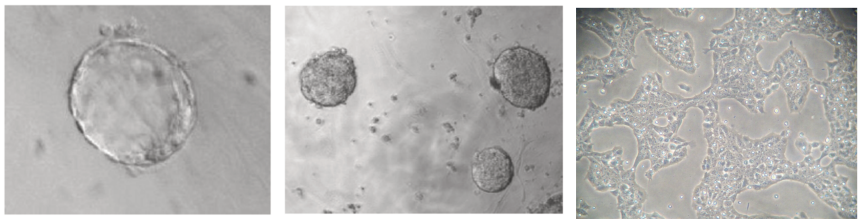


Figure 3. Loss of ZBTB24 functions lead to developmental delay and morphological defects. Pictures of WT, Z24+/* and Z24*/* embryos at 9.5 and 11.5 dpc. The arrows point to the head and limb buds malformations observed in Z24*/* embryos at E11.5.

A



B

Genotype	Males	Females
WT	2	0
Z24$^{+/*}$	5	3
Z24$^{*/*}$	2	0

C

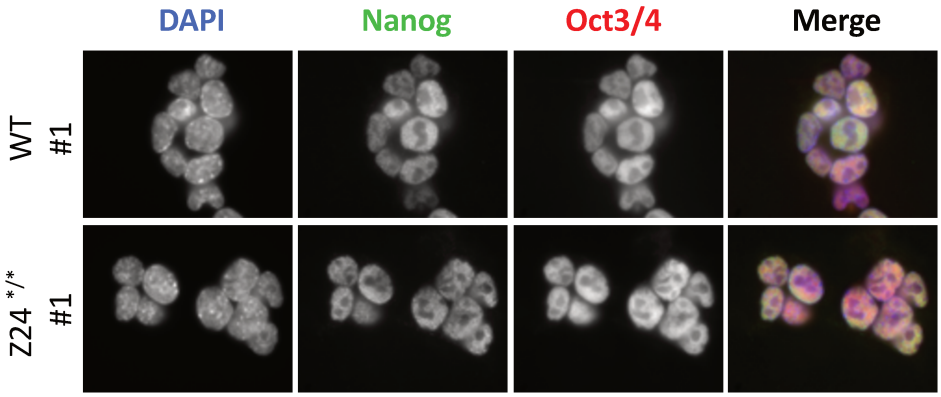


Figure 4. Derivation and characterisation of ZBTB24 mutant mouse embryonic stem cells. *(A)* Blastocysts were isolated at 3.5 dpc. Z24$^{*/*}$ appeared undistinguishable from WT and Z24$^{+/*}$ blastocysts of the same litter. We maintained mES cells in culture in 2i condition, in absence of feeder cells. *(B)* Table summarising the mES clones obtained. The genotype and the sex are also shown. *(C)* Homogeneous expression of the pluripotency factors NANOG and OCT3/4 in WT clone #1 and Z24$^{*/*}$ clone #1 cells, analysed by immunofluorescence.

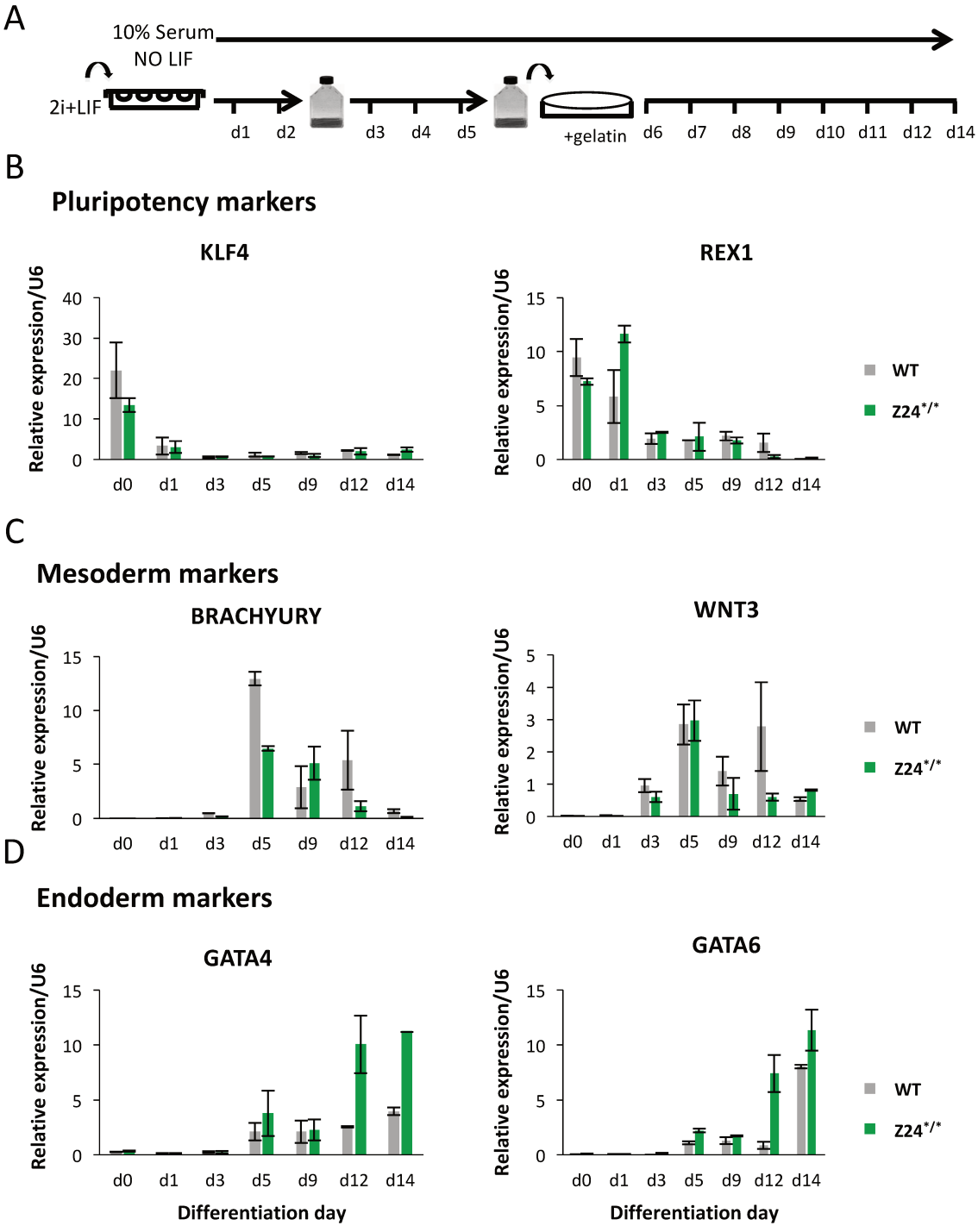


Figure 5. ZBTB24 functions are dispensable for *in vitro* differentiation in mesoderm and endoderm cells. *(A)* Schematic overview of the embryoid bodies differentiation protocol, using the hanging-drop technique *(B, C, D)* Expression levels of the pluripotency factors KLF4, REX1, NANOG and OCT3/4, of the mesoderm markers BRACHYURY and WNT3 and of the endoderm markers GATA4 and GATA6 were assessed by RT-qPCR, normlised to U6 small nuclear RNA. The graphs represent the mean of three experiments using WT clone #1 and Z24$^{*/*}$ clone #1 cells. Error bars represent the standard error of the mean (SEM).

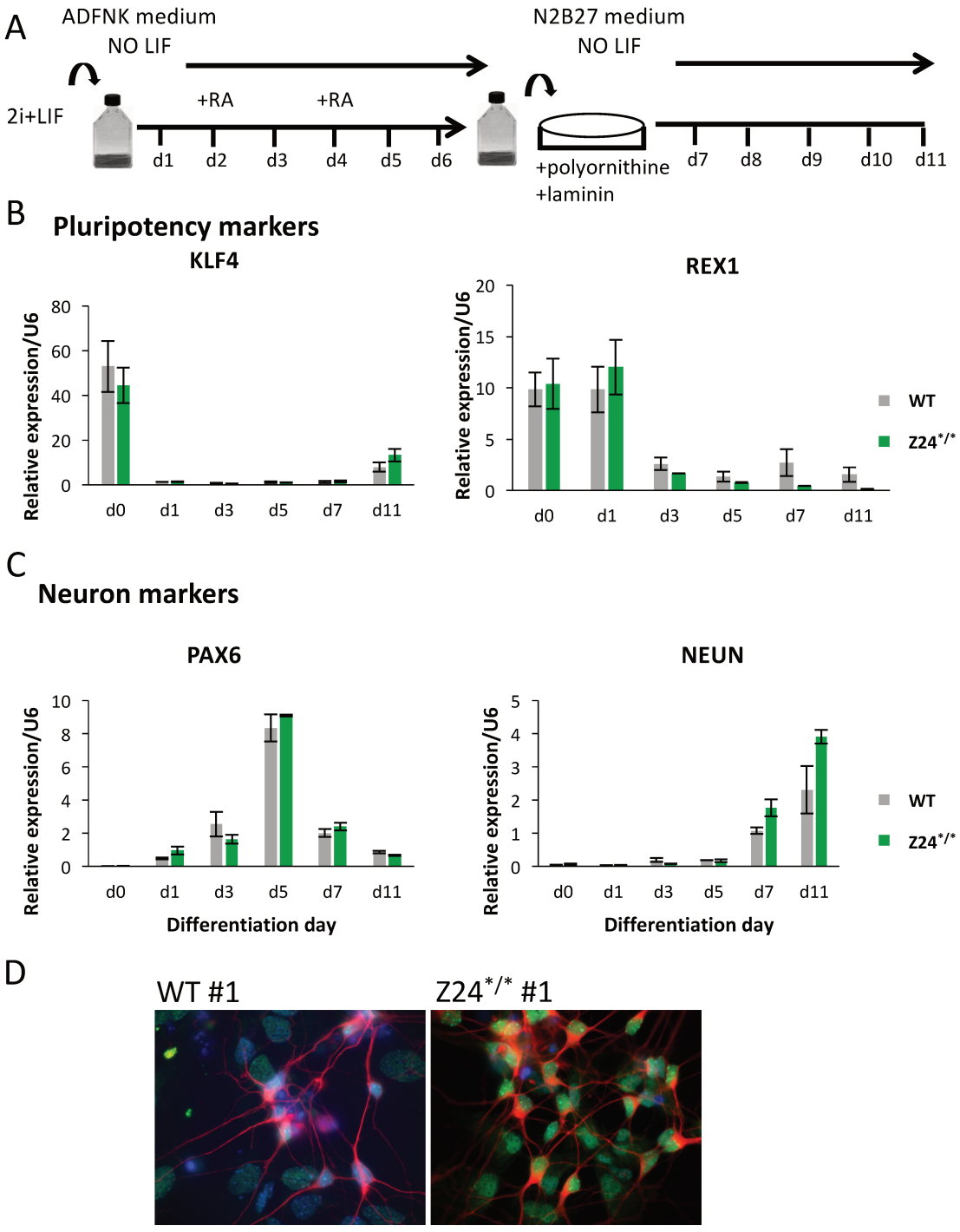


Figure 6. ZBTB24 functions are dispensable for *in vitro* differentiation in cortical neurons. *(A)* Schematic overview of the neuron specific differentiation protocol. *(B)* and *(C)* Expression levels of the pluripotency markers KLF4, REX1 and the neurons markers PAX6 and NEUN were assessed by RT-qPCR, normalised to U6 small nuclear RNA. The graphs represent the mean of three experiments using WT clone #1 and Z24$^{*/*}$ clone #1 cells. Error bars represent the standard error of the mean (SEM). *(D)* Expression pattern of heterochromatin marker H3K9me3 (green) and the neuron-specific marker **β**III-tubulin (red) were assessed by immunofluorescence in WT clone #1 and Z24*/* clone #1 neurons at day 11 of differentiation.

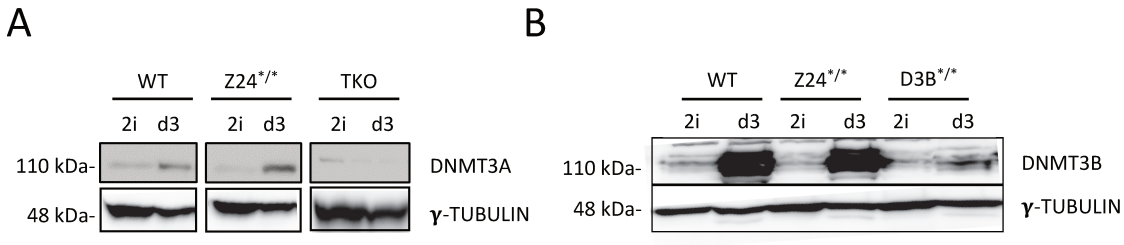


Figure 7. The loss of ZBTB24 functions does not impact the expression of DNMT3A and DNMT3B during mES differentiation. The expression of *(A)* DNMT3A and *(B) DNMT3B*, in 2i condition and after 3 days of differentiation (d3) in WT and Z24$^{*/*}$ cells, analysed by western blot. **γ**-tubulin was used as a loading control. A molecular weight scale in indicated on the left of the panel.

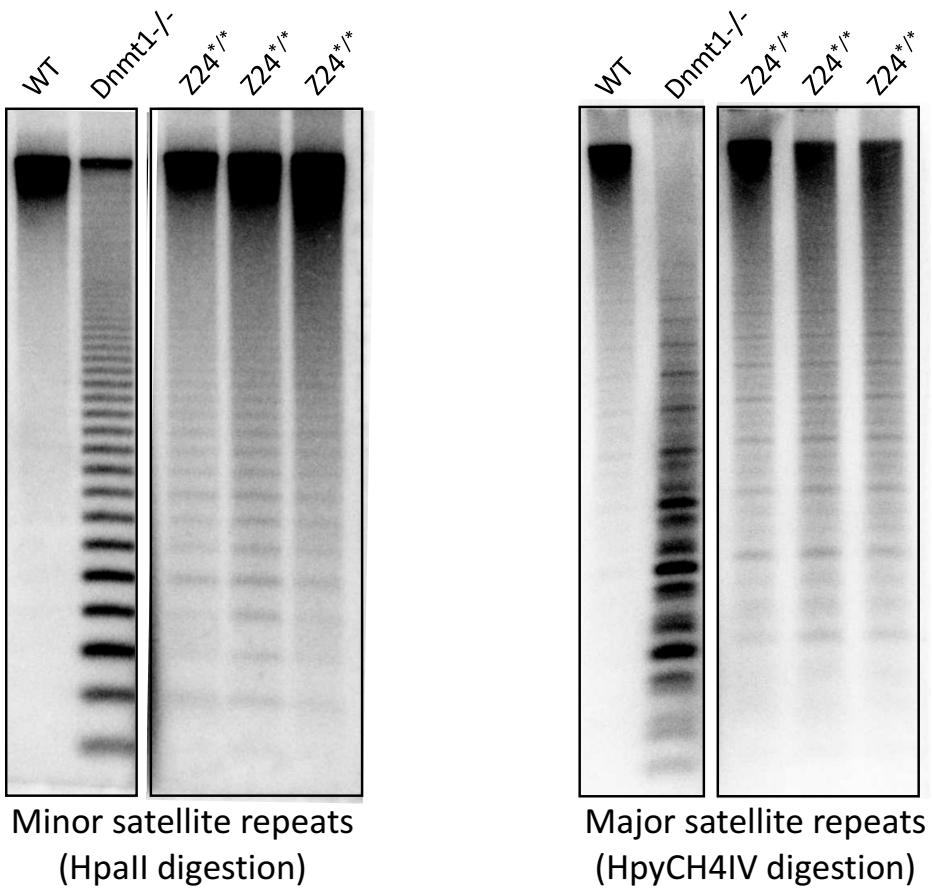


Figure 8. Loss of ZBTB24 functions impairs DNA methylation at centromeric and pericentromeric repeats. Southern blot analysis of DNA methylation at minor satellite (HpaII) and major satellite (HpyCH4IV) repeats on genomic DNA extracted from $Z24^{*/*}$ embryos at 11.5 dpc. Genomic DNA from WT and DNMT1 knock-out primary MEF were used as negative and positive controls, respectively.

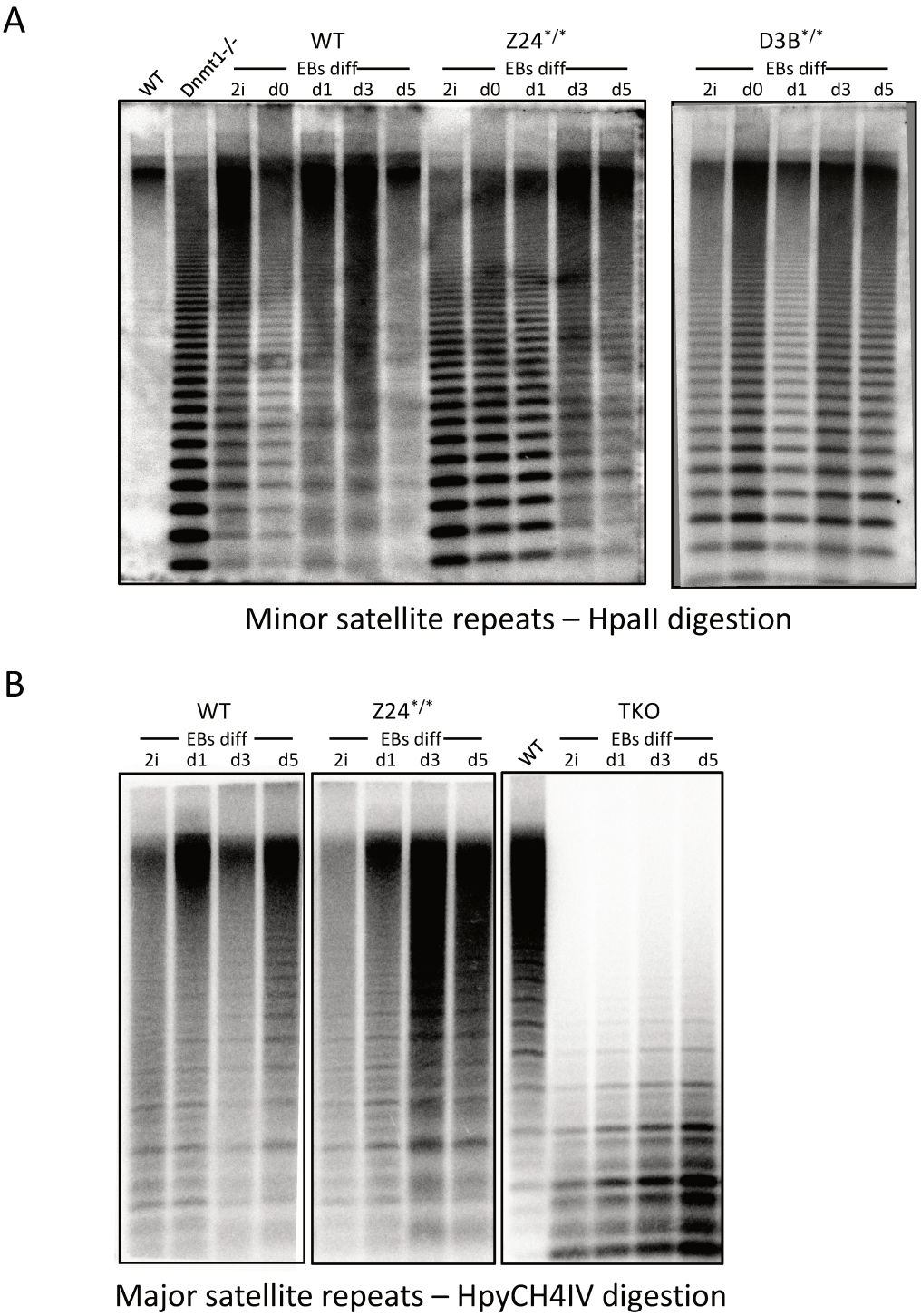
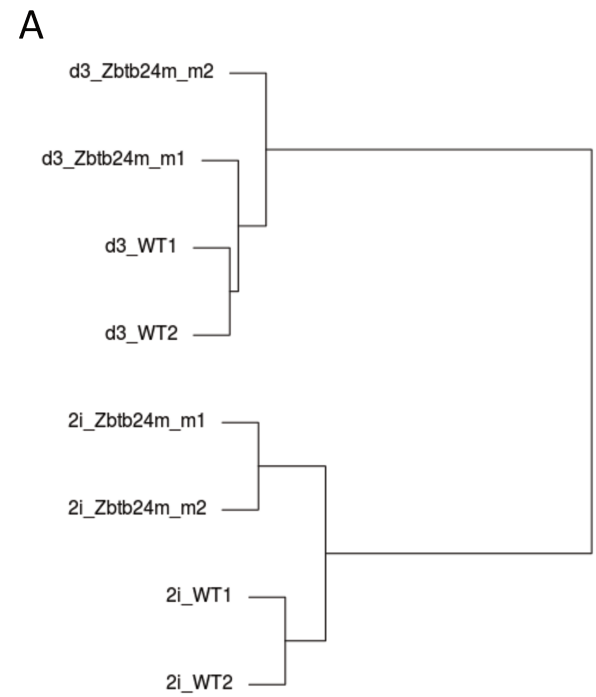


Figure 9. ZBTB24 functions are partially required for the establishment of DNA methylation at centromeric repeats but are dispensable at pericentromeric repeats Southern blot analysis of DNA methylation at *(A)* minor satellite (HpaII) and *(B)* major satellite (HpyCH4IV) repeats on genomic DNA extracted from WT, Z24$^{*/*}$ and D3B$^{*/*}$ or DNMT triple knock-out (TKO) mES during embryoid bodies differentiation (EBs). Genomic DNA from WT and Dnmt1 knock-out primary MEF were used as negative and positive controls, respectively.

Sample	Z24$^{*/*}$ #2 d3	Z24$^{*/*}$ #1 d3	WT #2 d3	WT #1 d3	Z24$^{*/*}$ #2 2i	Z24$^{*/*}$ #1 2i	WT #2 2i	WT #1 2i
Pair-end reads	35,994,541	32,099,618	40,274,608	41,150,098	31,848,001	30,853,286	38,076,349	34,416,403
Mapping efficinecy (%)	86.92	87.61	88.52	85.03	91.81	91.89	89.21	87.66
Reads with unique hints (%)	69.9	70.64	72.44	69.01	73.96	76.61	73.9	71.25
Conversion rate (%)	99.13	98.98	98.92	98.9	99.86	99.73	99.78	99.77
Median Sequencing depth	71x	65x	79x	79x	76x	67x	80x	71x
CpGx x1	1,749,283	1,723,732	1,729,088	1,770,798	1,556,658	1,475,544	1,488,889	1,517,320
CpGs x8	1,439,007	1,405,011	1,394,335	1,432,585	1,319,954	1,196,604	1,187,882	1,210,928

Table 1. Features of RRBS methylomes in WT and Z24$^{*/*}$ mES in 2i and after 3 days of embryoid bodies differentiation. Table summarising the sequenced RRBS libraries. The last two columns indicate the number of CpGs sequenced at least 1x and 8x obtained from reads that mapped uniquely in the genome. The bisulfite conversion rate was estimated by calculating the C to T conversion at non-CpG sites.



C

	Z24*/* vs WT	
	2i	d3
Total DMRs	10,947	199
Hypomethylated regions	10,940	195
Hypermethylated regios	7	4

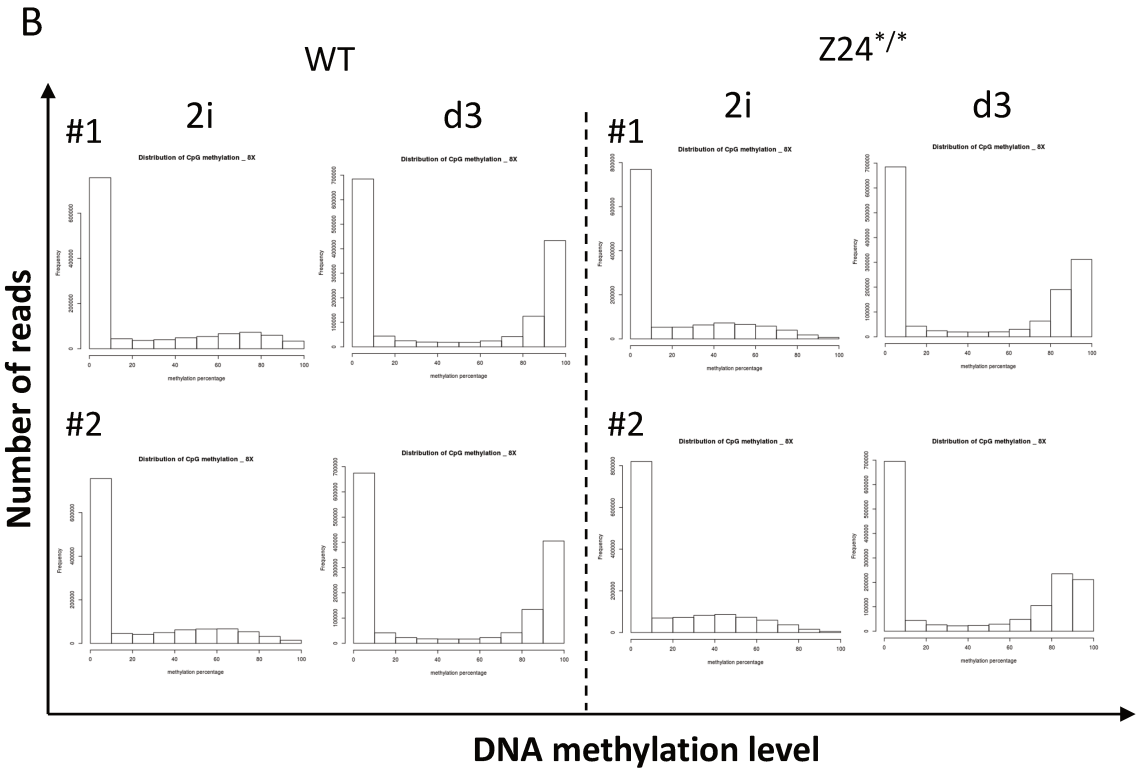
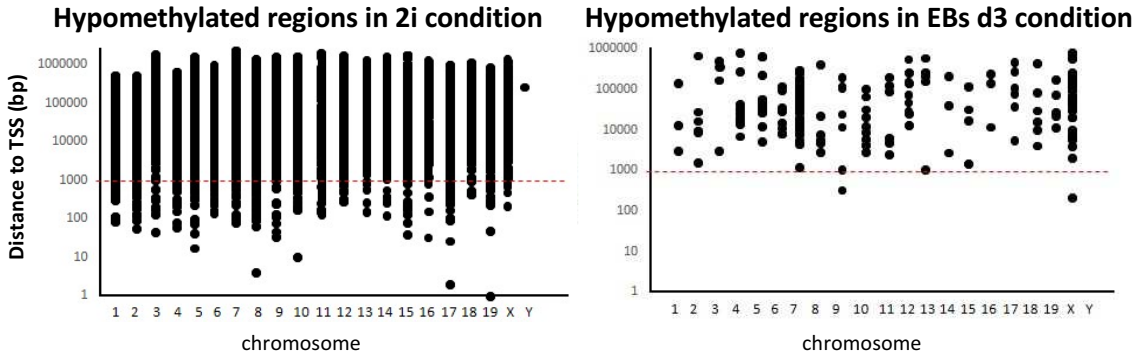


Figure 10. Loss of ZBTB24 functions affects DNA methylation mostly in mouse embryonic stem cells. *(A)* Clustering of WT and Z24$^{*/*}$ mES cells (clones #1 and #2) in 2i and after 3 days of differentiation (d3), based on DNA methylation values originated by RRBS. *(B)* Cumulative frequency plots showing the DNA methylation level of sequenced regions for each WT and Z24$^{*/*}$ clone in 2i condition and at day 3 of differentiation (d3). The DNA methylation level is represented on the X-axis and the frequencies (counts) on the Y-axis. *(C)* Table summarising the number of differentially methylated regions (DMRs). The number of hypomethylated and hypermethylated regions is also showed.

A

	2i (Z24$^{*/*}$ vs WT)	d3 (Z24$^{*/*}$ vs WT)
Hypomethylated regions in CpG island	13	2

B



C

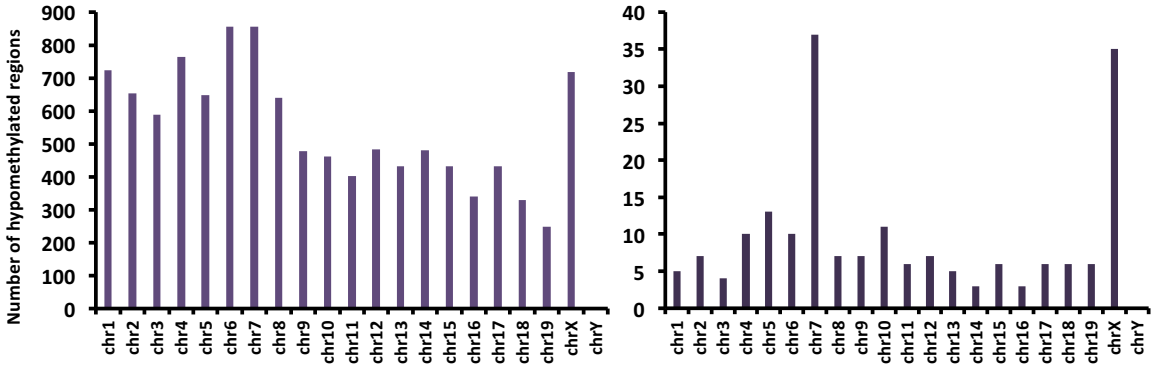


Figure 11. Genomic and chromosomic features of hypomethylated regions. *(A)* Table summarising the number of hypomethylated regions in CpG islands in 2i and at day 3 of differentiation in WT and Z24$^{*/*}$ cells. *(B)* Dot plot representing the distance of each hypomethylated region from the nearest TSS on each chromosome in 2i and at day 3 of differentiation (d3). The dashed line represent 1000 bp distance from TSS. *(C)* chromosomic repartition of hypomethylated regions in 2i and at d3.

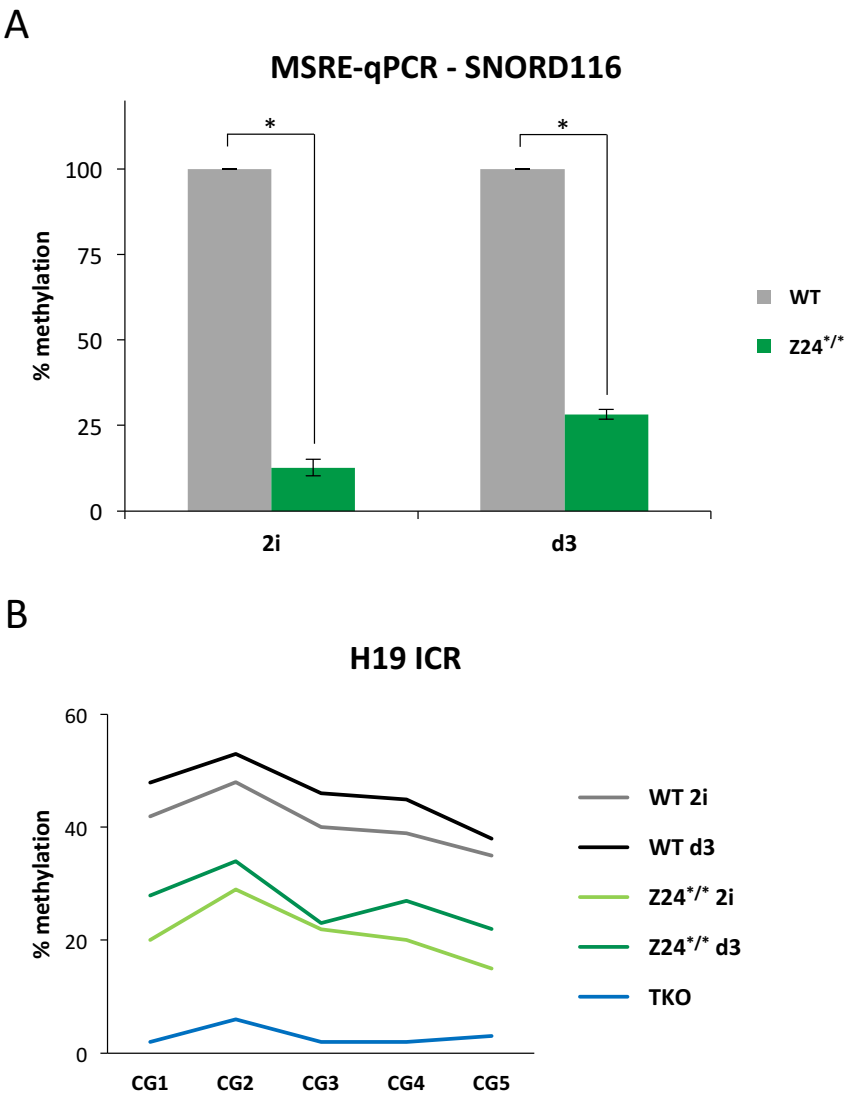
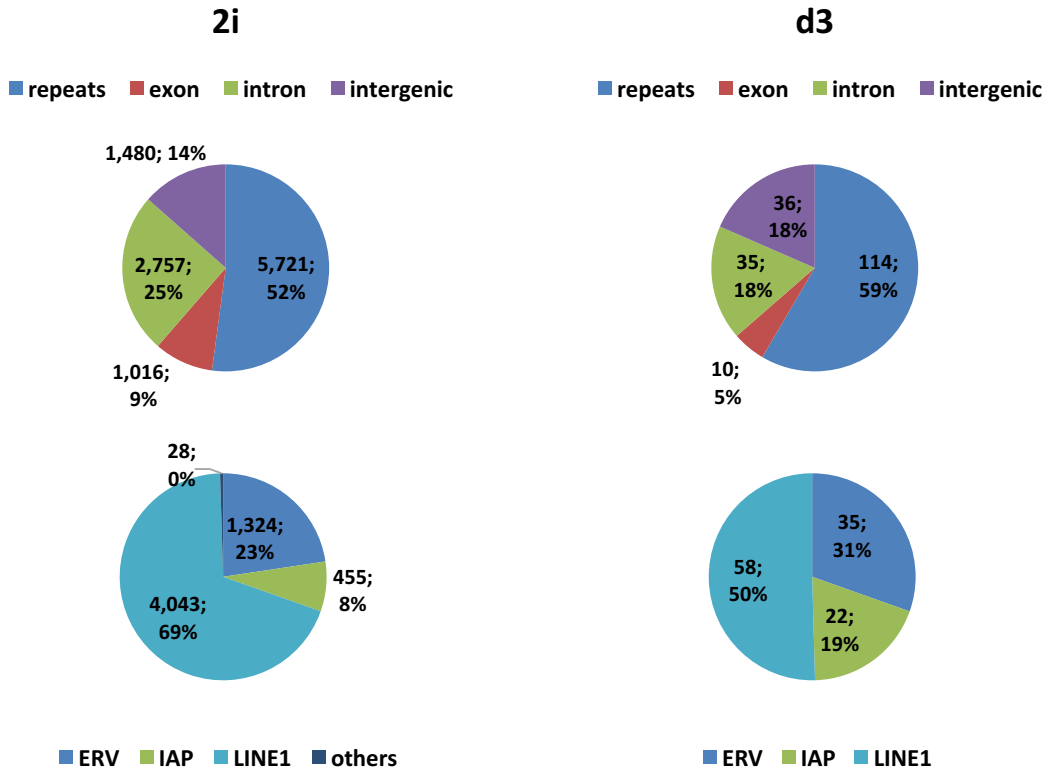


Figure 12. Loss of ZBTB24 functions impaired DNA methylation at the imprinted loci. *(A)* DNA methylation levels were assessed by MSRE-qPCR at the cluster of snoRNA SNORD116 . Error bars represent the standard error of the mean (SEM). p-values (Mann-Whitney Test) comparing WT and Z24$^{*/*}$: *p < 0.05 *(B)* DNA methylation levels of CpGs within H19 ICR region using pyrosequencing. The methylation levels at each CpG were measured in WT and Z24$^{*/*}$ in 2i and at day 3 of differentiation (d3). TKO mES cells were used as negative control.

A



C

	LINE1	LINE1A	LINE1F	LINE1G	LINE1T	IAP	ERV	no repeats	others	total
2i	294	830	2,362	43	495	452	1,322	5,111	31	10,940
d3	6	15	28	0	9	21	35	80	0	194

Figure 13. Loss of ZBTB24 functions impair DNA methylation mainly at DNA repeats. *(A)* Pie charts representing the genomic repartition of hypomethylated regions in Z24$^{*/*}$ cells in 2i and at say 3 of differentiation (d3). DNA repeats was further divided into the main classes of repetitive elements *(B)*. *(C)* table summarising the total number of hypomethylated regions in each class of repetitive elements. The total number of hypomethylated regions as well as the number in non-repetitive elements are also shown.

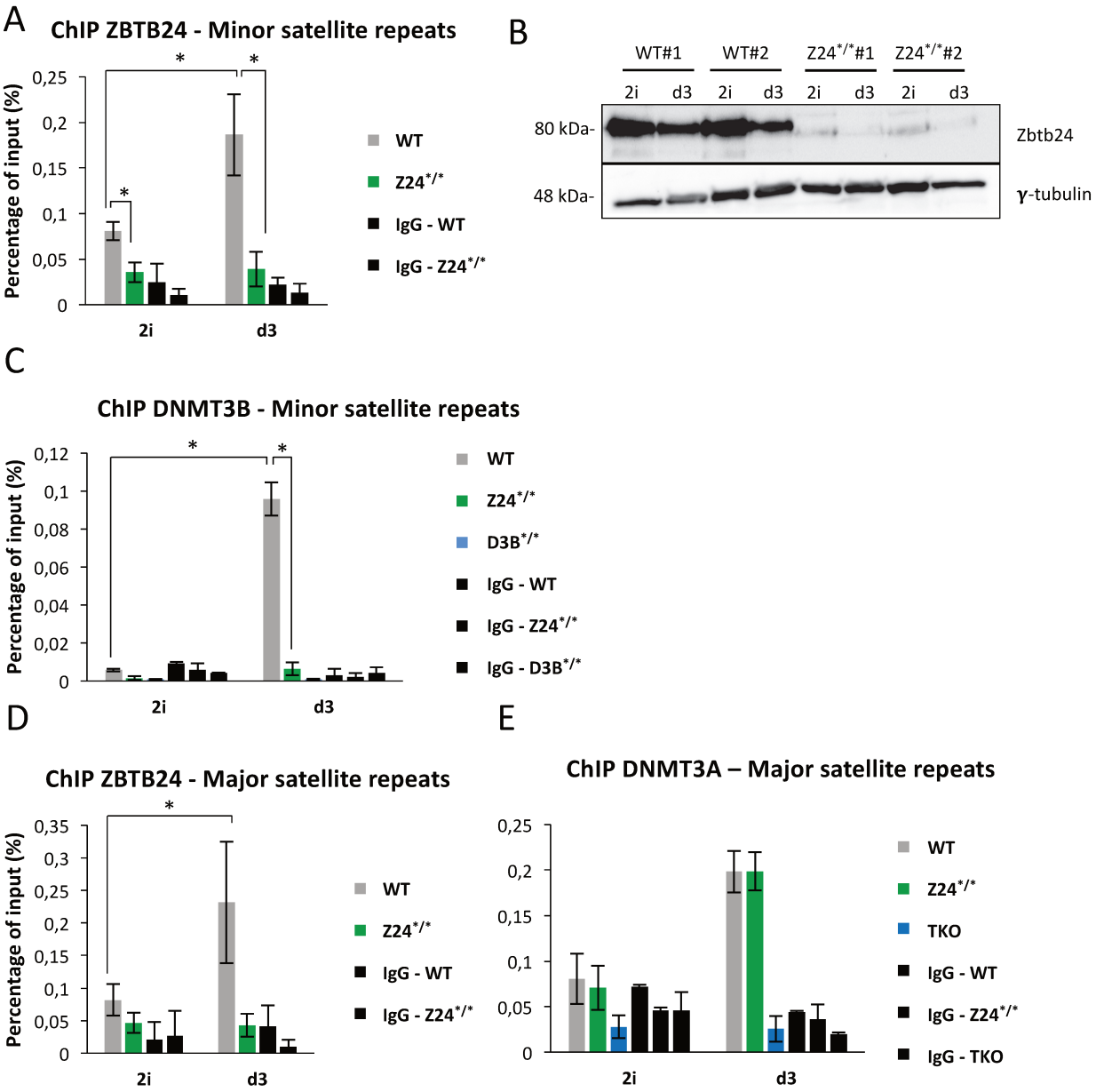


Figure 14. ZBTB24 functions are required for the the recruitment of DNMT3B at minor satellites but are dispensable for DNMT3A recruitment at major satellites. ChIP-qPCR was performed to analyse *(A)* the enrichment of ZBTB24 at minor satellite repeats; *(D)* the enrichment of ZBTB24 at major satellite repeats; *(C)* the enrichment of DNMT3B at minor satellite repeats; *(E)* the enrichment of DNMT3A at major satellite repeats. Data were normalized to the input. The graphs represent the mean of two experiments using both WT clones #1 and #2 and Z24$^{*/*}$ clones #1 and #2. Error bars represent the standard error of the mean (SEM). p-values (Mann-Whitney Test) comparing WT and Z24$^{*/*}$: *p < 0.05. The expression of *(B)* ZBTB24, in 2i condition and after 3 days of differentiation (d3) in WT and Z24$^{*/*}$ cells, analysed by western blot. **γ**-tubulin was used as a loading control. A molecular weight scale in indicated on the left of the panel.

A

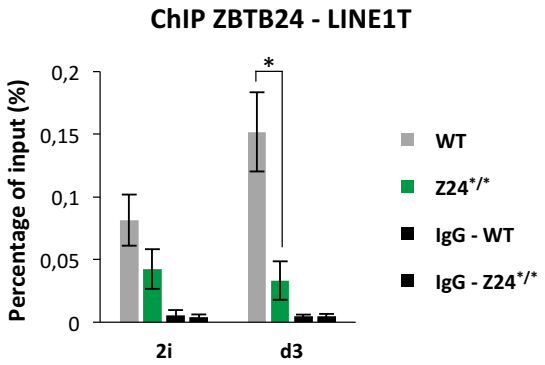
Coverage with DNMT3AKO (8.5 dpc)

	LINE1	LINE1A	LINE1F	LINE1G	LINE1T	IAP	ERV	no repeats	others	total
2i vs 3aKO	76 (25.9%)	234 (28.2%)	432 (18.3%)	11 (25.6%)	126 (25.5%)	48 (10.4%)	253 (19.3%)	1,873 (37.4%)	10 (32.2%)	3,063 (28%)
d3 vs 3aKO	4 (66.7%)	2 (13.3%)	12 (42.9%)	0	5 (55.6%)	4 (19%)	5 (14.3%)	45 (56.3%)	0	77 (39.7%)

Coverage with DNMT3BKO (8.5 dpc)

	LINE1	LINE1A	LINE1F	LINE1G	LINE1T	IAP	ERV	no repeats	others	Total
2i vs 3bKO	251 (85.4%)	708 (85.3%)	1,937 (82%)	43 (74.4%)	411 (83%)	173 (38.3%)	901 (68.2%)	4,239 (84.6%)	27 (87%)	8,679 (79.3%)
d3 vs 3bKO	6 (100%)	12 (80%)	24 (85.7%)	0	9 (100%)	12 (57.1%)	23 (65.7%)	70 (87.5%)	0	156 (80.4%)

B



C

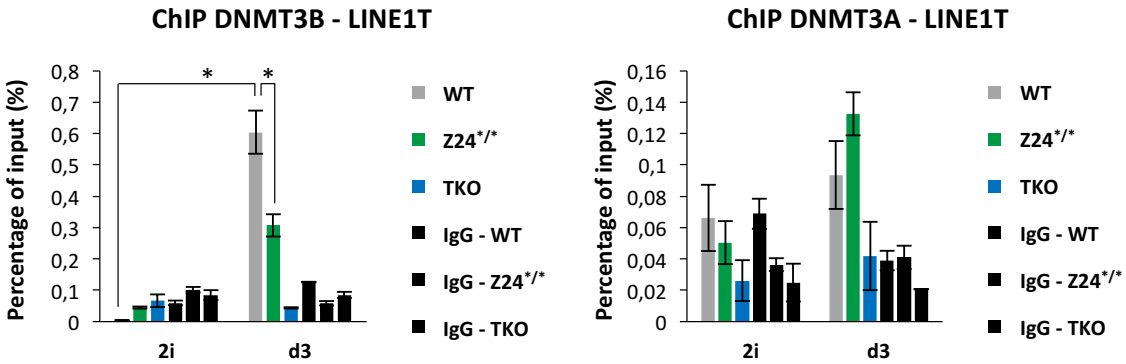


Figure 15. ZBTB24 functions are partially required for DNMT3B recruitment at LINE1T promoter. *(A)* tables summarising the results of the coverage of the Z24$^{*/*}$ RRBS with published RRBS performed in embryos at 8.5 dpc DNMT3A and DNMT3B knock-out (Auclair et al., 2014). The percentage of the regions covered by the analysis is indicated for each category. *(B)* ZBTB24, *(C)* DNMT3B and DNMT3A enrichment at LINE1T promoter were analysed by ChIP-qPCR. Data were normalized to the input. The graphs represent the mean of two experiments using both WT clones #1 and #2 and Z24$^{*/*}$ clones #1 and #2. Error bars represent the standard error of the mean (SEM). p-values (Mann-Whitney Test) comparing WT and Z24$^{*/*}$: *p < 0.05.

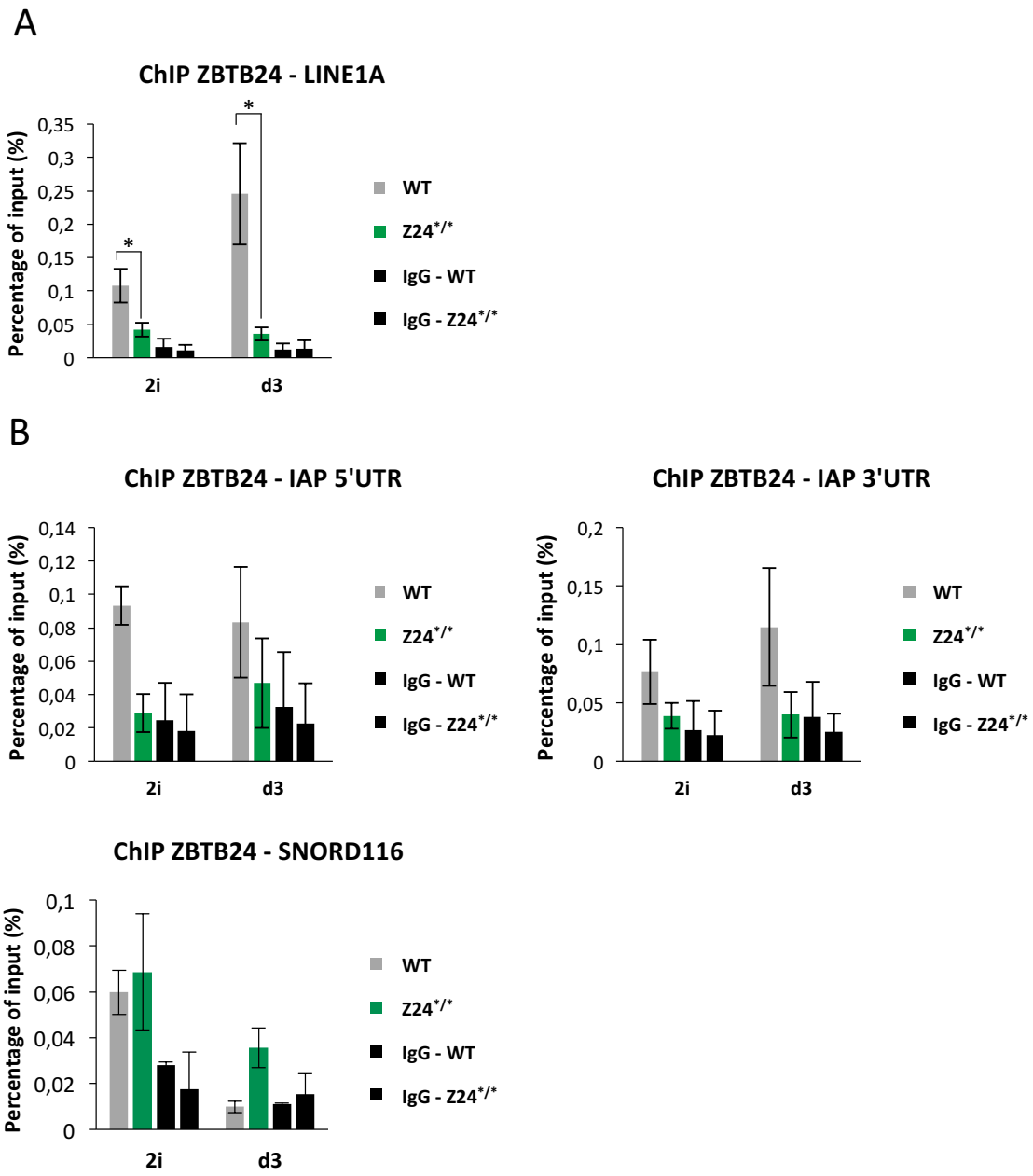


Figure 16. ZBTB24 binds directly to LINE1A and LINE1T promoters. ChIP-qPCR was performed to analyse ZBTB24 enrichment *(A)* at LINE1A, *(B)* IAP 5'UTR and 3'UTR and at SNORD116 cluster. Data were normalized to the input. The graphs represent the mean of two experiments using both WT clones #1 and #2 and Z24$^{*/*}$ clones #1 and #2. Error bars represent the standard error of the mean (SEM). p-values (Mann-Whitney Test) comparing WT and Z24$^{*/*}$: *$p < 0.05$.

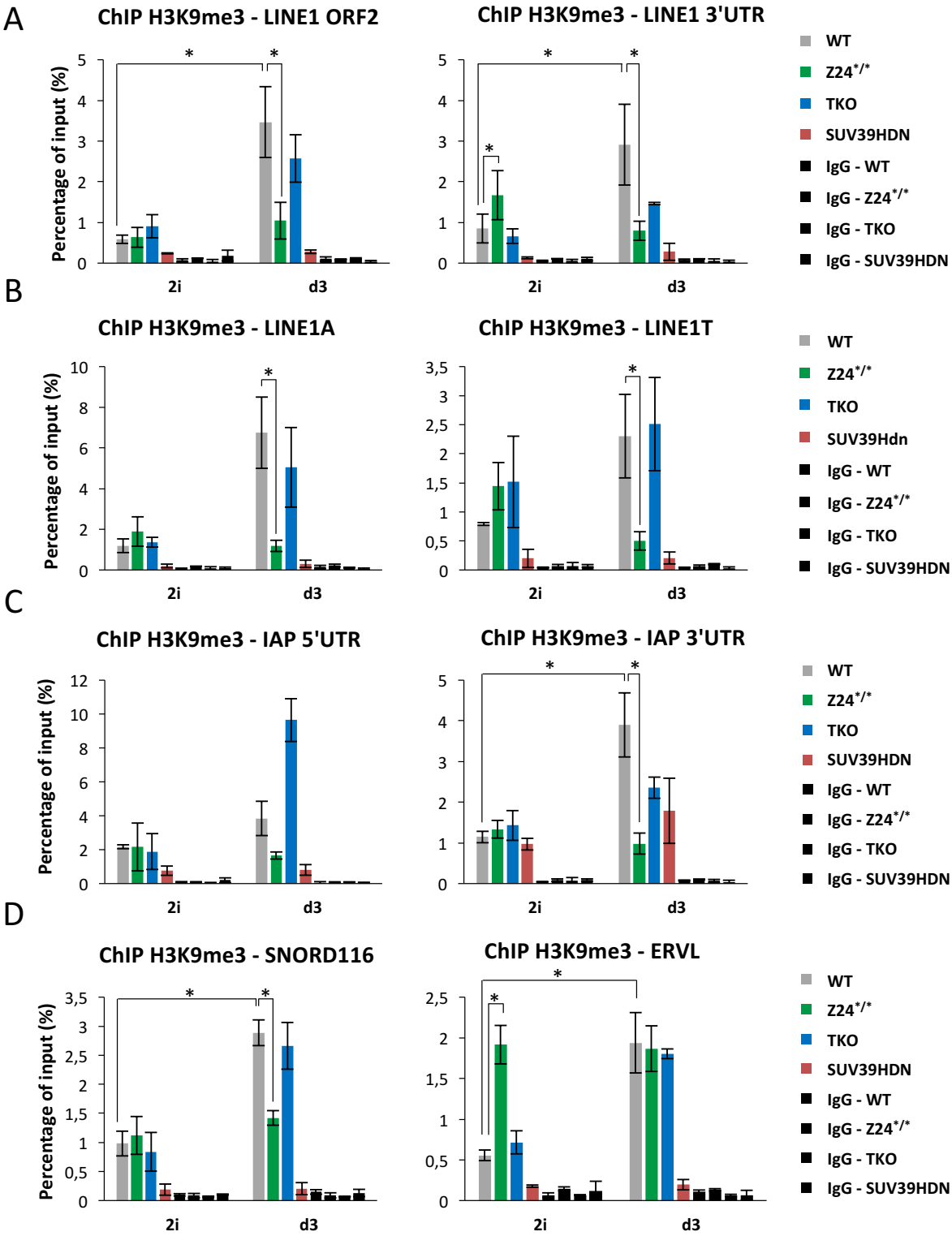


Figure 17. Loss of methylation at LINE1 and IAP repetitive elements and at the cluster of snoRNA SNORD116 correlates with a loss of H3K9me3, independently of absence of DNA methylation. H3K9me3 enrichment at *(A)* LINE1 ORF2 and 3'UTR regions, *(B)* LINE1A and LINE1T promoters, *(C)* IAP 5'UTR and 3'UTR and at *(D)* SNORD116 and ERVL regions was analysed by ChIP-qPCR. Data were normalized to the input. The graphs represent the mean of two experiments using both WT clones #1 and #2 and Z24$^{*/*}$ clones #1 and #2. Error bars represent the standard error of the mean (SEM). p-values (Mann-Whitney Test) comparing WT and Z24$^{*/*}$: *p < 0.05.

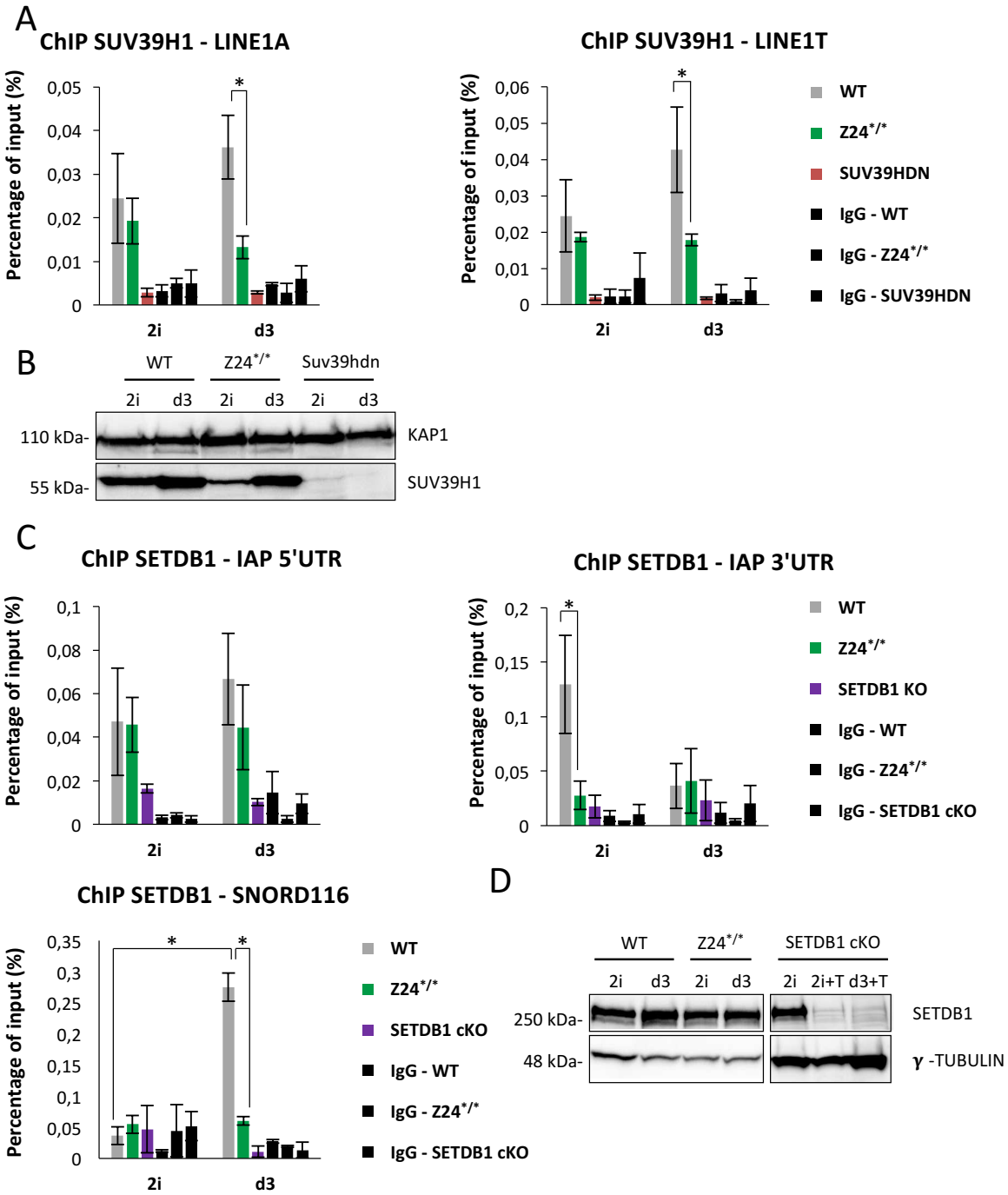


Figure 18. ZBTB24 is required to recruit SUV39H1 and SETDB1 at DNA repeats. *(A)* SUV39H1 enrichment at LINE1A and LINE1T promoters, was analysed by ChIP-qPCR. Data were normalized to the input. The graphs represent the mean of two experiments using both WT clones #1 and #2 and Z24$^{*/*}$ clones #1 and #2. Error bars represent the standard error of the mean (SEM). p-values (Mann-Whitney Test) comparing WT and Z24$^{*/*}$: *p < 0.05. *(B)* Western blot analysis for the detection SUV39H1 expression in 2i condition and after 3 days of differentiation (d3) in WT and Z24$^{*/*}$ cells. **γ**-TUBULIN was used as a loading control. A molecular weight scale in indicated on the left of the panel. *(C)* SETDB1 enrichment at IAP 5'UTR, IAP 3'UTR and SNORD116 cluster were assessed by ChIP-qPCR, as described in (A). *(D)* Western blot analysis for the detection SETDB1 expression, as described in (B). SETDB1 conditional knock-out (SETDB1 cKO) was induced treating cells with Tamoxifen (see Methods).

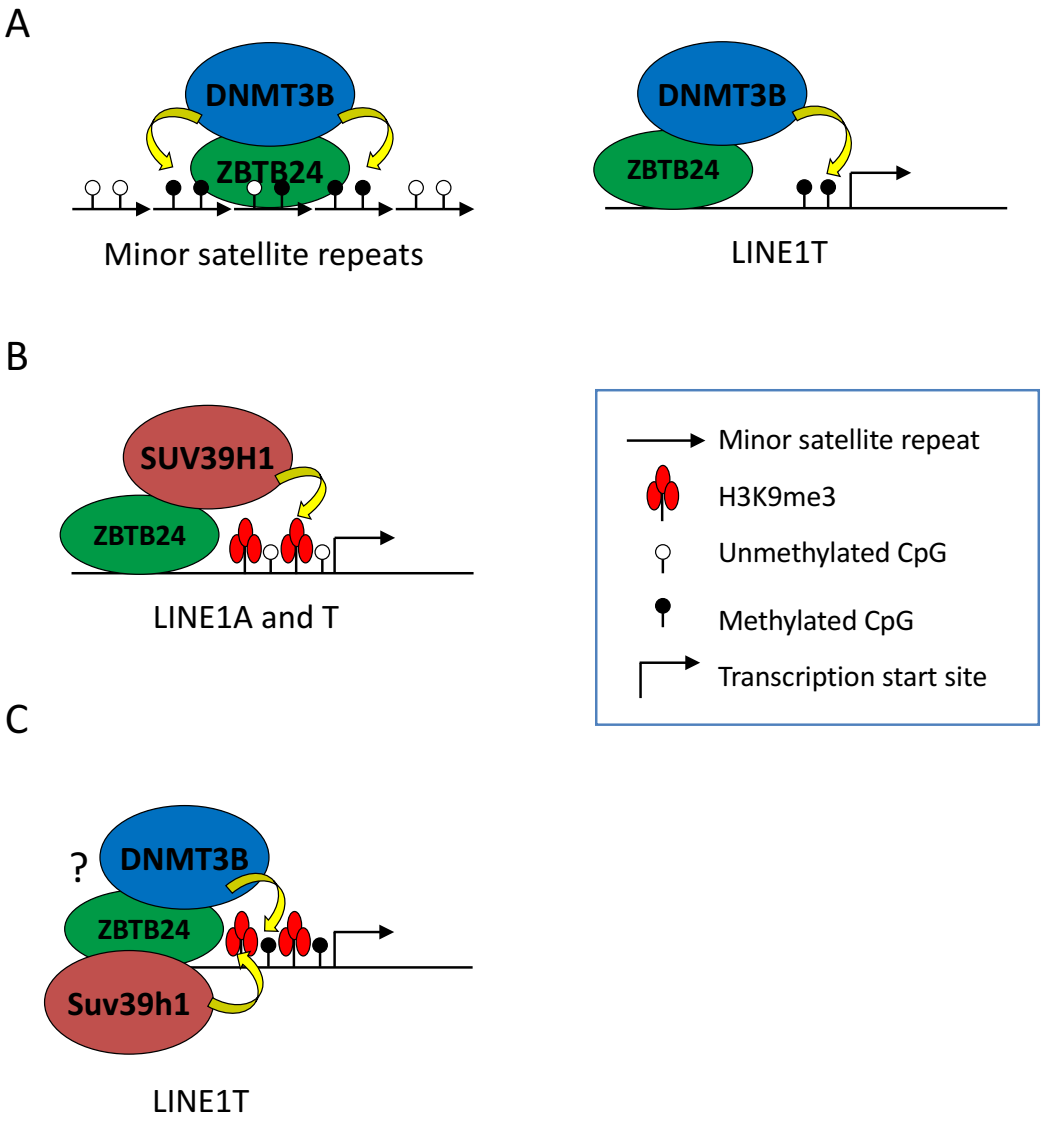


Figure 19. Proposed model of action of ZBTB24.

MATERIALS AND METHODS
(Unpublished results)

Generation of ZBTB24 mouse model

ZBTB24 mutant mice were generated at the Mouse Clinical Institute – Institut Clinique de la Souris, Illkirch, France (http://www-mci.u-strasbg.fr). The targeting construct was designed to replace Exon 2 by mutated sequences and a floxed PGK-neoR cassette. After transfection of C57BL/ 6J -derived P1 ES cells and selection with geneticin, homologous recombinants were identified by Southern blot analysis using 5′ and 3′ external probes. ES clones harboring the modified exon 2 were injected into C57BL/ 6J blastocysts to generate chimeric mice. Chimeric mice were interbred to generate the heterozygous mouse lines. Homozygous mice were generated by intercrossing heterozygous mice. Generation and use of these mice has been approved by the French Ministry of Research and Technologies and received agreement number 5314.

To genotype mice, genomic DNA was extracted from the yolk sac or the tail of embryos with the REDExtract-N-Amp Tissue PCR kit (Sigma) following the manufacturer's instructions. A combination of two pairs of specific primers, indicated below in the primer list (**Table 2**), was used to establish the genotype of each embryo by PCR (**Fig. 1C**).

Mouse embryonic stem cells derivation, culture and differentiation

Derivation and culture of mouse embryonic stem cells

WT and Z24$^{*/*}$ mES cells were derived as described in Czechanski et al., (2014). All mES cells used in this study were maintained in absence of feeder cells, on plates coated with 1X Laminin (SIGMA, L2020) as described in Tosolini and Jouneau, (2016), in 2i media. For the preparation of 2i media, DMEM/F12 Glutamax (Life Technologies) and Neurobasal MED SFM (Life Technologies) were supplemented with 1X N2 and B27 Supplements (both from Life Technologies), 1X Penicilline/Streptomycine (ThermoFisher), 1X non-essential amino acids (PAA), 100nM 2-mercaptoethanol (ThermoFisher), 2.5mM CHIR99201 and PD032552 inhibitors (both from Cell Guidance Systems) and 2x10^5 units of Leukemia Inhibitory Factor (Merck Millipore). Cell culture media was changed every two days and cells were dissociated in 0.05% acutase for passaging when confluent. DNMT3B*/* cells were derived in the lab (m3/m3 cells, Velasco et al., 2010). SUV39HDN cells were kindly provided by Antoine Peters and Alice Jouneau. Inducible knock-out SETDB1 mES cells were provided by Ait-Si-Ali's team and were described in Matsui et al., (2010) (33#6 cell line). 33#6 cell line was maintained in 2i for several weeks, to guarantee the same culture condition for all the cells. To induce the knock-out of SETDB1, Tamoxifen was added to cells in 2i at the final concentration of 800nM. After two days

cells a part of the cells was recovered (SETDB1 cKO 2i+T) and the other part was used to induce the embryoid bodies differentiation, keeping tamoxifen in the media at the same concentration as before.

Embyoid body Differentiation

Embryoid body differentiation was performed using the 'hanging drop' method as follows: Isolated mES cells were prepared as above (following acutase removal). Cells were counted and resuspended at a concentration of 3.34 x10^4 cells/ml in LIF-free media containing 10% Fetal Bovine Serum (FBS). From this suspension, droplets of 30µl (containing 1000 cells) were plated onto the lid of a petri dish (humidified with 10ml of PBS in the base) and embryoid bodies were left to form over the period of 48 hours. Embryoid bodies were subsequently collected by centrifugation and a grown in suspension for a further period of 72 hours. Following this, embryoid bodies were plated on gelatine-coated culture dishes and left to adhere. Embryoid bodies were harvested at various stages for RNA and genomic DNA extraction.

Neuron specific differentiation

mES cells were suspended in differentiation medium (ADFNK: 25% Advanced DMEM, 25% DMEM/F12 GlutaMAX, 50% Neurobasal medium, 1.3% of Penicilline/Streptavidin, 1X β-mercatoethanol and 11% of Knock out serum) (all products from Gibco). One million cells were grown in 10cm ultra-low culture dishes (from Corning) to allow their aggregation in suspension. Embryoid bodies were collected by centrifugation for 4 minutes at 700 rpm and resuspended in 10ml of fresh ADFNK medium. Retinoic acid was added to a final concentration of 1µM and 0.5µM respectively. Coating of the plates was prepared by a first incubation with poly-1-ornithin (0.001% final concentration) overnight at 37°C. After 3 washes in PBS, plates were coated on top with laminin (5µg/ml final concentration) by an overnight incubation at 37°C. At day six, embryoid bodies were dissociated and 5x105 – 8x105 cells/well were plated in 12-wells plates coated with poly- ornithine and laminin. From day 7 onwards, N2B27 medium was used: 50% GlutaMAX, 50% of Neurobasal medium, 0.5% of N2 supplement (containing Human Transferrin, recombinant insulin, full length progesterone, putrescine and selenite; Invitrogen), 0.1% of B27 supplement (containing the hormone Corticosterone among other products, which is an agonist of the Hedgehog pathway important for the formation of cortex neurons; Gibco).

RNA extraction and RT-qPCR analysis

Total RNA from cell lines was isolated using TRI Reagent® (Sigma) according to manufacturer's instructions. Contaminant genomic DNA was eliminated with TURBO DNAfree

kit (Ambion). Reverse transcription was carried out using 500ng DNA-free RNA and 50µM random hexamers (Roche), 20U of RNase inhibitor (New England Biolabs) and 20U of RevertAid Reverse Transcriptase (Fermentas). Complementary DNA reactions were used as templates for PCR reactions. Real-time PCR was performed using the LightCycler® DNA Master SYBR Green I mix (Roche) supplemented with 0.2µM specific primer pairs (Table 2). Real-time quantitative PCR were run on a light cycler rapid thermal system (LightCycler®480 2.0 Real time PCR system, Roche) with 20 sec of denaturation at 95°C, 20 sec of annealing at 60°C and 20 sec of extension at 72°C for all primers, and analysed by the comparative CT(ΔΔCT) method using U6 RNA as an invariant RNA. All the primers are listed in **Table 2**.

Chromatin Immunoprecipitation (ChIP)

Cells cultivated in 2i or at d3 of embryoid bodies differentiation were directly cross-linked with 1% formaldehyde for 10min at room temperature (RT) and the reaction was quenched by adding glycine at a final concentration of 0.125M for 5min at RT. Fixed cells were washed and harvested with PBS. Cell membranes were lysed in cell lysis buffer (5mM Pipes pH8.0, 85mM KCl, 0.5% NP40, Complete protease inhibitor cocktail (Roche)) for 20min, 4°C. Nuclei were pelleted by centrifugation and lysed in nuclei lysis buffer (50mM Tris HCl pH8.1, 150mM NaCl, 10mM EDTA, 1% SDS, Complete protease inhibitor cocktail (Roche)) for 20min, 4°C. Chromatin was subjected to sonication with Bioruptor Power-up (Diagenode) (3x[10× 15 s ON/45 s OFF]) yielding genomic DNA fragments with a bulk size of 150- 300 bp. Supernatant was diluted 10 times in IP dilution buffer (16.7 mM Tris pH8, 167 mM NaCl, 1.2 mM EDTA, 1.1% Triton X-100, 0.01% SDS, complete protease inhibitor cocktail (Roche)). IPs were performed overnight at 4°C using 20µg of chromatin for H3K9me3 and 40µg for factors and 2µg of specific antibodies against indicated proteins (**Table 3**). Immune complexes were recovered by adding pre-blocked protein G magnetic beads (Dynabeads, Life Technologies) and incubated for 2h at room temperature. Beads were washed once with Low salt buffer (0.1% SDS, 1% Triton, 2 mM EDTA, 20 mM Tris pH 8, 150 mM NaCl), once with High salt buffer (0.1% SDS, 1% Triton, 2 mM EDTA, 20 mM Tris pH 8; 500 mM NaCl), once with LiCl wash buffer (10 mM Tris pH 8.0, 1% Na-deoxycholate, 1% NP-40, 250 mM LiCl, 1 mM EDTA) and twice with TE (10mM Tris pH 8.0, 1mM EDTA pH 8.0, 50mM NaCl). Beads were eluted in Elution buffer (50mM Tris pH 8.0, 10mM EDTA pH 8.0, 1% SDS) at 65°C for 45min and cross-link was reversed O/N at 65°C after addition of NaCl to 0.2M final concentration. The eluted material was digested with 40µg of Proteinase K (Invitrogen) at 65°c for 2h then phenol/chloroform-extracted and ethanol-precipitated. DNA was resuspended in TE and q-PCR performed using LightCycler® DNA Master SYBR Green I mix (Roche) and analysed on a light cycler rapid thermal system (LightCycler®480 2.0 Real time PCR

system, Roche). ChIP-qPCR results were normalized on input signal (% input). Sequences of primers are listed in **Table 2**. Primers for interspersed DNA repeats were described in Zamudio et al., (2015). As a negative control, an additional mock ChIP using IgG was systematically added to each experiment to verify check that no more than 0.01% of the input was precipitated whatever the genomic region tested, including satellite repeats.

Genomic DNA extraction, DNA methylation analysis and validation of DNA methylation data

Genomic DNA was extracted and purified from cell lines, whole blood or buccal swabs using the NucleoSpin®Tissue kit (Macherey-Nagel) according to manufacturer's instructions. The DNA pellet was eluted in TE containing 20 µg/mL RNAse A (Euromedex).

Southern blot

Genomic DNA from MEFs (500 ng) was digested with 20 units of HpaII or HpyCH4IV enzymes (New England Biolabs) for 16 h to analyse the DNA methylation patterns of centromeric minor satellite repeats and pericentromeric major satellite repeats respectively. The digested DNA fragments were separated by electrophoresis using 1% agarose gels and transferred overnight to Hybond-N+ membranes (GE Healthcare) in 20 × SSC. After ultraviolet crosslink, the membranes were pre-hybridized in 6 × SSC, 5 × Denhardt and 0.1% SDS and then hybridized with 32P-labelled probe.

Minor satellite probe:

5′-ACATTCGTTGGAAACGGGATTTGTAGAACAGTGTATATCAATGAGTTAC AATGAGAAACAT-3′.

Major satellite probe:

5'-CACGTCCTACAGTGGACATTTCTAAATTTTCCACCTTTTTCAGTT-3'.

Reduced Resolution Bisulfite Sequencing (RRBS)

Preparation of RRBS libraries and processing of RRBS sequencing reads were performed as described in Auclair et al., (2014). Briefly, we digested 25 to 100 ng of genomic DNA for 5 h with MspI (Thermo Scientific, Waltham, MA, USA) followed by end-repair, A-tailing (with Klenow fragment, Thermo Scientific) and ligation to paired-end methylated adapters (with T4 DNA ligase, Thermo Scientific) in Tango 1X buffer. We purified fragments in the range 150 to 400 bp (insert plus adapter size) by electrophoresis on a 3% (w/v) agarose 0.5X TBE gel with the MinElute gel extraction kit (Qiagen). We then performed two rounds of bisulfite conversion with the EpiTect kit (Qiagen) according to the manufacturer's instructions. Final RRBS libraries were PCR

amplified with PfUTurbo Cx hotstart DNA polymerase (Agilent, Santa Clara, CA, USA) and indexed PE Illumina primers using the following PCR conditions: 95°C for 2 minutes, 14 to 16 cycles (95°C for 30 s, 65°C for 30 s, 72°C for 45 s), 72°C for 7 minutes. The libraries were purified with AMPure magnetic beads (Beckman Coulter, Brea, CA, USA), quantified with a Qubit fluorometer (Life Technologies, Carlsbad, CA, USA) and verified by loading 10 ng of the library on a 4-20% Criterion precast polyacrylamide gel (Biorad, Hercules, CA, USA) stained with SYBR Green. The libraries were paired-end sequenced (2 × 75 bp) in multiplex on an Illumina HiSeq2000 by Integragen SA (Evry, France) to generate an average of 30 million pairs of reads per sample. We performed quality control checks on sequencing reads with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Reads were trimmed with Trim Galore (v0.2.1) (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore) to remove adapter sequences and low-quality ends with a Phred score below 20. Trim Galore was run in --rrbs mode to remove two additional bases artificially introduced at the MspI restriction sites. We aligned sequencing reads to the mouse mm9 genome with BSMAP (v2.74) (https://code.google.com/p/bsmap) using the RRBS mode. For the mapping, we allowed a maximum of two mismatches and an insertion size for paired-end sequences of between 30 and 400 bp. We extracted methylation scores as the ratio of the number of Cs over the total number of Cs and Ts. We combined CpG methylation ratios from both strands and filtered for a minimum sequencing depth of 8×. We estimated the bisulfite conversion efficiency by calculating the C to T conversion at non-CpG sites, which was in most cases greater than 99.5%.

Validation of DNA methylation data

We used two different techniques, namely MSRE-qPCR (Methylation Sensitive Restriction Enzyme-based qPCR) and pyrosequencing. For MSRE-qPCR, we digested 300 ng of genomic DNA with 10 U of HpaII (New England BioLabs), followed by qPCR to measure the percentage of digestion relative to undigested control DNA (primers are listed in **Table 2**). Each data shown in qPCR analysis is the result of three independent experiments performed on three independent differentiation kinetics for WT clone #1 and $Z24^{*/*}$ clone #1. For pyrosequencing, bisulfite conversion of genomic DNA was performed with the EpiTect bisulfite kit (Qiagen). PCR amplification was performed using biotinylated reverse primers. PCR products were then converted to single- stranded DNA templates with the Pyromark Vacuum Prep Workstation (Qiagen) and sequencing was performed with primers listed in table 2 on a PyroMark Q24 System (Qiagen) according to manufacturer instructions. Data were extracted and analyzed using PyroMark Q24 2.0.6 software (Qiagen).

Coverage with published RRBS data

The coverage of our RRBS data was performed using Galaxy (https://usegalaxy.org). We used the published RRBS data obtained in embryos Dnmt3b and Dnmt3a knock-out at 8.5 dpc published by (Auclair et al., 2014) (GSE60334).

Protein extraction and western blot analysis

Cells were washed in PBS. After centrifugation, cell pellets were resuspendend in RIPA buffer (300 mM NaCl, 1% NP-40, 0.5% Na-deoxycholate, 0.1% SDS, 50 mM Tris pH 8.0, complete protease inhibitor cocktail). Samples were incubated for 30min on ice and were then sonicated (Bioruptor, Diagenode, 5× 30 s ON/30 s OFF). Cell debris were pelleted by centrifugation for 10 min at 16,000g and supernatant was collected. For Western blot analysis, 40µg of proteins were resolved by SDS–PAGE, transferred onto PVDF membrane (Thermo Fischer) and probed using the appropriate primary (**Table 3**) and secondary antibodies coupled to horse-radish peroxidase (Jackson Immunoresearch). Protein detection was performed with ECL reagents (Pierce).

Immunofluorescence

Cells were grown on glass coverslips until reaching 60% of confluence. Cells were then fixed with 4% paraformaldehyde in PBS (Electron Microscopy Sciences). After permeabilisation with 0.1% Triton X-100 in PBS, samples were blocked in 2% donkey serum (Jackson Immunoresearch) in PBS followed by incubation with primary antibodies (**Table 3**) in blocking buffer overnight at 4°C. After incubation with secondary antibodies conjugated to Alexa-Fluor 488 or 594 (Jackson ImmunoResearch), coverslips were mounted in Vectashield medium containing 2µg/ml DAPI (Vector Laboratories). Image acquisition was performed at room temperature on a fluorescence microscope (Axioplan 2; Zeiss) with a Plan-Neofluar 100X/1.3 NA oil immersion objective (Zeiss) using a digital cooled camera (CoolSNAP fx; Photometrics) and METAMORPH 7.04 software (Roper Scientific, Trenton, NJ). Images presented correspond to one focal plane.

Whole-mount *in situ* hybridization

The embryos at 7.5 dpc were fixed in 4% paraformaldehyde and subjected to whole-mount *in situ* hybridization as described in Hue et al., (2001) and Weisheit et al., (2002). Briefly, samples were permeabilised with proteinase K (10 µg/ml in PBS-Tween) for 10 mn before hybridization with Dig-labelled riboprobes. After incubation with anti-Dig antibody (Roche), samples were

incubated in BM purple (Roche) until the colour developed. The hybridized embryos were observed under an inverted microscope and photographed using a digital camera (Zeiss).

Statistical analysis

Mann-Whitney test was used to determine the statistical significance between values of qPCR, with a p-value <0.05 considered statistically significant.

Table 2. Primer list

Name	**Sequence (5' to 3')**	**Method**
Ef	GTACTATCTCCAGTGTGCAGATGAGG	*Genotyping*
Er	GGCCAATGCCAAGTCAGAGACAC	*Genotyping*
Lxr	AGTTATACTAGAGCGGCCGTTCACCG	*Genotyping*
Name	**Sequence (5' to 3')**	**Method**
BRACHYURY F	CACATCGTGAGAGTTGGGGG	*RT-qPCR*
BRACHYURY R	GGCATCAAGGAAGGCTTTAGC	*RT-qPCR*
GATA4 F	CACCCCAATCTCGATATGTTTGA	*RT-qPCR*
GATA4 R	GGTTGATGCCGTTCATCTTGT	*RT-qPCR*
GATA6 F	CCCACTTCTGTGTTCCCAATTG	*RT-qPCR*
GATA6 R	TTGGTCACGTGGTACAGGCG	*RT-qPCR*
KLF4 F	ACCCACACTTGTGACTATGCAG	*RT-qPCR*
KLF4 R	CCCAGTCACAGTGGTAAGGTTT	*RT-qPCR*
NEUN F	AACTAGCTCAGATGCTGACCG	*RT-qPCR*
NEUN R	ATATGGGTTCCCAGGCTTCTT	*RT-qPCR*
PAX6 F	GGACCCATTATCCAGATGTGTT	*RT-qPCR*
PAX6 R	GCCCTTCGATTAGAAAACCATA	*RT-qPCR*
REX1 F	CAGCTCCTGCACACAGAAGA	*RT-qPCR*
REX1 R	ACTGATCCGCAAACACCTG	*RT-qPCR*
U6 F	CTCGCTTCGGCAGCAGA	*RT-qPCR*
U6 R	AACGCTTCACGAATTTGCGT	*RT-qPCR*
WNT3 F	ACAAAGCCACCCGTGAATC	*RT-qPCR*
WNT3 R	TGGCCCCTTATGATGTGAGTC	*RT-qPCR*
Name	**Sequence (5' to 3')**	**Method**
IAP 3'UTR F	GCACATGCGCAGATTATTTGTT	*ChIP-qPCR*
IAP 3'UTR R	CCACATTCGCCGTTACAAGAT	*ChIP-qPCR*
IAP 5'UTR F	CGGGTCGCGGTAATAAAGGT	*ChIP-qPCR*
IAP 5'UTR R	ACTCTCGTTCCCCAGCTGAA	*ChIP-qPCR*
LINE1 3'UTR F	ATGGACCATGTAGAGACTGCCA	*ChIP-qPCR*
LINE1 3'UTR R	CAATGGTGTCAGCGTTTGGA	*ChIP-qPCR*
LINE1 ORF2 F	GGAGGGACACTTCATTCTCATCA	*ChIP-qPCR*
LINE1 ORF2 R	GCTGCTCTTGTATTTGGAGCATAGA	*ChIP-qPCR*

LINE1A F	GGATTCCACACGTGATCCTAA	*ChIP-qPCR*
LINE1A R	TCCTCTATGAGCAGACCTGGA	*ChIP-qPCR*
LINE1T F	CAGCGGTCGCCATCTTG	*ChIP-qPCR*
LINE1T R	CACCCTCTCACCTGTTCAGACTAA	*ChIP-qPCR*
Minor satellite repeat F	GAACATATTAGATGAGTGAGTTAC	*ChIP-qPCR*
Minor satellite repeat R	GTTCTACAAATCCCGTTTCCAAC	*ChIP-qPCR*
Major satellite repeat F	GACGACTTGAAAAATGACGAAATC	*ChIP-qPCR*
Major satellite repeat R	CATATTCCAGGTCCTTCAGTGTGC	*ChIP-qPCR*
SNORD116 F	TCCATTGCTCTACCCAAAGC	*ChIP-qPCR* *MSRE-qPCR*
SNORD116 R	CCAAGATGCTGGACCTCAGT	*ChIP-qPCR* *MSRE-qPCR*
Name	**Sequence (5' to 3')**	**Method**
H19 F	GGGTTTTTTTGGTTATTGAATTTTAA	*Pyrosequencing*
H19 R (biotin)	AATACACACATCTTACCACCCCTATA	
H19 sequencing	TGTTATGTGTAATAAGGGAA	

Table 3. Antibody list

Name	Reference	Method
βIII-TUBULIN	SIGMA – T8578	*IF*
DNMT3A	GeneTex - GTX129125	*Western blot – ChIP*
DNMT3B	Abcam - ab1360	*Western blot – ChIP*
γ-TUBULIN	SIGMA – T6557	*Western blot*
H3K9me3	Millipore – 07-523	*IF – ChIP*
IgG	SIGMA – I5006	*ChIP*
KAP1	Abcam - ab10484	*Western blot*
NANOG	Cosmo Bio Co. Ltd - REC-RCAB0002P-F	*IF*
OCT3/4	Santa Cruz Biotechnology – N-19 (sc-8628)	*IF*
SETDB1 (ESET)	Santa Cruz Biotechnology – H-300 (sc-66884)	*Western blot – ChIP*
SUV39H1	Cell Signaling – 8729 (D11B6)	*Western blot – ChIP*
ZBTB24	MBL life science – PM085	*Western blot – ChIP*

DISCUSSION AND PERSPECTIVES

DISCUSSION

The discovery of mutations in factors devoid of DNA methyltransferase activity as the cause of the ICF syndrome, and with a virtually unknown connection with DNA methylation with the exception of HELLS in the mouse (Huang et al., 2004; Myant and Stancheva, 2008; Termanis et al., 2016), raised the question of their role in DNA methylation, their genomic targets and their functional interplay, together and with DNMT3B. This discovery called for a comparative analysis of the perturbed methylation landscapes in ICF subtypes, to evaluate their similarities and differences depending on genotype and to document their genomic and epigenomic contexts in normal cells.

Strikingly, our high-throughput methylation analysis clearly distinguished two groups of ICF patients: ICF1 patients with DNMT3B mutations and the group of ICF2, ICF3 and ICF4 patients with no mutation in DNMT3B. This result is in agreement with the heterogeneity already described in ICF patients with regard to the methylation of satellite repeats. Indeed, patients without mutation in DNMT3B show hypomethylation of centromeric Satα in addition to hypomethylation of pericentromeric SatII and SatIII repeats that is a hallmark of all patients (Jiang et al., 2005; Wijmenga et al., 2000). This result is also consistent with distinguishing clinical signs and previous genotype-phenotype association studies showing that ICF1 patients suffer from a more severe immunodeficiency, whereas ICF2, ICF3 and ICF4 patients are more affected by intellectual disabilities (Sterlin et al., 2016; Weemaes et al., 2013).

Unifying features of ICF patients: ZBTB24, CDCA7 and HELLS as platform for DNMT3B recruitment?

Since DNA methylation is not randomly distributed on the genome and since DNMTs need to be guided or targeted to specific regions of the genome, DNMTs need to be guided to their genomic targets. Several studies showed that CpG content or chromatin environment can favour their recruitment at different genomic loci (Illingworth and Bird, 2009; Schübeler, 2015). For example, H3K36me3 in actively transcribed gene bodies plays roles in targeting DNA methylation machinery, through direct interaction with the PWWP domain of DNMT3B (Baubec et al., 2015; Neri et al., 2017). Moreover, transcription factors, owing to their ability to bind to consensus DNA sequences, were shown to shape DNA methylation patterns mediating DNMTs recruitment (Marchal and Miotto, 2015). For example, the repressor E2F6 was shown to be required for DNA methylation (Pohlers et al., 2005) and for the recruitment of DNMT3B at CGIs in the promoter region of a subset of germline genes (Velasco et al., 2010). Similarly, the structural maintenance

of chromosomes (SMC) family SMCHD1 was shown to be required for the recruitment of DNMT3B at CGIs on the inactive X chromosome and at X-linked promoters (Gendrel et al., 2012).

Despite the fact that our genome-wide analysis identified two groups of ICF patients, we identified a few hundreds of genomic loci whose methylation was affected in all four genotypes. The common hypomethylated regions mapped at intergenic regions and open sea CpGs context. Some genes were commonly affected, in particular olfactory receptor genes and protocadherin genes (PCDH). The PCDHβ gene cluster was hypomethylated in all ICF patients, although more prominently in ICF1. At least at these shared loci ZBTB24, CDCA7 and HELLS may serve as a platform for DNMT3B recruitment. In addition, we provide evidence, despite the biased representation of DNA repeats on the HM450K array, that centromeric satellite repeats on chromosome 8 and pericentromeric satellite repeats 12 were hypomethylated in all ICF patients, in addition to the repeats found on chromosome 1, 9 and 16. These results suggest that, at least at these genomic regions, ZBTB24, CDCA7 and HELLS may function as a platform for DNMT3B recruitment. Since ZBTB24 and CDCA7 can bind directly to DNA, they could act as platforms for the recruitment of the two other factors. Because nucleosomes are thought to constitute a barrier to DNA methylation (Felle et al., 2011), the helicase activity of HELLS could facilitate the accessibility and the methylation activity of DNMT3B, as proposed previously (Myant and Stancheva, 2008).

DNMT3B prefers CpG islands, inactive X-chromosome and germline gene promoters

Many studies have provided clues into DNMT3B function and preferential genomic and CpG context. DNMT3B was shown to actively methylate transcribed gene bodies (Baubec et al., 2015; Neri et al., 2017; Yang et al., 2014), and, in agreement with this observation, we could observe such a preference despite the design of the HM450K array on which CpGs in gene bodies are poorly represented. However, our analysis underlined that mutations in DNMT3B caused hypomethylation of CGIs, suggesting a role for DNMT3B in establishing DNA methylation at this CpG context. We report that the loss of DNMT3B activity caused hypomethylation of CGI on the inactive X chromosome in females, which is in line with previous studies performed directly on ICF1 patients cells (Bourc'his et al., 1999; Hansen et al., 2000; Holger Heyn et al., 2012) or on DNMT3B knock-out mouse (Auclair et al., 2014). In addition, we found hypomethylated CGI especially at the promoter of germline genes, which is in agreement with previous studies performed on ICF patients (Simo-Riudalbas et al., 2015; Velasco et al., 2014) or in the mouse (Auclair et al., 2014; Borgel et al., 2010; Hackett et al., 2012; Velasco et al., 2010; Walton et al., 2011). This subset of germline genes constitutes a molecular signature for the diagnosis of ICF1 patients (Velasco et al., 2014). We extended this analysis providing an extensive list of germline

genes starting from a list of almost 2,000 genes ((Djureinovic et al., 2014); http://www.proteinatlas.org/humanproteome/testis) at which DNMT3B is required for their methylated state in human somatic cells. Since our results have been obtained in uncultured cells, and validated in two different somatic tissues, these genes represent a strong signature for the identification of DNMT3B dysfunctions. This signature can be applied to identify DNMT3B dysfunctions in other diseases like cancer, that is characterized by the presence of inactive splice variants of DNMT3B (Gopalakrishnan et al., 2009; Gordon et al., 2013; Ostler et al., 2007; Saito et al., 2002).

ZBTB24, CDCA7 and HELLS are required for DNA methylation at CpG-poor regions, interspersed DNA repeats and monoallelically expressed genes

The clustering as well as the high correlation between methylation profiles of ICF2, ICF3 and ICF4 patients led us to investigate the intricacy between ZBTB24, CDCA7 and HELLS coding-genes. We showed that ZBTB24 is a master regulator of CDCA7, that is supported by the finding that ZBTB24 binds directly to CDCA7 promoter to promote its express (Wu et al., 2016). Moreover, we have evidence that ZBTB24, CDCA7 and HELLS belong to common protein complexes that promote their stability, suggesting that they may act in concert, at least, at same genomic regions, where the binding domains of ZBTB24 and CDCA7 would promote the direct binding of the complex to DNA, as I mentioned above for the PCDHβ gene cluster.

In contrast to CGI promoters of germline and inactive X-linked genes that are mainly affected by DNMT3B loss of function, mutations of ZBTB24, CDCA7 and HELLS mainly lead to hypomethylation of CpG-poor regions in intergenic regions, on the X chromosome in female patients and at CpG-poor promoters of protein-coding and non-coding genes. These findings are in agreement with previous studies of HELLS, also known as LSH (Lymphoid-Specific Helicase) in the mouse, showing hypomethylation of CpG-poor intergenic and promoter regions in mouse embryonic fibroblasts (Yu et al., 2014). Methylome analysis of ICF4 patient suggests that the function of HELLS at CpG-poor regions is conserved between mouse and humans. Similarly, our results obtained on ZBTB24 mutant mouse suggest that the function of ZBTB24 at CpG-poor regions is also conserved between human and mouse.

Our methylome analysis showed a more prominent hypomethylation of satellite repeats in ICF1 patients. In contrast, ICF2, ICF3 and ICF4 patients showed hypomethylation of interspersed DNA repeats, including LINE1 and LTR elements. The same families of DNA repeats were found to be hypomethylated in HELLS knock-out mouse (Dennis et al., 2001; Dunican et al., 2013; Huang et al., 2004) as well as in our $Z24^{*/*}$ mouse. Thus, the function of HELLS and ZBTB24 seem to be also conserved between human and mouse at interspersed DNA repeats.

We have extended the ZBTB24, CDCA7 and HELLS connection to single-copy genes that are affected in ICF2, 3 and 4 but not in ICF1 patients. We found the remarkable presence of a number of coding and non-coding genes that are organized in clusters, and expressed in a monoallelic manner, resulting from either random or imprinted choice (Chess, 2016; Eckersley-Maslin and Spector, 2014; Magklara and Lomvardas, 2013). Among those, we found several clusters of small non-coding RNAs: SNORD115-116 on chromosome 15, the primate-specific microRNA gene cluster C19MC on chromosome 19, the SNORD114 and the miRNA cluster C14MC on chromosome 14. C19MC and C14MC are large microRNA imprinted clusters whose expression is restricted to the placenta and undifferentiated cells (Landgraf et al., 2007; Noguer-Dance et al., 2010; Seitz et al., 2004). SNORD115 and 116 (Small Nucleolar RNA, C/D Box 115 and 116 Clusters) clusters of small ncRNAs are hosted in the imprinted locus on chromosome 15q11-q13 which deletion is the genetic cause of the Prader-Willi syndrome (PWS, OMIM#176270), a neurobehavioral disorder (Cassidy et al., 2012). Interestingly, together with the suspected role of C19MC and C14MC in brain functions (Laufer and Singh, 2012), it is tempting to speculate that hypomethylation of these clusters of non-coding RNA clusters that we reported in ICF2, ICF3 and ICF4 patients may contribute to the perturbed neurological functions that characterize these ICF sub-types. This speculation is reinforced by the hypomethylation of genes playing important functions in the central nervous system (CNS) specifically in ICF2, ICF3 and ICF4 patients. This observation provides biomarkers particularly relevant for neurodevelopmental defects that affect ICF2, ICF3 and ICF4 patients, as exemplified by studies reporting that this type of blood-based biomarkers can reflect DNA methylation changes in various tissues including in the CNS (Davies et al., 2012; Walton et al., 2016).

At the regions specifically hypomethylated in ICF2, ICF3 and ICF4 patients we can speculate that the methylation is controlled independently of DNMT3B. These three factors may therefore function as platforms for DNMT3A to establish DNA methylation, favouring its recruitment at these loci. Alternatively, ZBTB24, CDCA7 and HELLS may be required for the maintenance activity catalysed by DNMT1, favouring its targeting directly at the replication fork or in site-specific manner, as I will discuss below.

Are ZBTB24, CDCA7 and HELLS implicated in establishment of DNA methylation?

All the defects in DNA methylation discussed so far may originate from defects either in the establishment or in the maintenance of DNA methylation profiles. Knowing the role of DNMT3B in establishing DNA methylation at the first steps of development in mouse, it is likely that the defects in DNA methylation observed in ICF1 patients are inherited from the early steps of human embryogenesis. However, we cannot exclude that some defects may arise from a failure in DNA

methylation maintenance, in line with the documented role of DNMT3B in this pathway (Jones and Liang, 2009; Walton et al., 2011).

It will be difficult to test a potential role of these factors in the establishment of DNA methylation directly in humans. However, thanks to a mouse model for the ICF2 syndrome, we could show that mutations in ZBTB24 caused defects in DNA methylation in $Z24^{*/*}$ mES which apparently persisted in ZBTB24 mutant embryos at 11.5 dpc, suggesting that ZBTB24 might be required for the establishment of DNA methylation. By studying the DNA methylation patterns in the first steps of mES cells differentiation we identified defects during the recruitment of *de novo* DNA methyltransferases at their genomic targets. Well-established targets of DNMT3B in the mouse are centromeric minor satellite repeats (Okano et al., 1999). Our ChIP-qPCR experiments suggested that ZBTB24 binds directly to minor satellite repeats and is required for the recruitment of DNMT3B (**Fig. 19A unpublished results**). This result provided one of the few examples of a factor implicated in the recruitment of DNMT3B at murine centromeric repeats. Another mechanism previously proposed based on experiments in human cells, suggested that DNMT3B was recruited at SatII and Satα repeats through its interaction with the centromeric protein CENP-C (Gopalakrishnan et al., 2009).

Genome-wide analysis during differentiation of $Z24^{*/*}$ mES cells showed that a few hundreds of regions remained hypomethylated from 2i conditions to 3 days of differentiation, suggesting that DNA methylation is globally established in absence of ZBTB24, with the exception of interspersed DNA repeats, and more specifically LINE1 and IAP elements. The comparison with DNA methylation data obtained in DNMT3B knock-out embryos at 8.5 dpc (Auclair et al., 2014) showed that most of the regions that were found hypomethylated in $Z24^{*/*}$ cells were also hypomethylated in DNMT3B knock-out embryos, suggesting that ZBTB24 could also function as a platform for the recruitment of DNMT3B at interspersed DNA repeats. ChIP-qPCR experiments with an anti-DNMT3B antibody suggested that not only LINE1T promoter is a target region of DNMT3B, but also that ZBTB24 was required for the recruitment of DNMT3B at this genomic target (**Fig. 19A unpublished results**). These observations lead to propose ZBTB24 as a new epigenetic guardian of a subset of repetitive elements.

We did not observe defects in the establishment of DNA methylation at major satellite repeats. The loss of ZBTB24 functions did not prevent the recruitment of DNMT3A, the enzyme known to establish DNA methylation at these regions in murine embryos (Okano et al., 1999), despite ZBTB24 binds directly to major satellite repeats. However, the comparison of our genome-wide methylation data with datasets obtained in DNMT3A knock-out embryos (Auclair et al., 2014) suggested that ZBTB24 could be required for DNMT3A recruitment at few genomic

regions. For this reason, we cannot completely exclude a functional interaction between ZBTB24 and DNMT3A at some other genomic loci.

ZBTB24, CDCA7 and HELLS are required for DNA methylation maintenance

As I mentioned before, we cannot exclude that the defects of DNA methylation observed in patients carrying mutations in ZBTB24, CDCA7 and HELLS may derive from profiles of DNA methylation that are not maintained, although correctly established. We showed that knock-down of ZBTB24, CDCA7 and HELLS in mouse primary fibroblasts, where DNA methylation is already established, resulted in hypomethylation of minor satellite repeats. These results strongly suggested that ZBTB24, CDCA7 and HELLS are implicated in maintenance of DNA methylation, at least at these repeats. The implication of ZBTB24 in the maintenance of DNA methylation is further suggested by the observation that major satellite repeats are hypomethylated in $Z24^{*/*}$ embryos at 11.5 dpc.

DNA methylation is maintained during DNA replication by DNMT1 that localise at the replication fork through its interaction with PCNA and UHRF1 (Bostick et al., 2007; Chuang et al., 1997; Sharif et al., 2007). Although the localisation of ZBTB24, CDCA7 and HELLS at the replication fork has not been tested, we can speculate that their presence could be important for the correct maintenance of DNA methylation, in particular at regions, like minor satellite repeats or heterochromatin regions, that are known to be replicated in late S phase (Guenatri et al., 2004). ZBTB24 and CDCA7 may be required for the localisation of HELLS at the replication fork, which, as a helicase, may favour the process of DNA replication and DNA methylation maintenance. Alternatively, it has been proposed that DNA methylation maintenance could also occur in a site-specific manner, uncoupled from DNA replication, by the recruitment of DNMT1 at genomic regions, including gene clusters like the HOXA cluster, through the interaction with transcription factors (Hervouet et al., 2010). We cannot exclude that ZBTB24, CDCA7 and HELLS could be implicated in this site-specific mechanism through a possible direct interaction with DNMT1.

ZBTB24, CDCA7 and HELLS control DNA methylation by promoting H3K9me3 establishment

Up to now, we have considered the possible implication of ZBTB24, CDCA7 and HELLS in the direct recruitment of DNMTs to contribute to the non-random repartition of DNA methylation on the genome. However, we cannot exclude that these three factors may be implicated in the establishment of determinants of DNA methylation. It is known that DNA methylation landscapes result from a complex interplay between methylation, demethylation and protection from methylation/demethylation, as well as from a complex interplay between DNA

methylation machinery and histone modifications (Rose and Klose, 2014). For example, H3K4me3, that marks active promoters, appears to be mutually exclusive with DNA methylation. In contrast, H3K36me3 in actively transcribed gene bodies has been involved in targeting the DNA methylation machinery, through direct interaction with the PWWP domain of DNMT3B (Baubec et al., 2015; Neri et al., 2017). In contrast, the repressive histone mark H3K9me3 has been involved in targeting of DNA methylation in heterochromatin and at parent-of-origin-specific imprints (Rose and Klose, 2014). The remarkable hypomethylation of CpGs in ICF2, ICF3 and ICF4 patients lying within chromatin domains with heterochromatin marks in normal cells suggested that loss of DNA methylation at these regions may result from the loss of H3K9me repressive histone marks. This hypothesis is reinforced by the demonstration that HELLS and G9A cooperate to promote DNA methylation at a subset of gene promoters (Myant et al., 2011). In addition, SUV39H1, one of the major enzyme responsible for the deposition of H3K9me3 has been identified as a partner of ZBTB24, by a two-hybrid screen (Weimann et al., 2013). During differentiation of $Z24^{*/*}$ mES, we observed that loss of DNA methylation at LINE1 and IAP repeats was accompanied by a loss of H3K9me3 caused by the loss of recruitment of the KMTs SUV39H1 and SETDB1, that are responsible of the establishment of this mark. As previously described, the loss of DNA methylation did not prevent the enrichment in H3K9me3 at these regions (Karimi et al., 2011; Martens et al., 2005), suggesting that the loss of H3K9me3 observed in $Z24^{*/*}$ may directly depend on the loss of ZBTB24 functions. ZBTB24 binds directly to LINE1A and LINE1T promoters and is required for the recruitment of SUV39H1. The further establishment of H3K9me3 favours DNA methylation at these regions (**Fig. 19B unpublished results**). Moreover, the observation of the same dynamics at LINE1 repeats conserved in most of LINE elements (LINE1 ORF2 and 3'UTR) strongly suggest that the impact of the loss of ZBTB24 is not restricted to specific promoters of LINE1 elements. Interestingly, we observed that ZBTB24 was important for the recruitment of both DNMT3B and SUV39H1 at LINE1T promoter, leading to the hypothesis that the two enzymes may be recruited simultaneously (**Fig. 19C unpublished results**). The loss of DNA methylation and H3K9me3 at IAP elements is particularly interesting since these regions are known to be roboustly methylated both in gametes and in mES cells (Lane et al., 2003; Papin et al., 2017). Although ZBTB24 did not seem to bind directly to these elements, the loss of its functions prevent the binding of SETDB1, the other enzyme responsible for H3K9me3 deposition, and further establishment of H3K9me3. These results strongly suggest that ZBTB24 can be considered as a factor with a crucial role in the formation of heterochromatin, at least at LINE1 and IAP classes of interspersed DNA repeats.

We extended our mechanistic analysis to the SNORD116 cluster, contained in the PWS locus, enriched in DNA methylation and H3K9me3 (Cruvinel et al., 2014). A protein complex

containing the zinc finger protein ZNF274 and SETDB1 was recently involved in DNA methylation of the PWS imprinting centre (PWS-IC) that controls the transcriptional repression of SNORD116 cluster (Cruvinel et al., 2014; Liu et al., 2014). Our data adds a new zinc finger protein, ZBTB24, whose loss of function specifically alters DNA methylation at SNORD clusters without affecting the PWS-IC, through the regulation of SETDB1 recruitment and further H3K9me3 deposition. These results suggest that ZBTB24 controls DNA methylation *via* H3K9me3 deposition not only at interspersed repeats.

Does the loss of DNA methylation and H3K9me3 in absence of ZBTB24 favour the establishment of another histone mark?

The loss of heterochromatin marks in ZBTB24 mES cells may suggest that other histone marks could replace H3K9me3 in absence of DNA methylation. For example, the loss of SMCHD1 was shown to cause the loss of H3K9me3 and DNA methylation at D4Z4 repeats, that were replaced by H3K4me3 (Balog et al., 2012). We cannot exclude that a similar mechanism could take place with the loss of ZBTB24 functions, DNA methylation or H3K9me3 at LINE1 and IAP repeats. This hypothesis is reinforced by the identification of the H3K4 demethylase LSD1 as partner of ZBTB24, by a two-hybrid screen (Weimann et al., 2013). We can speculate that the absence of ZBTB24 may prevent the recruitment of LSD1 at some genomic targets. In this scenario, the loss of H3K9me3 and DNA methylation could cause an increase in the establishment of H3K4me3, which could be erased by LSD1. As a consequence, regions that are generally silenced, may switch to a potentially active transcriptional state. The transcription of retrotransposition-competent elements, like IAP and some LINE1, may cause their mobilization in the genome that may lead to insertional mutagenesis. These events could be deleterious for cells, and may lead to cancer development and subsequent progression, as previously described (Tufarelli et al., 2013).

Recently, in contrast to the enrichment in chromatin marks associated with permissive transcription, loss of DNA methylation at transposable elements was shown to cause the deposition of H3K27me3 to maintain the silent state of these elements (Walter et al., 2016). At LINE1, for example, the loss of DNA methylation was associated with persistence of H3K9me3 at the promoter region and increase in H3K27me3 at the ORF and 3'UTR regions (Walter et al., 2016: category A of elements). Since we observed a loss of H3K9me3 we can speculate that H3K27me3 may spread and also cover the promoter region, maintaining the repression of LINE1 elements. In contrast to LINE1 elements, IAP elements are not characterized by increase in H3K27me3 upon DNA demethylation but only by persistence of H3K9me3 (Walter et al., 2016: category B of elements). In absence of ZBTB24 we observed a loss of both DNA methylation and H3K9me3 at

IAP elements in mES cells as well as at d3 of differentiation. In this context, it is difficult to say whether H3K27me3 could be established at these regions. At other genomic regions, including major satellite repeats, the loss of DNA methylation and H3K9me3 cause an increase of H3K27me3, favoured by BEND3, that directly bind major satellite repeats in absence of DNA methylation and H3K9me3 (Lehnertz et al., 2003; Peters et al., 2003; Saksouk et al., 2015). We cannot exclude that loss of ZBTB24 functions, DNA methylation and H3K9me3 may favour the binding of BEND3, or other methylation-sensitive factors, which may favour the recruitment of PRC2 and the further establishment of H3K27me3.

Does ZBTB24 protect from DNA demethylation?

In mammals, after fertilization, DNA methylation is globally reprogrammed. A genome-wide demethylation occurs at the parental genomes by active demethylation on the paternal genome and passive demethylation on the maternal genome (Mayer et al., 2000; Oswald et al., 2000). Only a few genomic regions resist to this reprogramming, like ICRs, in both human and mouse, and IAP elements in mouse (Auclair et al., 2014; Borgel et al., 2010; Lane et al., 2003; Smallwood et al., 2011; Smith et al., 2012). During the mechanism of demethylation, resistant regions are protected from active or inactive demethylation. The hypomethylation of resistant regions may not only be the consequence of a defect in the establishment of DNA methylation, but could also be caused by a defect in their protection. For example, previous studies in mouse described PGC7/Stella as a key factor required to protect the paternally expressed genes PEG1, PEG3 and PEG10 as well as the IAP elements from DNA demethylation (Nakamura et al., 2007). More recently, ZFP57 was shown to directly bind and protect ICRs from DNA demethylation, including DLK1-DIO3 and H19-IGF2 ICRs (Quenneville et al., 2011; Strogantsev et al., 2015). Surprisingly, our RRBS results showed that mutations in ZBTB24 have a strong impact on DNA methylation in the 2i conditions, where the levels of DNA methylation are already low compared to differentiated cells or mES cultivated in Serum+LIF (Habibi et al., 2013). Among the hypomethylated regions, we identified IAP regions and H19 ICR. Since ZBTB24 binds to DNA sequences independently of their DNA methylation status (Nitta et al., 2013), we cannot exclude that ZBTB24 may bind directly to H19 ICR on the paternal allele, that is methylated (Bartolomei et al., 1993; Ferguson-Smith et al., 1993) and protect it from DNA demethylation. This protection may depend either on the occupancy of ZBTB24 blocking the action of TETs and the conversion of 5mC to 5hmC and further demethylation. Since ZBTB24 does not bind directly to IAP elements, the mechanism could be more complicated depending, for example, on the presence of a factor whose expression or stability depends on ZBTB24.

Mutations in ZBTB24 may alter indirectly the DNA methylation patterns

Since ZBTB24 is implicated in the organization of heterochromatin, we cannot thus exclude that indirect mechanisms may lead to large-scale alterations of DNA methylation observed in ICF patients and in $Z24^{*/*}$ mouse. Alterations in the organization of the chromatin were associated with modification of the replication timing, mostly at satellite repeats, that could impact the patterns on DNA methylation, presumably on its maintenance at other genomic loci (Jørgensen et al., 2007). Perturbations to pericentromeric heterochromatin compartments as a consequence of loss of methylation at the underlying DNA repeats may have long range impact on the epigenome as well as on transcription of certain genomic regions (Matarazzo et al., 2007). Similarly, studies directly on ICF patients cells already showed that the hypomethylation of pericentromeric satellite repeats caused heterochromatin decondensation and may alter the intra-nuclear organization of heterochromatin and alter higher-order chromatin associations (Jefferson et al., 2010). These alterations affected the positioning of SatII, from the nuclear periphery to the more internal locations in the nucleus, associated with an increase in the transcription of few genes close to SatII, and were proposed to cause extensive changes on the cellular epigenome (Jefferson et al., 2010).

Since ZBTB24 was shown to localise at chromocenters in murine cells and its localisation does not depend on the methylation status of underlying sequences (Nitta et al., 2013), it is tempting to speculate that ZBTB24 could orchestrate the recruitment of DNMTs at chromocenters that contribute to heterochromatin formation. Mutations altering or disrupting the binding of ZBTB24 may lead to alterations of the recruitment of DNMTs and in heterochromatin formation, resulting in aberrant profiles of DNA methylation and heterochromatin decondensation. We cannot exclude that these alterations may alter the nuclear organization and lead to further defects in DNA methylation and heterochromatin formation at other genomic loci.

Finally, all our results point to the fact that the existence of monogenic diseases affecting DNA methylation machinery of landscapes offers a unique opportunity to uncover actors and mechanisms and infer genomic regions controlled by these factors. How perturbed methylation at DNA repeats translates into physiopathological manifestations is unknown but the study of this rare disease opens new avenues relevant to other physiopathological conditions characterized by chromosomal instability or perturbed DNA methylation landscapes.

PERSPECTIVES

Does ZBTB24 act as platform for the recruitment of DNMT3A?

As I discuss above, we could not detect a defects in the establishment of DNA methylation at major satellite repeats. The recruitment of DNMT3A is not prevented by the loss of ZBTB24 functions at these repeats. However, we identified some regions commonly hypomethylated in $Z24^{*/*}$ cells and DNMT3A knock-out embyos at 8.5 dpc (dataset from Auclair et al., 2014), suggesting that at some regions ZBTB24 may contribute to the recruitment of DNMT3A. To test this hypothesis, we should analyse later embryonic stages, like after 10.5 dpc, which is the embryonal stage at which DNMT3A starts to be clearly detectable by immunostaining on total embryos (Watanabe et al., 2002). Interestingly, the regions that are commonly hypomethylated in $Z24^{*/*}$ cells and DNMT3A knock-out embryos were found outside DNA repeats. It is thus possible that ZBTB24 could be important for the recruitment of DNMT3A at coding and non-coding genes, more than at DNA repeats.

Does ZBTB24 protect from DNA demethylation?

Nowadays, a powerful tool to study the kinetic of DNA demethylation is based on the culture of mES cells in the presence of vitamin C, that is known to activate TETs enzyme, promote 5hmC and DNA (Blaschke et al., 2013). It would be interesting to compare the DNA methylation profiles of ZBTB24 mutant to WT mES cells during kinetics of demethylation. The identification of DNA sequences protected from DNA demethylation in absence of ZBTB24 could give more insights into a potential role of ZBTB24 in DNA demethylation.

ZBTB24: where and with whom?

Our results strongly support a role of ZBTB24 in DNA methylation, and more broadly in the organization of heterochromatin by the orchestration of different epigenetic modifications. Although our study gave some insights into the genomic regions at which ZBTB24 binds and is required for DNA methylation as well as chromatin organization, an overview of its DNA binding sites by ChIP-Seq analysis is still missing. This analysis could be then complemented by the identification of ZBTB24 partners, since only few of them have been described so far (Weimann et al., 2013). These data will provide further clues of the crucial role of ZBTB24 in the chromatin organization as well as in the epigenetic machinery.

ICF syndrome, aging and cancer: new guardians of the genome

ICF syndrome is characterized by alterations of DNA methylation profiles, in particular at (peri)centromeric repeats and chromosomal instability, which are also observed in other disease like cancer or inflammatory diseases (Ehrlich et al., 2006). This observation suggests that the study of this disease is relevant, not only for ICF patients, but also for other diseases characterized by alterations of DNA methylation profiles, to understand the causality relationship between alterations of DNA methylation profiles, chromosomal instability and perturbed cellular phenotypes. The hypomethylation of pericentromeric repeats is also a peculiar feature of cellular senescence (Cruickshanks et al., 2013), which was shown to be induced either by the activation of an oncogene, by the process called oncogene-induced senescence (OIS), or by the inactivation of a tumour-suppressor gene (Kuilman et al., 2010; Pal and Tyler, 2016). The effects of mutations in ZBTB24, CDCA7 and HELLS resemble the effect of the inactivation of a tumour-suppressor gene, like p53. Although it is difficult to evaluate the predisposition of ICF patients to cancer, partially due to the low number of patients and their short life span, we can speculate that the ICF factors may constitute new tumour-suppressors or, more generally, new guardians of the integrity of the mammalian genome, by controlling the DNA methylation patterns at (peri)centromeric and interspersed DNA repeats. It would be interesting to test this hypothesis performing inactivation of ZBTB24, CDCA7 or HELLS, by simple knock-down or CRISPR/Cas9 technology, and to analyse the phenotype of cells, like senescence, apoptosis, etc. Forcing the further expression of an oncogene, like RAS, could constitute the second hit to promote cellular transformation. Testing proliferation and cellular invasion, by soft-agar colony forming assay could be very informative on a putative role of ZBTB24, CDCA7 or HELLS as new tumour-suppressors.

REFERENCES

Aapola, U., Shibuya, K., Scott, H.S., Ollila, J., Vihinen, M., Heino, M., Shintani, A., Kawasaki, K., Minoshima, S., Krohn, K., Antonarakis, S.E., Shimizu, N., Kudoh, J., Peterson, P., 2000. Isolation and Initial Characterization of a Novel Zinc Finger Gene, DNMT3L, on 21q22.3, Related to the Cytosine-5- Methyltransferase 3 Gene Family. Genomics 65, 293–298. doi:10.1006/geno.2000.6168

Ahmad, K., Henikoff, S., 2002. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Mol. Cell 9, 1191–1200.

Allfrey, V.G., Mirsky, A.E., 1964. Structural Modifications of Histones and their Possible Role in the Regulation of RNA Synthesis. Science 144, 559. doi:10.1126/science.144.3618.559

Allsopp, R.C., Harley, C.B., 1995. Evidence for a critical telomere length in senescent human fibroblasts. Exp. Cell Res. 219, 130–136. doi:10.1006/excr.1995.1213

Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., Zoghbi, H.Y., 1999. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat. Genet. 23, 185–188.

Antequera, F., Bird, A., 1993. Number of CpG islands and genes in human and mouse. Proc. Natl. Acad. Sci. 90, 11995–11999.

Anzai, T., Osanai, M., Hamada, M., Fujiwara, H., 2005. Functional roles of 3'-terminal structures of template RNA during in vivo retrotransposition of non-LTR retrotransposon, R1Bm. Nucleic Acids Res. 33, 1993–2002. doi:10.1093/nar/gki347

Arand, J., Spieler, D., Karius, T., Branco, M.R., Meilinger, D., Meissner, A., Jenuwein, T., Xu, G., Leonhardt, H., Wolf, V., Walter, J., 2012. In Vivo Control of CpG and Non-CpG DNA Methylation by DNA Methyltransferases. PLoS Genet. 8, e1002750. doi:10.1371/journal.pgen.1002750

Aravind, L., 2001. The WWE domain: a common interaction module in protein ubiquitination and ADP ribosylation. Trends Biochem. Sci. 26, 273–275.

Arita, K., Isogai, S., Oda, T., Unoki, M., Sugita, K., Sekiyama, N., Kuwata, K., Hamamoto, R., Tochio, H., Sato, M., Ariyoshi, M., Shirakawa, M., 2012. Recognition of modification status on a histone H3 tail by linked histone reader modules of the epigenetic regulator UHRF1. Proc. Natl. Acad. Sci. 109, 12950–12955. doi:10.1073/pnas.1203701109

Arnoult, N., Van Beneden, A., Decottignies, A., 2012. Telomere length regulates TERRA levels through increased trimethylation of telomeric H3K9 and HP1α. Nat. Struct. Mol. Biol. 19, 948–956. doi:10.1038/nsmb.2364

Ashe, A., Morgan, D.K., Whitelaw, N.C., Bruxner, T.J., Vickaryous, N.K., Cox, L.L., Butterfield, N.C., Wicking, C., Blewitt, M.E., Wilkins, S.J., Anderson, G.J., Cox, T.C., Whitelaw, E., 2008. A genome-wide screen for modifiers of transgene variegation identifies genes with critical roles in development. Genome Biol. 9, R182. doi:10.1186/gb-2008-9-12-r182

Auclair, G., Borgel, J., Sanz, L.A., Vallet, J., Guibert, S., Dumas, M., Cavelier, P., Girardot, M., Forné, T., Feil, R., Weber, M., 2016. EHMT2 directs DNA methylation for efficient gene silencing in mouse embryos. Genome Res. 26, 192–202. doi:10.1101/gr.198291.115

Auclair, G., Guibert, S., Bender, A., Weber, M., 2014. Ontogeny of CpG island methylation and specificity

of DNMT3 methyltransferases during embryonic development in the mouse. Genome Biol. 15. doi:10.1186/s13059-014-0545-5

Aulie, R.P., 1961. Caspar Friedrich Wolff and his “Theoria generationis”, 1759. J. Hist. Med. Allied Sci. 16, 124–144.

Avery, O.T., MacLeod, C.M., McCarty, M., 1999. 1944 Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types. Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III. Microbiol. Centen. Perspect. 116.

Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., Lovell-Badge, R., 2003. Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev. 17, 126–140. doi:10.1101/gad.224503

Avner, P., Heard, E., 2001. X-chromosome inactivation: counting, choice and initiation. Nat. Rev. Genet. 2, 59–67. doi:10.1038/35047580

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., Fisher, A.G., 2006. Chromatin signatures of pluripotent cell lines. Nat. Cell Biol. 8, 532–538. doi:10.1038/ncb1403

Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E., Lingner, J., 2007. Telomeric repeat–containing RNA and RNA surveillance factors at mammalian chromosome ends. Science 318, 798–801.

Bailey, J.A., Carrel, L., Chakravarti, A., Eichler, E.E., 2000. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. Proc. Natl. Acad. Sci. U. S. A. 97, 6634–6639.

Ball, M.P., Li, J.B., Gao, Y., Lee, J.-H., LeProust, E.M., Park, I.-H., Xie, B., Daley, G.Q., Church, G.M., 2009. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat. Biotechnol. 27, 361–368. doi:10.1038/nbt.1533

Balog, J., Thijssen, P.E., de Greef, J.C., Shah, B., van Engelen, B.G.M., Yokomori, K., Tapscott, S.J., Tawil, R., van der Maarel, S.M., 2012. Correlation analysis of clinical parameters with epigenetic modifications in the DUX4 promoter in FSHD. Epigenetics 7, 579–584. doi:10.4161/epi.20001

Bao, S., Tang, F., Li, X., Hayashi, K., Gillich, A., Lao, K., Surani, M.A., 2009. Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. Nature 461, 1292–1295. doi:10.1038/nature08534

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Hérault, Y., Guillou, F., Bourc’his, D., 2016. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science 354, 909–912.

Barr, M.L., Bertram, E.G., 1949. A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. Nature 163, 676.

Barreto, G., Schäfer, A., Marhold, J., Stach, D., Swaminathan, S.K., Handa, V., Döderlein, G., Maltry, N., Wu, W., Lyko, F., Niehrs, C., 2007. Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. Nature 445, 671–675. doi:10.1038/nature05515

Bartolomé, C., Maside, X., Charlesworth, B., 2002. On the abundance and distribution of transposable

elements in the genome of Drosophila melanogaster. Mol. Biol. Evol. 19, 926–937.

Bartolomei, M.S., Webber, A.L., Brunkow, M.E., Tilghman, S.M., 1993. Epigenetic mechanisms underlying the imprinting of the mouse H19 gene. Genes Dev. 7, 1663–1673.

Bartolomei, M.S., Zemel, S., Tilghman, S.M., 1991. Parental imprinting of the mouse H19 gene. Nature 351, 153–155. doi:10.1038/351153a0

Bascom, G., Schlick, T., 2017. Linking Chromatin Fibers to Gene Folding by Hierarchical Looping. Biophys. J. 112, 434–445. doi:10.1016/j.bpj.2017.01.003

Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., Schübeler, D., 2015. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature 520, 243–247. doi:10.1038/nature14176

Baylin, S.B., Jones, P.A., 2011. A decade of exploring the cancer epigenome - biological and translational implications. Nat. Rev. Cancer 11, 726–734. doi:10.1038/nrc3130

Beadle, G.W., Tatum, E.L., 1941. Genetic Control of Biochemical Reactions in Neurospora. Proc. Natl. Acad. Sci. U. S. A. 27, 499–506.

Bedzhov, I., Graham, S.J.L., Leung, C.Y., Zernicka-Goetz, M., 2014. Developmental plasticity, cell fate specification and morphogenesis in the early mouse embryo. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 369. doi:10.1098/rstb.2013.0538

Bejerano, G., Lowe, C.B., Ahituv, N., King, B., Siepel, A., Salama, S.R., Rubin, E.M., Kent, W.J., Haussler, D., 2006. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. Nature 441, 87–90. doi:10.1038/nature04696

Bell, O., Tiwari, V.K., Thomä, N.H., Schübeler, D., 2011. Determinants and dynamics of genome accessibility. Nat. Rev. Genet. 12, 554–564. doi:10.1038/nrg3017

Bellefroid, E.J., Poncelet, D.A., Lecocq, P.J., Revelant, O., Martial, J.A., 1991. The evolutionarily conserved Krüppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. Proc. Natl. Acad. Sci. U. S. A. 88, 3608–3612.

Berdyshev, G.D., Korotaev, G.K., Boiarskikh, G.V., Vaniushin, B.F., 1967. [Nucleotide composition of DNA and RNA from somatic tissues of humpback and its changes during spawning]. Biokhimiia Mosc. Russ. 32, 988–993.

Berget, S.M., Sharp, P.A., 1977. A spliced sequence at the 5'-terminus of adenovirus late mRNA. Brookhaven Symp. Biol. 332–344.

Berndsen, C.E., Wolberger, C., 2014. New insights into ubiquitin E3 ligase mechanism. Nat. Struct. Mol. Biol. 21, 301–307. doi:10.1038/nsmb.2780

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S.L., Lander, E.S., 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125, 315–326. doi:10.1016/j.cell.2006.02.041

Bernstein, E., Duncan, E.M., Masui, O., Gil, J., Heard, E., Allis, C.D., 2006. Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin.

Mol. Cell. Biol. 26, 2560–2569. doi:10.1128/MCB.26.7.2560-2569.2006

Bestor, T., Laudano, A., Mattaliano, R., Ingram, V., 1988. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. J. Mol. Biol. 203, 971–983.

Bibikova, M., Fan, J.-B., 2009. GoldenGate assay for DNA methylation profiling. Methods Mol. Biol. Clifton NJ 507, 149–163. doi:10.1007/978-1-59745-522-0_12

Bird, A.P., Southern, E.M., 1978. Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from Xenopus laevis. J. Mol. Biol. 118, 27–47.

Black, J.C., Van Rechem, C., Whetstine, J.R., 2012. Histone lysine methylation dynamics: establishment, regulation, and biological impact. Mol. Cell 48, 491–507. doi:10.1016/j.molcel.2012.11.006

Blanco-Betancourt, C.E., Moncla, A., Milili, M., Jiang, Y.L., Viegas-Péquignot, E.M., Roquelaure, B., Thuret, I., Schiff, C., 2004. Defective B-cell-negative selection and terminal differentiation in the ICF syndrome. Blood 103, 2683–2690. doi:10.1182/blood-2003-08-2632

Blaschke, K., Ebata, K.T., Karimi, M.M., Zepeda-Martínez, J.A., Goyal, P., Mahapatra, S., Tam, A., Laird, D.J., Hirst, M., Rao, A., Lorincz, M.C., Ramalho-Santos, M., 2013. Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. Nature 500, 222–226. doi:10.1038/nature12362

Blasco, M.A., 2007. Telomere length, stem cells and aging. Nat. Chem. Biol. 3, 640–649. doi:10.1038/nchembio.2007.38

Boeke, J.D., Stoye, J.P., 1997. Retrotransposons, Endogenous Retroviruses, and the Evolution of Retroelements, in: Coffin, J.M., Hughes, S.H., Varmus, H.E. (Eds.), Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).

Bollati, V., Schwartz, J., Wright, R., Litonjua, A., Tarantini, L., Suh, H., Sparrow, D., Vokonas, P., Baccarelli, A., 2009. Decline in genomic DNA methylation through aging in a cohort of elderly subjects. Mech. Ageing Dev. 130, 234–239. doi:10.1016/j.mad.2008.12.003

Booth, M.J., Branco, M.R., Ficz, G., Oxley, D., Krueger, F., Reik, W., Balasubramanian, S., 2012. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Science 336, 934–937. doi:10.1126/science.1220671

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., Forné, T., Weber, M., 2010. Targets and dynamics of promoter DNA methylation during early mouse development. Nat. Genet. 42, 1093–1100. doi:10.1038/ng.708

Bostick, M., Kim, J.K., Estève, P.-O., Clark, A., Pradhan, S., Jacobsen, S.E., 2007. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science 317, 1760–1764. doi:10.1126/science.1147939

Bourc'his, D., Bestor, T.H., 2004. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature 431, 96–99. doi:10.1038/nature02886

Bourc'his, D., Miniou, P., Jeanpierre, M., Molina Gomes, D., Dupont, J., De Saint-Basile, G., Maraschio, P., Tiepolo, L., Viegas-Péquignot, E., 1999. Abnormal methylation does not prevent X inactivation in

ICF patients. Cytogenet. Cell Genet. 84, 245–252. doi:15269

Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B., Bestor, T.H., 2001. Dnmt3L and the establishment of maternal genomic imprints. Science 294, 2536–2539. doi:10.1126/science.1065848

Bourillot, P.-Y., Savatier, P., 2010. Krüppel-like transcription factors and control of pluripotency. BMC Biol. 8, 125.

Bouzinba-Segard, H., Guais, A., Francastel, C., 2006. Accumulation of small murine minor satellite transcripts leads to impaired centromeric architecture and function. Proc. Natl. Acad. Sci. 103, 8709–8714.

Boyes, J., Bird, A., 1992. Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. EMBO J. 11, 327.

Boyle, A.L., Ballard, S.G., Ward, D.C., 1990. Differential distribution of long and short interspersed element sequences in the mouse genome: chromosome karyotyping by fluorescence in situ hybridization. Proc. Natl. Acad. Sci. U. S. A. 87, 7757–7761.

Brady, T., Lee, Y.N., Ronen, K., Malani, N., Berry, C.C., Bieniasz, P.D., Bushman, F.D., 2009. Integration target site selection by a resurrected human endogenous retrovirus. Genes Dev. 23, 633–642. doi:10.1101/gad.1762309

Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A., Cedar, H., 1994. Sp1 elements protect a CpG island from de novo methylation. Nature 371, 435–438. doi:10.1038/371435a0

Breathnach, R., Mandel, J.L., Chambon, P., 1977. Ovalbumin gene is split in chicken DNA. Nature 270, 314–319.

Brien, G.L., Valerio, D.G., Armstrong, S.A., 2016. Exploiting the Epigenome to Control Cancer-Promoting Gene-Expression Programs. Cancer Cell 29, 464–476. doi:10.1016/j.ccell.2016.03.007

Briggs, R., King, T.J., 1952. Transplantation of living nuclei from blastula cells into enucleated frogs' eggs. Proc. Natl. Acad. Sci. 38, 455–463.

Brinkman, A.B., Gu, H., Bartels, S.J.J., Zhang, Y., Matarese, F., Simmer, F., Marks, H., Bock, C., Gnirke, A., Meissner, A., Stunnenberg, H.G., 2012. Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome Res. 22, 1128–1138. doi:10.1101/gr.133728.111

Brock, G.J., Charlton, J., Bird, A., 1999. Densely methylated sequences that are preferentially localized at telomere-proximal regions of human chromosomes. Gene 240, 269–277.

Brons, I.G.M., Smithers, L.E., Trotter, M.W.B., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A., Vallier, L., 2007. Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature 448, 191–195. doi:10.1038/nature05950

Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V., Kazazian, H.H., 2003. Hot L1s account for the bulk of retrotransposition in the human population. Proc. Natl. Acad. Sci. U. S. A. 100, 5280–5285. doi:10.1073/pnas.0831042100

Brown, S.W., 1966. Heterochromatin. Science 151, 417–425.

Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., Allis, C.D., 1996. Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell 84, 843–851.

Buchwitz, B.J., Ahmad, K., Moore, L.L., Roth, M.B., Henikoff, S., 1999. A histone-H3-like protein in C. elegans. Nature 401, 547–548. doi:10.1038/44062

Buck-Koehntop, B.A., Defossez, P.-A., 2013. On how mammalian transcription factors recognize methylated DNA. Epigenetics 8, 131–137. doi:10.4161/epi.23632

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., Wysocka, J., 2014. Reorganization of enhancer patterns in transition from naive to primed pluripotency. Cell Stem Cell 14, 838–853. doi:10.1016/j.stem.2014.04.003

Buehr, M., Meek, S., Blair, K., Yang, J., Ure, J., Silva, J., McLay, R., Hall, J., Ying, Q.-L., Smith, A., 2008. Capture of Authentic Embryonic Stem Cells from Rat Blastocysts. Cell 135, 1287–1298. doi:10.1016/j.cell.2008.12.007

Bulut-Karslioglu, A., De La Rosa-Velázquez, I.A., Ramirez, F., Barenboim, M., Onishi-Seebacher, M., Arand, J., Galán, C., Winter, G.E., Engist, B., Gerle, B., O'Sullivan, R.J., Martens, J.H.A., Walter, J., Manke, T., Lachner, M., Jenuwein, T., 2014. Suv39h-Dependent H3K9me3 Marks Intact Retrotransposons and Silences LINE Elements in Mouse Embryonic Stem Cells. Mol. Cell 55, 277–290. doi:10.1016/j.molcel.2014.05.029

Burton, A., Muller, J., Tu, S., Padilla-Longoria, P., Guccione, E., Torres-Padilla, M.-E., 2013. Single-Cell Profiling of Epigenetic Modifiers Identifies PRDM14 as an Inducer of Cell Fate in the Mammalian Embryo. Cell Rep. 5, 687–701. doi:10.1016/j.celrep.2013.09.044

Buschbeck, M., Hake, S.B., 2017. Variants of core histones and their roles in cell fate decisions, development and cancer. Nat. Rev. Mol. Cell Biol. doi:10.1038/nrm.2016.166

Buschbeck, M., Uribesalgo, I., Wibowo, I., Rué, P., Martin, D., Gutierrez, A., Morey, L., Guigó, R., López-Schier, H., Di Croce, L., 2009. The histone variant macroH2A is an epigenetic regulator of key developmental genes. Nat. Struct. Mol. Biol. 16, 1074–1079. doi:10.1038/nsmb.1665

Cai, K.Q., Capo-Chichi, C.D., Rula, M.E., Yang, D.-H., Xu, X.-X., 2008. Dynamic GATA6 expression in primitive endoderm formation and maturation in early mouse embryogenesis. Dev. Dyn. Off. Publ. Am. Assoc. Anat. 237, 2820–2829. doi:10.1002/dvdy.21703

Campisi, J., 2013. Aging, Cellular Senescence, and Cancer. Annu. Rev. Physiol. 75, 685–705. doi:10.1146/annurev-physiol-030212-183653

Camus, A., Perea-Gomez, A., Moreau, A., Collignon, J., 2006. Absence of Nodal signaling promotes precocious neural differentiation in the mouse embryo. Dev. Biol. 295, 743–755. doi:10.1016/j.ydbio.2006.03.047

Carone, D.M., Longo, M.S., Ferreri, G.C., Hall, L., Harris, M., Shook, N., Bulazel, K.V., Carone, B.R., Obergfell, C., O'Neill, M.J., O'Neill, R.J., 2009. A new class of retroviral and satellite encoded small RNAs emanates from mammalian centromeres. Chromosoma 118, 113–125. doi:10.1007/s00412-008-

0181-5

Carpenter, N.J., Filipovich, A., Blaese, R.M., Carey, T.L., Berkel, A.I., 1988. Variable immunodeficiency with abnormal condensation of the heterochromatin of chromosomes 1, 9, and 16. J. Pediatr. 112, 757–760.

Caspary, T., Cleary, M.A., Baker, C.C., Guan, X.J., Tilghman, S.M., 1998. Multiple mechanisms regulate imprinting of the mouse distal chromosome 7 gene cluster. Mol. Cell. Biol. 18, 3466–3474.

Cassidy, S.B., Schwartz, S., Miller, J.L., Driscoll, D.J., 2012. Prader-Willi syndrome. Genet. Med. Off. J. Am. Coll. Med. Genet. 14, 10–26. doi:10.1038/gim.0b013e31822bead0

Cerbone, M., Wang, J., Van der Maarel, S.M., D'Amico, A., D'Agostino, A., Romano, A., Brunetti-Pierri, N., 2012. Immunodeficiency, centromeric instability, facial anomalies (ICF) syndrome, due to *ZBTB24* mutations, presenting with large cerebral cyst. Am. J. Med. Genet. A. 158A, 2043–2046. doi:10.1002/ajmg.a.35486

Chaligné, R., Heard, E., 2014. X-chromosome inactivation in development and cancer. FEBS Lett. 588, 2514–2522. doi:10.1016/j.febslet.2014.06.023

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., Smith, A., 2003. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell 113, 643–655.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., Smith, A., 2007. Nanog safeguards pluripotency and mediates germline development. Nature 450, 1230–1234. doi:10.1038/nature06403

Chan, A.O.-O., Broaddus, R.R., Houlihan, P.S., Issa, J.-P.J., Hamilton, S.R., Rashid, A., 2002. CpG island methylation in aberrant crypt foci of the colorectum. Am. J. Pathol. 160, 1823–1830.

Chang, Y., Sun, L., Kokura, K., Horton, J.R., Fukuda, M., Espejo, A., Izumi, V., Koomen, J.M., Bedford, M.T., Zhang, X., Shinkai, Y., Fang, J., Cheng, X., 2011. MPP8 mediates the interactions between DNA methyltransferase Dnmt3a and H3K9 methyltransferase GLP/G9a. Nat. Commun. 2, 533. doi:10.1038/ncomms1549

Changolkar, L.N., Singh, G., Cui, K., Berletch, J.B., Zhao, K., Disteche, C.M., Pehrson, J.R., 2010. Genome-wide distribution of macroH2A1 histone variants in mouse liver chromatin. Mol. Cell. Biol. 30, 5473–5483. doi:10.1128/MCB.00518-10

Chapman, V., Forrester, L., Sanford, J., Hastie, N., Rossant, J., 1984. Cell lineage-specific undermethylation of mouse repetitive DNA. Nature 307, 284–286.

Chappell, G., Pogribny, I.P., Guyton, K.Z., Rusyn, I., 2016. Epigenetic alterations induced by genotoxic occupational and environmental human chemical carcinogens: A systematic literature review. Mutat. Res. Mutat. Res. 768, 27–45. doi:10.1016/j.mrrev.2016.03.004

Chaumeil, J., Okamoto, I., Guggiari, M., Heard, E., 2002. Integrated kinetics of X chromosome inactivation in differentiating embryonic stem cells. Cytogenet. Genome Res. 99, 75–84. doi:71577

Chazaud, C., Yamanaka, Y., Pawson, T., Rossant, J., 2006. Early Lineage Segregation between Epiblast and Primitive Endoderm in Mouse Blastocysts through the Grb2-MAPK Pathway. Dev. Cell 10, 615–

624. doi:10.1016/j.devcel.2006.02.020

Cheetham, S.W., Gruhl, F., Mattick, J.S., Dinger, M.E., 2013. Long noncoding RNAs and the genetics of cancer. Br. J. Cancer 108, 2419–2425. doi:10.1038/bjc.2013.233

Chen, E.S., Zhang, K., Nicolas, E., Cam, H.P., Zofall, M., Grewal, S.I.S., 2008. Cell cycle control of centromeric repeat transcription and heterochromatin assembly. Nature 451, 734–737. doi:10.1038/nature06561

Chen, J., Xu, X., 2010. Diet, epigenetic, and cancer prevention. Adv. Genet. 71, 237–255. doi:10.1016/B978-0-12-380864-6.00008-0

Chen, J.X., Zheng, Y., West, M., Tang, M.S., 1998. Carcinogens preferentially bind at methylated CpG in the p53 mutational hot spots. Cancer Res. 58, 2070–2075.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., Loh, Y.-H., Yeo, H.C., Yeo, Z.X., Narang, V., Govindarajan, K.R., Leong, B., Shahab, A., Ruan, Y., Bourque, G., Sung, W.-K., Clarke, N.D., Wei, C.-L., Ng, H.-H., 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell 133, 1106–1117. doi:10.1016/j.cell.2008.04.043

Chess, A., 2016. Monoallelic Gene Expression in Mammals. Annu. Rev. Genet. 50, 317–327. doi:10.1146/annurev-genet-120215-035120

Choo, K.H., 1997. Centromere DNA dynamics: latent centromeres and neocentromere formation. Am. J. Hum. Genet. 61, 1225–1233. doi:10.1086/301657

Chouery, E., Abou-Ghoch, J., Corbani, S., El Ali, N., Korban, R., Salem, N., Castro, C., Klayme, S., Azoury-Abou Rjeily, M., Khoury-Matar, R., Debo, G., Germanos-Haddad, M., Delague, V., Lefranc, G., Mégarbané, A., 2012. A novel deletion in *ZBTB24* in a Lebanese family with immunodeficiency, centromeric instability, and facial anomalies syndrome type 2. Clin. Genet. 82, 489–493. doi:10.1111/j.1399-0004.2011.01783.x

Chow, L.T., Gelinas, R.E., Broker, T.R., Roberts, R.J., 1977. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. Cell 12, 1–8.

Christman, J.K., 2002. 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. Oncogene 21, 5483–5495. doi:10.1038/sj.onc.1205699

Chuang, L.S., Ian, H.I., Koh, T.W., Ng, H.H., Xu, G., Li, B.F., 1997. Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science 277, 1996–2000.

Ciccone, D.N., Su, H., Hevi, S., Gay, F., Lei, H., Bajko, J., Xu, G., Li, E., Chen, T., 2009. KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. Nature 461, 415–418. doi:10.1038/nature08315

Ciferri, C., Lander, G.C., Maiolica, A., Herzog, F., Aebersold, R., Nogales, E., 2012. Molecular architecture of human polycomb repressive complex 2. eLife 1, e00005. doi:10.7554/eLife.00005

Cimmino, L., Aifantis, I., 2017. Alternative roles for oxidized mCs and TETs. Curr. Opin. Genet. Dev. 42, 1–7. doi:10.1016/j.gde.2016.11.003

Cobb, B.S., Nesterova, T.B., Thompson, E., Hertweck, A., O'Connor, E., Godwin, J., Wilson, C.B., Brockdorff, N., Fisher, A.G., Smale, S.T., Merkenschlager, M., 2005. T cell lineage choice and differentiation in the absence of the RNase III enzyme Dicer. J. Exp. Med. 201, 1367–1373. doi:10.1084/jem.20050572

Coit, P., Jeffries, M., Altorok, N., Dozmorov, M.G., Koelsch, K.A., Wren, J.D., Merrill, J.T., McCune, W.J., Sawalha, A.H., 2013. Genome-wide DNA methylation study suggests epigenetic accessibility and transcriptional poising of interferon-regulated genes in naïve CD4+ T cells from lupus patients. J. Autoimmun. 43, 78–84. doi:10.1016/j.jaut.2013.04.003

Condic, M.L., 2014. Totipotency: What It Is and What It Is Not. Stem Cells Dev. 23, 796–812. doi:10.1089/scd.2013.0364

Cooney, C.A., Dave, A.A., Wolff, G.L., 2002. Maternal methyl supplements in mice affect epigenetic variation and DNA methylation of offspring. J. Nutr. 132, 2393S–2400S.

Cooney, C.A., Matthews, H.R., 1985. The isolation of satellite DNA by density gradient centrifugation. Methods Mol. Biol. Clifton NJ 2, 21–29. doi:10.1385/0-89603-064-4:21

Cooper, S., Dienstbier, M., Hassan, R., Schermelleh, L., Sharif, J., Blackledge, N.P., De Marco, V., Elderkin, S., Koseki, H., Klose, R., Heger, A., Brockdorff, N., 2014. Targeting Polycomb to Pericentric Heterochromatin in Embryonic Stem Cells Reveals a Role for H2AK119u1 in PRC2 Recruitment. Cell Rep. 7, 1456–1470. doi:10.1016/j.celrep.2014.04.012

Costanzi, C., Pehrson, J.R., 1998. Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. Nature 393, 599–601. doi:10.1038/31275

Costanzi, C., Stein, P., Worrad, D.M., Schultz, R.M., Pehrson, J.R., 2000. Histone macroH2A1 is concentrated in the inactive X chromosome of female preimplantation mouse embryos. Dev. Camb. Engl. 127, 2283–2289.

Coulondre, C., Miller, J.H., Farabaugh, P.J., Gilbert, W., 1978. Molecular basis of base substitution hotspots in Escherichia coli. Nature 274, 775–780.

Couronné, L., Bastard, C., Bernard, O.A., 2012. TET2 and DNMT3A mutations in human T-cell lymphoma. N. Engl. J. Med. 366, 95–96.

Coviello-McLaughlin, G.M., Prowse, K.R., 1997. Telomere length regulation during postnatal development and ageing in Mus spretus. Nucleic Acids Res. 25, 3051–3058.

Crick, F.H., 1958. On protein synthesis. Symp. Soc. Exp. Biol. 12, 138–163.

Crider, K.S., Yang, T.P., Berry, R.J., Bailey, L.B., 2012. Folate and DNA Methylation: A Review of Molecular Mechanisms and the Evidence for Folate's Role. Adv. Nutr. Int. Rev. J. 3, 21–38. doi:10.3945/an.111.000992

Cross, S., Lindsey, J., Fantes, J., McKay, S., McGill, N., Cooke, H., 1990. The structure of a subterminal repeated sequence present on many human chromosomes. Nucleic Acids Res. 18, 6649–6657.

Crow, M.K., 2014. Type I Interferon in the Pathogenesis of Lupus. J. Immunol. 192, 5459–5468. doi:10.4049/jimmunol.1002795

Cruickshanks, H.A., McBryan, T., Nelson, D.M., VanderKraats, N.D., Shah, P.P., van Tuyn, J., Rai, T.S.,

Brock, C., Donahue, G., Dunican, D.S., Drotar, M.E., Meehan, R.R., Edwards, J.R., Berger, S.L., Adams, P.D., 2013. Senescent cells harbour features of the cancer epigenome. Nat. Cell Biol. 15, 1495–1506. doi:10.1038/ncb2879

Cruvinel, E., Budinetz, T., Germain, N., Chamberlain, S., Lalande, M., Martins-Taylor, K., 2014. Reactivation of maternal SNORD116 cluster via SETDB1 knockdown in Prader-Willi syndrome iPSCs. Hum. Mol. Genet. 23, 4674–4685. doi:10.1093/hmg/ddu187

Csankovszki, G., Nagy, A., Jaenisch, R., 2001. Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. J. Cell Biol. 153, 773–784.

Csankovszki, G., Panning, B., Bates, B., Pehrson, J.R., Jaenisch, R., 1999. Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. Nat. Genet. 22, 323–324. doi:10.1038/11887

Cubeñas-Potts, C., Matunis, M.J., 2013. SUMO: a multifaceted modifier of chromatin structure and function. Dev. Cell 24, 1–12. doi:10.1016/j.devcel.2012.11.020

Czechanski, A., Byers, C., Greenstein, I., Schrode, N., Donahue, L.R., Hadjantonakis, A.-K., Reinholdt, L.G., 2014. Derivation and characterization of mouse embryonic stem cells from permissive and nonpermissive strains. Nat. Protoc. 9, 559–574. doi:10.1038/nprot.2014.030

Dahm, R., 2005. Friedrich Miescher and the discovery of DNA. Dev. Biol. 278, 274–288. doi:10.1016/j.ydbio.2004.11.028

Dai, Q., Ren, A., Westholm, J.O., Serganov, A.A., Patel, D.J., Lai, E.C., 2013. The BEN domain is a novel sequence-specific DNA-binding domain conserved in neural transcriptional repressors. Genes Dev. 27, 602–614. doi:10.1101/gad.213314.113

Daniel, J.M., Spring, C.M., Crawford, H.C., Reynolds, A.B., Baig, A., 2002. The p120(ctn)-binding partner Kaiso is a bi-modal DNA-binding protein that recognizes both a sequence-specific consensus and methylated CpG dinucleotides. Nucleic Acids Res. 30, 2911–2919.

Daury, L., Chailleux, C., Bonvallet, J., Trouche, D., 2006. Histone H3.3 deposition at E2F-regulated genes is linked to transcription. EMBO Rep. 7, 66–71. doi:10.1038/sj.embor.7400561

Davies, M.N., Volta, M., Pidsley, R., Lunnon, K., Dixit, A., Lovestone, S., Coarfa, C., Harris, R.A., Milosavljevic, A., Troakes, C., Al-Sarraj, S., Dobson, R., Schalkwyk, L.C., Mill, J., 2012. Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. Genome Biol. 13, R43. doi:10.1186/gb-2012-13-6-r43

de Lange, T., Shiue, L., Myers, R.M., Cox, D.R., Naylor, S.L., Killery, A.M., Varmus, H.E., 1990. Structure and variability of human chromosome ends. Mol. Cell. Biol. 10, 518–527.

De Los Angeles, A., Loh, Y.-H., Tesar, P.J., Daley, G.Q., 2012. Accessing naïve human pluripotency. Curr. Opin. Genet. Dev. 22, 272–282. doi:10.1016/j.gde.2012.03.001

de Napoles, M., Mermoud, J.E., Wakao, R., Tang, Y.A., Endoh, M., Appanah, R., Nesterova, T.B., Silva, J., Otte, A.P., Vidal, M., Koseki, H., Brockdorff, N., 2004. Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. Dev. Cell 7, 663–676. doi:10.1016/j.devcel.2004.10.005

De Smet, C., Lurquin, C., Lethé, B., Martelange, V., Boon, T., 1999. DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. Mol. Cell. Biol. 19, 7327–7335.

de Greef, J.C., Wang, J., Balog, J., den Dunnen, J.T., Frants, R.R., Straasheijm, K.R., Aytekin, C., van der Burg, M., Duprez, L., Ferster, A., Gennery, A.R., Gimelli, G., Reisli, I., Schuetz, C., Schulz, A., Smeets, D.F.C.M., Sznajer, Y., Wijmenga, C., van Eggermond, M.C., van Ostaijen-ten Dam, M.M., Lankester, A.C., van Tol, M.J.D., van den Elsen, P.J., Weemaes, C.M., van der Maarel, S.M., 2011. Mutations in ZBTB24 Are Associated with Immunodeficiency, Centromeric Instability, and Facial Anomalies Syndrome Type 2. Am. J. Hum. Genet. 88, 796–804. doi:10.1016/j.ajhg.2011.04.018

Deininger, P.L., Batzer, M.A., 2002. Mammalian retroelements. Genome Res. 12, 1455–1465. doi:10.1101/gr.282402

DeLange, R.J., Smith, E.L., 1971. Histones: structure and function. Annu. Rev. Biochem. 40, 279–314. doi:10.1146/annurev.bi.40.070171.001431

Delhommeau, F., Dupont, S., Della Valle, V., James, C., Trannoy, S., Massé, A., Kosmider, O., Le Couedic, J.-P., Robert, F., Alberdi, A., Lécluse, Y., Plo, I., Dreyfus, F.J., Marzac, C., Casadevall, N., Lacombe, C., Romana, S.P., Dessen, P., Soulier, J., Viguié, F., Fontenay, M., Vainchenker, W., Bernard, O.A., 2009. Mutation in TET2 in myeloid cancers. N. Engl. J. Med. 360, 2289–2301. doi:10.1056/NEJMoa0810069

Deng, Z., Campbell, A.E., Lieberman, P.M., 2010. TERRA, CpG methylation, and telomere heterochromatin: Lessons from ICF syndrome cells. Cell Cycle 9, 69–74. doi:10.4161/cc.9.1.10358

Deng, Z., Norseen, J., Wiedmer, A., Riethman, H., Lieberman, P.M., 2009. TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. Mol. Cell 35, 403–413. doi:10.1016/j.molcel.2009.06.025

Dennis, K., Fan, T., Geiman, T., Yan, Q., Muegge, K., 2001. Lsh, a member of the SNF2 family, is required for genome-wide methylation. Genes Dev. 15, 2940–2944. doi:10.1101/gad.929101

Desai, M.A., Webb, H.D., Sinanan, L.M., Scarsdale, J.N., Walavalkar, N.M., Ginder, G.D., Williams, D.C., 2015. An intrinsically disordered region of methyl-CpG binding domain protein 2 (MBD2) recruits the histone deacetylase core of the NuRD complex. Nucleic Acids Res. 43, 3100–3113. doi:10.1093/nar/gkv168

Dewannieux, M., Heidmann, T., 2005. Role of poly(A) tail length in Alu retrotransposition. Genomics 86, 378–381. doi:10.1016/j.ygeno.2005.05.009

Dhalluin, C., Carlson, J.E., Zeng, L., He, C., Aggarwal, A.K., Zhou, M.M., 1999. Structure and ligand of a histone acetyltransferase bromodomain. Nature 399, 491–496. doi:10.1038/20974

Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R.Z., Ragozin, S., Jeltsch, A., 2010. The Dnmt3a PWWP Domain Reads Histone 3 Lysine 36 Trimethylation and Guides DNA Methylation. J. Biol. Chem. 285, 26114–26120. doi:10.1074/jbc.M109.089433

Diaz, M.O., Barsacchi-Pilone, G., Mahon, K.A., Gall, J.G., 1981. Transcripts from both strands of a satellite DNA occur on lampbrush chromosome loops of the newt Notophthalmus. Cell 24, 649–659.

Dietrich, J.-E., Hiiragi, T., 2007. Stochastic patterning in the mouse pre-implantation embryo. Dev. Camb. Engl. 134, 4219–4231. doi:10.1242/dev.003798

Djureinovic, D., Fagerberg, L., Hallström, B., Danielsson, A., Lindskog, C., Uhlén, M., Pontén, F., 2014. The human testis-specific proteome defined by transcriptomics and antibody-based profiling. Mol. Hum. Reprod. 20, 476–488. doi:10.1093/molehr/gau018

Do, D.V., Ueda, J., Messerschmidt, D.M., Lorthongpanich, C., Zhou, Y., Feng, B., Guo, G., Lin, P.J., Hossain, M.Z., Zhang, W., Moh, A., Wu, Q., Robson, P., Ng, H.H., Poellinger, L., Knowles, B.B., Solter, D., Fu, X.-Y., 2013. A genetic and developmental pathway from STAT3 to the OCT4-NANOG circuit is essential for maintenance of ICM lineages in vivo. Genes Dev. 27, 1378–1390. doi:10.1101/gad.221176.113

Dodge, J.E., Kang, Y.-K., Beppu, H., Lei, H., Li, E., 2004. Histone H3-K9 Methyltransferase ESET Is Essential for Early Development. Mol. Cell. Biol. 24, 2478–2486. doi:10.1128/MCB.24.6.2478-2486.2004

Dolinoy, D.C., Huang, D., Jirtle, R.L., 2007. Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. Proc. Natl. Acad. Sci. U. S. A. 104, 13056–13061. doi:10.1073/pnas.0703739104

Dong, H., Bonala, R.R., Suzuki, N., Johnson, F., Grollman, A.P., Shibutani, S., 2004. Mutagenic potential of benzo[a]pyrene-derived DNA adducts positioned in codon 273 of the human P53 gene. Biochemistry (Mosc.) 43, 15922–15928. doi:10.1021/bi0482194

Du, Q., Luu, P.-L., Stirzaker, C., Clark, S.J., 2015. Methyl-CpG-binding domain proteins: readers of the epigenome. Epigenomics 7, 1051–1073. doi:10.2217/epi.15.39

Dunham, I., *et al*., 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74. doi:10.1038/nature11247

Dunican, D.S., Cruickshanks, H.A., Suzuki, M., Semple, C.A., Davey, T., Arceci, R.J., Greally, J., Adams, I.R., Meehan, R.R., 2013. Lsh regulates LTR retrotransposon repression independently of Dnmt3b function. Genome Biol. 14, R146. doi:10.1186/gb-2013-14-12-r146

Dunleavy, E.M., Almouzni, G., Karpen, G.H., 2011. H3.3 is deposited at centromeres in S phase as a placeholder for newly assembled CENP-A in G_1 phase. Nucleus 2, 146–157. doi:10.4161/nucl.2.2.15211

Eads, C.A., Danenberg, K.D., Kawakami, K., Saltz, L.B., Blake, C., Shibata, D., Danenberg, P.V., Laird, P.W., 2000. MethyLight: a high-throughput assay to measure DNA methylation. Nucleic Acids Res. 28, E32.

Earnshaw, W.C., Rothfield, N., 1985. Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. Chromosoma 91, 313–321.

Eberl, H.C., Spruijt, C.G., Kelstrup, C.D., Vermeulen, M., Mann, M., 2013. A Map of General and Specialized Chromatin Readers in Mouse Tissues Generated by Label-free Interaction Proteomics. Mol. Cell 49, 368–378. doi:10.1016/j.molcel.2012.10.026

Eckersley-Maslin, M.A., Spector, D.L., 2014. Random monoallelic expression: regulating gene expression

one allele at a time. Trends Genet. TIG 30, 237–244. doi:10.1016/j.tig.2014.03.003

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V.K., Attwood, J., Burger, M., Burton, J., Cox, T.V., Davies, R., Down, T.A., Haefliger, C., Horton, R., Howe, K., Jackson, D.K., Kunde, J., Koenig, C., Liddle, J., Niblett, D., Otto, T., Pettett, R., Seemann, S., Thompson, C., West, T., Rogers, J., Olek, A., Berlin, K., Beck, S., 2006. DNA methylation profiling of human chromosomes 6, 20 and 22. Nat. Genet. 38, 1378–1385. doi:10.1038/ng1909

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D.G., Buetow, K.H., Gingeras, T.R., Misteli, T., Meshorer, E., 2008. Global Transcription in Pluripotent Embryonic Stem Cells. Cell Stem Cell 2, 437–447. doi:10.1016/j.stem.2008.03.021

Ehrlich, M., 2009. DNA hypomethylation in cancer cells. Epigenomics 1, 239–259. doi:10.2217/epi.09.33

Ehrlich, M., 2002. DNA methylation in cancer: too much, but also too little. Oncogene 21, 5400–5413. doi:10.1038/sj.onc.1205651

Ehrlich, M., Buchanan, K.L., Tsien, F., Jiang, G., Sun, B., Uicker, W., Weemaes, C.M., Smeets, D., Sperling, K., Belohradsky, B.H., Tommerup, N., Misek, D.E., Rouillard, J.M., Kuick, R., Hanash, S.M., 2001. DNA methyltransferase 3B mutations linked to the ICF syndrome cause dysregulation of lymphogenesis genes. Hum. Mol. Genet. 10, 2917–2931.

Ehrlich, M., Jackson, K., Weemaes, C., 2006. Immunodeficiency, centromeric region instability, facial anomalies syndrome (ICF). Orphanet J. Rare Dis. 1, 2. doi:10.1186/1750-1172-1-2

Elgin, S.C.R., Grewal, S.I.S., 2003. Heterochromatin: silence is golden. Curr. Biol. CB 13, R895-898.

ENCODE Project Consortium, 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74. doi:10.1038/nature11247

Epsztejn-Litman, S., Feldman, N., Abu-Remaileh, M., Shufaro, Y., Gerson, A., Ueda, J., Deplus, R., Fuks, F., Shinkai, Y., Cedar, H., Bergman, Y., 2008. De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. Nat. Struct. Mol. Biol. 15, 1176–1183. doi:10.1038/nsmb.1476

Ergun, S., Buschmann, C., Heukeshoven, J., Dammann, K., Schnieders, F., Lauke, H., Chalajour, F., Kilic, N., Stratling, W.H., Schumann, G.G., 2004. Cell Type-specific Expression of LINE-1 Open Reading Frames 1 and 2 in Fetal and Adult Human Tissues. J. Biol. Chem. 279, 27753–27763. doi:10.1074/jbc.M312985200

Escamilla-Del-Arenal, M., da Rocha, S.T., Heard, E., 2011. Evolutionary diversity and developmental regulation of X-chromosome inactivation. Hum. Genet. 130, 307–327. doi:10.1007/s00439-011-1029-2

Eskeland, R., Leeb, M., Grimes, G.R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A.I., Wutz, A., Bickmore, W.A., 2010. Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. Mol. Cell 38, 452–464. doi:10.1016/j.molcel.2010.02.032

Esteller, M., 2008. Epigenetics in cancer. N. Engl. J. Med. 358, 1148–1159. doi:10.1056/NEJMra072067

Eymery, A., Horard, B., El Atifi-Borel, M., Fourel, G., Berger, F., Vitte, A.-L., Van den Broeck, A., Brambilla, E., Fournier, A., Callanan, M., Gazzeri, S., Khochbin, S., Rousseaux, S., Gilson, E., Vourc'h,

C., 2009. A transcriptomic analysis of human centromeric and pericentric sequences in normal and tumor cells. Nucleic Acids Res. 37, 6340–6354. doi:10.1093/nar/gkp639

Fachinetti, D., Diego Folco, H., Nechemia-Arbely, Y., Valente, L.P., Nguyen, K., Wong, A.J., Zhu, Q., Holland, A.J., Desai, A., Jansen, L.E.T., Cleveland, D.W., 2013. A two-step mechanism for epigenetic specification of centromere identity and function. Nat. Cell Biol. 15, 1056–1066. doi:10.1038/ncb2805

Fadloun, A., Eid, A., Torres-Padilla, M.-E., 2013. Mechanisms and Dynamics of Heterochromatin Formation During Mammalian Development, in: Current Topics in Developmental Biology. Elsevier, pp. 1–45. doi:10.1016/B978-0-12-416027-9.00001-2

Fang, J., Chen, T., Chadwick, B., Li, E., Zhang, Y., 2004. Ring1b-mediated H2A ubiquitination associates with inactive X chromosomes and is involved in initiation of X inactivation. J. Biol. Chem. 279, 52812–52815. doi:10.1074/jbc.C400493200

Farthing, C.R., Ficz, G., Ng, R.K., Chan, C.-F., Andrews, S., Dean, W., Hemberger, M., Reik, W., 2008. Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. PLoS Genet. 4, e1000116. doi:10.1371/journal.pgen.1000116

Fatemi, M., Hermann, A., Pradhan, S., Jeltsch, A., 2001. The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. J. Mol. Biol. 309, 1189–1199. doi:10.1006/jmbi.2001.4709

Faulkner, G.J., Carninci, P., 2009. Altruistic functions for selfish DNA. Cell Cycle 8, 2895–2900. doi:10.4161/cc.8.18.9536

Feil, R., Fraga, M.F., 2012. Epigenetics and the environment: emerging patterns and implications. Nat. Rev. Genet. doi:10.1038/nrg3142

Feinberg, A.P., Vogelstein, B., 1983. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature 301, 89–92.

Feldmann, A., Ivanek, R., Murr, R., Gaidatzis, D., Burger, L., Schübeler, D., 2013. Transcription Factor Occupancy Can Mediate Active Turnover of DNA Methylation at Regulatory Regions. PLoS Genet. 9, e1003994. doi:10.1371/journal.pgen.1003994

Felle, M., Hoffmeister, H., Rothammer, J., Fuchs, A., Exler, J.H., Längst, G., 2011. Nucleosomes protect DNA from DNA methylation in vivo and in vitro. Nucleic Acids Res. 39, 6956–6969. doi:10.1093/nar/gkr263

Felsenfeld, G., 2014. A Brief History of Epigenetics. Cold Spring Harb. Perspect. Biol. 6, a018200–a018200. doi:10.1101/cshperspect.a018200

Feng, J., Zhou, Y., Campbell, S.L., Le, T., Li, E., Sweatt, J.D., Silva, A.J., Fan, G., 2010. Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. Nat. Neurosci. 13, 423–430. doi:10.1038/nn.2514

Feng, Q., Moran, J.V., Kazazian, H.H., Boeke, J.D., 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell 87, 905–916.

Feng, Q., Schumann, G., Boeke, J.D., 1998. Retrotransposon R1Bm endonuclease cleaves the target

sequence. Proc. Natl. Acad. Sci. U. S. A. 95, 2083–2088.

Feng, Q., Zhang, Y., 2001. The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. Genes Dev. 15, 827–832. doi:10.1101/gad.876201

Ferguson-Smith, A.C., Sasaki, H., Cattanach, B.M., Surani, M.A., 1993. Parental-origin-specific epigenetic modification of the mouse H19 gene. Nature 362, 751–755. doi:10.1038/362751a0

Ferguson-Smith, A.C., Surani, M.A., 2001. Imprinting and the epigenetic asymmetry between parental genomes. Science 293, 1086–1089. doi:10.1126/science.1064020

Ferreira, D., Meles, S., Escudeiro, A., Mendes-da-Silva, A., Adega, F., Chaves, R., 2015. Satellite non-coding RNAs: the emerging players in cells, cellular pathways and cancer. Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol. 23, 479–493. doi:10.1007/s10577-015-9482-8

Ferri, F., Bouzinba-Segard, H., Velasco, G., Hube, F., Francastel, C., 2009. Non-coding murine centromeric transcripts associate with and potentiate Aurora B kinase. Nucleic Acids Res. 37, 5071–5080. doi:10.1093/nar/gkp529

Feschotte, C., Pritham, E.J., 2007. DNA Transposons and the Evolution of Eukaryotic Genomes. Annu. Rev. Genet. 41, 331–368. doi:10.1146/annurev.genet.40.110405.090448

Filion, G.J.P., Zhenilo, S., Salozhin, S., Yamada, D., Prokhortchouk, E., Defossez, P.-A., 2006. A Family of Human Zinc Finger Proteins That Bind Methylated DNA and Repress Transcription. Mol. Cell. Biol. 26, 169–181. doi:10.1128/MCB.26.1.169-181.2006

Finch, J.T., Klug, A., 1976. Solenoidal model for superstructure in chromatin. Proc. Natl. Acad. Sci. U. S. A. 73, 1897–1901.

Finch, J.T., Lutter, L.C., Rhodes, D., Brown, R.S., Rushton, B., Levitt, M., Klug, A., 1977. Structure of nucleosome core particles of chromatin. Nature 269, 29–36.

Fischle, W., Tseng, B.S., Dormann, H.L., Ueberheide, B.M., Garcia, B.A., Shabanowitz, J., Hunt, D.F., Funabiki, H., Allis, C.D., 2005. Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. Nature 438, 1116–1122. doi:10.1038/nature04219

Francastel, C., Schübeler, D., Martin, D.I., Groudine, M., 2000. Nuclear compartmentalization and gene activity. Nat. Rev. Mol. Cell Biol. 1, 137–143. doi:10.1038/35040083

Francis, N.J., Kingston, R.E., Woodcock, C.L., 2004. Chromatin compaction by a polycomb group protein complex. Science 306, 1574–1577. doi:10.1126/science.1100576

Friedenstein, A.J., Chailakhyan, R.K., Latsinik, N.V., Panasyuk, A.F., Keiliss-Borok, I.V., 1974. Stromal cells responsible for transferring the microenvironment of the hemopoietic tissues. Cloning in vitro and retransplantation in vivo. Transplantation 17, 331–340.

Fryns, J.P., Azou, M., Jaeken, J., Eggermont, E., Pedersen, J.C., Van den Berghe, H., 1981. Centromeric instability of chromosomes 1, 9, and 16 associated with combined immunodeficiency. Hum. Genet. 57, 108–110.

Fukagawa, T., Nogami, M., Yoshikawa, M., Ikeno, M., Okazaki, T., Takami, Y., Nakayama, T., Oshimura, M., 2004. Dicer is essential for formation of the heterochromatin structure in vertebrate cells. Nat. Cell

Biol. 6, 784–791. doi:10.1038/ncb1155

Fuks, F., 2003. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Res. 31, 2305–2312. doi:10.1093/nar/gkg332

Furano, A.V., Robb, S.M., Robb, F.T., 1988. The structure of the regulatory region of the rat L1 (L1Rn, long interspersed repeated) DNA family of transposable elements. Nucleic Acids Res. 16, 9215–9231.

Gamble, M.J., Kraus, W.L., 2010. Multiple facets of the unique histone variant macroH2A: from genomics to cell biology. Cell Cycle Georget. Tex 9, 2568–2574. doi:10.4161/cc.9.13.12144

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., Reinberg, D., 2012. PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Mol. Cell 45, 344–356. doi:10.1016/j.molcel.2012.01.002

García-Cao, M., O'Sullivan, R., Peters, A.H.F.M., Jenuwein, T., Blasco, M.A., 2004. Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. Nat. Genet. 36, 94–99. doi:10.1038/ng1278

Gaudet, F., Rideout, W.M., Meissner, A., Dausman, J., Leonhardt, H., Jaenisch, R., 2004. Dnmt1 Expression in Pre- and Postimplantation Embryogenesis and the Maintenance of IAP Silencing. Mol. Cell. Biol. 24, 1640–1648. doi:10.1128/MCB.24.4.1640-1648.2004

Geiduschek, E.P., Kassavetis, G.A., 2001. The RNA polymerase III transcription apparatus. J. Mol. Biol. 310, 1–26. doi:10.1006/jmbi.2001.4732

Geijsen, N., Horoschak, M., Kim, K., Gribnau, J., Eggan, K., Daley, G.Q., 2004. Derivation of embryonic germ cells and male gametes from embryonic stem cells. Nature 427, 148–154. doi:10.1038/nature02247

Gendrel, A.-V., Apedaile, A., Coker, H., Termanis, A., Zvetkova, I., Godwin, J., Tang, Y.A., Huntley, D., Montana, G., Taylor, S., Giannoulatou, E., Heard, E., Stancheva, I., Brockdorff, N., 2012. Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. Dev. Cell 23, 265–279. doi:10.1016/j.devcel.2012.06.011

Gilbert, S.L., Pehrson, J.R., Sharp, P.A., 2000. XIST RNA associates with specific regions of the inactive X chromatin. J. Biol. Chem. 275, 36491–36494. doi:10.1074/jbc.C000409200

Goldman, M.A., Holmquist, G.P., Gray, M.C., Caston, L.A., Nag, A., 1984. Replication timing of genes and middle repetitive sequences. Science 224, 686–692.

Goll, M.G., Bestor, T.H., 2005. Eukaryotic cytosine methyltransferases. Annu. Rev. Biochem. 74, 481–514. doi:10.1146/annurev.biochem.74.010904.153721

Gopalakrishnan, S., Sullivan, B.A., Trazzi, S., Della Valle, G., Robertson, K.D., 2009. DNMT3B interacts with constitutive centromere protein CENP-C to modulate DNA methylation and the histone code at centromeric regions. Hum. Mol. Genet. 18, 3178–3193. doi:10.1093/hmg/ddp256

Gordon, C.A., Hartono, S.R., Chédin, F., 2013. Inactive DNMT3B splice variants modulate de novo DNA methylation. PloS One 8, e69486. doi:10.1371/journal.pone.0069486

Gowher, H., Jeltsch, A., 2002. Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases. J. Biol. Chem. 277, 20409–20414. doi:10.1074/jbc.M202148200

Graur, D., Zheng, Y., Price, N., Azevedo, R.B.R., Zufall, R.A., Elhaik, E., 2013. On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. Genome Biol. Evol. 5, 578–590. doi:10.1093/gbe/evt028

Graziano, V., Gerchman, S.E., Schneider, D.K., Ramakrishnan, V., 1994. Histone H1 is located in the interior of the chromatin 30-nm filament. Nature 368, 351–354. doi:10.1038/368351a0

Greer, E.L., Shi, Y., 2012. Histone methylation: a dynamic mark in health, disease and inheritance. Nat. Rev. Genet. 13, 343–357. doi:10.1038/nrg3173

Greger, V., Passarge, E., Höpping, W., Messmer, E., Horsthemke, B., 1989. Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. Hum. Genet. 83, 155–158.

Grewal, S.I.S., Elgin, S.C.R., 2007. Transcription and RNA interference in the formation of heterochromatin. Nature 447, 399–406. doi:10.1038/nature05914

Grønbaek, K., Hother, C., Jones, P.A., 2007. Epigenetic changes in cancer. APMIS Acta Pathol. Microbiol. Immunol. Scand. 115, 1039–1059. doi:10.1111/j.1600-0463.2007.apm_636.xml.x

Groudine, M., Holtzer, H., Scherrer, K., Therwath, A., 1974. Lineage-dependent transcription of globin genes. Cell 3, 243–247.

Grove, G.W., Zweidler, A., 1984. Regulation of nucleosomal core histone variant levels in differentiating murine erythroleukemia cells. Biochemistry (Mosc.) 23, 4436–4443.

Gu, T.-P., Guo, F., Yang, H., Wu, H.-P., Xu, G.-F., Liu, W., Xie, Z.-G., Shi, L., He, X., Jin, S., Iqbal, K., Shi, Y.G., Deng, Z., Szabó, P.E., Pfeifer, G.P., Li, J., Xu, G.-L., 2011. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature 477, 606–610. doi:10.1038/nature10443

Guenatri, M., Bailly, D., Maison, C., Almouzni, G., 2004. Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. J. Cell Biol. 166, 493–505. doi:10.1083/jcb.200403109

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., Young, R.A., 2007. A chromatin landmark and transcription initiation at most promoters in human cells. Cell 130, 77–88. doi:10.1016/j.cell.2007.05.042

Guo, G., Yang, J., Nichols, J., Hall, J.S., Eyres, I., Mansfield, W., Smith, A., 2009. Klf4 reverts developmentally programmed restriction of ground state pluripotency. Dev. Camb. Engl. 136, 1063–1069. doi:10.1242/dev.030957

Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., Tian, C., Wang, J., Cong, Y., Xu, Y., 2015. Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature 517, 640–644. doi:10.1038/nature13899

Gurard-Levin, Z.A., Almouzni, G., 2014. Histone modifications and a choice of variant: a language that helps the genome express itself. F1000Prime Rep. 6. doi:10.12703/P6-76

Gurdon, J.B., 1962a. The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. J. Embryol. Exp. Morphol. 10, 622–640.

Gurdon, J.B., 1962b. Multiple genetically identical frogs. J. Hered. 53, 5–9.

Gurdon, J.B., 1962c. The transplantation of nuclei between two species of Xenopus. Dev. Biol. 5, 68–83.

Gusel'nikova, V.V., Korzhevskiy, D.E., 2015. NeuN As a Neuronal Nuclear Antigen and Neuron

Differentiation Marker. Acta Naturae 7, 42–47.

Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H.D., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C., Benedetti, R., Altucci, L., Jansen, J.H., Walter, J., Gut, I.G., Marks, H., Stunnenberg, H.G., 2013. Whole-Genome Bisulfite Sequencing of Two Distinct Interconvertible DNA Methylomes of Mouse Embryonic Stem Cells. Cell Stem Cell 13, 360–369. doi:10.1016/j.stem.2013.06.002

Hackett, J.A., Reddington, J.P., Nestor, C.E., Dunican, D.S., Branco, M.R., Reichmann, J., Reik, W., Surani, M.A., Adams, I.R., Meehan, R.R., 2012. Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. Dev. Camb. Engl. 139, 3623–3632. doi:10.1242/dev.081661

Hagarman, J.A., Motley, M.P., Kristjansdottir, K., Soloway, P.D., 2013. Coordinate Regulation of DNA Methylation and H3K27me3 in Mouse Embryonic Stem Cells. PLoS ONE 8, e53880. doi:10.1371/journal.pone.0053880

Hagleitner, M.M., Lankester, A., Maraschio, P., Hulten, M., Fryns, J.P., Schuetz, C., Gimelli, G., Davies, E.G., Gennery, A., Belohradsky, B.H., de Groot, R., Gerritsen, E.J.A., Mattina, T., Howard, P.J., Fasth, A., Reisli, I., Furthner, D., Slatter, M.A., Cant, A.J., Cazzola, G., van Dijken, P.J., van Deuren, M., de Greef, J.C., van der Maarel, S.M., Weemaes, C.M.R., 2007. Clinical spectrum of immunodeficiency, centromeric instability and facial dysmorphism (ICF syndrome). J. Med. Genet. 45, 93–99. doi:10.1136/jmg.2007.053397

Hahn, M.A., Wu, X., Li, A.X., Hahn, T., Pfeifer, G.P., 2011. Relationship between gene body DNA methylation and intragenic H3K9me3 and H3K36me3 chromatin marks. PloS One 6, e18844. doi:10.1371/journal.pone.0018844

Hake, S.B., Allis, C.D., 2006. Histone H3 variants and their potential role in indexing mammalian genomes: the “H3 barcode hypothesis.” Proc. Natl. Acad. Sci. U. S. A. 103, 6428–6435. doi:10.1073/pnas.0600803103

Hall, I.M., Shankaranarayana, G.D., Noma, K.-I., Ayoub, N., Cohen, A., Grewal, S.I.S., 2002. Establishment and maintenance of a heterochromatin domain. Science 297, 2232–2237. doi:10.1126/science.1076466

Han, M., Grunstein, M., 1988. Nucleosome loss activates yeast downstream promoters in vivo. Cell 55, 1137–1145.

Hansen, R.S., Stöger, R., Wijmenga, C., Stanek, A.M., Canfield, T.K., Luo, P., Matarazzo, M.R., D’Esposito, M., Feil, R., Gimelli, G., Weemaes, C.M., Laird, C.D., Gartler, S.M., 2000. Escape from gene silencing in ICF syndrome: evidence for advanced replication time as a major determinant. Hum. Mol. Genet. 9, 2575–2587.

Hansen, R.S., Wijmenga, C., Luo, P., Stanek, A.M., Canfield, T.K., Weemaes, C.M., Gartler, S.M., 1999. The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proc. Natl. Acad. Sci. U. S. A. 96, 14412–14417.

Hänzelmann, S., Beier, F., Gusmao, E.G., Koch, C.M., Hummel, S., Charapitsa, I., Joussen, S., Benes, V.,

Brümmendorf, T.H., Reid, G., Costa, I.G., Wagner, W., 2015. Replicative senescence is associated with nuclear reorganization and with DNA methylation at specific transcription factor binding sites. Clin. Epigenetics 7, 19. doi:10.1186/s13148-015-0057-5

Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M., Tilghman, S.M., 2000. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature 405, 486–489. doi:10.1038/35013106

Hartwell, L.H., Kastan, M.B., 1994. Cell cycle control and cancer. Science 266, 1821–1828.

Hata, K., Okano, M., Lei, H., Li, E., 2002. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Dev. Camb. Engl. 129, 1983–1993.

Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S., Antosiewicz-Bourget, J., Ye, Z., Espinoza, C., Agarwahl, S., Shen, L., Ruotti, V., Wang, W., Stewart, R., Thomson, J.A., Ecker, J.R., Ren, B., 2010. Distinct Epigenomic Landscapes of Pluripotent and Lineage-Committed Human Cells. Cell Stem Cell 6, 479–491. doi:10.1016/j.stem.2010.03.018

Hayashi, K., Lopes, S.M.C. de S., Tang, F., Surani, M.A., 2008. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. Cell Stem Cell 3, 391–401. doi:10.1016/j.stem.2008.07.027

He, Z., Li, J., Zhen, C., Feng, L., Ding, X., 2006. Effect of leukemia inhibitory factor on embryonic stem cell differentiation: implications for supporting neuronal differentiation. Acta Pharmacol. Sin. 27, 80–90. doi:10.1111/j.1745-7254.2006.00254.x

Heard, E., 2004. Recent advances in X-chromosome inactivation. Curr. Opin. Cell Biol. 16, 247–255. doi:10.1016/j.ceb.2004.03.005

Heard, E., Disteche, C.M., 2006. Dosage compensation in mammals: fine-tuning the expression of the X chromosome. Genes Dev. 20, 1848–1867. doi:10.1101/gad.1422906

Heard, E., Rougeulle, C., Arnaud, D., Avner, P., Allis, C.D., Spector, D.L., 2001. Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. Cell 107, 727–738.

Hédouin, S., Grillo, G., Ivkovic, I., Velasco, G., Francastel, C., 2017. CENP-A chromatin disassembly in stressed and senescent murine cells. Sci. Rep. 7, 42520. doi:10.1038/srep42520

Hendrich, B., Bird, A., 1998. Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol. Cell. Biol. 18, 6538–6547.

Hendrich, B., Tweedie, S., 2003. The methyl-CpG binding domain and the evolving role of DNA methylation in animals. Trends Genet. 19, 269–277. doi:10.1016/S0168-9525(03)00080-5

Henikoff, S., 2000. Heterochromatin function in complex genomes. Biochim. Biophys. Acta 1470, O1-8.

Hergeth, S.P., Schneider, R., 2015. The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. EMBO Rep. 16, 1439–1453. doi:10.15252/embr.201540749

Hernández-Muñoz, I., Lund, A.H., van der Stoop, P., Boutsma, E., Muijrers, I., Verhoeven, E., Nusinow, D.A., Panning, B., Marahrens, Y., van Lohuizen, M., 2005. Stable X chromosome inactivation involves the PRC1 Polycomb complex and requires histone MACROH2A1 and the CULLIN3/SPOP ubiquitin

E3 ligase. Proc. Natl. Acad. Sci. U. S. A. 102, 7635–7640. doi:10.1073/pnas.0408918102

Hershko, A., Ciechanover, A., 1998. The ubiquitin system. Annu. Rev. Biochem. 67, 425–479. doi:10.1146/annurev.biochem.67.1.425

Hervouet, E., Debien, E., Campion, L., Charbord, J., Menanteau, J., Vallette, F.M., Cartron, P.-F., 2009. Folate Supplementation Limits the Aggressiveness of Glioma via the Remethylation of DNA Repeats Element and Genes Governing Apoptosis and Proliferation. Clin. Cancer Res. 15, 3519–3529. doi:10.1158/1078-0432.CCR-08-2062

Hervouet, E., Vallette, F.M., Cartron, P.-F., 2010. Dnmt1/Transcription factor interactions: an alternative mechanism of DNA methylation inheritance. Genes Cancer 1, 434–443. doi:10.1177/1947601910373794

Heun, P., Erhardt, S., Blower, M.D., Weiss, S., Skora, A.D., Karpen, G.H., 2006. Mislocalization of the Drosophila centromere-specific histone CID promotes formation of functional ectopic kinetochores. Dev. Cell 10, 303–315. doi:10.1016/j.devcel.2006.01.014

Heyn, H., Li, N., Ferreira, H.J., Moran, S., Pisano, D.G., Gomez, A., Diez, J., Sanchez-Mut, J.V., Setien, F., Carmona, F.J., Puca, A.A., Sayols, S., Pujana, M.A., Serra-Musach, J., Iglesias-Platas, I., Formiga, F., Fernandez, A.F., Fraga, M.F., Heath, S.C., Valencia, A., Gut, I.G., Wang, J., Esteller, M., 2012. Distinct DNA methylomes of newborns and centenarians. Proc. Natl. Acad. Sci. 109, 10522–10527. doi:10.1073/pnas.1120658109

Heyn, H., Vidal, E., Sayols, S., Sanchez-Mut, J.V., Moran, S., Medina, I., Sandoval, J., Simó-Riudalbas, L., Szczesna, K., Huertas, D., Gatto, S., Matarazzo, M.R., Dopazo, J., Esteller, M., 2012. Whole-genome bisulfite DNA sequencing of a DNMT3B mutant patient. Epigenetics 7, 542–550. doi:10.4161/epi.20523

Hirai, H., Karian, P., Kikyo, N., 2011. Regulation of embryonic stem cell self-renewal and pluripotency by leukaemia inhibitory factor. Biochem. J. 438, 11–23. doi:10.1042/BJ20102152

Hirasawa, R., Chiba, H., Kaneda, M., Tajima, S., Li, E., Jaenisch, R., Sasaki, H., 2008. Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. Genes Dev. 22, 1607–1616. doi:10.1101/gad.1667008

Hirasawa, R., Feil, R., 2008. A KRAB Domain Zinc Finger Protein in Imprinting and Disease. Dev. Cell 15, 487–488. doi:10.1016/j.devcel.2008.09.006

Hirasawa, R., Sasaki, H., 2009. Dynamic transition of Dnmt3b expression in mouse pre- and early post-implantation embryos. Gene Expr. Patterns GEP 9, 27–30. doi:10.1016/j.gep.2008.09.002

Hirota, T., Lipp, J.J., Toh, B.-H., Peters, J.-M., 2005. Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin. Nature 438, 1176–1180. doi:10.1038/nature04254

Hirst, M., Marra, M.A., 2009. Epigenetics and human disease. Int. J. Biochem. Cell Biol. 41, 136–146. doi:10.1016/j.biocel.2008.09.011

Hoagland, M.B., Stephenson, M.L., Scott, J.F., Hecht, L.I., Zamecnik, P.C., 1958. A soluble ribonucleic acid intermediate in protein synthesis. J. Biol. Chem. 231, 241–257.

Hoagland, M.B., Zamecnik, P.C., Stephenson, M.L., 1957. Intermediate reactions in protein biosynthesis.

Biochim. Biophys. Acta 24, 215–216.

Holliday, R., 2006. Epigenetics: a historical overview. Epigenetics 1, 76–80.

Holliday, R., Pugh, J.E., 1975. DNA modification mechanisms and gene activity during development. Science 187, 226–232.

Hottiger, M.O., 2011. ADP-ribosylation of histones by ARTD1: an additional module of the histone code? FEBS Lett. 585, 1595–1599. doi:10.1016/j.febslet.2011.03.031

Howard, P.J., Lewis, I.J., Harris, F., Walker, S., 1985. Centromeric instability of chromosomes 1 and 16 with variable immune deficiency: a new syndrome. Clin. Genet. 27, 501–505.

Howlett, S.K., Reik, W., 1991. Methylation levels of maternal and paternal genomes during preimplantation development. Dev. Camb. Engl. 113, 119–127.

Howman, E.V., Fowler, K.J., Newson, A.J., Redward, S., MacDonald, A.C., Kalitsis, P., Choo, K.H., 2000. Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. Proc. Natl. Acad. Sci. U. S. A. 97, 1148–1153.

Hsieh, C.L., 1994. Dependence of transcriptional repression on CpG methylation density. Mol. Cell. Biol. 14, 5487–5494.

Huang, J., Fan, T., Yan, Q., Zhu, H., Fox, S., Issaq, H.J., Best, L., Gangi, L., Munroe, D., Muegge, K., 2004. Lsh, an epigenetic guardian of repetitive elements. Nucleic Acids Res. 32, 5019–5028. doi:10.1093/nar/gkh821

Hue, I., Renard, J.P., Viebahn, C., 2001. Brachyury is expressed in gastrulating bovine embryos well ahead of implantation. Dev. Genes Evol. 211, 157–159.

Huff, J.T., Plocik, A.M., Guthrie, C., Yamamoto, K.R., 2010. Reciprocal intronic and exonic histone modification regions in humans. Nat. Struct. Mol. Biol. 17, 1495–1499. doi:10.1038/nsmb.1924

Illingworth, R.S., Bird, A.P., 2009. CpG islands - "A rough guide." FEBS Lett. 583, 1713–1720. doi:10.1016/j.febslet.2009.04.012

Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., Bird, A.P., 2010. Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome. PLoS Genet. 6, e1001134. doi:10.1371/journal.pgen.1001134

Irizarry, R.A., Wu, H., Feinberg, A.P., 2009. A species-generalized probabilistic model-based definition of CpG islands. Mamm. Genome Off. J. Int. Mamm. Genome Soc. 20, 674–680. doi:10.1007/s00335-009-9222-5

Issa, J.-P., 2004. CpG island methylator phenotype in cancer. Nat. Rev. Cancer 4, 988–993. doi:10.1038/nrc1507

Issa, J.-P.J., 2007. DNA Methylation as a Therapeutic Target in Cancer. Clin. Cancer Res. 13, 1634–1637. doi:10.1158/1078-0432.CCR-06-2076

Issa, J.-P.J., Byrd, J.C., 2005. Decitabine in chronic leukemias. Semin. Hematol. 42, S43-49.

Issa, J.-P.J., Kantarjian, H.M., 2009. Targeting DNA Methylation. Clin. Cancer Res. 15, 3938–3946. doi:10.1158/1078-0432.CCR-08-2783

Iyengar, S., Farnham, P.J., 2011. KAP1 Protein: An Enigmatic Master Regulator of the Genome. J. Biol.

Chem. 286, 26267–26276. doi:10.1074/jbc.R111.252569

Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., Dausman, J., Lee, P., Wilson, C., Lander, E., Jaenisch, R., 2001. Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat. Genet. 27, 31–39. doi:10.1038/83730

Jaco, I., Canela, A., Vera, E., Blasco, M.A., 2008. Centromere mitotic recombination in mammalian cells. J. Cell Biol. 181, 885–892. doi:10.1083/jcb.200803042

Jacob, F., Monod, J., 1961. Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. 3, 318–356.

Jacob, R.A., Gretz, D.M., Taylor, P.C., James, S.J., Pogribny, I.P., Miller, B.J., Henning, S.M., Swendseid, M.E., 1998. Moderate folate depletion increases plasma homocysteine and decreases lymphocyte DNA methylation in postmenopausal women. J. Nutr. 128, 1204–1212.

Janic, A., Mendizabal, L., Llamazares, S., Rossell, D., Gonzalez, C., 2010. Ectopic expression of germline genes drives malignant brain tumor growth in Drosophila. Science 330, 1824–1827. doi:10.1126/science.1195481

Jansen, L.E.T., Black, B.E., Foltz, D.R., Cleveland, D.W., 2007. Propagation of centromeric chromatin requires exit from mitosis. J. Cell Biol. 176, 795–805. doi:10.1083/jcb.200701066

Javierre, B.M., Fernandez, A.F., Richter, J., Al-Shahrour, F., Martin-Subero, J.I., Rodriguez-Ubreva, J., Berdasco, M., Fraga, M.F., O'Hanlon, T.P., Rider, L.G., Jacinto, F.V., Lopez-Longo, F.J., Dopazo, J., Forn, M., Peinado, M.A., Carreno, L., Sawalha, A.H., Harley, J.B., Siebert, R., Esteller, M., Miller, F.W., Ballestar, E., 2010. Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. Genome Res. 20, 170–179. doi:10.1101/gr.100289.109

Jeanpierre, M., Turleau, C., Aurias, A., Prieur, M., Ledeist, F., Fischer, A., Viegas-Pequignot, E., 1993. An embryonic-like methylation pattern of classical satellite DNA is observed in ICF syndrome. Hum. Mol. Genet. 2, 731–735.

Jefferson, A., Colella, S., Moralli, D., Wilson, N., Yusuf, M., Gimelli, G., Ragoussis, J., Volpi, E.V., 2010. Altered intra-nuclear organisation of heterochromatin and genes in ICF syndrome. PloS One 5, e11364. doi:10.1371/journal.pone.0011364

Jeffries, M.A., Sawalha, A.H., 2015. Autoimmune disease in the epigenetic era: how has epigenetics changed our understanding of disease and how can we expect the field to evolve? Expert Rev. Clin. Immunol. 11, 45–58. doi:10.1586/1744666X.2015.994507

Ji, W., Hernandez, R., Zhang, X.Y., Qu, G.Z., Frady, A., Varela, M., Ehrlich, M., 1997. DNA demethylation and pericentromeric rearrangements of chromosome 1. Mutat. Res. 379, 33–41.

Jiang, Y., Lev-Lehman, E., Bressler, J., Tsai, T.F., Beaudet, A.L., 1999. Genetics of Angelman syndrome. Am. J. Hum. Genet. 65, 1–6.

Jiang, Y.L., Rigolet, M., Bourc'his, D., Nigon, F., Bokesoy, I., Fryns, J.P., Hultén, M., Jonveaux, P., Maraschio, P., Mégarbané, A., Moncla, A., Viegas-Péquignot, E., 2005. DNMT3B mutations and DNA methylation defect define two types of ICF syndrome: DNMT3B AND TWO ICF SYNDROME TYPES. Hum. Mutat. 25, 56–63. doi:10.1002/humu.20113

Jin, B., Li, Y., Robertson, K.D., 2011. DNA Methylation: Superior or Subordinate in the Epigenetic Hierarchy? Genes Cancer 2, 607–617. doi:10.1177/1947601910393957

Jin, F., Dowdy, S.C., Xiong, Y., Eberhardt, N.L., Podratz, K.C., Jiang, S.-W., 2005. Up-regulation of DNA methyltransferase 3B expression in endometrial cancers. Gynecol. Oncol. 96, 531–538. doi:10.1016/j.ygyno.2004.10.039

Jin, S.-G., Guo, C., Pfeifer, G.P., 2008. GADD45A Does Not Promote DNA Demethylation. PLoS Genet. 4, e1000013. doi:10.1371/journal.pgen.1000013

Jintaridth, P., Mutirangura, A., 2010. Distinctive patterns of age-dependent hypomethylation in interspersed repetitive sequences. Physiol. Genomics 41, 194–200. doi:10.1152/physiolgenomics.00146.2009

Johnson, D.G., Dent, S.Y.R., 2013. Chromatin: Receiver and Quarterback for Cellular Signals. Cell 152, 685–689. doi:10.1016/j.cell.2013.01.017

Johnson, T.B., Coghill, R.D., 1925. RESEARCHES ON PYRIMIDINES. C111. THE DISCOVERY OF 5-METHYL-CYTOSINE IN TUBERCULINIC ACID, THE NUCLEIC ACID OF THE TUBERCLE BACILLUS [1]. J. Am. Chem. Soc. 47, 2838–2844. doi:10.1021/ja01688a030

Jolly, C., Metz, A., Govin, J., Vigneron, M., Turner, B.M., Khochbin, S., Vourc'h, C., 2004. Stress-induced transcription of satellite III repeats. J. Cell Biol. 164, 25–33. doi:10.1083/jcb.200306104

Jones, P.A., 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. 13, 484–492. doi:10.1038/nrg3230

Jones, P.A., Liang, G., 2009. Rethinking how DNA methylation patterns are maintained. Nat. Rev. Genet. 10, 805–811. doi:10.1038/nrg2651

Jones, P.A., Taylor, S.M., 1980. Cellular differentiation, cytidine analogs and DNA methylation. Cell 20, 85–93.

Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J., Wolffe, A.P., 1998. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat. Genet. 19, 187–191. doi:10.1038/561

Jørgensen, H.F., Azuara, V., Amoils, S., Spivakov, M., Terry, A., Nesterova, T., Cobb, B.S., Ramsahoye, B., Merkenschlager, M., Fisher, A.G., 2007. The impact of chromatin modifiers on the timing of locus replication in mouse embryonic stem cells. Genome Biol. 8, R169. doi:10.1186/gb-2007-8-8-r169

Jørgensen, S., Schotta, G., Sørensen, C.S., 2013. Histone H4 lysine 20 methylation: key player in epigenetic regulation of genomic integrity. Nucleic Acids Res. 41, 2797–2806. doi:10.1093/nar/gkt012

Joshi, A.A., Struhl, K., 2005. Eaf3 Chromodomain Interaction with Methylated H3-K36 Links Histone Deacetylation to Pol II Elongation. Mol. Cell 20, 971–978. doi:10.1016/j.molcel.2005.11.021

Jost, J.P., 1993. Nuclear extracts of chicken embryos promote an active demethylation of DNA by excision repair of 5-methyldeoxycytidine. Proc. Natl. Acad. Sci. U. S. A. 90, 4684–4688.

Jubier-Maurin, V., Cuny, G., Laurent, A.M., Paquereau, L., Roizes, G., 1992. A new 5' sequence associated with mouse L1 elements is representative of a major class of L1 termini. Mol. Biol. Evol. 9, 41–55.

Jung, M., Pfeifer, G.P., 2015. Aging and DNA methylation. BMC Biol. 13, 7. doi:10.1186/s12915-015-0118-4

Jurka, J., 1997. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc. Natl. Acad. Sci. U. S. A. 94, 1872–1877.

Kacem, S., Feil, R., 2009. Chromatin mechanisms in genomic imprinting. Mamm. Genome 20, 544–556. doi:10.1007/s00335-009-9223-4

Kaer, K., Speek, M., 2013. Retroelements in human disease. Gene 518, 231–241. doi:10.1016/j.gene.2013.01.008

Kalitsis, P., Griffiths, B., Choo, K.H.A., 2006. Mouse telocentric sequences reveal a high rate of homogenization and possible role in Robertsonian translocation. Proc. Natl. Acad. Sci. U. S. A. 103, 8786–8791. doi:10.1073/pnas.0600250103

Kalitsis, P., MacDonald, A.C., Newson, A.J., Hudson, D.F., Choo, K.H., 1998. Gene structure and sequence analysis of mouse centromere proteins A and C. Genomics 47, 108–114. doi:10.1006/geno.1997.5109

Kanai, Y., Ushijima, S., Nakanishi, Y., Sakamoto, M., Hirohashi, S., 2003. Mutation of the DNA methyltransferase (DNMT) 1 gene in human colorectal cancers. Cancer Lett. 192, 75–82.

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., Sasaki, H., 2004. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature 429, 900–903. doi:10.1038/nature02633

Kanellopoulou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M., Rajewsky, K., 2005. Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes Dev. 19, 489–501. doi:10.1101/gad.1248505

Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., Tang, J.X., Shinkai, Y., Mager, D.L., Jones, S., Hirst, M., Lorincz, M.C., 2011. DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in mESCs. Cell Stem Cell 8, 676–687. doi:10.1016/j.stem.2011.04.004

Kashiwabara, S.-I., Noguchi, J., Zhuang, T., Ohmura, K., Honda, A., Sugiura, S., Miyamoto, K., Takahashi, S., Inoue, K., Ogura, A., Baba, T., 2002. Regulation of spermatogenesis by testis-specific, cytoplasmic poly(A) polymerase TPAP. Science 298, 1999–2002. doi:10.1126/science.1074632

Kato, H., Goto, D.B., Martienssen, R.A., Urano, T., Furukawa, K., Murakami, Y., 2005. RNA polymerase II is required for RNAi-dependent heterochromatin assembly. Science 309, 467–469. doi:10.1126/science.1114955

Kazazian, H.H., 2004. Mobile elements: drivers of genome evolution. Science 303, 1626–1632. doi:10.1126/science.1089670

Kelsen, J., Dawany, N., Conrad, M., Sullivan, K., Behrens, E., Devoto, M., 2017. O-001 A Diagnostic Approach of Immune Dysregulation on Very Early-Onset IBD. Inflamm. Bowel Dis. 23 Suppl 1, S1. doi:10.1097/MIB.0000000000001005

Kidder, B.L., Hu, G., Cui, K., Zhao, K., 2017. SMYD5 regulates H4K20me3-marked heterochromatin to safeguard ES cell self-renewal and prevent spurious differentiation. Epigenetics Chromatin 10. doi:10.1186/s13072-017-0115-7

Kidder, B.L., Yang, J., Palmer, S., 2008. Stat3 and c-Myc genome-wide promoter occupancy in embryonic

stem cells. PloS One 3, e3932. doi:10.1371/journal.pone.0003932

Kim, J., Guermah, M., McGinty, R.K., Lee, J.-S., Tang, Z., Milne, T.A., Shilatifard, A., Muir, T.W., Roeder, R.G., 2009. RAD6-Mediated Transcription-Coupled H2B Ubiquitylation Directly Stimulates H3K4 Methylation in Human Cells. Cell 137, 459–471. doi:10.1016/j.cell.2009.02.027

Kim, J., Kollhoff, A., Bergmann, A., Stubbs, L., 2003. Methylation-sensitive binding of transcription factor YY1 to an insulator sequence within the paternally expressed imprinted gene, Peg3. Hum. Mol. Genet. 12, 233–245. doi:10.1093/hmg/ddg028

Kim, S., Namekawa, S.H., Niswander, L.M., Ward, J.O., Lee, J.T., Bardwell, V.J., Zarkower, D., 2007. A Mammal-Specific Doublesex Homolog Associates with Male Sex Chromatin and Is Required for Male Meiosis. PLoS Genet. 3, e62. doi:10.1371/journal.pgen.0030062

Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., Ren, B., 2005. A high-resolution map of active promoters in the human genome. Nature 436, 876–880. doi:10.1038/nature03877

Kim, U.J., Han, M., Kayne, P., Grunstein, M., 1988. Effects of histone H4 depletion on the cell cycle and transcription of Saccharomyces cerevisiae. EMBO J. 7, 2211–2219.

Kim, Y.-I., 2006. Folate: a magic bullet or a double edged sword for colorectal cancer prevention? Gut 55, 1387–1389. doi:10.1136/gut.2006.095463

Kimura, K., Wakamatsu, A., Suzuki, Y., Ota, T., Nishikawa, T., Yamashita, R., Yamamoto, J., Sekine, M., Tsuritani, K., Wakaguri, H., Ishii, S., Sugiyama, T., Saito, K., Isono, Y., Irie, R., Kushida, N., Yoneyama, T., Otsuka, R., Kanda, K., Yokoi, T., Kondo, H., Wagatsuma, M., Murakawa, K., Ishida, S., Ishibashi, T., Takahashi-Fujii, A., Tanase, T., Nagai, K., Kikuchi, H., Nakai, K., Isogai, T., Sugano, S., 2006. Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. Genome Res. 16, 55–65. doi:10.1101/gr.4039406

Kiyonari, H., Kaneko, M., Abe, S., Aizawa, S., 2010. Three inhibitors of FGF receptor, ERK, and GSK3 establishes germline-competent embryonic stem cells of C57BL/6N mouse strain with high efficiency and stability. genesis NA-NA. doi:10.1002/dvg.20614

Klein, C.J., Botuyan, M.-V., Wu, Y., Ward, C.J., Nicholson, G.A., Hammans, S., Hojo, K., Yamanishi, H., Karpf, A.R., Wallace, D.C., Simon, M., Lander, C., Boardman, L.A., Cunningham, J.M., Smith, G.E., Litchy, W.J., Boes, B., Atkinson, E.J., Middha, S., B Dyck, P.J., Parisi, J.E., Mer, G., Smith, D.I., Dyck, P.J., 2011. Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss. Nat. Genet. 43, 595–600. doi:10.1038/ng.830

Kleinsmith, L.J., Allfrey, V.G., Mirsky, A.E., 1966. Phosphorylation of nuclear protein early in the course of gene activation in lymphocytes. Science 154, 780–781.

Kloc, A., Zaratiegui, M., Nora, E., Martienssen, R., 2008. RNA interference guides histone modification during the S phase of chromosomal replication. Curr. Biol. CB 18, 490–495. doi:10.1016/j.cub.2008.03.016

Klug, A., Rhodes, D., Smith, J., Finch, J.T., Thomas, J.O., 1980. A low resolution structure for the histone core of the nucleosome. Nature 287, 509–516.

Knezetic, J.A., Luse, D.S., 1986. The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. Cell 45, 95–104.

Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R., Liu, X.S., Aravind, L., Agarwal, S., Maciejewski, J.P., Rao, A., 2010. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature 468, 839–843. doi:10.1038/nature09586

Koerner, M.V., Pauler, F.M., Huang, R., Barlow, D.P., 2009. The function of non-coding RNAs in genomic imprinting. Development 136, 1771–1783. doi:10.1242/dev.030403

Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T., Wutz, A., 2004. A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. PLoS Biol. 2, E171. doi:10.1371/journal.pbio.0020171

Kokura, K., Sun, L., Bedford, M.T., Fang, J., 2010. Methyl-H3K9-binding protein MPP8 mediates E-cadherin gene silencing and promotes tumour cell motility and invasion. EMBO J. 29, 3673–3687. doi:10.1038/emboj.2010.239

Kolosha, V.O., Martin, S.L., 1997. In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition. Proc. Natl. Acad. Sci. U. S. A. 94, 10155–10160.

Kondo, T., Bobek, M.P., Kuick, R., Lamb, B., Zhu, X., Narayan, A., Bourc'his, D., Viegas-Péquignot, E., Ehrlich, M., Hanash, S.M., 2000. Whole-genome methylation scan in ICF syndrome: hypomethylation of non-satellite DNA repeats D4Z4 and NBL2. Hum. Mol. Genet. 9, 597–604.

Korenberg, J.R., Rykowski, M.C., 1988. Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. Cell 53, 391–400.

Kornberg, R.D., 1974. Chromatin structure: a repeating unit of histones and DNA. Science 184, 868–871.

Kossel, A., 2009. Ueber einen peptonartigen Bestandtheil des Zellkerns. Z. Für Physiol. Chem. 8, 511–515. doi:10.1515/bchm1.1884.8.6.511

Kouzarides, T., 2007. Chromatin modifications and their function. Cell 128, 693–705. doi:10.1016/j.cell.2007.02.005

Kriaucionis, S., Heintz, N., 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 324, 929–930. doi:10.1126/science.1169786

Kriaucionis, S., Tahiliani, M., 2014. Expanding the epigenetic landscape: novel modifications of cytosine in genomic DNA. Cold Spring Harb. Perspect. Biol. 6, a018630. doi:10.1101/cshperspect.a018630

Kriwacki, R.W., Schultz, S.C., Steitz, T.A., Caradonna, J.P., 1992. Sequence-specific recognition of DNA by zinc-finger peptides derived from the transcription factor Sp1. Proc. Natl. Acad. Sci. U. S. A. 89, 9759–9763.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., Adli, M., Kasif, S., Ptaszek, L.M., Cowan, C.A., Lander, E.S., Koseki, H., Bernstein, B.E., 2008. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet. 4, e1000242. doi:10.1371/journal.pgen.1000242

Kubota, T., Furuumi, H., Kamoda, T., Iwasaki, N., Tobita, N., Fujiwara, N., Goto, Y., Matsui, A., Sasaki, H., Kajii, T., 2004. ICF syndrome in a girl with DNA hypomethylation but without detectable DNMT3B mutation. Am. J. Med. Genet. 129A, 290–293. doi:10.1002/ajmg.a.30135

Kuilman, T., Michaloglou, C., Mooi, W.J., Peeper, D.S., 2010. The essence of senescence. Genes Dev. 24, 2463–2479. doi:10.1101/gad.1971610

Kunarso, G., Chia, N.-Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.-S., Ng, H.-H., Bourque, G., 2010. Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat. Genet. 42, 631–634. doi:10.1038/ng.600

Kunath, T., Saba-El-Leil, M.K., Almousailleakh, M., Wray, J., Meloche, S., Smith, A., 2007. FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. Development 134, 2895–2902. doi:10.1242/dev.02880

Kurosawa, H., 2007. Methods for inducing embryoid body formation: in vitro differentiation system of embryonic stem cells. J. Biosci. Bioeng. 103, 389–398. doi:10.1263/jbb.103.389

Kutluğ, S., Ogur, G., Yilmaz, A., Thijssen, P.E., Abur, U., Yildiran, A., 2016. Vesicourethral reflux-induced renal failure in a patient with ICF syndrome due to a novel DNMT3B mutation. Am. J. Med. Genet. A. 170, 3253–3257. doi:10.1002/ajmg.a.37866

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., Jenuwein, T., 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature 410, 116–120. doi:10.1038/35065132

Lacoste, N., Woolfe, A., Tachiwana, H., Garea, A.V., Barth, T., Cantaloube, S., Kurumizaka, H., Imhof, A., Almouzni, G., 2014. Mislocalization of the centromeric histone variant CenH3/CENP-A in human cells depends on the chaperone DAXX. Mol. Cell 53, 631–644. doi:10.1016/j.molcel.2014.01.018

Laget, S., Joulie, M., Le Masson, F., Sasai, N., Christians, E., Pradhan, S., Roberts, R.J., Defossez, P.-A., 2010. The Human Proteins MBD5 and MBD6 Associate with Heterochromatin but They Do Not Bind Methylated DNA. PLoS ONE 5, e11982. doi:10.1371/journal.pone.0011982

Lako, M., Lindsay, S., Lincoln, J., Cairns, P.M., Armstrong, L., Hole, N., 2001. Characterisation of Wnt gene expression during the differentiation of murine embryonic stem cells in vitro: role of Wnt3 in enhancing haematopoietic differentiation. Mech. Dev. 103, 49–59.

Lander, *et al.* (International Human Genome Sequencing Consortium), 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921. doi:10.1038/35057062

Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A.O., Landthaler, M., Lin, C., Socci, N.D., Hermida, L., Fulci, V., Chiaretti, S., Foà, R., Schliwka, J., Fuchs, U., Novosel, A., Müller, R.-U., Schermer, B., Bissels, U., Inman, J., Phan, Q., Chien, M., Weir, D.B., Choksi, R., De Vita, G., Frezzetti, D., Trompeter, H.-I., Hornung, V., Teng, G., Hartmann, G., Palkovits, M., Di Lauro, R., Wernet, P., Macino, G., Rogler, C.E., Nagle, J.W., Ju, J., Papavasiliou, F.N., Benzing, T., Lichter, P., Tam, W., Brownstein, M.J., Bosio, A., Borkhardt, A., Russo, J.J., Sander, C., Zavolan, M., Tuschl, T., 2007. A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 129, 1401–1414. doi:10.1016/j.cell.2007.04.040

Lane, N., Dean, W., Erhardt, S., Hajkova, P., Surani, A., Walter, J., Reik, W., 2003. Resistance of IAPs to

methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. Genes. N. Y. N 2000 35, 88–93. doi:10.1002/gene.10168

Lanzuolo, C., Roure, V., Dekker, J., Bantignies, F., Orlando, V., 2007. Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. Nat. Cell Biol. 9, 1167–1174. doi:10.1038/ncb1637

Laufer, B.I., Singh, S.M., 2012. A macro role for imprinted clusters of microRNAs in the brain. MicroRNA Shariqah United Arab Emir. 1, 59–64.

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Kin Sung, K.W., Rigoutsos, I., Loring, J., Wei, C.-L., 2010. Dynamic changes in the human methylome during differentiation. Genome Res. 20, 320–331. doi:10.1101/gr.101907.109

Lee, D.-H., Singh, P., Tsai, S.Y., Oates, N., Spalla, A., Spalla, C., Brown, L., Rivas, G., Larson, G., Rauch, T.A., Pfeifer, G.P., Szabó, P.E., 2010. CTCF-Dependent Chromatin Bias Constitutes Transient Epigenetic Memory of the Mother at the H19-Igf2 Imprinting Control Region in Prospermatogonia. PLoS Genet. 6, e1001224. doi:10.1371/journal.pgen.1001224

Lee, J.-H., Skalnik, D.G., 2005. CpG-binding Protein (CXXC Finger Protein 1) Is a Component of the Mammalian Set1 Histone H3-Lys4 Methyltransferase Complex, the Analogue of the Yeast Set1/COMPASS Complex. J. Biol. Chem. 280, 41725–41731. doi:10.1074/jbc.M508312200

Lee, S.-U., Maeda, T., 2012. POK/ZBTB proteins: an emerging family of proteins that regulate lymphoid development and function. Immunol. Rev. 247, 107–119. doi:10.1111/j.1600-065X.2012.01116.x

Lehnertz, B., Ueda, Y., Derijck, A.A.H.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., Peters, A.H.F.M., 2003. Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr. Biol. CB 13, 1192–1200.

Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., Surani, M.A., Hajkova, P., 2013. Naive pluripotency is associated with global DNA hypomethylation. Nat. Struct. Mol. Biol. 20, 311–316. doi:10.1038/nsmb.2510

Leung, D., Du, T., Wagner, U., Xie, W., Lee, A.Y., Goyal, P., Li, Y., Szulwach, K.E., Jin, P., Lorincz, M.C., Ren, B., 2014. Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1. Proc. Natl. Acad. Sci. 111, 6690–6695. doi:10.1073/pnas.1322273111

Li, E., Beard, C., Jaenisch, R., 1993. Role for DNA methylation in genomic imprinting. Nature 366, 362–365. doi:10.1038/366362a0

Li, E., Bestor, T.H., Jaenisch, R., 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69, 915–926.

Li, E., Zhang, Y., 2014. DNA methylation in mammals. Cold Spring Harb. Perspect. Biol. 6, a019133. doi:10.1101/cshperspect.a019133

Li, F., Sonbuchner, L., Kyes, S.A., Epp, C., Deitsch, K.W., 2008. Nuclear non-coding RNAs are transcribed from the centromeres of Plasmodium falciparum and are associated with centromeric chromatin. J. Biol.

Chem. 283, 5692–5698. doi:10.1074/jbc.M707344200

Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B.E., Reinberg, D., 2010. Jarid2 and PRC2, partners in regulating gene expression. Genes Dev. 24, 368–380. doi:10.1101/gad.1886410

Li, H., 2006. The Histone Methyltransferase SETDB1 and the DNA Methyltransferase DNMT3A Interact Directly and Localize to Promoters Silenced in Cancer Cells. J. Biol. Chem. 281, 19489–19500. doi:10.1074/jbc.M513249200

Li, M., Fang, Y., 2015. Histone variants: the artists of eukaryotic chromatin. Sci. China Life Sci. 58, 232–239. doi:10.1007/s11427-015-4817-4

Li, P., Tong, C., Mehrian-Shai, R., Jia, L., Wu, N., Yan, Y., Maxson, R.E., Schulze, E.N., Song, H., Hsieh, C.-L., Pera, M.F., Ying, Q.-L., 2008. Germline Competent Embryonic Stem Cells Derived from Rat Blastocysts. Cell 135, 1299–1310. doi:10.1016/j.cell.2008.12.006

Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., Ferguson-Smith, A.C., 2008. A Maternal-Zygotic Effect Gene, Zfp57, Maintains Both Maternal and Paternal Imprints. Dev. Cell 15, 547–557. doi:10.1016/j.devcel.2008.08.014

Li, Y., Seah, M.K.Y., O'Neill, C., 2016. Mapping global changes in nuclear cytosine base modifications in the early mouse embryo. Reproduction 151, 83–95. doi:10.1530/REP-15-0207

Liang, G., Chan, M.F., Tomigahara, Y., Tsai, Y.C., Gonzales, F.A., Li, E., Laird, P.W., Jones, P.A., 2002. Cooperativity between DNA Methyltransferases in the Maintenance Methylation of Repetitive Elements. Mol. Cell. Biol. 22, 480–491. doi:10.1128/MCB.22.2.480-491.2002

Liao, J., Karnik, R., Gu, H., Ziller, M.J., Clement, K., Tsankov, A.M., Akopian, V., Gifford, C.A., Donaghey, J., Galonska, C., Pop, R., Reyon, D., Tsai, S.Q., Mallard, W., Joung, J.K., Rinn, J.L., Gnirke, A., Meissner, A., 2015. Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. Nat. Genet. 47, 469–478. doi:10.1038/ng.3258

Lienert, F., Wirbelauer, C., Som, I., Dean, A., Mohn, F., Schübeler, D., 2011. Identification of genetic elements that autonomously determine DNA methylation states. Nat. Genet. 43, 1091–1097. doi:10.1038/ng.946

Lindroth, A.M., Park, Y.J., McLean, C.M., Dokshin, G.A., Persson, J.M., Herman, H., Pasini, D., Miró, X., Donohoe, M.E., Lee, J.T., Helin, K., Soloway, P.D., 2008. Antagonism between DNA and H3K27 methylation at the imprinted Rasgrf1 locus. PLoS Genet. 4, e1000145. doi:10.1371/journal.pgen.1000145

Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., Yu, M., Tonti-Filippini, J., Heyn, H., Hu, S., Wu, J.C., Rao, A., Esteller, M., He, C., Haghighi, F.G., Sejnowski, T.J., Behrens, M.M., Ecker, J.R., 2013. Global Epigenomic Reconfiguration During Mammalian Brain Development. Science 341, 1237905–1237905. doi:10.1126/science.1237905

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A.H., Thomson, J.A., Ren, B., Ecker, J.R., 2009. Human DNA methylomes at base resolution show widespread

epigenomic differences. Nature 462, 315–322. doi:10.1038/nature08514

Liu, L., van Groen, T., Kadish, I., Li, Y., Wang, D., James, S.R., Karpf, A.R., Tollefsbol, T.O., 2011. Insufficient DNA methylation affects healthy aging and promotes age-related health problems. Clin. Epigenetics 2, 349–360. doi:10.1007/s13148-011-0042-6

Liu, S., Brind'Amour, J., Karimi, M.M., Shirane, K., Bogutz, A., Lefebvre, L., Sasaki, H., Shinkai, Y., Lorincz, M.C., 2014. Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. Genes Dev. 28, 2041–2055. doi:10.1101/gad.244848.114

Liu, Y., Siegmund, K.D., Laird, P.W., Berman, B.P., 2012. Bis-SNP: combined DNA methylation and SNP calling for Bisulfite-seq data. Genome Biol. 13, R61. doi:10.1186/gb-2012-13-7-r61

Loeb, D.D., Padgett, R.W., Hardies, S.C., Shehee, W.R., Comer, M.B., Edgell, M.H., Hutchison, C.A., 1986. The sequence of a large L1Md element reveals a tandemly repeated 5' end and several features found in retrotransposons. Mol. Cell. Biol. 6, 168–182.

Long, H.K., Blackledge, N.P., Klose, R.J., 2013. ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. Biochem. Soc. Trans. 41, 727–740. doi:10.1042/BST20130028

López de Silanes, I., Stagno d'Alcontres, M., Blasco, M.A., 2010. TERRA transcripts are bound by a complex array of RNA-binding proteins. Nat. Commun. 1, 33. doi:10.1038/ncomms1032

Lorch, Y., LaPointe, J.W., Kornberg, R.D., 1987. Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. Cell 49, 203–210.

Lorincz, M.C., Schübeler, D., 2007. RNA Polymerase II: Just Stopping By. Cell 130, 16–18. doi:10.1016/j.cell.2007.06.040

Lu, J., Gilbert, D.M., 2007. Proliferation-dependent and cell cycle regulated transcription of mouse pericentric heterochromatin. J. Cell Biol. 179, 411–421. doi:10.1083/jcb.200706176

Luan, D.D., Eickbush, T.H., 1995. RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. Mol. Cell. Biol. 15, 3882–3891.

Luan, D.D., Korman, M.H., Jakubczak, J.L., Eickbush, T.H., 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 72, 595–605.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J., 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389, 251–260. doi:10.1038/38444

Lunyak, V.V., Burgess, R., Prefontaine, G.G., Nelson, C., Sze, S.-H., Chenoweth, J., Schwartz, P., Pevzner, P.A., Glass, C., Mandel, G., Rosenfeld, M.G., 2002. Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. Science 298, 1747–1752. doi:10.1126/science.1076469

Luzzati, V., Mathis, A., Masson, F., Witz, J., 1964. SUTURE TRANSITIONS OBSERVED IN DNA AND POLY A IN SOLUTION AS A FUNCTION OF TEMPERATURE AND PH. J. Mol. Biol. 10, 28–41.

Lynch, V.J., Leclerc, R.D., May, G., Wagner, G.P., 2011. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nat. Genet. 43, 1154–1159. doi:10.1038/ng.917

Lyon, M.F., 2000. LINE-1 elements and X chromosome inactivation: a function for “junk” DNA? Proc. Natl. Acad. Sci. U. S. A. 97, 6248–6249.

Lyon, M.F., 1961. Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature 190, 372–373.

Ma, D.K., Guo, J.U., Ming, G., Song, H., 2009. DNA excision repair proteins and Gadd45 as molecular players for active DNA demethylation. Cell Cycle 8, 1526–1531. doi:10.4161/cc.8.10.8500

Maatouk, D.M., Kellam, L.D., Mann, M.R.W., Lei, H., Li, E., Bartolomei, M.S., Resnick, J.L., 2006. DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. Dev. Camb. Engl. 133, 3411–3418. doi:10.1242/dev.02500

Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., Pfaff, S.L., 2012. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature 487, 57–63. doi:10.1038/nature11244

Macleod, D., Charlton, J., Mullins, J., Bird, A.P., 1994. Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. Genes Dev. 8, 2282–2292.

Maderspacher, F., 2008. Theodor Boveri and the natural experiment. Curr. Biol. 18, R279–R286.

Maestroni, L., Matmati, S., Coulon, S., 2017. Solving the Telomere Replication Problem. Genes 8, 55. doi:10.3390/genes8020055

Magklara, A., Lomvardas, S., 2013. Stochastic gene expression in mammals: lessons from olfaction. Trends Cell Biol. 23, 449–456. doi:10.1016/j.tcb.2013.04.005

Mahsoudi, B., Li, A., O’Neill, C., 2007. Assessment of the long-term and transgenerational consequences of perturbing preimplantation embryo development in mice. Biol. Reprod. 77, 889–896. doi:10.1095/biolreprod.106.057885

Maksakova, I.A., Romanish, M.T., Gagnier, L., Dunn, C.A., van de Lagemaat, L.N., Mager, D.L., 2006. Retroviral Elements and Their Hosts: Insertional Mutagenesis in the Mouse Germ Line. PLoS Genet. 2, e2. doi:10.1371/journal.pgen.0020002

Maraschio, P., Zuffardi, O., Dalla Fior, T., Tiepolo, L., 1988. Immunodeficiency, centromeric heterochromatin instability of chromosomes 1, 9, and 16, and facial anomalies: the ICF syndrome. J. Med. Genet. 25, 173–180.

Marchal, C., Miotto, B., 2015. Emerging concept in DNA methylation: role of transcription factors in shaping DNA methylation patterns. J. Cell. Physiol. 230, 743–751. doi:10.1002/jcp.24836

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury III, W.J., Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., Pirrotta, V., Reinberg, D., Gamblin, S.J., 2009. Role of the polycomb protein EED in the propagation of repressive histone marks. Nature 461, 762–767. doi:10.1038/nature08398

Margueron, R., Reinberg, D., 2010. Chromatin structure and the inheritance of epigenetic information. Nat. Rev. Genet. 11, 285–296. doi:10.1038/nrg2752

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A., Stunnenberg, H.G., 2012. The transcriptional and epigenomic foundations of ground state pluripotency. Cell 149, 590–604. doi:10.1016/j.cell.2012.03.026

Martens, J.H.A., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., Jenuwein, T., 2005. The profile of repeat-associated histone lysine methylation states in the mouse epigenome. EMBO J. 24, 800–812. doi:10.1038/sj.emboj.7600545

Martin, G.R., 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc. Natl. Acad. Sci. U. S. A. 78, 7634–7638.

Mastick, G.S., Andrews, G.L., 2001. Pax6 regulates the identity of embryonic diencephalic neurons. Mol. Cell. Neurosci. 17, 190–207. doi:10.1006/mcne.2000.0924

Matarazzo, M.R., Boyle, S., D'Esposito, M., Bickmore, W.A., 2007. Chromosome territory reorganization in a human disease with altered DNA methylation. Proc. Natl. Acad. Sci. 104, 16546–16551.

Mathias, S.L., Scott, A.F., Kazazian, H.H., Boeke, J.D., Gabriel, A., 1991. Reverse transcriptase encoded by a human transposable element. Science 254, 1808–1810.

Matsui, T., Leung, D., Miyashita, H., Maksakova, I.A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M.C., Shinkai, Y., 2010. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature 464, 927–931. doi:10.1038/nature08858

Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y., Turecki, G., Delaney, A., Varhol, R., Thiessen, N., Shchors, K., Heine, V.M., Rowitch, D.H., Xing, X., Fiore, C., Schillebeeckx, M., Jones, S.J.M., Haussler, D., Marra, M.A., Hirst, M., Wang, T., Costello, J.F., 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature 466, 253–257. doi:10.1038/nature09165

Mayer, W., Niveleau, A., Walter, J., Fundele, R., Haaf, T., 2000. Demethylation of the zygotic paternal genome. Nature 403, 501–502. doi:10.1038/35000654

McGrath, J., Solter, D., 1984. Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell 37, 179–183.

Medstrand, P., van de Lagemaat, L.N., Mager, D.L., 2002. Retroelement distributions in the human genome: variations associated with age and proximity to genes. Genome Res. 12, 1483–1495. doi:10.1101/gr.388902

Meehan, R.R., Lewis, J.D., Bird, A.P., 1992. Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. Nucleic Acids Res. 20, 5085–5092.

Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L., Bird, A.P., 1989. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell 58, 499–507.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., Gnirke, A., Jaenisch, R., Lander, E.S., 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature. doi:10.1038/nature07107

Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S., Heintz, N., 2012. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. Cell 151, 1417–1430. doi:10.1016/j.cell.2012.11.022

Melnikov, A.A., Gartenhaus, R.B., Levenson, A.S., Motchoulskaia, N.A., Levenson Chernokhvostov, V.V., 2005. MSRE-PCR for analysis of gene-specific DNA methylation. Nucleic Acids Res. 33, e93.

doi:10.1093/nar/gni092

Meneghini, M.D., Wu, M., Madhani, H.D., 2003. Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. Cell 112, 725–736.

Menezes, J.R., Luskin, M.B., 1994. Expression of neuron-specific tubulin defines a novel population in the proliferative layers of the developing telencephalon. J. Neurosci. 14, 5399–5416.

Menezo, Y.J., Russo, G., Tosti, E., Mouatassim, S.E., Benkhalifa, M., 2007. Expression profile of genes coding for DNA repair in human oocytes using pangenomic microarrays, with a special focus on ROS linked decays. J. Assist. Reprod. Genet. 24, 513–520. doi:10.1007/s10815-007-9167-0

Métivier, R., Gallais, R., Tiffoche, C., Le Péron, C., Jurkowska, R.Z., Carmouche, R.P., Ibberson, D., Barath, P., Demay, F., Reid, G., Benes, V., Jeltsch, A., Gannon, F., Salbert, G., 2008. Cyclical DNA methylation of a transcriptionally active promoter. Nature 452, 45–50. doi:10.1038/nature06544

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., Nusbaum, C., Lander, E.S., Bernstein, B.E., 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448, 553–560. doi:10.1038/nature06008

Miniou, P., Jeanpierre, M., Blanquet, V., Sibella, V., Bonneau, D., Herbelin, C., Fischer, A., Niveleau, A., Viegas-Péquignot, E., 1994. Abnormal methylation pattern in constitutive and facultative (X inactive chromosome) heterochromatin of ICF patients. Hum. Mol. Genet. 3, 2093–2102.

Miniou, P., Jeanpierre, M., Bourc'his, D., Coutinho Barbosa, A.C., Blanquet, V., Viegas-Péquignot, E., 1997. alpha-satellite DNA methylation in normal individuals and in ICF patients: heterogeneous methylation of constitutive heterochromatin in adult and fetal tissues. Hum. Genet. 99, 738–745.

Miousse, I.R., Chalbot, M.-C.G., Lumen, A., Ferguson, A., Kavouras, I.G., Koturbash, I., 2015. Response of transposable elements to environmental stressors. Mutat. Res. Mutat. Res. 765, 19–39. doi:10.1016/j.mrrev.2015.05.003

Mitchell, R.S., Beitzel, B.F., Schroder, A.R.W., Shinn, P., Chen, H., Berry, C.C., Ecker, J.R., Bushman, F.D., 2004. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol. 2, E234. doi:10.1371/journal.pbio.0020234

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., Yamanaka, S., 2003. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell 113, 631–642.

Miyanari, Y., Torres-Padilla, M.-E., 2012. Control of ground-state pluripotency by allelic regulation of Nanog. Nature 483, 470–473. doi:10.1038/nature10807

Moarefi, A.H., Chédin, F., 2011. ICF syndrome mutations cause a broad spectrum of biochemical defects in DNMT3B-mediated de novo DNA methylation. J. Mol. Biol. 409, 758–772. doi:10.1016/j.jmb.2011.04.050

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., Schübeler, D., 2008. Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential

of Neuronal Progenitors. Mol. Cell 30, 755–766. doi:10.1016/j.molcel.2008.05.007

Mohn, F., Weber, M., Schübeler, D., Roloff, T.-C., 2009. Methylated DNA immunoprecipitation (MeDIP). Methods Mol. Biol. Clifton NJ 507, 55–64. doi:10.1007/978-1-59745-522-0_5

Monk, M., Boubelik, M., Lehnert, S., 1987. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. Dev. Camb. Engl. 99, 371–382.

Moretti, P., Zoghbi, H.Y., 2006. MeCP2 dysfunction in Rett syndrome and related disorders. Curr. Opin. Genet. Dev. 16, 276–281. doi:10.1016/j.gde.2006.04.009

Mouse Genome Sequencing Consortium, *et al.*, 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420, 520–562. doi:10.1038/nature01262

Moyzis, R.K., Buckingham, J.M., Cram, L.S., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J., Ratliff, R.L., Wu, J.R., 1988. A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. Proc. Natl. Acad. Sci. U. S. A. 85, 6622–6626.

Murray, K., 1964. THE OCCURRENCE OF EPSILON-N-METHYL LYSINE IN HISTONES. Biochemistry (Mosc.) 3, 10–15.

Musacchio, A., Desai, A., 2017. A Molecular View of Kinetochore Assembly and Function. Biology 6, 5. doi:10.3390/biology6010005

Myant, K., Stancheva, I., 2008. LSH cooperates with DNA methyltransferases to repress transcription. Mol. Cell. Biol. 28, 215–226. doi:10.1128/MCB.01073-07

Myant, K., Termanis, A., Sundaram, A.Y.M., Boe, T., Li, C., Merusi, C., Burrage, J., de Las Heras, J.I., Stancheva, I., 2011. LSH and G9a/GLP complex are required for developmentally programmed DNA methylation. Genome Res. 21, 83–94. doi:10.1101/gr.108498.110

Naas, T.P., DeBerardinis, R.J., Moran, J.V., Ostertag, E.M., Kingsmore, S.F., Seldin, M.F., Hayashizaki, Y., Martin, S.L., Kazazian, H.H., 1998. An actively retrotransposing, novel subfamily of mouse L1 elements. EMBO J. 17, 590–597. doi:10.1093/emboj/17.2.590

Nakagawa, H., Opitz, O.G., 2007. Inducing cellular senescence using defined genetic elements. Methods Mol. Biol. Clifton NJ 371, 167–178. doi:10.1007/978-1-59745-361-5_13

Nakamura, T., Arai, Y., Umehara, H., Masuhara, M., Kimura, T., Taniguchi, H., Sekimoto, T., Ikawa, M., Yoneda, Y., Okabe, M., Tanaka, S., Shiota, K., Nakano, T., 2007. PGC7/Stella protects against DNA demethylation in early embryogenesis. Nat. Cell Biol. 9, 64–71. doi:10.1038/ncb1519

Nakano, M., Okamoto, Y., Ohzeki, J., Masumoto, H., 2003. Epigenetic assembly of centromeric chromatin at ectopic alpha-satellite sites on human chromosomes. J. Cell Sci. 116, 4021–4034. doi:10.1242/jcs.00697

Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., Grewal, S.I., 2001. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science 292, 110–113. doi:10.1126/science.1060118

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., Bird, A., 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase

complex. Nature 393, 386–389. doi:10.1038/30764

Narita, M., Nũnez, S., Heard, E., Narita, M., Lin, A.W., Hearn, S.A., Spector, D.L., Hannon, G.J., Lowe, S.W., 2003. Rb-mediated heterochromatin formation and silencing of E2F target genes during cellular senescence. Cell 113, 703–716.

Nekrasov, M., Wild, B., Müller, J., 2005. Nucleosome binding and histone methyltransferase activity of Drosophila PRC2. EMBO Rep. 6, 348–353. doi:10.1038/sj.embor.7400376

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., Oliviero, S., 2017. Intragenic DNA methylation prevents spurious transcription initiation. Nature 543, 72–77. doi:10.1038/nature21373

Neumann, B., Kubicka, P., Barlow, D.P., 1995. Characteristics of imprinted genes. Nat. Genet. 9, 12–13. doi:10.1038/ng0195-12

Ng, H.H., Zhang, Y., Hendrich, B., Johnson, C.A., Turner, B.M., Erdjument-Bromage, H., Tempst, P., Reinberg, D., Bird, A., 1999. MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. Nat. Genet. 23, 58–61. doi:10.1038/12659

Nguyen, M.V.C., Du, F., Felice, C.A., Shan, X., Nigam, A., Mandel, G., Robinson, J.K., Ballas, N., 2012. MeCP2 Is Critical for Maintaining Mature Neuronal Networks and Global Brain Anatomy during Late Stages of Postnatal Brain Development and in the Mature Adult Brain. J. Neurosci. 32, 10021–10034. doi:10.1523/JNEUROSCI.1316-12.2012

Nichols, J., Jones, K., Phillips, J.M., Newland, S.A., Roode, M., Mansfield, W., Smith, A., Cooke, A., 2009. Validated germline-competent embryonic stem cell lines from nonobese diabetic mice. Nat. Med. 15, 814–818. doi:10.1038/nm.1996

Nichols, J., Silva, J., Roode, M., Smith, A., 2009. Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. Development 136, 3215–3222. doi:10.1242/dev.038893

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H., Smith, A., 1998. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell 95, 379–391.

Nielsen, A.L., Ortiz, J.A., You, J., Oulad-Abdelghani, M., Khechumian, R., Gansmuller, A., Chambon, P., Losson, R., 1999. Interaction with members of the heterochromatin protein 1 (HP1) family and histone deacetylation are differentially involved in transcriptional silencing by members of the TIF1 family. EMBO J. 18, 6385–6395. doi:10.1093/emboj/18.22.6385

Nishibuchi, G., Déjardin, J., 2017. The molecular basis of the organization of repetitive DNA-containing constitutive heterochromatin in mammals. Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol. 25, 77–87. doi:10.1007/s10577-016-9547-3

Nishioka, Y., Leder, P., 1979. The complete sequence of a chromosomal mouse alpha--globin gene reveals elements conserved throughout vertebrate evolution. Cell 18, 875–882.

Nitta, H., Unoki, M., Ichiyanagi, K., Kosho, T., Shigemura, T., Takahashi, H., Velasco, G., Francastel, C., Picard, C., Kubota, T., Sasaki, H., 2013. Three novel ZBTB24 mutations identified in Japanese and Cape Verdean type 2 ICF syndrome patients. J. Hum. Genet. 58, 455–460. doi:10.1038/jhg.2013.56

Niwa, H., Burdon, T., Chambers, I., Smith, A., 1998. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Genes Dev. 12, 2048–2060.

Noguer-Dance, M., Abu-Amero, S., Al-Khtib, M., Lefevre, A., Coullin, P., Moore, G.E., Cavaille, J., 2010. The primate-specific microRNA gene cluster (C19MC) is imprinted in the placenta. Hum. Mol. Genet. 19, 3566–3582. doi:10.1093/hmg/ddq272

Oei, S.-L., Babich, V.S., Kazakov, V.I., Usmanova, N.M., Kropotov, A.V., Tomilin, N.V., 2004. Clusters of regulatory signals for RNA polymerase II transcription associated with Alu family repeats and CpG islands in human promoters. Genomics 83, 873–882. doi:10.1016/j.ygeno.2003.11.001

Ohno, S., Kaplan, W.D., Kinosita, R., 1959. Formation of the sex chromatin by a single X-chromosome in liver cells of Rattus norvegicus. Exp. Cell Res. 18, 415–418.

Ohshima, K., Hamada, M., Terai, Y., Okada, N., 1996. The 3' ends of tRNA-derived short interspersed repetitive elements are derived from the 3' ends of long interspersed repetitive elements. Mol. Cell. Biol. 16, 3756–3764.

Okada, N., 1991. SINEs. Curr. Opin. Genet. Dev. 1, 498–504.

Okamoto, I., Otte, A.P., Allis, C.D., Reinberg, D., Heard, E., 2004. Epigenetic dynamics of imprinted X inactivation during early mouse development. Science 303, 644–649. doi:10.1126/science.1092727

Okano, M., Bell, D.W., Haber, D.A., Li, E., 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99, 247–257.

Okano, M., Xie, S., Li, E., 1998. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat. Genet. 19, 219–220. doi:10.1038/890

Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D., Cheng, X., Bestor, T.H., 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448, 714–717. doi:10.1038/nature05987

Ord, M.G., Stocken, L.A., 1967. Phosphate and thiol groups in histone f3 from rat liver and thymus nuclei. Biochem. J. 102, 631–636.

Orkin, S.H., Hochedlinger, K., 2011. Chromatin connections to pluripotency and cellular reprogramming. Cell 145, 835–850. doi:10.1016/j.cell.2011.05.019

Osanai, M., Takahashi, H., Kojima, K.K., Hamada, M., Fujiwara, H., 2004. Essential motifs in the 3' untranslated region required for retrotransposition and the precise start of reverse transcription in non-long-terminal-repeat retrotransposon SART1. Mol. Cell. Biol. 24, 7902–7913. doi:10.1128/MCB.24.18.7902-7913.2004

Osterhage, J.L., Friedman, K.L., 2009. Chromosome End Maintenance by Telomerase. J. Biol. Chem. 284, 16061–16065. doi:10.1074/jbc.R900011200

Ostler, K.R., Davis, E.M., Payne, S.L., Gosalia, B.B., Exposito-Cespedes, J., Le Beau, M.M., Godley, L.A., 2007. Cancer cells express aberrant DNMT3B transcripts encoding truncated proteins. Oncogene 26, 5553–5563.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., Walter, J., 2000. Active demethylation of the paternal genome in the mouse zygote. Curr. Biol. CB 10, 475–478.

Ouyang, B., Baxter, C.S., Lam, H.-M., Yeramaneni, S., Levin, L., Haynes, E., Ho, S., 2012. Hypomethylation of Dual Specificity Phosphatase 22 Promoter Correlates With Duration of Service in Firefighters and Is Inducible by Low-Dose Benzo[a]Pyrene: J. Occup. Environ. Med. 54, 774–780. doi:10.1097/JOM.0b013e31825296bc

Padgett, R.W., Hutchison, C.A., Edgell, M.H., 1988. The F-type 5' motif of mouse L1 elements: a major class of L1 termini similar to the A-type in organization but unrelated in sequence. Nucleic Acids Res. 16, 739–749.

Pal, S., Tyler, J.K., 2016. Epigenetics and aging. Sci. Adv. 2, e1600584–e1600584. doi:10.1126/sciadv.1600584

Palade, G.E., 1955. A small particulate component of the cytoplasm. J. Biophys. Biochem. Cytol. 1, 59–68.

Palmer, D.K., O'Day, K., Trong, H.L., Charbonneau, H., Margolis, R.L., 1991. Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proc. Natl. Acad. Sci. U. S. A. 88, 3734–3738.

Palmer, D.K., O'Day, K., Wener, M.H., Andrews, B.S., Margolis, R.L., 1987. A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. J. Cell Biol. 104, 805–815.

Papin, C., Ibrahim, A., Le Gras, S., Velt, A., Jost, B., Stoll, I., Menoni, H., Bronner, C., Dimitrov, S., Hamiche, A., 2017. Combinatorial DNA methylation codes at repetitive elements. Genome Res. doi:10.1101/gr.213983.116

Park, H.-J., Yu, E., Shim, Y.-H., 2006. DNA methyltransferase expression and DNA hypermethylation in human hepatocellular carcinoma. Cancer Lett. 233, 271–278. doi:10.1016/j.canlet.2005.03.017

Passarge, E., 1979. Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago. Am. J. Hum. Genet. 31, 106–115.

Pastor, W.A., Aravind, L., Rao, A., 2013. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nat. Rev. Mol. Cell Biol. 14, 341–356. doi:10.1038/nrm3589

Pauler, F.M., Sloane, M.A., Huang, R., Regha, K., Koerner, M.V., Tamir, I., Sommer, A., Aszodi, A., Jenuwein, T., Barlow, D.P., 2008. H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. Genome Res. 19, 221–233. doi:10.1101/gr.080861.108

Pavri, R., Zhu, B., Li, G., Trojer, P., Mandal, S., Shilatifard, A., Reinberg, D., 2006. Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. Cell 125, 703–717. doi:10.1016/j.cell.2006.04.029

Paweletz, N., 2001. Walther Flemming: pioneer of mitosis research. Nat. Rev. Mol. Cell Biol. 2, 72–75. doi:10.1038/35048077

Peaston, A.E., Evsikov, A.V., Graber, J.H., de Vries, W.N., Holbrook, A.E., Solter, D., Knowles, B.B., 2004. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev. Cell 7, 597–606. doi:10.1016/j.devcel.2004.09.004

Pehrson, J.R., Fuji, R.N., 1998. Evolutionary conservation of histone macroH2A subtypes and domains.

Nucleic Acids Res. 26, 2837–2842.

Peng, J.C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., Wysocka, J., 2009. Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. Cell 139, 1290–1302. doi:10.1016/j.cell.2009.12.002

Penn, N.W., Suwalski, R., O'Riley, C., Bojanowski, K., Yura, R., 1972. The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. Biochem. J. 126, 781–790.

Perea-Gomez, A., Vella, F.D.J., Shawlot, W., Oulad-Abdelghani, M., Chazaud, C., Meno, C., Pfister, V., Chen, L., Robertson, E., Hamada, H., Behringer, R.R., Ang, S.-L., 2002. Nodal antagonists in the anterior visceral endoderm prevent the formation of multiple primitive streaks. Dev. Cell 3, 745–756.

Perez-Torrado, R., Yamada, D., Defossez, P.-A., 2006. Born to bind: the BTB protein–protein interaction domain. BioEssays 28, 1194–1202. doi:10.1002/bies.20500

Peters, A.H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schöfer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A., Opravil, S., Doyle, M., Sibilia, M., Jenuwein, T., 2001. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell 107, 323–337.

Peters, A.H., Schübeler, D., 2005. Methylation of histones: playing memory with DNA. Curr. Opin. Cell Biol. 17, 230–238. doi:10.1016/j.ceb.2005.02.006

Peters, A.H.F.M., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A.H.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., Martens, J.H.A., Jenuwein, T., 2003. Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Mol. Cell 12, 1577–1589.

Peters, A.H.F.M., Mermoud, J.E., O'Carroll, D., Pagani, M., Schweizer, D., Brockdorff, N., Jenuwein, T., 2002. Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. Nat. Genet. 30, 77–80. doi:10.1038/ng789

Pezer, Z., Ugarković, D., 2008. RNA Pol II promotes transcription of centromeric satellite DNA in beetles. PloS One 3, e1594. doi:10.1371/journal.pone.0001594

Pinheiro, I., Margueron, R., Shukeir, N., Eisold, M., Fritzsch, C., Richter, F.M., Mittler, G., Genoud, C., Goyama, S., Kurokawa, M., Son, J., Reinberg, D., Lachner, M., Jenuwein, T., 2012. Prdm3 and Prdm16 are H3K9me1 methyltransferases required for mammalian heterochromatin integrity. Cell 150, 948–960. doi:10.1016/j.cell.2012.06.048

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., Zhang, Y., 2003. Role of histone H3 lysine 27 methylation in X inactivation. Science 300, 131–135. doi:10.1126/science.1084274

Pohlers, M., Truss, M., Frede, U., Scholz, A., Strehle, M., Kuban, R.-J., Hoffmann, B., Morkel, M., Birchmeier, C., Hagemeier, C., 2005. A role for E2F6 in the restriction of male-germ-cell-specific gene expression. Curr. Biol. CB 15, 1051–1057. doi:10.1016/j.cub.2005.04.060

Pontes, T.B., Chen, E.S., Gigek, C.O., Calcagno, D.Q., Wisnieski, F., Leal, M.F., Demachki, S., Assumpção, P.P., Artigiani, R., Lourenço, L.G., Burbano, R.R., Arruda Cardoso Smith, M., 2014.

Reduced mRNA expression levels of MBD2 and MBD3 in gastric carcinogenesis. Tumour Biol. J. Int. Soc. Oncodevelopmental Biol. Med. 35, 3447–3453. doi:10.1007/s13277-013-1455-y

Porro, A., Feuerhahn, S., Delafontaine, J., Riethman, H., Rougemont, J., Lingner, J., 2014. Functional characterization of the TERRA transcriptome at damaged telomeres. Nat. Commun. 5, 5379. doi:10.1038/ncomms6379

Porro, A., Feuerhahn, S., Reichenbach, P., Lingner, J., 2010. Molecular dissection of telomeric repeat-containing RNA biogenesis unveils the presence of distinct and multiple regulatory pathways. Mol. Cell. Biol. 30, 4808–4817. doi:10.1128/MCB.00460-10

Pradhan, S., Bacolla, A., Wells, R.D., Roberts, R.J., 1999. Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. J. Biol. Chem. 274, 33002–33010.

Prokhortchouk, A., Hendrich, B., Jørgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A., Prokhortchouk, E., 2001. The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. Genes Dev. 15, 1613–1618. doi:10.1101/gad.198501

Pronier, E., Almire, C., Mokrani, H., Vasanthakumar, A., Simon, A., da Costa Reis Monte Mor, B., Masse, A., Le Couedic, J.-P., Pendino, F., Carbonne, B., Larghero, J., Ravanat, J.-L., Casadevall, N., Bernard, O.A., Droin, N., Solary, E., Godley, L.A., Vainchenker, W., Plo, I., Delhommeau, F., 2011. Inhibition of TET2-mediated conversion of 5-methylcytosine to 5-hydroxymethylcytosine disturbs erythroid and granulomonocytic differentiation of human hematopoietic progenitors. Blood 118, 2551–2555. doi:10.1182/blood-2010-12-324707

Quénet, D., Dalal, Y., 2014. A long non-coding RNA is required for targeting centromeric protein A to the human centromere. eLife 3, e03254. doi:10.7554/eLife.03254

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P.V., Grimaldi, G., Riccio, A., Trono, D., 2011. In Embryonic Stem Cells, ZFP57/KAP1 Recognize a Methylated Hexanucleotide to Affect Chromatin and DNA Methylation of Imprinting Control Regions. Mol. Cell 44, 361–372. doi:10.1016/j.molcel.2011.08.032

Rahman, A., Isenberg, D.A., 2008. Systemic lupus erythematosus. N. Engl. J. Med. 358, 929–939. doi:10.1056/NEJMra071297

Raj, K., Mufti, G.J., 2006. Azacytidine (Vidaza(R)) in the treatment of myelodysplastic syndromes. Ther. Clin. Risk Manag. 2, 377–388.

Ralston, A., Rossant, J., 2008. Cdx2 acts downstream of cell polarization to cell-autonomously promote trophectoderm fate in the early mouse embryo. Dev. Biol. 313, 614–629. doi:10.1016/j.ydbio.2007.10.054

Rampersaud, G.C., Kauwell, G.P., Hutson, A.D., Cerda, J.J., Bailey, L.B., 2000. Genomic DNA methylation decreases in response to moderate folate depletion in elderly women. Am. J. Clin. Nutr. 72, 998–1003.

Rasmussen, K.D., Helin, K., 2016. Role of TET enzymes in DNA methylation, development, and cancer. Genes Dev. 30, 733–750. doi:10.1101/gad.276568.115

Ratnakumar, K., Duarte, L.F., LeRoy, G., Hasson, D., Smeets, D., Vardabasso, C., Bönisch, C., Zeng, T., Xiang, B., Zhang, D.Y., Li, H., Wang, X., Hake, S.B., Schermelleh, L., Garcia, B.A., Bernstein, E., 2012. ATRX-mediated chromatin association of histone variant macroH2A1 regulates α-globin expression. Genes Dev. 26, 433–438. doi:10.1101/gad.179416.111

Rauch, T.A., Pfeifer, G.P., 2009. The MIRA method for DNA methylation analysis. Methods Mol. Biol. Clifton NJ 507, 65–75. doi:10.1007/978-1-59745-522-0_6

Ray-Gallet, D., Quivy, J.-P., Scamps, C., Martini, E.M.-D., Lipinski, M., Almouzni, G., 2002. HIRA is critical for a nucleosome assembly pathway independent of DNA synthesis. Mol. Cell 9, 1091–1100.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., Jenuwein, T., 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature 406, 593–599. doi:10.1038/35020506

Rechavi, E., Lev, A., Eyal, E., Barel, O., Kol, N., Barhom, S.F., Pode-Shakked, B., Anikster, Y., Somech, R., Simon, A.J., 2016. A Novel Mutation in a Critical Region for the Methyl Donor Binding in DNMT3B Causes Immunodeficiency, Centromeric Instability, and Facial Anomalies Syndrome (ICF). J. Clin. Immunol. 36, 801–809. doi:10.1007/s10875-016-0340-z

Redon, C., Pilch, D., Rogakou, E., Sedelnikova, O., Newrock, K., Bonner, W., 2002. Histone H2A variants H2AX and H2AZ. Curr. Opin. Genet. Dev. 12, 162–169.

Reik, W., Collick, A., Norris, M.L., Barton, S.C., Surani, M.A., 1987. Genomic imprinting determines methylation of parental alleles in transgenic mice. Nature 328, 248–251. doi:10.1038/328248a0

Reik, W., Dean, W., Walter, J., 2001. Epigenetic reprogramming in mammalian development. Science 293, 1089–1093. doi:10.1126/science.1063443

Reilly, M.T., Faulkner, G.J., Dubnau, J., Ponomarev, I., Gage, F.H., 2013. The Role of Transposable Elements in Health and Diseases of the Central Nervous System. J. Neurosci. 33, 17577–17586. doi:10.1523/JNEUROSCI.3369-13.2013

Ren, J., Singh, B.N., Huang, Q., Li, Z., Gao, Y., Mishra, P., Hwa, Y.L., Li, J., Dowdy, S.C., Jiang, S.-W., 2011. DNA hypermethylation as a chemotherapy target. Cell. Signal. 23, 1082–1093. doi:10.1016/j.cellsig.2011.02.003

Ribatti, D., 2014. The chemical nature of the factor responsible for embryonic induction: An historical overview. Organogenesis 10, 38–43. doi:10.4161/org.28026

Ribet, D., Louvet-Vallée, S., Harper, F., de Parseval, N., Dewannieux, M., Heidmann, O., Pierron, G., Maro, B., Heidmann, T., 2008. Murine endogenous retrovirus MuERV-L is the progenitor of the "orphan" epsilon viruslike particles of the early mouse embryo. J. Virol. 82, 1622–1625. doi:10.1128/JVI.02097-07

Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D., Klug, A., 1984. Structure of the nucleosome core particle at 7 A resolution. Nature 311, 532–537.

Riethman, H., Ambrosini, A., Paul, S., 2005. Human subtelomere structure and variation. Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol. 13, 505–515. doi:10.1007/s10577-005-0998-1

Riggs, A.D., 1975. X inactivation, differentiation, and DNA methylation. Cytogenet. Cell Genet. 14, 9–25.

Ringrose, L., Paro, R., 2004. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. Annu. Rev. Genet. 38, 413–443. doi:10.1146/annurev.genet.38.072902.091907

Rizzi, N., Denegri, M., Chiodi, I., Corioni, M., Valgardsdottir, R., Cobianchi, F., Riva, S., Biamonti, G., 2004. Transcriptional activation of a constitutive heterochromatic domain of the human genome in response to heat shock. Mol. Biol. Cell 15, 543–551. doi:10.1091/mbc.E03-07-0487

Robertson, K.D., 2005. DNA methylation and human disease. Nat. Rev. Genet. 6, 597–610. doi:10.1038/nrg1655

Robertson, K.D., Uzvolgyi, E., Liang, G., Talmadge, C., Sumegi, J., Gonzales, F.A., Jones, P.A., 1999. The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors. Nucleic Acids Res. 27, 2291–2298.

Robertson, M.A., Staden, R., Tanaka, Y., Catterall, J.F., O'malley, B.W., Brownlee, G.G., 1979. Sequence of three introns in the chick ovalbumin gene. Nature 278, 370–372. doi:10.1038/278370a0

Rodić, N., Oka, M., Hamazaki, T., Murawski, M.R., Jorgensen, M., Maatouk, D.M., Resnick, J.L., Li, E., Terada, N., 2005. DNA methylation is required for silencing of ant4, an adenine nucleotide translocase selectively expressed in mouse embryonic stem cells and germ cells. Stem Cells Dayt. Ohio 23, 1314–1323. doi:10.1634/stemcells.2005-0119

Rogakou, E.P., Sekeri–Pataryas, K.E., 1999. Histone variants of H2A and H3 families are regulated during in vitro aging in the same manner as during differentiation. Exp. Gerontol. 34, 741–754. doi:10.1016/S0531-5565(99)00046-7

Rogers, M.B., Hosler, B.A., Gudas, L.J., 1991. Specific expression of a retinoic acid-regulated, zinc-finger gene, Rex-1, in preimplantation embryos, trophoblast and spermatocytes. Development 113, 815–824.

Rollins, R.A., Haghighi, F., Edwards, J.R., Das, R., Zhang, M.Q., Ju, J., Bestor, T.H., 2006. Large-scale structure of genomic methylation patterns. Genome Res. 16, 157–163. doi:10.1101/gr.4362006

Rondelet, G., Dal Maso, T., Willems, L., Wouters, J., 2016. Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. J. Struct. Biol. 194, 357–367. doi:10.1016/j.jsb.2016.03.013

Rose, N.R., Klose, R.J., 2014. Understanding the relationship between DNA methylation and histone lysine methylation. Biochim. Biophys. Acta BBA - Gene Regul. Mech. 1839, 1362–1372. doi:10.1016/j.bbagrm.2014.02.007

Rothbart, S.B., Dickson, B.M., Ong, M.S., Krajewski, K., Houliston, S., Kireev, D.B., Arrowsmith, C.H., Strahl, B.D., 2013. Multivalent histone engagement by the linked tandem Tudor and PHD domains of UHRF1 is required for the epigenetic inheritance of DNA methylation. Genes Dev. 27, 1288–1298. doi:10.1101/gad.220467.113

Roussel-Gervais, A., Naciri, I., Kirsh, O., Kasprzyk, L., Velasco, G., Grillo, G., Dubus, P., Defossez, P.-A., 2017. Loss of the Methyl-CpG–Binding Protein ZBTB4 Alters Mitotic Checkpoint, Increases Aneuploidy, and Promotes Tumorigenesis. Cancer Res. 77, 62–73. doi:10.1158/0008-5472.CAN-16-1181

Rudert, F., Bronner, S., Garnier, J.M., Dollé, P., 1995. Transcripts from opposite strands of gamma satellite DNA are differentially expressed during mouse development. Mamm. Genome Off. J. Int. Mamm. Genome Soc. 6, 76–83.

Ryan, R.F., Schultz, D.C., Ayyanathan, K., Singh, P.B., Friedman, J.R., Fredericks, W.J., Rauscher, F.J., 1999. KAP-1 corepressor protein interacts and colocalizes with heterochromatic and euchromatic HP1 proteins: a potential role for Krüppel-associated box-zinc finger proteins in heterochromatin-mediated gene silencing. Mol. Cell. Biol. 19, 4366–4378.

Sadikovic, B., Rodenhiser, D., 2006. Benzopyrene exposure disrupts DNA methylation and growth dynamics in breast cancer cells. Toxicol. Appl. Pharmacol. 216, 458–468. doi:10.1016/j.taap.2006.06.012

Sagie, S., Toubiana, S., Hartono, S.R., Katzir, H., Tzur-Gilat, A., Havazelet, S., Francastel, C., Velasco, G., Chédin, F., Selig, S., 2017. Telomeres in ICF syndrome cells are vulnerable to DNA damage due to elevated DNA:RNA hybrids. Nat. Commun. 8, 14015. doi:10.1038/ncomms14015

Saito, Y., Kanai, Y., Sakamoto, M., Saito, H., Ishii, H., Hirohashi, S., 2002. Overexpression of a splice variant of DNA methyltransferase 3b, DNMT3b4, associated with DNA hypomethylation on pericentromeric satellite regions during human hepatocarcinogenesis. Proc. Natl. Acad. Sci. U. S. A. 99, 10060–10065. doi:10.1073/pnas.152121799

Saitou, M., Kagiwada, S., Kurimoto, K., 2012. Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. Dev. Camb. Engl. 139, 15–31. doi:10.1242/dev.050849

Sakaue, M., Ohta, H., Kumaki, Y., Oda, M., Sakaide, Y., Matsuoka, C., Yamagiwa, A., Niwa, H., Wakayama, T., Okano, M., 2010. DNA Methylation Is Dispensable for the Growth and Survival of the Extraembryonic Lineages. Curr. Biol. 20, 1452–1457. doi:10.1016/j.cub.2010.06.050

Saksouk, N., Simboeck, E., Déjardin, J., 2015. Constitutive heterochromatin formation and transcription in mammals. Epigenetics Chromatin 8. doi:10.1186/1756-8935-8-3

Samaco, R.C., 2004. Multiple pathways regulate MeCP2 expression in normal brain development and exhibit defects in autism-spectrum disorders. Hum. Mol. Genet. 13, 629–639. doi:10.1093/hmg/ddh063

Sandmeyer, S.B., Clemens, K.A., 2010. Function of a retrotransposon nucleocapsid protein. RNA Biol. 7, 642–654.

Sandoval, J., Heyn, H., Moran, S., Serra-Musach, J., Pujana, M.A., Bibikova, M., Esteller, M., 2011. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics 6, 692–702. doi:10.4161/epi.6.6.16196

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchison, C.A., Slocombe, P.M., Smith, M., 1977. Nucleotide sequence of bacteriophage |[phi]|X174 DNA. Nature 265, 687–695. doi:10.1038/265687a0

Santini, V., Kantarjian, H.M., Issa, J.P., 2001. Changes in DNA methylation in neoplasia: pathophysiology and therapeutic implications. Ann. Intern. Med. 134, 573–586.

Sanz, L.A., Kota, S.K., Feil, R., 2010. Genome-wide DNA demethylation in mammals. Genome Biol. 11, 110. doi:10.1186/gb-2010-11-3-110

Sapienza, C., Peterson, A.C., Rossant, J., Balling, R., 1987. Degree of methylation of transgenes is dependent on gamete of origin. Nature 328, 251–254. doi:10.1038/328251a0

Sasai, N., Nakao, M., Defossez, P.-A., 2010. Sequence-specific recognition of methylated DNA by human zinc-finger proteins. Nucleic Acids Res. 38, 5015–5022. doi:10.1093/nar/gkq280

Sassaman, D.M., Dombroski, B.A., Moran, J.V., Kimberland, M.L., Naas, T.P., DeBerardinis, R.J., Gabriel, A., Swergold, G.D., Kazazian, H.H., 1997. Many human L1 elements are capable of retrotransposition. Nat. Genet. 16, 37–43. doi:10.1038/ng0597-37

Scanlan, M.J., Gure, A.O., Jungbluth, A.A., Old, L.J., Chen, Y.-T., 2002. Cancer/testis antigens: an expanding family of targets for cancer immunotherapy. Immunol. Rev. 188, 22–32.

Scheffer, I., Brett, E.M., Wilson, J., Baraitser, M., 1990. Angelman's syndrome. J. Med. Genet. 27, 275.

Scheibe, M., Arnoult, N., Kappei, D., Buchholz, F., Decottignies, A., Butter, F., Mann, M., 2013. Quantitative interaction screen of telomeric repeat-containing RNA reveals novel TERRA regulators. Genome Res. 23, 2149–2157. doi:10.1101/gr.151878.112

Scheibye-Knudsen, M., 2016. Neurodegeneration in accelerated aging. Dan. Med. J. 63.

Schichman, S.A., Adey, N.B., Edgell, M.H., Hutchison, C.A., 1993. L1 A-monomer tandem arrays have expanded during the course of mouse L1 evolution. Mol. Biol. Evol. 10, 552–570.

Schmelz, K., Sattler, N., Wagner, M., Lübbert, M., Dörken, B., Tamm, I., 2004. Induction of gene expression by 5-Aza-2′-deoxycytidine in acute myeloid leukemia (AML) and myelodysplastic syndrome (MDS) but not epithelial cells by DNA-methylation-dependent and -independent mechanisms. Leukemia. doi:10.1038/sj.leu.2403552

Schmidt, W.M., Sedivy, R., Forstner, B., Steger, G.G., Zöchbauer-Müller, S., Mader, R.M., 2007. Progressive up-regulation of genes encoding DNA methyltransferases in the colorectal adenoma-carcinoma sequence. Mol. Carcinog. 46, 766–772. doi:10.1002/mc.20307

Schmitz, J., Brosius, J., 2011. Exonization of transposed elements: A challenge and opportunity for evolution. Biochimie 93, 1928–1934. doi:10.1016/j.biochi.2011.07.014

Schoeftner, S., Blasco, M.A., 2009. A "higher order" of telomere regulation: telomere heterochromatin and telomeric RNAs. EMBO J. 28, 2323–2336. doi:10.1038/emboj.2009.197

Schoeftner, S., Blasco, M.A., 2008. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. Nat. Cell Biol. 10, 228–236. doi:10.1038/ncb1685

Schoenherr, C.J., Levorse, J.M., Tilghman, S.M., 2002. CTCF maintains differential methylation at the Igf2/H19 locus. Nat. Genet. 33, 66–69. doi:10.1038/ng1057

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., Jenuwein, T., 2004. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev. 18, 1251–1262. doi:10.1101/gad.300704

Schotta, G., Sengupta, R., Kubicek, S., Malin, S., Kauer, M., Callén, E., Celeste, A., Pagani, M., Opravil, S., De La Rosa-Velazquez, I.A., Espejo, A., Bedford, M.T., Nussenzweig, A., Busslinger, M., Jenuwein, T., 2008. A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. Genes Dev. 22, 2048–2061. doi:10.1101/gad.476008

Schramm, L., Hernandez, N., 2002. Recruitment of RNA polymerase III to its target promoters. Genes Dev. 16, 2593–2620. doi:10.1101/gad.1018902

Schübeler, D., 2015. Function and information content of DNA methylation. Nature 517, 321–326. doi:10.1038/nature14192

Schueler, M.G., Sullivan, B.A., 2006. Structural and functional dynamics of human centromeric chromatin. Annu. Rev. Genomics Hum. Genet. 7, 301–313. doi:10.1146/annurev.genom.7.080505.115613

Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., Rauscher, F.J., 2002. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. Genes Dev. 16, 919–932.

Schultz, D.C., Friedman, J.R., Rauscher, F.J., 2001. Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD. Genes Dev. 15, 428–443. doi:10.1101/gad.869501

Schultz, J., 1941. THE EVIDENCE OF THE NUCLEOPROTEIN NATURE OF THE GENE. Cold Spring Harb. Symp. Quant. Biol. 9, 55–65. doi:10.1101/SQB.1941.009.01.008

Schulz, E.G., Meisig, J., Nakamura, T., Okamoto, I., Sieber, A., Picard, C., Borensztein, M., Saitou, M., Blüthgen, N., Heard, E., 2014. The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network. Cell Stem Cell 14, 203–216. doi:10.1016/j.stem.2013.11.022

Scott, A.F., Schmeckpeper, B.J., Abdelrazik, M., Comey, C.T., O'Hara, B., Rossiter, J.P., Cooley, T., Heath, P., Smith, K.D., Margolet, L., 1987. Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. Genomics 1, 113–125.

Scourzic, L., Mouly, E., Bernard, O.A., 2015. TET proteins and the control of cytosine demethylation in cancer. Genome Med. 7, 9. doi:10.1186/s13073-015-0134-6

Seitz, H., Royo, H., Bortolin, M.-L., Lin, S.-P., Ferguson-Smith, A.C., Cavaillé, J., 2004. A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. Genome Res. 14, 1741–1748. doi:10.1101/gr.2743304

Shahbazian, M.D., Zoghbi, H.Y., 2001. Molecular genetics of Rett syndrome and clinical spectrum of MECP2 mutations. Curr. Opin. Neurol. 14, 171–176.

Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J.R., Wu, C.T., Bender, W., Kingston, R.E., 1999. Stabilization of chromatin structure by PRC1, a Polycomb complex. Cell 98, 37–46. doi:10.1016/S0092-8674(00)80604-2

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., Tajima, S., Mitsuya, K., Okano, M., Koseki, H., 2007. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature 450, 908–912. doi:10.1038/nature06397

Shiio, Y., Eisenman, R.N., 2003. Histone sumoylation is associated with transcriptional repression. Proc. Natl. Acad. Sci. U. S. A. 100, 13225–13230. doi:10.1073/pnas.1735528100

Siggs, O., Beutler, B., 2012. The BTB-ZF transcription factors. Cell Cycle 11, 3358–3369.

doi:10.4161/cc.21277

Simon, J.A., Kingston, R.E., 2009. Mechanisms of polycomb gene silencing: knowns and unknowns. Nat. Rev. Mol. Cell Biol. 10, 697–708. doi:10.1038/nrm2763

Simo-Riudalbas, L., Diaz-Lagares, A., Gatto, S., Gagliardi, M., Crujeiras, A.B., Matarazzo, M.R., Esteller, M., Sandoval, J., 2015. Genome-Wide DNA Methylation Analysis Identifies Novel Hypomethylated Non-Pericentromeric Genes with Potential Clinical Implications in ICF Syndrome. PloS One 10, e0132517. doi:10.1371/journal.pone.0132517

Singer, M.F., 1982. SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. Cell 28, 433–434.

Smallwood, S.A., Tomizawa, S.-I., Krueger, F., Ruf, N., Carli, N., Segonds-Pichon, A., Sato, S., Hata, K., Andrews, S.R., Kelsey, G., 2011. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat. Genet. 43, 811–814. doi:10.1038/ng.864

Smith, A.D., Kim, Y.-I., Refsum, H., 2008. Is folic acid good for everyone? Am. J. Clin. Nutr. 87, 517–533.

Smith, A.G., Heath, J.K., Donaldson, D.D., Wong, G.G., Moreau, J., Stahl, M., Rogers, D., 1988. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature 336, 688–690. doi:10.1038/336688a0

Smith, F.M., Garfield, A.S., Ward, A., 2006. Regulation of growth and metabolism by imprinted genes. Cytogenet. Genome Res. 113, 279–291. doi:10.1159/000090843

Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A., Meissner, A., 2012. A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature 484, 339–344. doi:10.1038/nature10960

Song, J., Teplova, M., Ishibe-Murakami, S., Patel, D.J., 2012. Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science 335, 709–712.

Sookdeo, A., Hepp, C.M., McClure, M.A., Boissinot, S., 2013. Revisiting the evolution of mouse LINE-1 in the genomic era. Mob. DNA 4, 3. doi:10.1186/1759-8753-4-3

Sorek, R., Ast, G., Graur, D., 2002. Alu-containing exons are alternatively spliced. Genome Res. 12, 1060–1067. doi:10.1101/gr.229302

Spencer, D.H., Russler-Germain, D.A., Ketkar, S., Helton, N.M., Lamprecht, T.L., Fulton, R.S., Fronick, C.C., O'Laughlin, M., Heath, S.E., Shinawi, M., Westervelt, P., Payton, J.E., Wartman, L.D., Welch, J.S., Wilson, R.K., Walter, M.J., Link, D.C., DiPersio, J.F., Ley, T.J., 2017. CpG Island Hypermethylation Mediated by DNMT3A Is a Consequence of AML Progression. Cell 168, 801–816.e13. doi:10.1016/j.cell.2017.01.021

Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W.T.C., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., Eberl, H.C., Mensinga, A., Brinkman, A.B., Lephikov, K., Müller, U., Walter, J., Boelens, R., van Ingen, H., Leonhardt, H., Carell, T., Vermeulen, M., 2013. Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives. Cell 152, 1146–1159. doi:10.1016/j.cell.2013.02.004

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., Tiwari, V.K., Schübeler, D., 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature 480, 490–495. doi:10.1038/nature10716

Stamhuis, I.H., 2003. The reactions on Hugo de Vries's Intracellular pangenesis: the discussion with August Weismann. J. Hist. Biol. 36, 119–152.

Stanley, J.S., Griffin, J.B., Zempleni, J., 2001. Biotinylation of histones in human cells. Effects of cell proliferation. Eur. J. Biochem. 268, 5424–5429.

Statham, A.L., Robinson, M.D., Song, J.Z., Coolen, M.W., Stirzaker, C., Clark, S.J., 2012. Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. Genome Res. 22, 1120–1127. doi:10.1101/gr.132076.111

Stedman, E., 1950. Cell specificity of histones. Nature 166, 780–781.

Sterlin, D., Velasco, G., Moshous, D., Touzot, F., Mahlaoui, N., Fischer, A., Suarez, F., Francastel, C., Picard, C., 2016. Genetic, Cellular and Clinical Features of ICF Syndrome: a French National Survey. J. Clin. Immunol. 36, 149–159. doi:10.1007/s10875-016-0240-2

Sterner, D.E., Berger, S.L., 2000. Acetylation of histones and transcription-related factors. Microbiol. Mol. Biol. Rev. 64, 435–459.

Stimpson, K.M., Sullivan, B.A., 2010. Epigenomics of centromere assembly and function. Curr. Opin. Cell Biol. 22, 772–780. doi:10.1016/j.ceb.2010.07.002

Stirzaker, C., Taberlay, P.C., Statham, A.L., Clark, S.J., 2014. Mining cancer methylomes: prospects and challenges. Trends Genet. TIG 30, 75–84. doi:10.1016/j.tig.2013.11.004

Stock, J.K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A.G., Pombo, A., 2007. Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nat. Cell Biol. 9, 1428–1435. doi:10.1038/ncb1663

Stoler, S., Keith, K.C., Curnick, K.E., Fitzgerald-Hayes, M., 1995. A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. Genes Dev. 9, 573–586.

Storre, J., Elsässer, H.-P., Fuchs, M., Ullmann, D., Livingston, D.M., Gaubatz, S., 2002. Homeotic transformations of the axial skeleton that accompany a targeted deletion of E2f6. EMBO Rep. 3, 695–700. doi:10.1093/embo-reports/kvf141

Storre, J., Schäfer, A., Reichert, N., Barbero, J.L., Hauser, S., Eilers, M., Gaubatz, S., 2005. Silencing of the meiotic genes SMC1beta and STAG3 in somatic cells by E2F6. J. Biol. Chem. 280, 41380–41386. doi:10.1074/jbc.M506797200

Strahl, B.D., Allis, C.D., 2000. The language of covalent histone modifications. Nature 403, 41–45. doi:10.1038/47412

Stribinskis, V., 2006. Activation of Human Long Interspersed Nuclear Element 1 Retrotransposition by Benzo(a)pyrene, an Ubiquitous Environmental Carcinogen. Cancer Res. 66, 2616–2620. doi:10.1158/0008-5472.CAN-05-3478

Strogantsev, R., Krueger, F., Yamazawa, K., Shi, H., Gould, P., Goldman-Roberts, M., McEwen, K., Sun,

B., Pedersen, R., Ferguson-Smith, A.C., 2015. Allele-specific binding of ZFP57 in the epigenetic regulation of imprinted and non-imprinted monoallelic expression. Genome Biol. 16. doi:10.1186/s13059-015-0672-7

Sunahori, K., Nagpal, K., Hedrich, C.M., Mizui, M., Fitzgerald, L.M., Tsokos, G.C., 2013. The Catalytic Subunit of Protein Phosphatase 2A (PP2Ac) Promotes DNA Hypomethylation by Suppressing the Phosphorylated Mitogen-activated Protein Kinase/Extracellular Signal-regulated Kinase (ERK) Kinase (MEK)/Phosphorylated ERK/DNMT1 Protein Pathway in T-cells from Controls and Systemic Lupus Erythematosus Patients. J. Biol. Chem. 288, 21936–21944. doi:10.1074/jbc.M113.467266

Surani, M.A., Barton, S.C., Norris, M.L., 1984. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature 308, 548–550.

Suzuki, M.M., Bird, A., 2008. DNA methylation landscapes: provocative insights from epigenomics. Nat. Rev. Genet. 9, 465–476. doi:10.1038/nrg2341

Tachibana, M., Sugimoto, K., Fukushima, T., Shinkai, Y., 2001. Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. J. Biol. Chem. 276, 25309–25317. doi:10.1074/jbc.M101914200

Tagami, H., Ray-Gallet, D., Almouzni, G., Nakatani, Y., 2004. Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell 116, 51–61.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., Rao, A., 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324, 930–935. doi:10.1126/science.1170116

Takahashi, K., Chen, E.S., Yanagida, M., 2000. Requirement of Mis6 centromere connector for localizing a CENP-A-like protein in fission yeast. Science 288, 2215–2219.

Takahashi, K., Yamanaka, S., 2006. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell 126, 663–676. doi:10.1016/j.cell.2006.07.024

Takahashi, T., Lord, B., Schulze, P.C., Fryer, R.M., Sarang, S.S., Gullans, S.R., Lee, R.T., 2003. Ascorbic acid enhances differentiation of embryonic stem cells into cardiac myocytes. Circulation 107, 1912–1916. doi:10.1161/01.CIR.0000064899.53876.A3

Tam, P.P., Behringer, R.R., 1997. Mouse gastrulation: the formation of a mammalian body plan. Mech. Dev. 68, 3–25.

Tanay, A., O'Donnell, A.H., Damelin, M., Bestor, T.H., 2007. Hyperconserved CpG domains underlie Polycomb-binding sites. Proc. Natl. Acad. Sci. U. S. A. 104, 5521–5526. doi:10.1073/pnas.0609746104

Tang, W., Levin, L., Talaska, G., Cheung, Y.Y., Herbstman, J., Tang, D., Miller, R.L., Perera, F., Ho, S.-M., 2012. Maternal Exposure to Polycyclic Aromatic Hydrocarbons and 5′-CpG Methylation of Interferon-γ in Cord White Blood Cells. Environ. Health Perspect. 120, 1195–1200. doi:10.1289/ehp.1103744

Taruscio, D., Mantovani, A., 2004. Factors regulating endogenous retroviral sequences in human and mouse. Cytogenet. Genome Res. 105, 351–362. doi:10.1159/000078208

Tatton-Brown, K., Seal, S., Ruark, E., Harmer, J., Ramsay, E., del Vecchio Duarte, S., Zachariou, A.,

Hanks, S., O'Brien, E., Aksglaede, L., Baralle, D., Dabir, T., Gener, B., Goudie, D., Homfray, T., Kumar, A., Pilz, D.T., Selicorni, A., Temple, I.K., Van Maldergem, L., Yachelevich, N., van Montfort, R., Rahman, N., 2014. Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. Nat. Genet. 46, 385–388. doi:10.1038/ng.2917

Taunton, J., Hassig, C.A., Schreiber, S.L., 1996. A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science 272, 408–411.

Taylor, S.M., 1993. 5-Aza-2'-deoxycytidine: cell differentiation and DNA methylation. Leukemia 7 Suppl 1, 3–8.

Teneng, I., Montoya-Durango, D.E., Quertermous, J.L., Lacy, M.E., Ramos, K.S., 2011. Reactivation of L1 retrotransposon by benzo(a)pyrene involves complex genetic and epigenetic regulation. Epigenetics 6, 355–367. doi:10.4161/epi.6.3.14282

Termanis, A., Torrea, N., Culley, J., Kerr, A., Ramsahoye, B., Stancheva, I., 2016. The SNF2 family ATPase LSH promotes cell-autonomous de novo DNA methylation in somatic cells. Nucleic Acids Res. 44, 7592–7604. doi:10.1093/nar/gkw424

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., McKay, R.D.G., 2007. New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature 448, 196–199. doi:10.1038/nature05972

Thijssen, P.E., Ito, Y., Grillo, G., Wang, J., Velasco, G., Nitta, H., Unoki, M., Yoshihara, M., Suyama, M., Sun, Y., Lemmers, R.J.L.F., de Greef, J.C., Gennery, A., Picco, P., Kloeckener-Gruissem, B., Güngör, T., Reisli, I., Picard, C., Kebaili, K., Roquelaure, B., Iwai, T., Kondo, I., Kubota, T., van Ostaijen-Ten Dam, M.M., van Tol, M.J.D., Weemaes, C., Francastel, C., van der Maarel, S.M., Sasaki, H., 2015. Mutations in CDCA7 and HELLS cause immunodeficiency–centromeric instability–facial anomalies syndrome. Nat. Commun. 6, 7870. doi:10.1038/ncomms8870

Thomas, J.O., 1999. Histone H1: location and role. Curr. Opin. Cell Biol. 11, 312–317. doi:10.1016/S0955-0674(99)80042-8

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R.W., Deaton, A., Andrews, R., James, K.D., Turner, D.J., Illingworth, R., Bird, A., 2010. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature 464, 1082–1086. doi:10.1038/nature08924

Thorvaldsen, J.L., Duran, K.L., Bartolomei, M.S., 1998. Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. Genes Dev. 12, 3693–3702.

Tiepolo, L., Maraschio, P., Gimelli, G., Cuoco, C., Gargani, G.F., Romano, C., 1979. Multibranched chromosomes 1, 9, and 16 in a patient with combined IgA and IgE deficiency. Hum. Genet. 51, 127–137.

Ting, D.T., Lipson, D., Paul, S., Brannigan, B.W., Akhavanfard, S., Coffman, E.J., Contino, G., Deshpande, V., Iafrate, A.J., Letovsky, S., others, 2011. Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. Science 331, 593–596.

Tiwari, V.K., McGarvey, K.M., Licchesi, J.D.F., Ohm, J.E., Herman, J.G., Schübeler, D., Baylin, S.B.,

2008. PcG proteins, DNA methylation, and gene repression by chromatin looping. PLoS Biol. 6, 2911–2927. doi:10.1371/journal.pbio.0060306

Tommasi, S., Zheng, A., Yoon, J.-I., Besaratinia, A., 2014. Epigenetic targeting of the Nanog pathway and signaling networks during chemical carcinogenesis. Carcinogenesis 35, 1726–1736. doi:10.1093/carcin/bgu026

Tong, W.-M., Lee, M.-K., Galendo, D., Wang, Z.-Q., Sabapathy, K., 2006. Aflatoxin-B exposure does not lead top53 mutations but results in enhanced liver cancer of Hupki (human p53 knock-in) mice. Int. J. Cancer 119, 745–749. doi:10.1002/ijc.21890

Topp, C.N., Zhong, C.X., Dawe, R.K., 2004. Centromere-encoded RNAs are integral components of the maize kinetochore. Proc. Natl. Acad. Sci. U. S. A. 101, 15986–15991.

Tosolini, M., Jouneau, A., 2016. Acquiring Ground State Pluripotency: Switching Mouse Embryonic Stem Cells from Serum/LIF Medium to 2i/LIF Medium. Methods Mol. Biol. Clifton NJ 1341, 41–48. doi:10.1007/7651_2015_207

Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K., Niwa, H., 2008. Identification and characterization of subpopulations in undifferentiated ES cell culture. Dev. Camb. Engl. 135, 909–918. doi:10.1242/dev.017400

Traiffort, E., Angot, E., Ruat, M., 2010. Sonic Hedgehog signaling in the mammalian brain. J. Neurochem. 113, 576–590. doi:10.1111/j.1471-4159.2010.06642.x

Trelogan, S.A., Martin, S.L., 1995. Tightly regulated, developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis. Proc. Natl. Acad. Sci. 92, 1520–1524.

Tretyakova, N., Guza, R., Matter, B., 2008. Endogenous cytosine methylation and the formation of carcinogen carcinogen-DNA adducts. Nucleic Acids Symp. Ser. 2004 49–50. doi:10.1093/nass/nrn025

Trojer, P., Reinberg, D., 2007. Facultative Heterochromatin: Is There a Distinctive Molecular Signature? Mol. Cell 28, 1–13. doi:10.1016/j.molcel.2007.09.011

Tsurumi, A., Li, W., 2012. Global heterochromatin loss: A unifying theory of aging? Epigenetics 7, 680–688. doi:10.4161/epi.20540

Tuck-Muller, C.M., Narayan, A., Tsien, F., Smeets, D.F., Sawyer, J., Fiala, E.S., Sohn, O.S., Ehrlich, M., 2000. DNA hypomethylation and unusual chromosome instability in cell lines from ICF syndrome patients. Cytogenet. Cell Genet. 89, 121–128. doi:15590

Tufarelli, C., Cruickshanks, H.A., Meehan, R.R., 2013. LINE-1 activation and epigenetic silencing of suppressor genes in cancer. Mob. Genet. Elem. 3. doi:10.4161/mge.26832

Turner, B.M., 2000. Histone acetylation and an epigenetic code. BioEssays News Rev. Mol. Cell. Dev. Biol. 22, 836–845. doi:10.1002/1521-1878(200009)22:9<836::AID-BIES9>3.0.CO;2-X

Ueda, K., Omachi, A., Kawaichi, M., Hayaishi, O., 1975. Natural occurrence of poly(ADP-ribosyl) histones in rat liver. Proc. Natl. Acad. Sci. U. S. A. 72, 205–209.

Ueda, Y., Okano, M., Williams, C., Chen, T., Georgopoulos, K., Li, E., 2006. Roles for Dnmt3b in mammalian development: a mouse model for the ICF syndrome. Dev. Camb. Engl. 133, 1183–1192. doi:10.1242/dev.02293

Underhill, C., Qutob, M.S., Yee, S.P., Torchia, J., 2000. A novel nuclear receptor corepressor complex, N-CoR, contains components of the mammalian SWI/SNF complex and the corepressor KAP-1. J. Biol. Chem. 275, 40463–40470. doi:10.1074/jbc.M007864200

Valgardsdottir, R., Chiodi, I., Giordano, M., Rossi, A., Bazzini, S., Ghigna, C., Riva, S., Biamonti, G., 2008. Transcription of Satellite III non-coding RNAs is a general stress response in human cells. Nucleic Acids Res. 36, 423–434. doi:10.1093/nar/gkm1056

Valkova, G., Ghenev, E., Tzancheva, M., 1987. Centromeric instability of chromosomes 1, 9 and 16 with variable immune deficiency. Support of a new syndrome. Clin. Genet. 31, 119–124.

van den Boogaard, M.L., Lemmers, R.J.L.F., Balog, J., Wohlgemuth, M., Auranen, M., Mitsuhashi, S., van der Vliet, P.J., Straasheijm, K.R., van den Akker, R.F.P., Kriek, M., Laurense-Bik, M.E.Y., Raz, V., van Ostaijen-Ten Dam, M.M., Hansson, K.B.M., van der Kooi, E.L., Kiuru-Enari, S., Udd, B., van Tol, M.J.D., Nishino, I., Tawil, R., Tapscott, S.J., van Engelen, B.G.M., van der Maarel, S.M., 2016. Mutations in DNMT3B Modify Epigenetic Repression of the D4Z4 Repeat and the Penetrance of Facioscapulohumeral Dystrophy. Am. J. Hum. Genet. 98, 1020–1029. doi:10.1016/j.ajhg.2016.03.013

van den Boogaard, M.L., Thijssen, P.E., Aytekin, C., Licciardi, F., Kıykım, A.A., Spossito, L., Dalm, V. a. S.H., Driessen, G.J., Kersseboom, R., de Vries, F., van Ostaijen-Ten Dam, M.M., Ikinciogullari, A., Dogu, F., Oleastro, M., Bailardo, E., Daxinger, L., Nain, E., Baris, S., van Tol, M.J.D., Weemaes, C., van der Maarel, S.M., 2017. Expanding the mutation spectrum in ICF syndrome: Evidence for a gender bias in ICF2. Clin. Genet. doi:10.1111/cge.12979

van Holde, K., Zlatanova, J., 1995. Chromatin higher order structure: chasing a mirage? J. Biol. Chem. 270, 8373–8376.

Vansant, G., Reynolds, W.F., 1995. The consensus sequence of a major Alu subfamily contains a functional retinoic acid response element. Proc. Natl. Acad. Sci. U. S. A. 92, 8229–8233.

Varley, J.M., Macgregor, H.C., Nardi, I., Andrews, C., Erba, H.P., 1980. Cytological evidence of transcription of highly repeated DNA sequences during the lampbrush stage in Triturus cristatus carnifex. Chromosoma 80, 289–307.

Varriale, A., Bernardi, G., 2010. Distribution of DNA methylation, CpGs, and CpG islands in human isochores. Genomics 95, 25–28. doi:10.1016/j.ygeno.2009.09.006

Vasanthakumar, A., Lepore, J.B., Zegarek, M.H., Kocherginsky, M., Singh, M., Davis, E.M., Link, P.A., Anastasi, J., Le Beau, M.M., Karpf, A.R., Godley, L.A., 2013. Dnmt3b is a haploinsufficient tumor suppressor gene in Myc-induced lymphomagenesis. Blood 121, 2059–2063. doi:10.1182/blood-2012-04-421065

Veillard, A.-C., Marks, H., Bernardo, A.S., Jouneau, L., Laloë, D., Boulanger, L., Kaan, A., Brochard, V., Tosolini, M., Pedersen, R., Stunnenberg, H., Jouneau, A., 2014. Stable Methylation at Promoters Distinguishes Epiblast Stem Cells from Embryonic Stem Cells and the In Vivo Epiblasts. Stem Cells Dev. 23, 2014–2029. doi:10.1089/scd.2013.0639

Velasco, G., Hube, F., Rollin, J., Neuillet, D., Philippe, C., Bouzinba-Segard, H., Galvani, A., Viegas-Pequignot, E., Francastel, C., 2010. Dnmt3b recruitment through E2F6 transcriptional repressor

mediates germ-line gene silencing in murine somatic tissues. Proc. Natl. Acad. Sci. 107, 9281–9286. doi:10.1073/pnas.1000473107

Velasco, G., Walton, E.L., Sterlin, D., Hédouin, S., Nitta, H., Ito, Y., Fouyssac, F., Mégarbané, A., Sasaki, H., Picard, C., Francastel, C., 2014. Germline genes hypomethylation and expression define a molecular signature in peripheral blood of ICF patients: implications for diagnosis and etiology. Orphanet J. Rare Dis. 9, 56. doi:10.1186/1750-1172-9-56

Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W.M.P., van Schaik, F.M.A., Varier, R.A., Baltissen, M.P.A., Stunnenberg, H.G., Mann, M., Timmers, H.T.M., 2007. Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4. Cell 131, 58–69. doi:10.1016/j.cell.2007.08.016

Viré, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.-M., Bollen, M., Esteller, M., Di Croce, L., de Launoit, Y., Fuks, F., 2005. The Polycomb group protein EZH2 directly controls DNA methylation. Nature 439, 871–874. doi:10.1038/nature04431

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I.S., Martienssen, R.A., 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science 297, 1833–1837. doi:10.1126/science.1074973

von Bernuth, H., Ravindran, E., Du, H., Fröhler, S., Strehl, K., Krämer, N., Issa-Jahns, L., Amulic, B., Ninnemann, O., Xiao, M.-S., others, 2014. Combined immunodeficiency develops with age in Immunodeficiency-centromeric instability-facial anomalies syndrome 2 (ICF2). Orphanet J. Rare Dis. 9, 116.

Vourc'h, C., Biamonti, G., 2011. Transcription of Satellite DNAs in Mammals. Prog. Mol. Subcell. Biol. 51, 95–118. doi:10.1007/978-3-642-16502-3_5

Waddington, C.H., 2012. The epigenotype. 1942. Int. J. Epidemiol. 41, 10–13. doi:10.1093/ije/dyr184

Wagner, W., Fernandez-Rebollo, E., Frobel, J., 2016. DNA-methylation changes in replicative senescence and aging: two sides of the same coin? Epigenomics 8, 1–3. doi:10.2217/epi.15.100

Walter, M., Teissandier, A., Pérez-Palacios, R., Bourc'his, D., 2016. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. Elife 5, e11418.

Walton, E., Francastel, C., Velasco, G., 2014. Dnmt3b Prefers Germ Line Genes and Centromeric Regions: Lessons from the ICF Syndrome and Cancer and Implications for Diseases. Biology 3, 578–605. doi:10.3390/biology3030578

Walton, E., Hass, J., Liu, J., Roffman, J.L., Bernardoni, F., Roessner, V., Kirsch, M., Schackert, G., Calhoun, V., Ehrlich, S., 2016. Correspondence of DNA Methylation Between Blood and Brain Tissue and Its Application to Schizophrenia Research. Schizophr. Bull. 42, 406–414. doi:10.1093/schbul/sbv074

Walton, E.L., Francastel, C., Velasco, G., 2011. Maintenance of DNA methylation: Dnmt3b joins the dance. Epigenetics 6, 1373–1377. doi:10.4161/epi.6.11.17978

Wang, H., An, W., Cao, R., Xia, L., Erdjument-Bromage, H., Chatton, B., Tempst, P., Roeder, R.G., Zhang,

Y., 2003. mAM facilitates conversion by ESET of dimethyl to trimethyl lysine 9 of histone H3 to cause transcriptional repression. Mol. Cell 12, 475–487.

Wang, H., Dey, S.K., 2006. Roadmap to embryo implantation: clues from mouse models. Nat. Rev. Genet. 7, 185–199. doi:10.1038/nrg1808

Wang, H., Maurano, M.T., Qu, H., Varley, K.E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R., Thurman, R.E., Kaul, R., Myers, R.M., Stamatoyannopoulos, J.A., 2012. Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome Res. 22, 1680–1688. doi:10.1101/gr.136101.111

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., Zhang, Y., 2004. Role of histone H2A ubiquitination in Polycomb silencing. Nature 431, 873–878. doi:10.1038/nature02985

Wang, J., Emadali, A., Le Bescont, A., Callanan, M., Rousseaux, S., Khochbin, S., 2011. Induced malignant genome reprogramming in somatic cells by testis-specific factors. Biochim. Biophys. Acta 1809, 221–225. doi:10.1016/j.bbagrm.2011.04.003

Wang, J., Hevi, S., Kurash, J.K., Lei, H., Gay, F., Bajko, J., Su, H., Sun, W., Chang, H., Xu, G., Gaudet, F., Li, E., Chen, T., 2009. The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. Nat. Genet. 41, 125–129. doi:10.1038/ng.268

Wang, X.Q., Ng, R.K., Ming, X., Zhang, W., Chen, L., Chu, A.C.Y., Pang, R., Lo, C.M., Tsao, S.W., Liu, X., Poon, R.T.P., Fan, S.T., 2013. Epigenetic regulation of pluripotent genes mediates stem cell features in human hepatocellular carcinoma and cancer cell lines. PloS One 8, e72435. doi:10.1371/journal.pone.0072435

Wasylyk, B., Chambon, P., 1979. Transcription by eukaryotic RNA polymerases A and B of chromatin assembled in vitro. Eur. J. Biochem. 98, 317–327.

Watanabe, D., Suetake, I., Tada, T., Tajima, S., 2002. Stage-and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. Mech. Dev. 118, 187–190.

Watanabe, D., Uchiyama, K., Hanaoka, K., 2006. Transition of mouse de novo methyltransferases expression from Dnmt3b to Dnmt3a during neural progenitor cell development. Neuroscience 142, 727–737. doi:10.1016/j.neuroscience.2006.07.053

Waterson, P.E., Older Gray, M.T., Clegg, C.W., 2002. A sociotechnical method for designing work systems. Hum. Factors 44, 376–391. doi:10.1518/0018720024497628

Watkins, A.J., Papenbrock, T., Fleming, T.P., 2008. The preimplantation embryo: handle with care. Semin. Reprod. Med. 26, 175–185. doi:10.1055/s-2008-1042956

Watson, J.D., Crick, F.H., 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171, 737–738.

Watson, P., Black, G., Ramsden, S., Barrow, M., Super, M., Kerr, B., Clayton-Smith, J., 2001. Angelman syndrome phenotype associated with mutations inMECP2, a gene encoding a methyl CpG binding protein. J. Med. Genet. 38, 224–228.

Weaver, J.R., Bartolomei, M.S., 2014. Chromatin regulators of genomic imprinting. Biochim. Biophys. Acta BBA - Gene Regul. Mech. 1839, 169–177. doi:10.1016/j.bbagrm.2013.12.002

Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., Schübeler, D., 2005. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat. Genet. 37, 853–862. doi:10.1038/ng1598

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Pääbo, S., Rebhan, M., Schübeler, D., 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat. Genet. 39, 457–466. doi:10.1038/ng1990

Weber, M., Schübeler, D., 2007. Genomic patterns of DNA methylation: targets and function of an epigenetic mark. Curr. Opin. Cell Biol. 19, 273–280. doi:10.1016/j.ceb.2007.04.011

Weemaes, C.M.R., van Tol, M.J.D., Wang, J., van Ostaijen-ten Dam, M.M., van Eggermond, M.C.J.A., Thijssen, P.E., Aytekin, C., Brunetti-Pierri, N., van der Burg, M., Graham Davies, E., Ferster, A., Furthner, D., Gimelli, G., Gennery, A., Kloeckener-Gruissem, B., Meyn, S., Powell, C., Reisli, I., Schuetz, C., Schulz, A., Shugar, A., van den Elsen, P.J., van der Maarel, S.M., 2013. Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects. Eur. J. Hum. Genet. EJHG 21, 1219–1225. doi:10.1038/ejhg.2013.40

Wei, W., Gilbert, N., Ooi, S.L., Lawler, J.F., Ostertag, E.M., Kazazian, H.H., Boeke, J.D., Moran, J.V., 2001. Human L1 retrotransposition: cis preference versus trans complementation. Mol. Cell. Biol. 21, 1429–1439. doi:10.1128/MCB.21.4.1429-1439.2001

Wei, Y., Mizzen, C.A., Cook, R.G., Gorovsky, M.A., Allis, C.D., 1998. Phosphorylation of histone H3 at serine 10 is correlated with chromosome condensation during mitosis and meiosis in Tetrahymena. Proc. Natl. Acad. Sci. U. S. A. 95, 7480–7484.

Weimann, M., Grossmann, A., Woodsmith, J., Özkan, Z., Birth, P., Meierhofer, D., Benlasfer, N., Valovka, T., Timmermann, B., Wanker, E.E., Sauer, S., Stelzl, U., 2013. A Y2H-seq approach defines the human protein methyltransferase interactome. Nat. Methods 10, 339–342. doi:10.1038/nmeth.2397

Weintraub, H., Groudine, M., 1976. Chromosomal subunits in active genes have an altered conformation. Science 193, 848–856.

Weisenberger, D.J., 2005. Analysis of repetitive element DNA methylation by MethyLight. Nucleic Acids Res. 33, 6823–6836. doi:10.1093/nar/gki987

Weisenberger, D.J., Romano, L.J., 1999. Cytosine methylation in a CpG sequence leads to enhanced reactivity with Benzo[a]pyrene diol epoxide that correlates with a conformational change. J. Biol. Chem. 274, 23948–23955.

Weisheit, G., Mertz, D., Schilling, K., Viebahn, C., 2002. An efficient in situ hybridization protocol for multiple tissue sections and probes on miniaturized slides. Dev. Genes Evol. 212, 403–406. doi:10.1007/s00427-002-0258-8

Weismann, A., Parker, W.N. (William N., Rönnfeldt, H., 1893. The germ-plasm; a theory of heredity. New York, Scribner's.

Wichterle, H., Peljto, M., Nedelec, S., 2009. Xenotransplantation of embryonic stem cell-derived motor neurons into the developing chick spinal cord. Methods Mol. Biol. Clifton NJ 482, 171–183. doi:10.1007/978-1-59745-060-7_11

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., Schulman, A.H., 2007. A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. 8, 973–982. doi:10.1038/nrg2165

Wijmenga, C., Hansen, R.S., Gimelli, G., Björck, E.J., Davies, E.G., Valentine, D., Belohradsky, B.H., van Dongen, J.J., Smeets, D.F., van den Heuvel, L.P., Luyten, J.A., Strengman, E., Weemaes, C., Pearson, P.L., 2000. Genetic variation in ICF syndrome: evidence for genetic heterogeneity. Hum. Mutat. 16, 509–517. doi:10.1002/1098-1004(200012)16:6<509::AID-HUMU8>3.0.CO;2-V

Wijmenga, C., van den Heuvel, L.P., Strengman, E., Luyten, J.A., van der Burgt, I.J., de Groot, R., Smeets, D.F., Draaisma, J.M., van Dongen, J.J., De Abreu, R.A., others, 1998. Localization of the ICF syndrome to chromosome 20 by homozygosity mapping. Am. J. Hum. Genet. 63, 803–809.

Williams, R.L., Hilton, D.J., Pease, S., Willson, T.A., Stewart, C.L., Gearing, D.P., Wagner, E.F., Metcalf, D., Nicola, N.A., Gough, N.M., 1988. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature 336, 684–687. doi:10.1038/336684a0

Wilson, V.L., Jones, P.A., 1984. Chemical carcinogen-mediated decreases in DNA 5-methylcytosine content of BALB/3T3 cells. Carcinogenesis 5, 1027–1031.

Wilson, V.L., Jones, P.A., 1983. Inhibition of DNA methylation by chemical carcinogens in vitro. Cell 32, 239–246.

Wilson, V.L., Smith, R.A., Ma, S., Cutler, R.G., 1987. Genomic 5-methyldeoxycytidine decreases with age. J. Biol. Chem. 262, 9948–9951.

Winkelmann, J., Lin, L., Schormair, B., Kornum, B.R., Faraco, J., Plazzi, G., Melberg, A., Cornelio, F., Urban, A.E., Pizza, F., Poli, F., Grubert, F., Wieland, T., Graf, E., Hallmayer, J., Strom, T.M., Mignot, E., 2012. Mutations in DNMT1 cause autosomal dominant cerebellar ataxia, deafness and narcolepsy. Hum. Mol. Genet. 21, 2205–2210. doi:10.1093/hmg/dds035

Winston, F., Allis, C.D., 1999. The bromodomain: a chromatin-targeting module? Nat. Struct. Biol. 6, 601–604. doi:10.1038/10640

Wojciechowski, M.F., Meehan, T., 1984. Inhibition of DNA methyltransferases in vitro by benzo[a]pyrene diol epoxide-modified substrates. J. Biol. Chem. 259, 9711–9716.

Wolf, S.F., Jolly, D.J., Lunnen, K.D., Friedmann, T., Migeon, B.R., 1984. Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation. Proc. Natl. Acad. Sci. 81, 2806–2810.

Wray, J., Kalkan, T., Gomez-Lopez, S., Eckardt, D., Cook, A., Kemler, R., Smith, A., 2011. Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. Nat. Cell Biol. 13, 838–845. doi:10.1038/ncb2267

Wray, J., Kalkan, T., Smith, A.G., 2010. The ground state of pluripotency. Biochem. Soc. Trans. 38, 1027–1032. doi:10.1042/BST0381027

Wright, S.I., Agrawal, N., Bureau, T.E., 2003. Effects of recombination rate and gene density on transposable element distributions in Arabidopsis thaliana. Genome Res. 13, 1897–1903. doi:10.1101/gr.1281503

Wu, H., Chen, X., Xiong, J., Li, Y., Li, H., Ding, X., Liu, S., Chen, S., Gao, S., Zhu, B., 2011. Histone methyltransferase G9a contributes to H3K27 methylation in vivo. Cell Res. 21, 365–367. doi:10.1038/cr.2010.157

Wu, H., Thijssen, P.E., de Klerk, E., Vonk, K.K.D., Wang, J., den Hamer, B., Aytekin, C., Maarel, V.D., M, S., Daxinger, L., 2016. Converging disease genes in ICF syndrome: ZBTB24 controls expression of CDCA7 in mammals. Hum. Mol. Genet. 25, 4041–4051. doi:10.1093/hmg/ddw243

Wu, S.C., Zhang, Y., 2010. Active DNA demethylation: many roads lead to Rome. Nat. Rev. Mol. Cell Biol. 11, 607–620. doi:10.1038/nrm2950

Wu, W.-H., Alami, S., Luk, E., Wu, C.-H., Sen, S., Mizuguchi, G., Wei, D., Wu, C., 2005. Swc2 is a widely conserved H2AZ-binding module essential for ATP-dependent histone exchange. Nat. Struct. Mol. Biol. 12, 1064–1071. doi:10.1038/nsmb1023

Wyatt, G.R., 1950. Occurrence of 5-methylcytosine in nucleic acids. Nature 166, 237–238.

Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P., Wu, C., Allis, C.D., 2006. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature. doi:10.1038/nature04815

Xie, D., Chen, C.-C., Ptaszek, L.M., Xiao, S., Cao, X., Fang, F., Ng, H.H., Lewin, H.A., Cowan, C., Zhong, S., 2010. Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. Genome Res. 20, 804–815. doi:10.1101/gr.100594.109

Xin, B., Cruz Marino, T., Szekely, J., Leblanc, J., Cechner, K., Sency, V., Wensel, C., Barabas, M., Therriault, V., Wang, H., 2017. Novel *DNMT3A* germline mutations are associated with inherited Tatton-Brown-Rahman syndrome: Novel *DNMT3A* mutations cause TBRS. Clin. Genet. 91, 623–628. doi:10.1111/cge.12878

Xin, H., Liu, D., Songyang, Z., 2008. The telosome/shelterin complex and its functions. Genome Biol. 9, 232. doi:10.1186/gb-2008-9-9-232

Xiong, Z., Laird, P.W., 1997. COBRA: a sensitive and quantitative DNA methylation assay. Nucleic Acids Res. 25, 2532–2534.

Xu, G.L., Bestor, T.H., Bourc'his, D., Hsieh, C.L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J., Viegas-Péquignot, E., 1999. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature 402, 187–191. doi:10.1038/46052

Xu, J., Zhu, W., Xu, W., Yao, W., Zhang, B., Xu, Y., Ji, S., Liu, C., Long, J., Ni, Q., Yu, X., 2013. Up-regulation of MBD1 promotes pancreatic cancer cell epithelial-mesenchymal transition and invasion by epigenetic down-regulation of E-cadherin. Curr. Mol. Med. 13, 387–400.

Yamagishi, Y., Sakuno, T., Shimura, M., Watanabe, Y., 2008. Heterochromatin links to centromeric protection by recruiting shugoshin. Nature 455, 251–255. doi:10.1038/nature07217

Yamashita, Y., Yuan, J., Suetake, I., Suzuki, H., Ishikawa, Y., Choi, Y.L., Ueno, T., Soda, M., Hamada, T., Haruta, H., others, 2010. Array-based genomic resequencing of human leukemia. Oncogene 29, 3723–3731.

Yang, L., Arias, L.R., Lane, T.S., Yancey, M.D., Mamouni, J., 2011. Real-time electrical impedance-based

measurement to distinguish oral cancer cells and non-cancer oral epithelial cells. Anal. Bioanal. Chem. 399, 1823–1833. doi:10.1007/s00216-010-4584-9

Yang, X., Han, H., De Carvalho, D.D., Lay, F.D., Jones, P.A., Liang, G., 2014. Gene body methylation can alter gene expression and is a therapeutic target in cancer. Cancer Cell 26, 577–590. doi:10.1016/j.ccr.2014.07.028

Yehezkel, S., Segev, Y., Viegas-Pequignot, E., Skorecki, K., Selig, S., 2008. Hypomethylation of subtelomeric regions in ICF syndrome is associated with abnormally short telomeres and enhanced transcription from telomeric regions. Hum. Mol. Genet. 17, 2776–2789. doi:10.1093/hmg/ddn177

Yehezkel, S., Shaked, R., Sagie, S., Berkovitz, R., Shachar-Bener, H., Segev, Y., Selig, S., 2013. Characterization and rescue of telomeric abnormalities in ICF syndrome type I fibroblasts. Front. Oncol. 3. doi:10.3389/fonc.2013.00035

Ying, Q.L., Nichols, J., Chambers, I., Smith, A., 2003. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell 115, 281–292.

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., Smith, A., 2008. The ground state of embryonic stem cell self-renewal. Nature 453, 519–523. doi:10.1038/nature06968

Yoder, J.A., Bestor, T.H., 1998. A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. Hum. Mol. Genet. 7, 279–284.

Yokochi, T., Robertson, K.D., 2002. Preferential methylation of unmethylated DNA by Mammalian de novo DNA methyltransferase Dnmt3a. J. Biol. Chem. 277, 11735–11745. doi:10.1074/jbc.M106590200

Yoo, C.B., Jones, P.A., 2006. Epigenetic therapy of cancer: past, present and future. Nat. Rev. Drug Discov. 5, 37–50. doi:10.1038/nrd1930

Yu, W., McIntosh, C., Lister, R., Zhu, I., Han, Y., Ren, J., Landsman, D., Lee, E., Briones, V., Terashima, M., Leighty, R., Ecker, J.R., Muegge, K., 2014. Genome-wide DNA methylation patterns in LSH mutant reveals de-repression of repeat elements and redundant epigenetic silencing pathways. Genome Res. 24, 1613–1623. doi:10.1101/gr.172015.114

Yun, M., Wu, J., Workman, J.L., Li, B., 2011. Readers of histone modifications. Cell Res. 21, 564–578.

Yunis, J.J., Yasmineh, W.G., 1971. Heterochromatin, satellite DNA, and cell function. Structural DNA of eucaryotes may support and protect genes and aid in speciation. Science 174, 1200–1209.

Zachariah, R.M., Rastegar, M., 2012. Linking Epigenetics to Human Disease and Rett Syndrome: The Emerging Novel and Challenging Concepts in MeCP2 Research. Neural Plast. 2012, 1–10. doi:10.1155/2012/415825

Zamudio, N., Barau, J., Teissandier, A., Walter, M., Borsos, M., Servant, N., Bourc'his, D., 2015. DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. Genes Dev. 29, 1256–1270. doi:10.1101/gad.257840.114

Zhang, H., Gayen, S., Xiong, J., Zhou, B., Shanmugam, A.K., Sun, Y., Karatas, H., Liu, L., Rao, R.C., Wang, S., Nesvizhskii, A.I., Kalantry, S., Dou, Y., 2016. MLL1 Inhibition Reprograms Epiblast Stem Cells to Naive Pluripotency. Cell Stem Cell 18, 481–494. doi:10.1016/j.stem.2016.02.004

Zhang, W., Xu, J., 2017. DNA methyltransferases and their roles in tumorigenesis. Biomark. Res. 5. doi:10.1186/s40364-017-0081-z

Zhang, W., Yi, C., Bao, W., Liu, B., Cui, J., Yu, H., Cao, X., Gu, M., Liu, M., Cheng, Z., 2005. The transcribed 165-bp CentO satellite is the major functional centromeric element in the wild rice species Oryza punctata. Plant Physiol. 139, 306–315. doi:10.1104/pp.105.064147

Zhao, S.-L., Zhu, S.-T., Hao, X., Li, P., Zhang, S.-T., 2011. Effects of DNA methyltransferase 1 inhibition on esophageal squamous cell carcinoma: The role of DNMT1 in ESCC. Dis. Esophagus 24, 601–610. doi:10.1111/j.1442-2050.2011.01199.x

Zhu, L., Hathcock, K.S., Hande, P., Lansdorp, P.M., Seldin, M.F., Hodes, R.J., 1998. Telomere length regulation in mice is linked to a novel chromosome locus. Proc. Natl. Acad. Sci. U. S. A. 95, 8648–8653.

Zhu, Q., Pao, G.M., Huynh, A.M., Suh, H., Tonnu, N., Nederlof, P.M., Gage, F.H., Verma, I.M., 2011. BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. Nature 477, 179–184. doi:10.1038/nature10371

Zijlmans, J.M., Martens, U.M., Poon, S.S., Raap, A.K., Tanke, H.J., Ward, R.K., Lansdorp, P.M., 1997. Telomeres in the mouse have large inter-chromosomal variations in the number of T2AG3 repeats. Proc. Natl. Acad. Sci. U. S. A. 94, 7423–7428.

Zuo, X., Sheng, J., Lau, H.-T., McDonald, C.M., Andrade, M., Cullen, D.E., Bell, F.T., Iacovino, M., Kyba, M., Xu, G., Li, X., 2012. Zinc Finger Protein ZFP57 Requires Its Co-factor to Recruit DNA Methyltransferases and Maintains DNA Methylation Imprint in Embryonic Stem Cells via Its Transcriptional Repression Domain. J. Biol. Chem. 287, 2107–2118. doi:10.1074/jbc.M111.322644

ANNEXES

ANNEXE 1

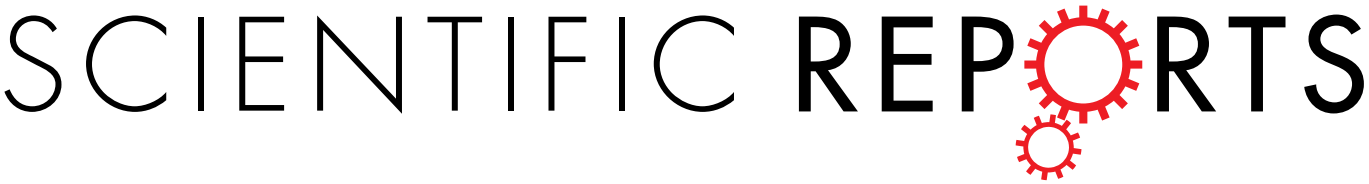
Centromeres are key chromosomal regions that assemble on short sequences repeated in tandem over large domains, which serve as an assembly platform for attachment of kinetochore and mitotic spindle. They are required for accurate chromosome segregation, and thus, their integrity is essential for genomic stability. My host team brought the first evidence that ncRNAs originating from murine centromeric repeats are essential for centromere identity and function. They participate in the correct spatio-temporal assembly of kinetochore at the onset of mitosis in cycling cells (Ferri et al., 2009), whereas their deregulated accumulation in various stress conditions is toxic to the cell and promotes chromosomal instability and perturbation of nuclear architecture (Bouzinba-Segard et al., 2006; Valgardsdottir et al., 2008). These features are also commonly observed in several human diseases (Ferreira et al., 2015), including cancers (Ting et al., 2011) and the ICF syndrome that was the focus of my PhD, raising the question of a causal link with perturbed functioning of the cell.

Focusing on cellular stress in cultured murine cells, our work revealed unexpected transcriptional and epigenetic perturbations to centromeres in genotoxic stress conditions, associated with permanent cell cycle arrest. Rapid and selective activated transcription of centromeric repeats is followed by a striking delocalization of the centromere-specific histone variant CENP-A away from its default location. Using SNAP-tag technology, we demonstrated that nucleosomal CENP-A is actively evicted from centromeric chromatin in a manner that is dependent on the DNA Damage Response (DDR) sensor ATM, activated transcription of centromeric repeats and chromatin remodelers/histone chaperones, and requires a WT p53 context. In absence of p53 checkpoint, CENP-A remains at centromeres, cells do not arrest and accumulate micronuclei after genotoxic stress indicative of genomic instability. Activated transcription and loss of centromeres identity are also features of permanently arrested senescent cells with persistent DDR activation, suggesting a cooperation between the DDR and loss of centromere identity as a safeguard mechanism to prevent genomic instability.

As a whole, our work uncovered a novel aspect of the stress response whereby activated transcription and disorganization of centromeres functions as an effector in response to genomic insults and provides a protective mechanism, to ensure or maintain proper growth arrest.

I contributed to this study performing experiment on WT and p53 cells. In particular, I participate in treating WT 3T3 cells with ATM inhibitors and observing minor satellite repeats transcription by RT-qPCR and delocalization of CENP-A by immunofluorescence.

This study has been published in February 2017, in *Scientific Reports*.



OPEN

CENP-A chromatin disassembly in stressed and senescent murine cells

Sabrine Hédouin, Giacomo Grillo, Ivana Ivkovic, Guillaume Velasco & Claire Francastel

Received: 12 September 2016
Accepted: 11 January 2017
Published: 10 February 2017

Centromeres are chromosomal domains essential for genomic stability. We report here the remarkable transcriptional and epigenetic perturbations at murine centromeres in genotoxic stress conditions. A strong and selective transcriptional activation of centromeric repeats is detected within hours. This is followed by disorganization of centromeres with striking delocalization of nucleosomal CENP-A, the key determinant of centromere identity and function, in a mechanism requiring active transcription of centromeric repeats, the DNA Damage Response (DDR) effector ATM and chromatin remodelers/histone chaperones. In the absence of p53 checkpoint, activated transcription of centromeric repeats and CENP-A delocalization do not occur and cells accumulate micronuclei indicative of genomic instability. In addition, activated transcription and loss of centromeres identity are features of permanently arrested senescent cells with persistent DDR activation. Together, these findings bring out cooperation between DDR effectors and loss of centromere integrity as a safeguard mechanism to prevent genomic instability in context of persistent DNA damage signalling.

Centromeres are specialized chromosomal regions that serve as an assembly platform for attachment of kinetochore and mitotic spindle, which are essential for faithful segregation of genetic material during cell division[1]. Hence, maintenance of centromere identity and function is tightly linked to maintenance of genome stability and integrity.

Most centromeres assemble on repeated sequences, yet no sequence conservation between species helps to genetically define their position. However, a centromere-specific variant of histone H3, CENP-A, serves as a key epigenetic determinant of centromere identity and kinetochore assembly through the generation of a unique chromatin organization[2,3]. Furthermore, centromeric transcripts are emerging as integral components of centromeric chromatin, participating in CENP-A deposition on chromatin and centromere function[4–10]. Their levels are tightly regulated during cell cycle[7] and their unscheduled accumulation has been observed in human diseases[11–13] and stress conditions[14–16]. Recently, we functionally linked this accumulation to perturbed centromere architecture and function leading to genome instability and aneuploidy in the mouse[5]. Hence, accumulation of centromeric transcripts is probably not a mere consequence of a physiopathological state and might represent a conserved feature of the cellular stress response.

Agents and processes that inflict damage to DNA and cause genotoxic stress are particularly deleterious since they severely compromise genome integrity. To counteract the adverse effects of DNA damage and their transmission to daughter cells, cells have developed sophisticated and coordinated surveillance mechanisms[17,18]. The multifactorial DNA damage response (DDR) is the central regulator of this network. It senses the DNA lesion and transmits the damage signal through the activation of signalling cascades to initiate DNA repair and stall damaged cells until DNA lesions are repaired. DDR is orchestrated by the ATM and ATR kinases, which phosphorylate a multitude of proteins to modulate cellular response depending on the type of damage, cellular context and intensity and duration of stress[19]. The appropriate response is triggered by effector pathways allowing DNA repair, cell cycle arrest, senescence, apoptosis or cell death, among which the p53 pathway is probably the main effector downstream of DNA strand breaks and activation of ATM/ATR[20].

Here, we aimed at establishing the kinetics of epigenetic and transcriptional perturbations that impact centromere identity in response to stress. We report that murine centromeric transcripts accumulate upon DNA damage within a few hours, in a manner that is dependent on the DDR effector p53. This is followed by disorganization of centromeric chromatin associated with the striking relocation of parental nucleosomal CENP-A, in a manner that also requires ATM-mediated signalling pathway and chromatin chaperones/remodelling factors, the most prominent being the FACT (facilitates chromatin transcription) complex. We found that perturbations to transcription and centromeric architecture are also hallmarks of senescent cells where the DDR is activated

Université Paris Diderot, Sorbonne Paris Cité, Epigenetics and Cell Fate, CNRS UMR7216, 75205 Paris cedex, France. Correspondence and requests for materials should be addressed to C.F. (email: claire.francastel@univ-paris-diderot.fr)

independently of the presence of exogenous genotoxic stressors[21]. As a whole, our data uncovered a novel cross-talk between DDR effectors and dynamics at centromeric chromatin, where a p53/ATM-dependent disruption of centromeric structure and identity may trigger safeguard mechanisms to prevent genomic instability in cases of persistent DNA damage signalling.

Results

Accumulation of DNA damage leads to CENP-A mislocalization. We treated murine NIH/3T3 cells with a representative panel of genotoxic agents under conditions known to promote various types of DNA damage (Table S1) as revealed by accumulation of phosphorylated histone variant H2A.X (γH2A.X) and stabilization of p53 (Figure S1A). We monitored the impact of various drug treatments on cell cycle by FACS (Figure S1B). Centromere architecture was assessed in single cells using immunofluorescence (IF) to follow CENP-A localization and DNA-FISH using probes specific for centromeric repeats termed minor satellites in the mouse. In untreated cells, CENP-A staining and minor satellite repeats adopted the typical punctate pattern in the vicinity of chromocenters[22], composed of pericentromeric major satellite repeats or visualized as dense DAPI staining (Fig. 1A and B; top rows). We first focused on Etoposide (ETOP), a potent inducer of DNA double strand breaks (DSB), as a paradigm for studying the impact of DNA damage on centromeres. We found that CENP-A became remarkably mislocalized away from its normal location and occupied the periphery and inside of the nucleolus marked by B23/nucleophosmin signal (Fig. 1A; bottom row). Likewise, ETOP treatment led to the disorganization of centromeric DNA, mainly characterized by stretched rather than punctiform minor satellite signals, while chromocenters remained globally less affected (Fig. 1B; two bottom rows).

To exclude an effect restricted to topoisomerase II inhibition by ETOP, we tested other genotoxic drugs that have various impacts on DNA (Table S1) and on cell cycle (Figure S1B). We found that Zeocin (ZEO), Mitomycin-C (MMC) and, to a lesser extent, Hydroxyurea (HU) in the conditions used, led to CENP-A relocation away from centromeric foci (Figure S2A). In contrast, forced cell-cycle arrest in G2/M after Nocodazole (NOC) treatment, a poison of microtubules, was not sufficient to promote CENP-A delocalization (Figure S2A).

Kinetics experiments following genotoxic stress showed that, while maintaining the typical punctate pattern nearby chromocenters during the first 4 hr of ETOP treatment, CENP-A became visibly mislocalized by 8 hr, and further occupied the nucleoplasm after 24 hr (Fig. 1C). In contrast, other centromeric proteins like CENP-B and CENP-C mainly displayed the typical centromeric punctate pattern, despite being slightly enriched in the nucleoplasm (Figure S2B and C). Although CENP-A delocalization seemed to be a relatively early event in the response to DNA damage, we followed the appearance of cleaved Caspase-3, one of the earliest molecular events at the onset of apoptosis. We did not detect cleaved Caspase-3 before 24 hr of ETOP treatment (Figure S2D). The signal was also undetectable in cells treated for 4 hr that have recovered from stress after a 24 hr period (Figure S2D). Hence, CENP-A relocation does not seem to be a direct consequence of the activation of pro-apoptotic signals. Rather, it may precede cellular changes.

Etoposide forms non-repairable DSB[23] while cells progressively accumulate in G2/M (Figure S1B). Interestingly, whereas a 4 hr ETOP treatment did not promote visible CENP-A delocalization (Fig. 1C), a 24 hr recovery period (4 + 24H) during which DNA damage signalling and γH2A.X signal persisted (Figure S2D) and most cells arrested in G2/M (Figure S1B) promoted CENP-A delocalization in more than 60% of the cells (Fig. 1C; lower row). Similar data were obtained with 4 hr ZEO or MMC treatments followed by a 24 hr recovery period (Figure S3A), time point at which γH2A.X is still detectable (Figure S3B). These data support a link between persistence of DNA damage and centromeres loss of identity through CENP-A delocalization. To confirm this hypothesis, we treated cells with lower doses of ETOP (2 μM instead of 10 μM) to allow cells to repair DNA damage, as shown by γH2A.X that returned to baseline levels after a 24 hr recovery from the 4 hr ETOP treatment (Figure S3B). In addition, cells were only transiently arrested in the cell cycle when treated with lower doses of ETOP. Indeed, in contrast to treatment with 10 μM ETOP that led to sustained cell cycle arrest, treatment with 2 μM ETOP allowed cells to resume cell cycle after the recovery period as shown by the increased number of cells in G1 after a 24 hr recovery from the 4 hr treatment (Figure S1B). Under these conditions, the percentage of cells with delocalized CENP-A decreased from nearly 70% to less than 40%, suggesting that CENP-A reoccupies its normal location when DNA damage is repaired (Figure S3A), although we cannot exclude that undamaged cells with or normal CENP-A location take over in the culture during the recovery period.

To confirm the observations made by microscopy, we used Chromatin Immunoprecipitation (ChIP) that showed a decreased CENP-A occupancy at centromeric repeats after ETOP treatment relative to untreated control cells, with a 2.4 fold after 24 hr and almost 4 fold after 4 hr ETOP+ 24 hr recovery (Fig. 1D) that parallels the percentage of cells with delocalized CENP-A (Fig. 1C). In contrast to CENP-A, centromere content in canonical histone H3 did not vary significantly following genotoxic stress (Fig. 1E), whereas content in histone H4 showed a reproducible 1.5-fold reduction after 24 hr ETOP treatment. ChIP experiments also showed that CENP-A did not accumulate at ectopic sites in the nearby pericentromeric major satellite repeats, nor in ribosomal DNA repeats (rDNA) despite its accumulation close to nucleoli (Fig. 1D).

DNA damage leads to nucleosomal CENP-A delocalization. The dynamics of parental nucleosomal and newly synthesized CENP-A is tightly regulated during cell cycle[24–27]. We took advantage of the SNAP-tag technology that allows distinguishing the fate of parental or newly synthesized histones[28]. We generated NIH/3T3 cell lines expressing the fused CENP-A-SNAP protein, and used quench-chase-pulse or pulse-chase imaging protocols (see the Methods section). Quenching of parental SNAP histones followed by a 24 hr chase of the non-fluorescent substrate in the presence or absence of ETOP, before labelling of newly synthesized CENP-A with the fluorescent substrate TMR-Star (Fig. 2A), showed that ectopic CENP-A was correctly incorporated in centromeric chromatin of untreated cells (Fig. 2B). In contrast, and consistent with cells arresting in G2 while CENP-A is incorporated in G1, newly synthesized CENP-A was not deposited in centromeres nor close to the

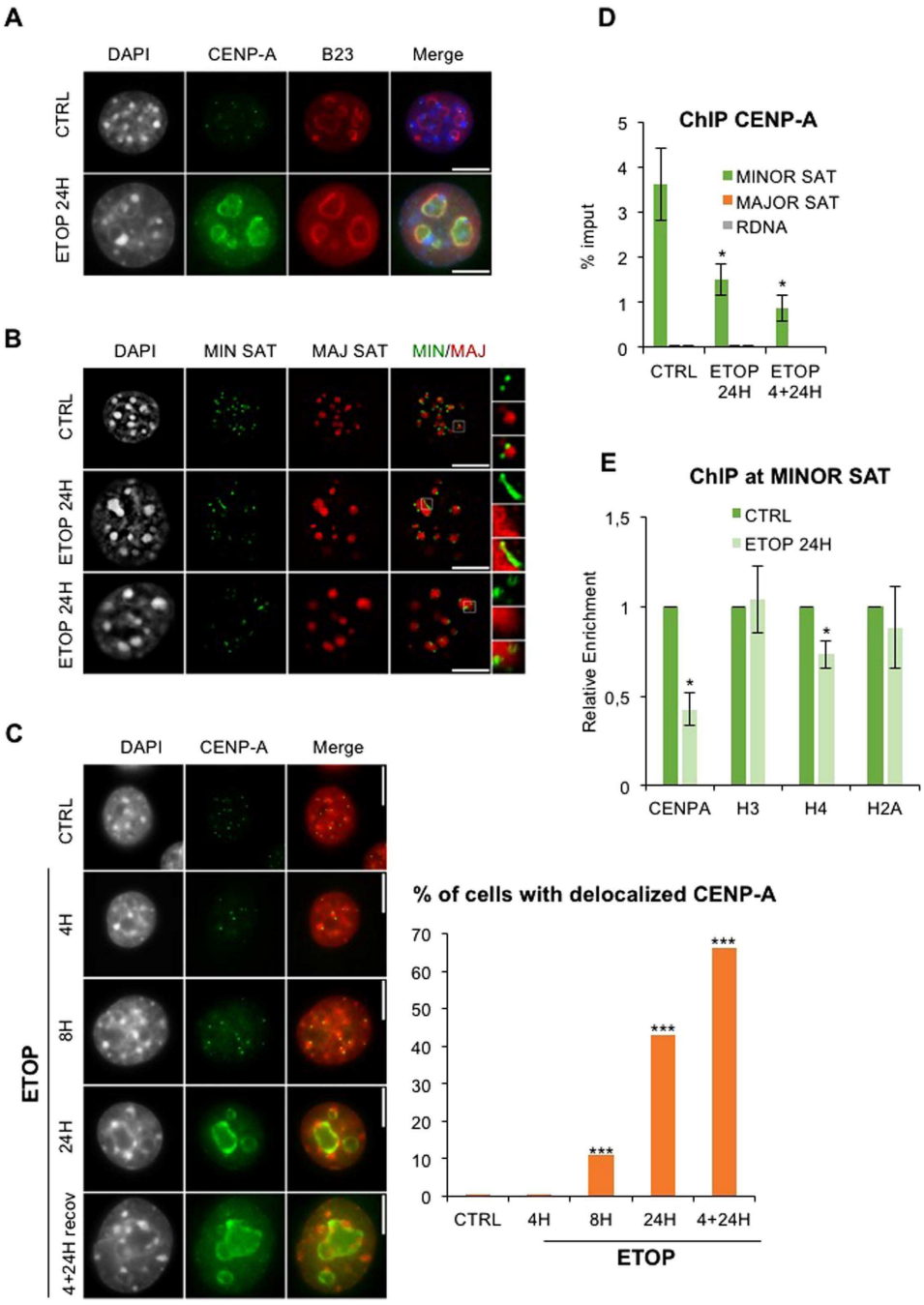


Figure 1. CENP-A is delocalized from centromeric repeats in response to genotoxic stress. (**A**) CENP-A localization in NIH/3T3 cells treated with 10 μM ETOP for 24 hr, analyzed by IF using anti-murine CENP-A and secondary anti-rabbit antibody conjugated to Alexa-488 (green). Antibodies against B23 (nucleophosmin) and anti-mouse antibody conjugated to Alexa-594 (red) were used to mark nucleoli. DAPI was used to stain nuclei (blue). (**B**) Architecture of minor and major satellites repeats in NIH/3T3 cells treated with 10 μM ETOP for 24 hr, analyzed by DNA-FISH with specific probes fluorescently labelled with Alexa 488- (green) or Cy3- (red) conjugated dUTP, respectively. Insets are magnified an additional 3X. (**C**) CENP-A localization analyzed in kinetics experiments on NIH/3T3 cells treated with 10 μM ETOP for 4, 8, 24 hr, or 4 hr followed by a 24 hr recovery period, using immunofluorescence as in (**A**). Percentage of cells that displayed delocalized CENP-A exemplified in (**A**), bottom panel, are indicated on the histogram on the right ($n > 200$ for each condition). P-values (Chi-square test) comparing control and ETOP-treated cells: ***$p < 0.001$. (**D**) CENP-A enrichment at minor satellite, major satellite and rDNA repeats after ETOP treatment, analyzed by ChIP-qPCR. Data are normalized to the input. (**E**) CENP-A and canonical histones enrichment at minor satellite repeats after ETOP treatment, analyzed by ChIP-qPCR. Data are normalized to the input then to the control condition (CTRL) arbitrarily set at 1. Error bars represent S.E.M of at least three independent experiments. P-values (unpaired t-test) comparing controls (CTRL) and drug treatments: *$p < 0.05$ (unpaired t-test). CTRL: untreated cells. Images represent one focal plan. Scale bar, 10 μm.

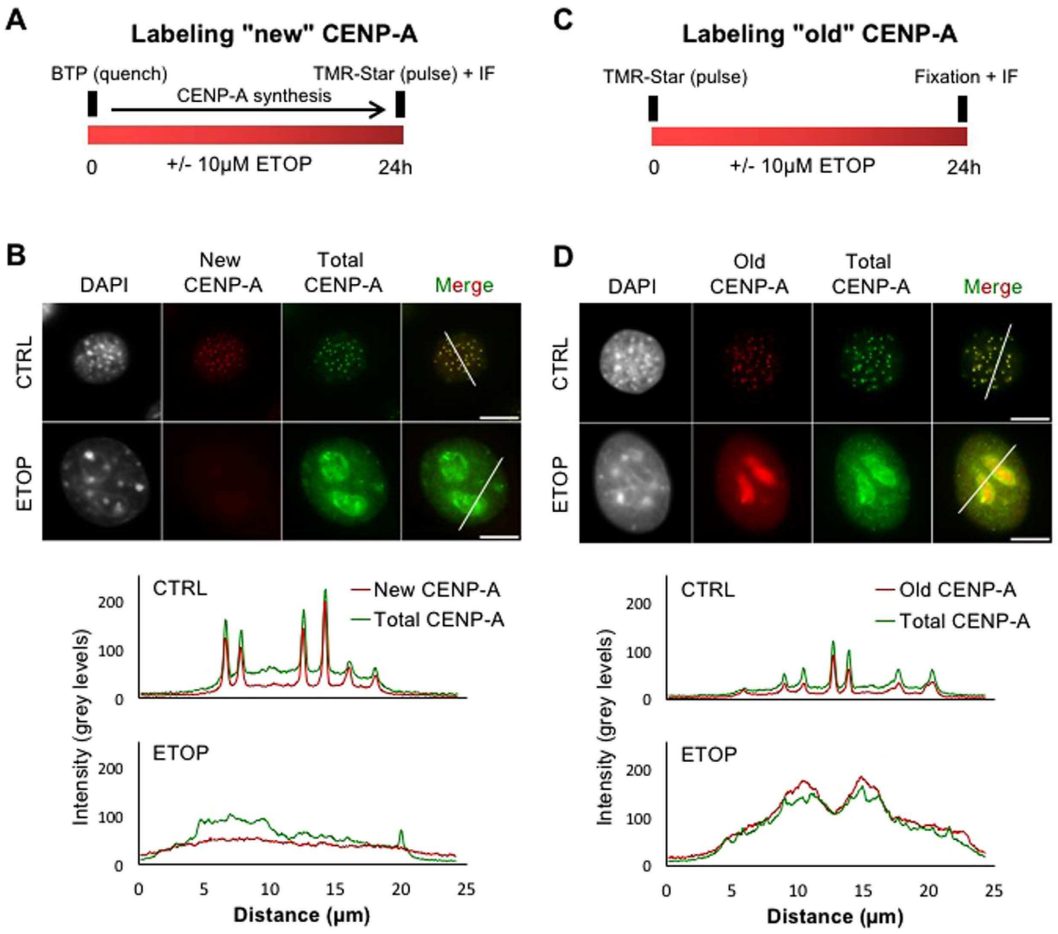


Figure 2. Parental CENP-A is delocalized from centromeres upon DNA damage. (**A**) Schematic representation of the assay used to monitor sub-nuclear localization of newly synthesized CENP-A-SNAP proteins after ETOP treatment using fluorescence microscopy on NIH/3T3 cells stably expressing SNAP-tagged CENP-A. Pre-existing SNAP-tagged CENP-A was quenched with a non-fluorescent substrate (BTP) so that only histones neo-synthetized during the chase period can be further labelled with tetramethylrhodamine (TMR-Star; red) during the pulse step. ETOP treatment was performed during the chase and cells were fixed after 24 hr followed by an IF as in Fig. 1A. (**B**) Detection of total CENP-A using anti-CENP-A antibodies (green) and tagged parental CENP-A (red) following protocol A. Fluorescence intensity line scan through each nucleus (white bar) are represented below images and show the broad nucleolar profiles of total (green line) CENP-A in response to ETOP whereas newly synthesized exogenous CENP-A (SNAP; red line), normally loaded in G1, cannot be incorporated nor accumulates in G2-arrested cells. (**C**) Schematic representation of the assay used to monitor sub-nuclear localization of pre-existing CENP-A proteins after ETOP treatment, using fluorescence microscopy on NIH/3T3 cells stably expressing SNAP-tagged CENP-A. Pre-existing CENP-A-SNAP proteins were labelled with TMR-Star followed by a chase in presence or absence of ETOP. Cells were fixed after 24 hr and IF was performed as in Fig. 1A. (**D**) Same as in (**B**) except that cells were processed following protocol C. Fluorescence intensity line scans through each nucleus show regular punctate pattern in control cells for both exogenous parental (SNAP; red line) and total (green line) CENP-A, whereas ETOP-treated cells show broader signals.

nucleolus of ETOP-treated cells (Fig. 2B). Fluorescent labelling of the SNAP-tagged CENP-A (Fig. 2C) before a 24 hr ETOP treatment showed that parental nucleosomal CENP-A histones were mostly relocated close to the nucleolus concomitantly with endogenous CENP-A, as shown by line scan analysis (Fig. 2D, bottom panels). These experiments suggested that CENP-A dispersion in response to DNA damage results from eviction of parental nucleosomal histones.

ATM is required for CENP-A delocalization following DNA damage. CENP-A mislocalization was observed in genotoxic stress conditions. Thus, we assessed the contribution of ATM and ATR kinase signalling cascades that play central roles in cellular response to DNA damage[19]. We found that inhibition of ATM (ATMi), but not that of ATR (ATRi), prevented stress-mediated CENP-A delocalization (Fig. 3A). While DNA damage persists in the presence of genotoxic stressor, its signalling is impaired by ATMi. Interestingly, the reduced

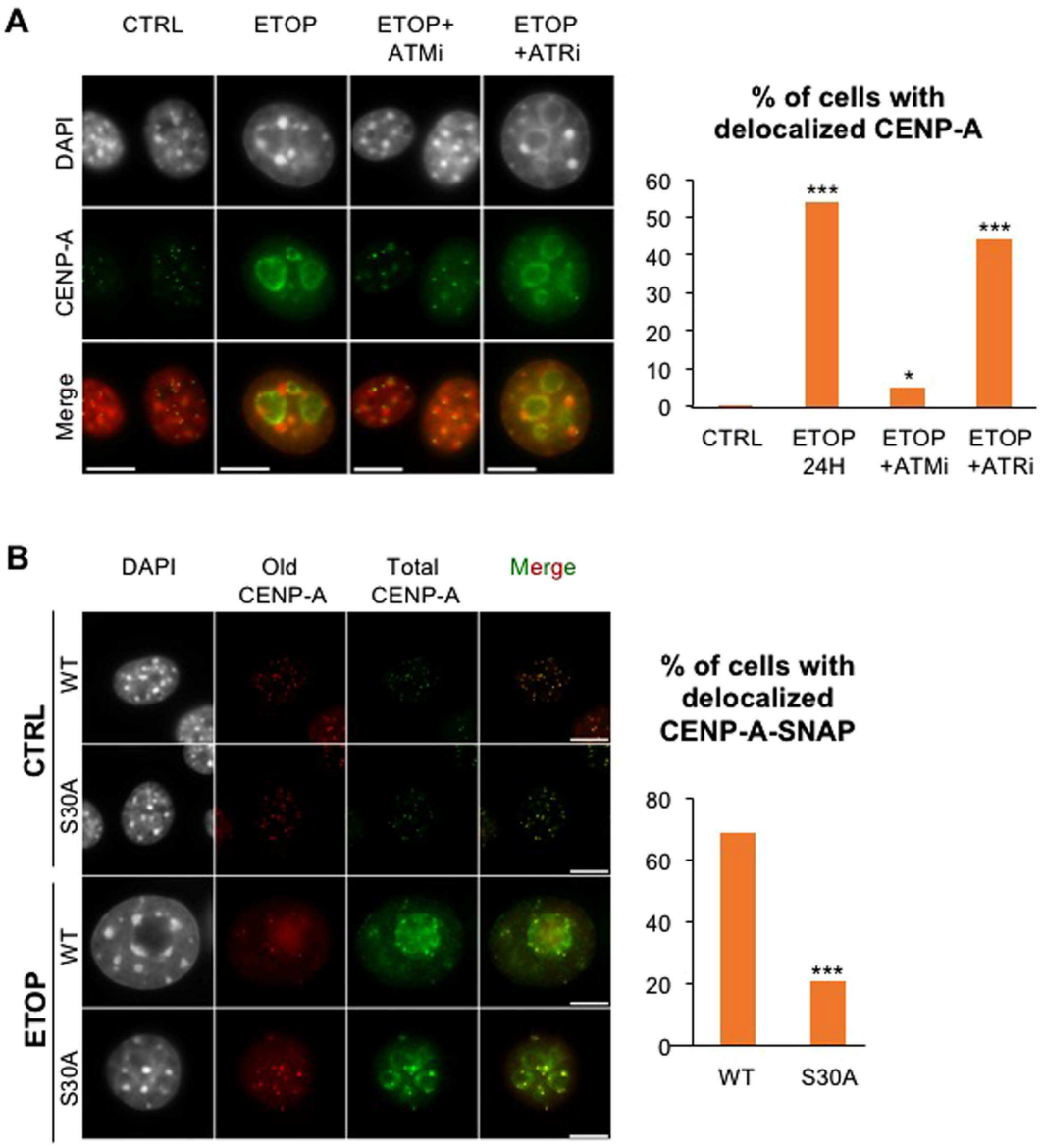


Figure 3. ATM signaling and a phosphorylatable S30 residue are required for CENP-A delocalization. (**A**) CENP-A localization in NIH/3T3 cells co-treated with ETOP and specific inhibitors of ATM (ATMi) or ATR (ATRi) for 24 hr, analyzed as in Fig. 1A. Percentage of cells that displayed CENP-A delocalization is indicated on the histogram on the right ($n > 300$ cells). (**B**) Sub-nuclear localization of pre-existing CENP-A-SNAP proteins, either WT or S30A mutant, after a 24 hr ETOP treatment was analyzed as in Fig. 2. Images represent one focal plan. Scale bar, 10 μm. P-values (Chi-square test) comparing controls (CTRL) and drug treatments: ***$p < 0.001$.

CENP-A delocalization observed with ATMi but not ATRi correlated with override of the G2/M cell-cycle block (Figure S3C).

We noticed the presence of a phosphorylation consensus site for ATM (SQ motif) at serine 30 of murine CENP-A. Mutation of this residue into a non-phosphorylatable alanine (S30A) did not prevent deposition of CENP-A-SNAP suggested by its presence at centromeric foci (Fig. 3B). In contrast, delocalization of the parental CENP-A-SNAP S30A mutant version following ETOP treatment was clearly reduced compared to that of the CENP-A-SNAP WT protein (Fig. 3B). Delocalization of total CENP-A upon DNA damage was equivalent between CENP-A-SNAP WT and S30A cell lines (50 and 60%, respectively) and comparable to that of endogenous CENP-A in untransfected cells (Fig. 1C). However, within the population of cells with delocalized total CENP-A, the percentage of cells with delocalized SNAP-tagged CENP-A dropped from 70% for the WT to 20% for the S30A mutant (Fig. 3B), suggesting that both the ATM signalling pathway and a phosphorylatable Ser30 are required for CENP-A eviction in response to DNA damage in association with cell-cycle arrest.

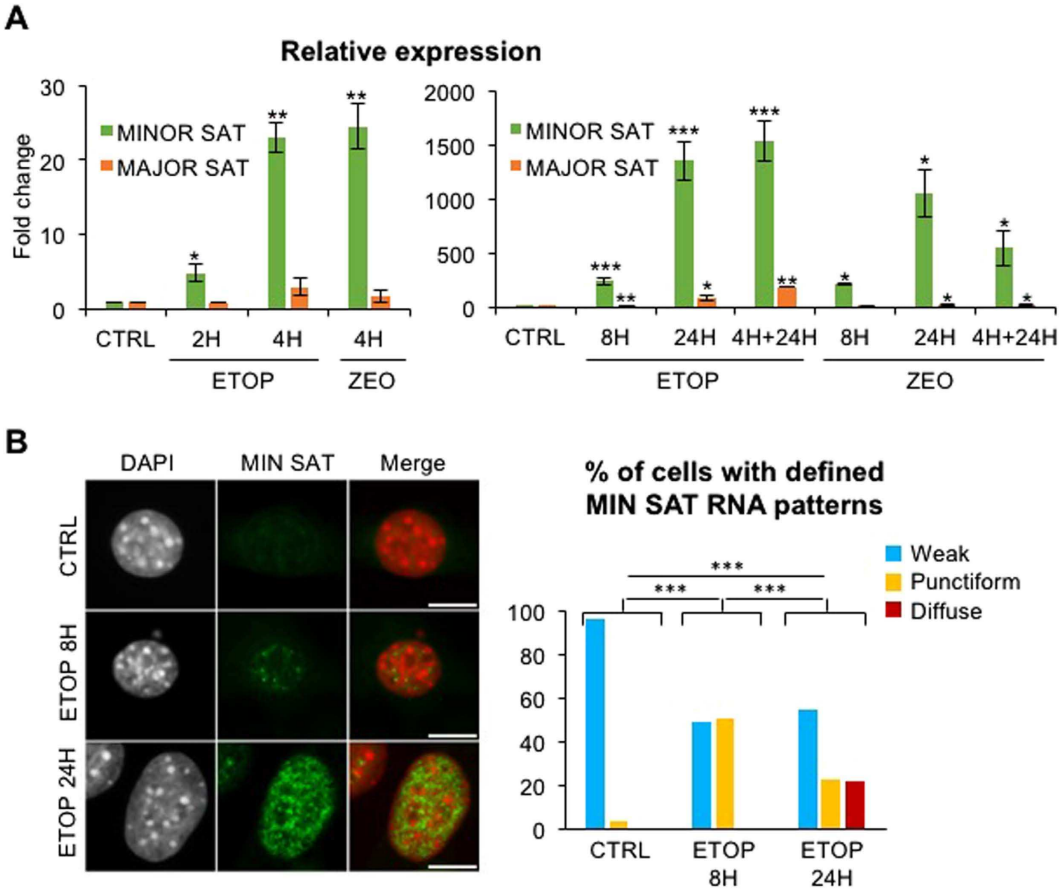


Figure 4. Transcriptional activation of minor satellite repeats in response to genotoxic stress. (**A**) Activated transcription of minor and major satellites in response to ETOP or ZEO treatments, analyzed by RT-qPCR. Short time-point series are shown on the left panel, and longer time-point series or a 4 hr treatment followed by a 24 hr recovery from stress (4 h + 24 h) are shown on the right panel. P-values (unpaired t-test) comparing controls (CTRL) and drug treatments: ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$. CTRL: untreated cells. (**B**) Subnuclear localization of centromeric transcripts in single cells after 8 or 24 hr ETOP treatment, analyzed by RNA-FISH as in Fig. 1B. Histogram shows percentage of cells presenting weak (like in CTRL lane), punctiform (like in ETOP 8 h lane) or diffuse (like in ETOP 24 h lane) minor satellite RNA patterns from RNA-FISH images shown on the left ($n > 200$ cells counted in each conditions). Error bars represent S.E.M of at least three independent experiments. Images represent one focal plan. Scale bar, 10 μm. P-values comparing the repartition in the three categories at different time points of ETOP treatment: ***$p < 0.001$ (chi-square test).

Activated transcription of centromeric repeats precedes CENP-A dispersal. CENP-A delocalization was detectable only after 8 hr of genotoxic stress (Fig. 1C). Therefore, we assessed the accumulation of centromeric transcripts, another feature of stressed cells reported in various stress conditions (Reviewed in refs 29,30). Various genotoxic conditions led to a strong increase in levels of centromeric transcripts as measured by RT-qPCR, relative to untreated cells (Figure S4A). This is in contrast to non-genotoxic stressors like heat shock (HS) or ethanol (EtOH) that had only modest effects in conditions commonly used (Figure S4A), suggesting that increased transcription of centromeric repeats requires a context of DNA damage as evidenced by accumulation of γ-H2A.X and accumulation of the cyclin-dependent kinase inhibitor 1A p21Cip/CDKN1A ($p21^{Cip}$) (Figure S4A).

Kinetic analysis in response to ETOP or ZEO revealed that levels of centromeric transcripts increased within a few hours, with a 5-fold increase by 2 hr of ETOP (Fig. 4A; left panel) reaching levels up to a thousand-fold in 24 hr (Fig. 4A; right panel). Tandemly repeated pericentromeric major satellite repeats were also induced although with a slower kinetics, with a 2 to 3-fold increase by 4 hr with either drug (Fig. 4A; left panel), and reaching apparent lower levels compared to minor satellites transcription, with a 100-fold increase at 24 hr (Fig. 4A; right panel). In contrast, telomeric or rDNA tandem repeats, interspersed repeats like long (LINE) or short interspersed nuclear elements (SINE) and transposable elements like intracisternal A-particle (IAP), did not show changes in their transcripts levels upon DNA damage (Figure S4B), suggesting the selective activation of centromeric repeats in response to genotoxic stress.

We performed RNA-FISH to assess accumulation of centromeric transcripts at the single cell level. In agreement with previous reports[5,7], centromeric transcripts were hardly detectable in control cycling cells (Fig. 4B). In

contrast, centromeric transcripts first accumulated into discrete dots typical of centromeric foci detected in the vicinity of chromocenters in almost half of the cell population after 8 hr of ETOP treatment, suggesting that they remain associated with their transcription sites. Spreading of the signal throughout the nucleoplasm appeared in 20% of the cells after 24 hr of treatment (Fig. 4B), raising the interesting possibility that the impact of increased levels of centromeric transcripts may not be restricted to local perturbations and may affect other nuclear compartments and functions.

Accumulation of murine minor satellites is regulated during cell cycle with a peak in G2/M phase[7]. However, levels reached in G2-arrested ETOP- or ZEO-treated cells were always 10-fold higher than in NOC-arrested cells (Figure S4A). Moreover, high levels of centromeric transcripts were also observed in cells treated with HU and MMC (Figure S4A) that accumulated in G1 or in S (Figure S1B). Hence, accumulation of centromeric transcripts is not a mere consequence of a cell-cycle arrest in G2 but rather the result of an increased transcription or stabilization of these transcripts in response to genotoxic insults. Indeed, inhibition of the transcriptional machinery using the RNA polymerase II (RNA Pol II) inhibitor Actinomycin D (ACTD) together with ETOP treatment abolished the observed strong accumulation of minor satellite transcripts (Figure S5A) as well as CENP-A delocalization (Figure S5B), suggesting that genotoxic stress promotes transcriptional activation of centromeric repeats and that this activated transcription is required for CENP-A relocation. However, forced expression of minor satellite repeats carried out as before[5] and leading, 24 hr post-transfection, to levels similar to that obtained with a 24 hr ETOP treatment, (Figure S5C) was not sufficient to affect CENP-A endogenous localization (Figure S5D).

A p53 WT context is required for activated transcription. One of the main effects of DNA damage is the activation of ATM/ATR signaling[17] and rapid stabilization of the transcription factor p53[20], which in turn activates downstream effectors for an appropriate cellular response to DNA damage. We found that neither ATMi nor ATRi affected levels of centromeric transcripts compared to ETOP treated cells (Figure S6A), suggesting that transcription of centromeric transcripts is activated in DNA damage conditions but independently of ATM/ATR signalling. It also confirmed that high levels of centromeric repeats are not sufficient to promote eviction of CENP-A since ATMi prevented CENP-A delocalization (Fig. 3A) but not accumulation of satellite repeat transcripts (Figure S6A).

We then subjected p53-null murine embryonic fibroblasts (MEF) to the two different doses of ETOP used previously, which led to increased levels of γH2A.X regardless the p53 context, indicating that ATM/ATR signalling is intact in these cells. We verified that the levels of p53 were undetectable (Figure S6B), and that transcription and protein levels of its downstream target $p21^{Cip}$ were activated at both doses in NIH/3T3 cells with wildtype-p53 (WT-p53) but not in p53 null cells ($p53^{-/-}$) (Figure S6B). We found that the levels of minor satellite transcripts did not increase in the absence of p53 after 4 hr of ETOP treatment (Fig. 5A and S6B) and 14-times less compared to WT-p53 cells at later time points (Fig. 5A), suggesting that minor satellite transcription was not delayed but rather severely impaired in a p53-null cellular context.

In non-genotoxic conditions of p53 stabilization, using the small molecule inhibitor of MDM2/p53 interaction Nutlin-3a, transcription from minor satellite repeats was not enhanced, although a slight effect was observed on p53-direct target gene *CDKN1A* transcription (Figure S6C). In contrast, stabilization of p53 by Nutlin-3 potentiated ETOP-mediated transcriptional activation of minor satellite repeats by 3-fold (Figure S6C). ChIP assays showed that, in ETOP-treated cells, p53 had a substantial ability to bind to a consensus single binding site found in minor satellite repeats, although to a lesser extent than to the strong double binding sites described in *CDNK1A* promoter (Fig. 5B). This binding increased by 2-fold in cells treated by both ETOP and Nutlin-3 in correlation with the observed increased transcription (Figs 5B and S6C). These results suggest that stabilized p53 may bind to non-canonical sites in minor satellite repeats, although this binding and the consequent activated transcription of centromeric repeats appears to require the DDR signalling.

CENP-A maintained its default localization in 80% of p53-null cells (Fig. 5C), consistent with both increased transcription of minor satellite transcripts and genotoxic stress signalling being required for CENP-A delocalization. Interestingly, while maintaining the normal CENP-A localization, cells with a defective p53 checkpoint showed an increased micronuclei formation (Figs 5C, arrows; and Fig. 5D) and accumulated in the culture with more than a 2n genomic content (Figure S6D) indicative of defective mitosis and genomic instability.

Histone chaperone FACT is required for DNA damage-induced CENP-A dispersal. Since activated centromeric transcription in stress conditions, but not centromeric transcripts themselves, was required for CENP-A delocalization, we further tested whether it could act through chromatin remodelling. We used siRNA-mediated knockdown of a panel of chromatin remodelers for which a number of studies have shown their role at centromeric and pericentromeric repeats (Table S2 and references included). We also focused on the FACT complex, an ATP-independent histone chaperone first discovered as promoting transcriptional elongation[31] through nucleosome destabilization[32]. In addition, FACT was shown to co-purify with CENP-A nucleosomes[33], and its subunit SSRP1 to be required for centromeric localization of CENP-A[34]. We found that the levels of centromeric transcripts in ETOP-treated cells were unaffected by reduced levels of the chromatin remodelers/chaperones tested (Fig. 6A). In conditions in which we were able to obtain more than 50% reduction in transcripts levels (Figure S7) the percentage of cells with mislocalized CENP-A after ETOP treatment was largely decreased depending on the remodelling/chaperone factor tested (Fig. 6B). The most striking result was obtained following knockdown of SSRP1 expression, confirmed by a strong decrease in the levels of both mRNA and protein (Fig. 6C), which resulted in less than 10% of cells with delocalized CENP-A after a 24 hr ETOP treatment compared to almost 50% in cells transfected with control siRNAs (Fig. 6B). As a whole, these data revealed cooperation between activated transcription and chromatin chaperones/remodelers to relocate CENP-A in DNA damage conditions.

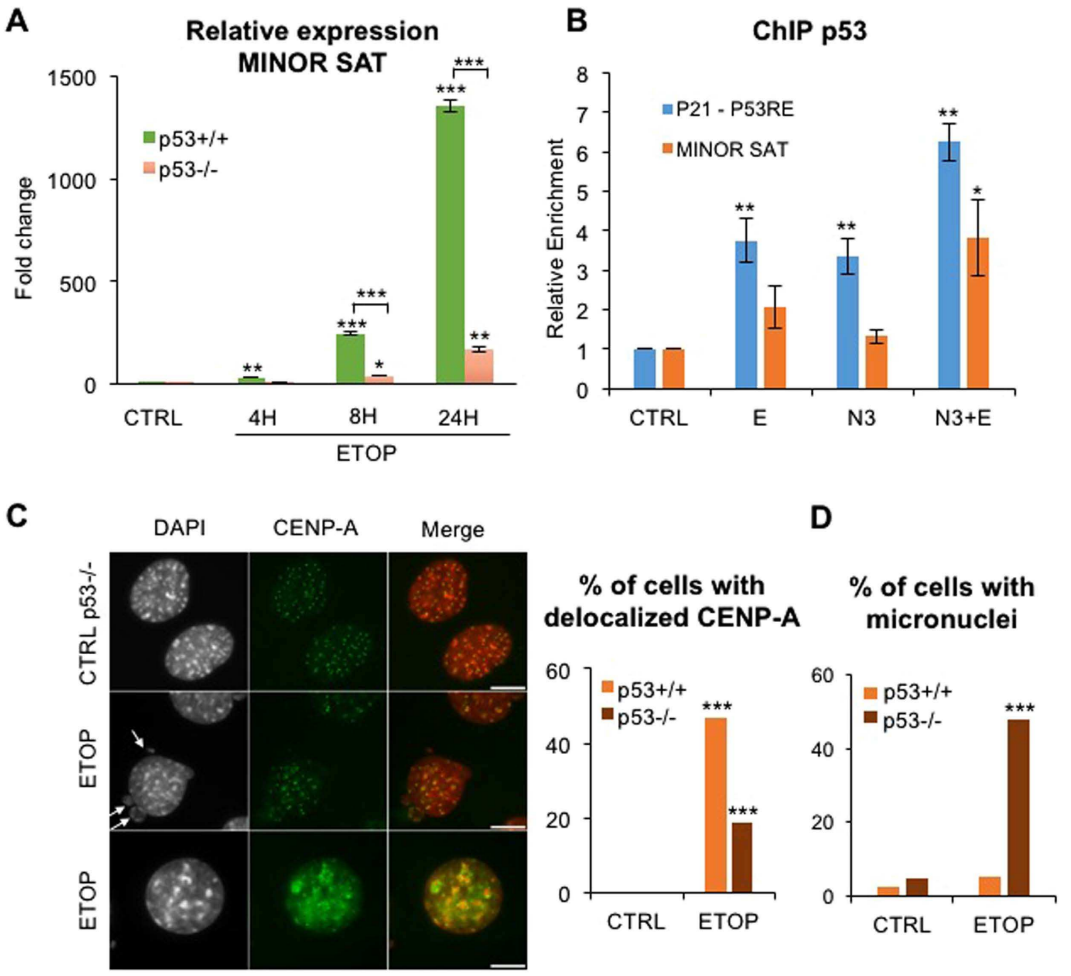


Figure 5. Transcription of centromeric repeats in response to genotoxic stress requires a wildtype p53 context. (**A**) Transcription of minor satellite repeats in WT (NIH3T3; $p53^{+/+}$) or null ($p53^{-/-}$) p53 contexts following ETOP treatment at the indicated time points, was analyzed as in Fig. 4A. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ (unpaired t-test). (**B**) Enrichment of endogenous p53 at the p53 responsive element found at $p21^{CIP}$ promoter (p21-p53RE) or at minor satellite repeats in NIH/3T3 cells after a 4 hr treatment with ETOP, Nutlin-3 (N3) or both (N3 + E), analyzed by ChIP-qPCR. Data are normalized to the input then to the control condition. *$p < 0.05$, **$p < 0.01$ (unpaired t-test). (**C**) CENP-A localization in $p53^{-/-}$ fibroblasts treated with ETOP for 24 hr, analyzed as in Fig. 1A. Percentage of cells with delocalized CENP-A in $p53^{-/-}$ cells compared to $p53^{+/+}$ cells are represented in the adjacent graph ($n > 300$ cells). Arrows point to the presence of micronuclei. ***$p < 0.001$ (chi-square test). (**D**) Percentage of cells with micronuclei in $p53^{-/-}$ cells compared to $p53^{+/+}$ cells ($n > 300$ cells). ***$p < 0.001$ (chi-square test). CTRL: untreated cells; E: ETOP; N3: Nutlin-3a. Error bars represent S.E.M of at least three independent experiments. Images represent one focal plan. Scale bar, 10 μm.

CENP-A dispersal is also a feature of permanently arrested senescent cells. Genotoxic conditions leading to CENP-A delocalization and activated transcription of minor satellite repeats led to transcriptional activation of $p21^{Cip}$, indicative of a pre-senescent state (Figure S4A). In primary murine embryonic fibroblasts (MEF), accumulation of minor satellite transcripts also correlated with CENP-A delocalization, although with apparent different kinetics and lower magnitude that may reflect differences of cellular context and the increased sensitivity of primary cells to genotoxic stress. Indeed, a statistically significant and reproducible 2-fold increased transcription of centromeric transcripts was detectable after 4 hr of ETOP treatment together with relocation of CENP-A. These events correlated with increased expression of the senescent marker β-galactosidase (β-gal) and of cell cycle inhibitors like Cyclin-Dependent Kinase Inhibitor 2 A Ink4/$p16^{INK4a}$ ($p16^{INK4a}$) that started being detectable after 24 hr ETOP treatment (Figure S8A,B). We thus assessed centromere integrity in a context of homogeneous populations of senescent cells, where cells are permanently arrested independently of a genotoxic stressor, while maintaining high levels of ATM signalling and stabilized p53[21].

We first used MEFs as a classical model system as they rapidly lose their proliferation potential after a few passages in culture. MEFs at passage 6 (p6) stained positive for senescence-associated endogenous β-gal in more than 80% of the cells (Figure S8C), and exhibited higher levels of $p21^{Cip}$, $p16^{INK4a}$ and stabilization of p53 (Figure S8D). We also induced senescence in primary cells through the known Nutlin-3a-forced stabilization

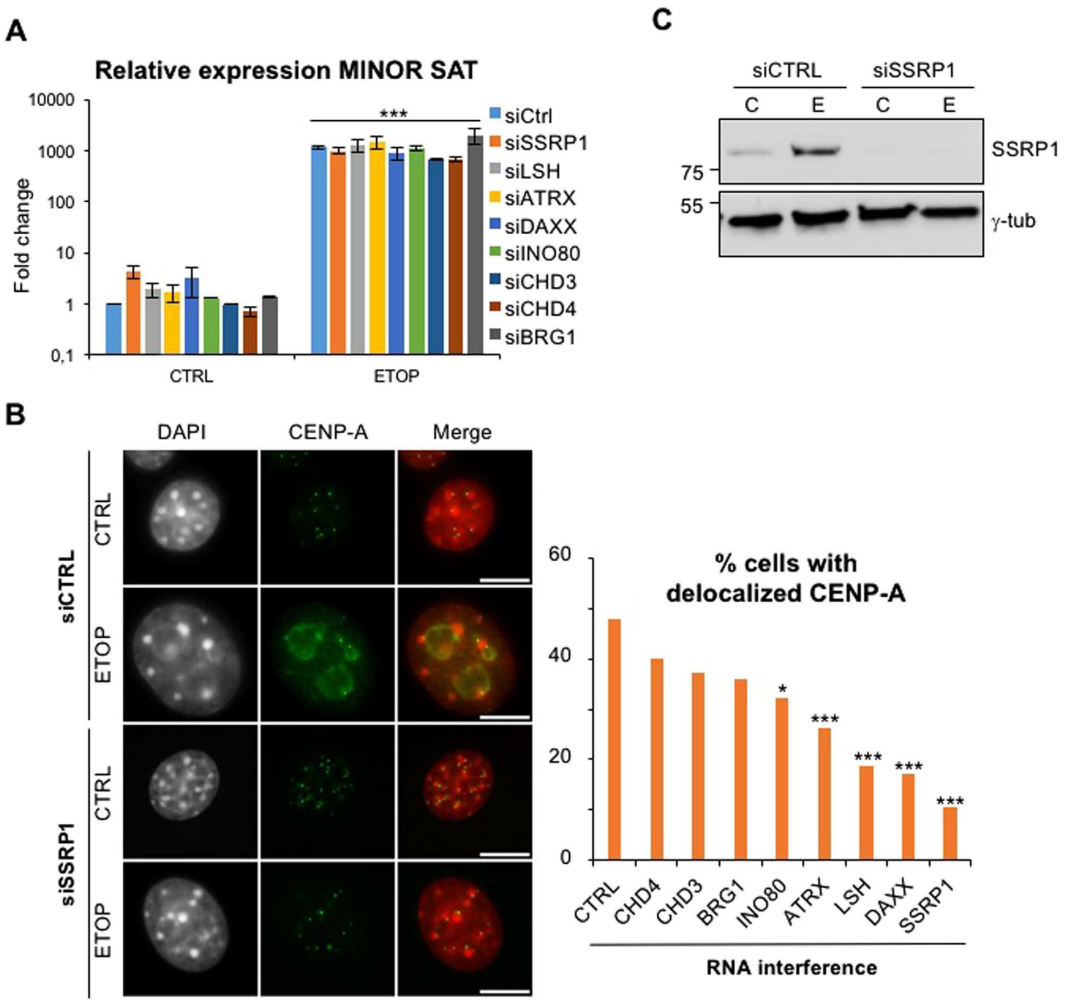


Figure 6. CENP-A delocalization is dependent on chromatin remodelers and the FACT complex. (**A**) Minor satellites transcription following knock down of SSRP1 or other chromatin remodelers/chaperones, was analyzed by qRT-PCR as in Fig. 4A. Histogram represents relative fold change values normalized to U6 RNA levels. Value in untreated cells (CTRL) transfected with control siRNA was arbitrarily set at one. Error bars represent S.E.M of at least 3 independent experiments. ***$p < 0.001$ (unpaired t-test). (**B**) Impact of SSRP1 knock down, compared to control siRNAs, on CENP-A delocalization in NIH/3T3 cells after a 24 hr ETOP treatment, analyzed as in Fig. 1A. Percentage of cells that displayed CENP-A delocalization in ETOP-treated cells after knock down of SSRP1 or other chromatin remodelers/chaperones are indicated on the histogram on the right ($n > 300$ cells). Images represent one focal plan. Scale bar, 10 μm. ***$p < 0.001$ (chi-square test). (**C**) Efficiency of SSRP1 protein knock down using specific compared to control siRNAs, analyzed by western blot. Levels of γ-tubulin served as a loading control.

of p53[35] that showed the same features as above (Figure S8E). Levels of centromeric transcripts in senescent MEFs-p6 or Nutlin-treated MEFs were 16 and 5-fold higher compared to the same cells at passage 2 (p2), respectively (Figure S9A,B). CENP-A was mislocalized in both cases of senescent cells (Figs 7 and S9C). Compared to punctiform foci observed in control cells at p2, CENP-A formed stretched and rather fragmented signals in 40% of the MEFs-p6 cells (Figure S9C, middle panel). An additional 20% of the cells showed both stretched signals and partial accumulation close to the nucleolus (Figure S9C, lower panel). This effect was even more pronounced following Nutlin-3-forced p53 stabilization with almost 60% of the cells showing aberrant CENP-A localization (Fig. 7). DNA-FISH showed that centromeric architecture was drastically perturbed since, in addition to the known perturbation of pericentromeric heterochromatin foci detected by dense DAPI staining or major satellite probes, centromeric minor satellite signals were also highly disorganized and stretched in senescent cells (Figure S9D).

Discussion

Our data revealed that murine centromeres are strikingly disassembled in genotoxic stress conditions in a manner that is dependent on the main DDR effectors. Centromeric repeats are selectively and rapidly transcriptionally activated in a p53-dependant but ATM-independent manner. Transcription of centromeric repeats

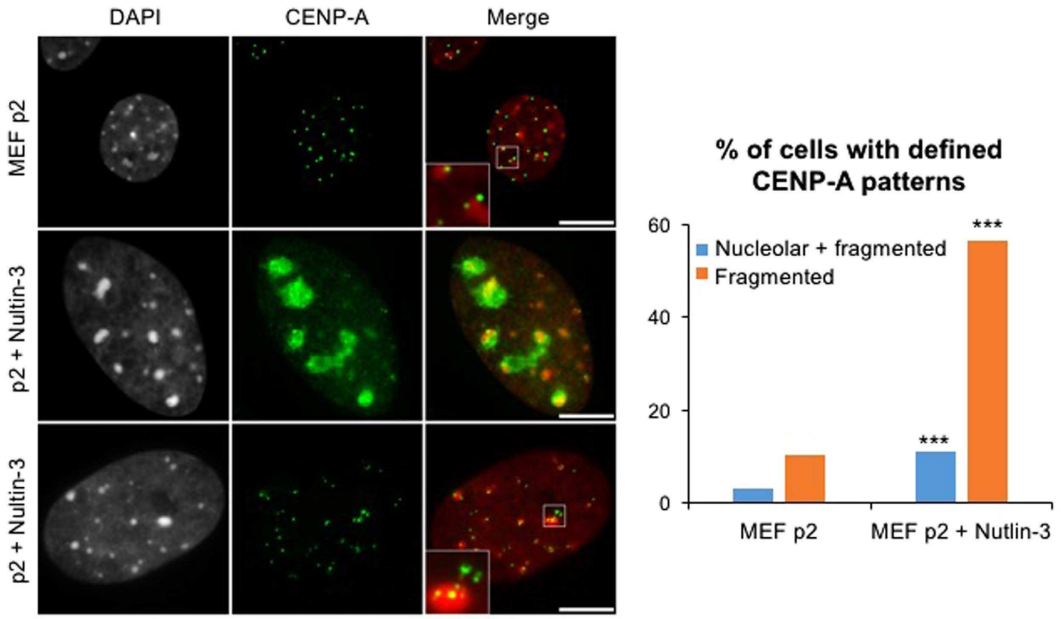


Figure 7. Perturbation of centromeres in senescent cells. CENP-A localization in Nutlin-3-induced senescent cells, analyzed as in Fig. 1A. Percentage of cells showing CENP-A nucleolar (as in middle panel) or fragmented (as in lower panel) patterns are represented on the histogram ($n > 200$ cells). Images represent one focal plan. Scale bar, 10 μm. ***$p < 0.001$ (chi-square test).

has been suggested to play structural role in centromere or kinetochore integrity in many species (Reviewed in ref. 29. However, we report here that, in combination with activated DDR, it leads to and is required for the drastic structural disassembly of centromeric chromatin in the form of ATM and chromatin remodelers/chaperones-dependent delocalization of its epigenetic mark CENP-A. Remarkably, these features are also hallmarks of permanently arrested senescent cells where the DDR is activated, consistent with the role of p53 and ATM in maintenance of the senescent state and suggestive of a causal link. Together, these data shed light on molecular mechanisms through which activated transcription of murine centromeric repeats may operate as an effector in conditions of sustained DNA damage signalling to trigger safeguard mechanisms and prevent cells from dividing to preserve genomic integrity.

In unstressed conditions, the transcription of murine centromeric transcripts is tightly regulated during the cell cycle[7]. Here, we showed that transcription of centromeric repeats is rapidly activated in response to genotoxic stress, within hours. Interestingly, this response seems to be selective for centromeric repeats, at least in the first hours of the response, since other repetitive elements, either in tandem or interspersed, were not activated. This is in contrast with studies on heat or osmotic shocks that revealed transcriptional activation of human pericentromeric satellite type III repeats, but not that of centromeric alpha satellites[14–16,36]. Although satellite repeats transcription seems to be a conserved feature of the stress response, and in absence of sequence conservation among species, the question of any selectivity depending on the type of stress or species is unclear. However, it is interesting to note that conditions where DDR activation is persistent, in cases of non repairable damage or in senescent MEF cells, are characterized by high levels of centromeric transcripts and sustained cell cycle arrest. In addition, a WT-p53 content, known to trigger stable cell-cycle arrest after DNA damage[37], is also required for transcriptional activation of minor satellite transcripts. Hence, it is tempting to speculate that it is the function of centromeres that is targeted under persistent stress signalling where stable cell cycle arrest is required whereas mild conditions that allow cells to recover from heat shock for instance preserve this function.

We showed that persistent stress signalling and senescence are also characterized by the striking dispersion of CENP-A away from its default localization at centromeric repeats. In fact, perturbed architecture of centromeric repeats seems to be a conserved feature of senescent cells in bovine[38], human[39] and murine cells (this study), and most likely relies on CENP-A depletion at centromeres. Indeed, reduced levels of CENP-A levels have been previously reported in human senescent cells, and their forced down regulation by RNA interference leads to premature senescence[40]. We reported here that, in the absence of p53, centromeric transcriptional activation and delocalization of CENP-A do not occur, while cells accumulate micronuclei. Therefore, CENP-A down regulation or relocation away from centromeres might act as a defence mechanism to maintain genomic stability and cell viability by preventing centromere-defective cells from undergoing cell division. Altogether, these data functionally link activated transcription and chromatin remodelling at centromeric regions with stable proliferative halt characteristic of senescent cells or cells subjected to persistent DNA damage, by altering centromere identity and function and hence, preventing cell division.

Transcriptional activation of centromeric repeats in genotoxic stress conditions was detected before CENP-A delocalization suggesting that it may promote or participate in chromatin remodelling process and CENP-A delocalization from centromeric regions at later time points. In support of this hypothesis, we found that

ectopically expressed minor satellite transcripts were not sufficient to induce CENP-A relocation. In addition, knock down of several chromatin remodelers and chaperones prevented CENP-A relocation without affecting centromeric repeats transcription. Among these factors, knock down of FACT complex subunit SSRP1 showed the most dramatic effects. The FACT complex emerged as a versatile factor in the control of chromatin dynamics at centromeres depending on cellular contexts, promoting CENP-A deposition during mitosis[34,41] or preventing pervasive deposition[42–44]. Other chromatin remodelers like LSH/HELLS, ATRX and DAXX, for which a role in centromere remodelling has also been reported[36,45,46], may cooperate in the eviction process since their loss of function also prevented CENP-A eviction, although to a lesser extent than that of SSRP1. Interestingly, this seems to be independent of the known role of FACT on transcription of centromeric repeats[31] since loss of function of these factors did not prevent their transcription, suggesting that other factors are implicated.

Chromatin is highly dynamic in response to DNA damage, including post-translational modification of histones and deposition of histone variants at sites of DNA damage as integral parts of the DDR to coordinate efficient signalling and repair[47]. The DDR kinases are involved in genome surveillance and in sensing DNA damage[19]. When activated in the presence of genome threatening insults, they rapidly phosphorylate several hundreds of target proteins[48]. As mentioned above, phosphorylation of histones, like the histone variant H2A.X, participates in chromatin dynamics at strand breaks and is one of the earliest events in the DDR[49]. We further showed that the DDR kinase ATM, but not ATR, was required to promote CENP-A delocalization in response to DNA damage caused by a panel of genotoxic agents. We showed that the rather late DDR-mediated relocation of CENP-A required the ATM kinase and a phosphorylatable consensus SQ motif at Ser30. Under these conditions, it is difficult to assess whether CENP-A is one of the hundreds of targets of ATM signalling pathway, but our data support the view that CENP-A delocalization is mediated by sustained ATM signalling in contexts of persistent DNA damage or senescent cells.

Given that CENP-A nucleosomes may represent only a small fraction of total centromeric nucleosomes in human cells[50], the slight but reproducible decrease in histone H4 content but not that of canonical H3 suggests that DNA damage and activated centromeric transcription trigger active eviction of a subset of CENP-A nucleosomes from centromeric repeats. This is remarkable since CENP-A nucleosomes are very stable once incorporated into centromeric sequences[51,52]. CENP-A deposition and maintenance at the centromere is orchestrated by numerous factors and so far, CENP-A depletion from centromeres has only been linked to perturbations of its deposition machinery[8,26,27,53]. Interestingly, active eviction and degradation of CENP-A has been shown in the case of its ectopic incorporation in yeast and fly chromosome arms[43,51,54,55]. However, in stressed murine cells, we did not find evidence for ectopic incorporation of CENP-A in other DNA repeats although the protein showed apparent accumulation nearby nucleoli. Then, why delocalized CENP-A is not degraded remains a puzzling question.

CENP-A was shown to be very rapidly and transiently recruited at sites of laser induced DNA damage in human and mouse cells[56], or in drosophila cells depleted for the histone fold protein Chrac14 after MMS treatment[57]. However, it is worth noting that the CENP-A relocation that we observed in response to DNA damage is a rather late event and therefore may not participate in DNA repair *per se* but rather, may be a consequence of prolonged stress and persistent DDR as in senescent cells, promoting or involved in maintenance of cell cycle arrest.

The functional intricacy between DDR and chromatin remodelling at specialized chromosomal domains has been long known in the context of telomeres[58,59]. Remarkably, non-canonical p53-binding sites in subtelomeric or centromeric regions confer enhancer-like functions for transcriptional activation at telomeric repeat-containing RNA (TERRA)[60] or centromeric repeats (this study), required to elicit chromatin changes that will further prevent the cell from dividing. Despite these similarities, in humans, proliferative arrest and replicative senescence are directly linked to telomeres length and function[58,61]. In contrast, murine cells enter senescence while maintaining long telomeres. In addition, murine CENP-A presents a consensus site for ATM phosphorylation. Our study uncovered an original link between DDR effectors and centromeres whereby centromere activated transcription and loss of identity is a prominent feature of cell cycle-arrested stressed and senescent murine cells. Hence, whether organisms evolved compensatory mechanisms to efficiently delay or halt the cell cycle to safeguard their genome remains a puzzling question.

Experimental Procedures

Full experimental procedures are provided in the Supplemental Information section.

Cell lines and transfections. Mouse embryonic fibroblasts (MEFs), NIH/3T3 cells and their CENP-A-SNAP derivatives were cultured and transfected as described[62]. CENP-A–SNAP was constructed by inserting a PCR-generated murine CENP-A cDNA into pSNAPf vector (New England Biolabs) in frame with the SNAP tag. Mutation of the Ser-30 in alanine (S30A) was generated using the Q5® Site-Directed Mutagenesis Kit according to the manufacturer's recommendations (New England Biolabs). The eukaryotic expression vector containing four minor satellite repeats units under the control of the CMV promoter was generated as described previously[5].

SNAP quench and pulse labelling. We followed previously described protocols for SNAP quench-chase-pulse or pulse-chase imaging[63]. SNAP-tagged histones were pulse-labelled with 2 μM SNAP-cell TMR-star (New England Biolabs) for 15 min or quenched with 2 μM SNAP-cell Block (BTP, New England Biolabs) for 30 min. After quenching or pulse labelling, cells were washed twice with PBS and then incubated in complete medium for 30 min to allow excess compound to diffuse from cells. Then, cells were washed twice with PBS and incubated for 24 h in complete medium containing or not 10 μM Etoposide. Labelling of total CENP-A was then monitored by immunofluorescence as described below.

Immunofluorescence. Immunofluorescence was performed as previously described[5].

DNA & RNA FISH. Cells were directly grown on Superfrost Plus microscope slides (Menzel–Glaser, Braunschweig, Germany) then fixed and permeabilized as described above. DNA-FISH and RNA-FISH were performed as previously described[5].

Microscopy. Image acquisition was performed at room temperature on a fluorescence microscope (Axioplan 2; Zeiss) with a Plan-Neofluar 100X/1.3 NA oil immersion objective (Zeiss) using a digital cooled camera (CoolSNAP fx; Photometrics) and METAMORPH 7.04 software (Roper Scientific, Trenton, NJ). Images presented correspond to one focal plane. Analysis was performed by scoring at least 150 cells in each experiment.

siRNA and plasmid transfection. siRNA purchased from Sigma or Eurofins MWG Operon (Table S5) were transfected into cells using Interferin (Polyplus transfection) following manufacturer's instructions.

RNA extraction and analysis of gene expression. Total RNA from cell lines was isolated using TRI Reagent® (Sigma) according to manufacturer's instructions. Contaminant genomic DNA was eliminated with TURBO DNA-free kit (Ambion). Real-time PCR was performed using the LightCycler® DNA Master SYBR Green I mix (Roche) supplemented with 0.2 μM specific primer pairs (Table S6) and analysed by the comparative CT($\Delta\Delta$CT) method using U6 RNA as an invariant RNA. Each data shown in RT-qPCR analysis is the result of at least three independent experiments performed on at least three independent RNA extractions.

Nuclear extracts and Western blot. Nuclear fractionation and western blot analysis was performed as before[62].

Chromatin Immunoprecipitation. ChIP was essentially performed as described[62]. Sequences of primers are listed in Table S6.

References

1. Allshire, R. C. & Karpen, G. H. Epigenetic regulation of centromeric chromatin: old dogs, new tricks? *Nat. Rev. Genet.* **9,** 923–937 (2008).
2. Sullivan, B. A. & Karpen, G. H. Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nat. Struct. Mol. Biol.* **11,** 1076–1083 (2004).
3. Earnshaw, W. C. Discovering centromere proteins: from cold white hands to the A, B, C of CENPs. *Nat. Rev. Mol. Cell. Biol.* **16,** 443–449 (2015).
4. Topp, C. N., Zhong, C. X. & Dawe, R. K. Centromere-encoded RNAs are integral components of the maize kinetochore. *Proc. Natl. Acad. Sci. USA* **101,** 15986–15991 (2004).
5. Bouzinba-Segard, H., Guais, A. & Francastel, C. Accumulation of small murine minor satellite transcripts leads to impaired centromeric architecture and function. *Proc. Natl. Acad. Sci. USA* **103,** 8709–8714 (2006).
6. Pezer, Z. & Ugarković, D. Transcription of pericentromeric heterochromatin in beetles—satellite DNAs as active regulatory elements. *Cytogenet. Genome Res.* **124,** 268–76 (2009).
7. Ferri, F., Bouzinba-Segard, H., Velasco, G., Hube, F. & Francastel, C. Non-coding murine centromeric transcripts associate with and potentiate Aurora B kinase. *Nucleic Acids Res.* **37,** 5071–5080 (2009).
8. Quénet, D. & Dalal, Y. A long non-coding RNA is required for targeting centromeric protein A to the human centromere. *Elife* **3,** e03254 (2014).
9. Rosic, S., Kohler, F. & Erhardt, S. Repetitive centromeric satellite RNA is essential for kinetochore formation and cell division. *J. Cell Biol.* **207,** 335–349 (2014).
10. Blower, M. D. Centromeric Transcription Regulates Aurora-B Localization and Activation. *Cell Rep.* **15,** 1624–1633 (2016).
11. Eymery, A. *et al.* A transcriptomic analysis of human centromeric and pericentric sequences in normal and tumor cells. *Nucleic Acids Res.* **37,** 6340–6354 (2009).
12. Ting, D. T. *et al.* Aberrant Overexpression of Satellite Repeats in Pancreatic and Other Epithelial Cancers. *Science* **331,** 593–596 (2011).
13. Zhu, Q. *et al.* BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. *Nature* **477,** 179–184 (2011).
14. Jolly, C. *et al.* Stress-induced transcription of satellite III repeats. *J. Cell Biol.* **164,** 25–33 (2004).
15. Rizzi, N. *et al.* Transcriptional activation of a constitutive heterochromatic domain of the human genome in response to heat shock. *Mol. Biol. Cell* **15,** 543–551 (2004).
16. Valgardsdottir, R. *et al.* Transcription of Satellite III non-coding RNAs is a general stress response in human cells. *Nucleic Acids Res.* **36,** 423–434 (2008).
17. Jackson, S. P. Sensing and repairing DNA double-strand breaks. *Carcinogenesis* **23,** 687–696 (2002).
18. Papamichos-Chronakis, M. & Peterson, C. L. Chromatin and the genome integrity network. *Nat. Rev. Genet.* **14,** 62–75 (2013).
19. Durocher, D. & Jackson, S. P. DNA-PK, ATM and ATR as sensors of DNA damage: variations on a theme? *Curr. Opin. Cell Biol.* **13,** 225–231 (2001).
20. Meek, D. W. The p53 response to DNA damage. *DNA Repair* **3,** 1049–1056 (2004).
21. Fumagalli, M., Rossiello, F., Mondello, C. & d'Adda di Fagagna, F. Stable cellular senescence is associated with persistent DDR activation. *PLoS One* **9,** e110969 (2014).
22. Guenatri, M., Bailly, D., Maison C. & Almouzni, G. Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *J. Cell Biol.* **166,** 493–505 (2004).
23. Burden, D. A. & Osheroff, N. Mechanism of action of eukaryotic topoisomerase II and drugs targeted to the enzyme. *Biochim. Biophys. Acta BBA - Gene Struct. Expr.* **1400,** 139–154 (1998).
24. Jansen, L. E. T., Black, B. E., Foltz, D. R. & Cleveland, D. W. Propagation of centromeric chromatin requires exit from mitosis. *J. Cell Biol.* **176,** 795–805 (2007).
25. Hemmerich, P. *et al.* Dynamics of inner kinetochore assembly and maintenance in living cells. *J. Cell Biol.* **180,** 1101–1114 (2008).
26. Dunleavy, E. M. *et al.* HJURP Is a Cell-Cycle-Dependent Maintenance and Deposition Factor of CENP-A at Centromeres. *Cell* **137,** 485–497 (2009).
27. Foltz, D. R. *et al.* Centromere-Specific Assembly of CENP-A Nucleosomes Is Mediated by HJURP. *Cell* **137,** 472–484 (2009).
28. Keppler, A. *et al.* A general method for the covalent labeling of fusion proteins with small molecules *in vivo*. *Nat. Biotechnol.* **21,** 86–89 (2003).
29. Chen, C. C. & Mellone, B. G. Chromatin assembly: Journey to the CENter of the chromosome. *J. Cell Biol.* **214,** 13–24 (2016).

30. Biscotti, M. A., Canapa, A., Forconi, M., Olmo, E. & Barucca, M. Transcription of tandemly repetitive DNA: functional roles. *Chromosome Res.* **23,** 463–477 (2015).
31. Orphanides, G., LeRoy, G., Chang, C. H., Luse, D. S. & Reinberg, D. FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell* **92,** 105–116 (1998).
32. Belotserkovskaya, R. *et al.* FACT facilitates transcription-dependent nucleosome alteration. *Science* **301,** 1090–1093 (2003).
33. Foltz, D. R. *et al.* The human CENP-A centromeric nucleosome-associated complex. *Nat. Cell Biol.* **8,** 458–469 (2006).
34. Okada, M., Okawa, K., Isobe, T. & Fukagawa, T. CENP-H–containing complex facilitates centromere deposition of CENP-A in cooperation with FACT and CHD1. *Mol. Biol. Cell* **20,** 3986–3995 (2009).
35. Efeyan, A. *et al.* Induction of p53-dependent senescence by the MDM2 antagonist nutlin-3a in mouse cells of fibroblast origin. *Cancer Res.* **67,** 7350–7357 (2007).
36. Morozov, V. M., Gavrilova, E. V., Ogryzko, V. V. & Ishov, A. M. Dualistic function of Daxx at centromeric and pericentromeric heterochromatin in normal and stress conditions. *Nucleus* **3,** 276–285 (2012).
37. Bunz, F. *et al.* Requirement for p53 and p21 to sustain G2 arrest after DNA damage. *Science.* **282,** 1497–501 (1998).
38. Pichugin, A., Beaujean, N., Vignon, X. & Vassetzky, Y. Ring-Like Distribution of Constitutive Heterochromatin in Bovine Senescent Cells. *PLoS One* **6,** e26844 (2011).
39. Swanson, E. C., Manning, B., Zhang, H. & Lawrence, J. B. Higher-order unfolding of satellite heterochromatin is a consistent and early event in cell senescence. *J. Cell Biol.* **203,** 929–942 (2013).
40. Maehara, K., Takahashi, K. & Saitoh, S. CENP-A Reduction Induces a p53-Dependent Cellular Senescence Response To Protect Cells from Executing Defective Mitoses. *Mol. Cell. Biol.* **30,** 2090–2104 (2010).
41. Chen, C.-C. *et al.* Establishment of Centromeric Chromatin by the CENP-A Assembly Factor CAL1 Requires FACT-Mediated Transcription. *Dev. Cell* **34,** 73–84 (2015).
42. Choi, E. S. *et al.* Factors that promote H3 chromatin integrity during transcription prevent promiscuous deposition of CENP-A(Cnp1) in fission yeast. *PLoS Genet*. **8,** e1002985 (2012).
43. Deyter, G. M. R. & Biggins, S. The FACT complex interacts with the E3 ubiquitin ligase Psh1 to prevent ectopic localization of CENP-A. *Genes Dev.* **28,** 1815–1826 (2014).
44. Jeronimo, C. & Robert, F. Histone chaperones FACT and Spt6 prevent histone variants from turning into histone deviants. *Bioessays.* **38,** 420–6 (2016).
45. Huang, J. *et al.* Lsh, an epigenetic guardian of repetitive elements. *Nucleic Acids Res*. **32,** 5019–5028 (2004).
46. Lacoste, N. *et al.* Mislocalization of the Centromeric Histone Variant CenH3/CENP-A in Human Cells Depends on the Chaperone DAXX. *Mol. Cell* **53,** 631–644 (2014).
47. Price, B. D. & D'Andrea, A. D. Chromatin Remodeling at DNA Double-Strand Breaks. *Cell* **152,** 1344–1354 (2013).
48. Matsuoka, S. *et al.* ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* **316,** 1160–1166 (2007).
49. Modesti, M. & Kanaar, R. DNA repair: spot (light) s on chromatin. *Curr. Biol.* **11,** R229–R232 (2001).
50. Bodor, D. L. *et al.* The quantitative architecture of centromeric chromatin. *eLife* **3,** e02137 (2014).
51. Collins, K. A., Furuyama, S. & Biggins, S. Proteolysis contributes to the exclusive centromere localization of the yeast Cse4/CENP-A histone H3 variant. *Curr. Biol.* **14,** 1968–1972 (2004).
52. Fachinetti, D. *et al.* A two-step mechanism for epigenetic specification of centromere identity and function. *Nat. Cell Biol.* **15,** 1056–1066 (2013).
53. Bergmann, J. H. *et al.* Epigenetic engineering shows H3K4me2 is required for HJURP targeting and CENP-A assembly on a synthetic human kinetochore. *EMBO J.* **30,** 328–340 (2011).
54. Moreno-Moreno, O., Torras-Llort, M. & Azorin, F. Proteolysis restricts localization of CID, the centromere-specific histone H3 variant of Drosophila, to centromeres. *Nucleic Acids Res.* **34,** 6247–6255 (2006).
55. Ranjitkar, P. *et al.* An E3 Ubiquitin Ligase Prevents Ectopic Localization of the Centromeric Histone H3 Variant via the Centromere Targeting Domain. *Mol. Cell* **40,** 455–464 (2010).
56. Zeitlin, S. G. *et al.* Double-strand DNA breaks recruit the centromeric histone CENP-A. *Proc. Natl. Acad. Sci. USA* **106,** 15762–15767 (2009).
57. Mathew, V. *et al.* The Histone-Fold Protein CHRAC14 Influences Chromatin Composition in Response to DNA Damage. *Cell Rep.* **7,** 321–330 (2014).
58. Cusanelli, E. & Chartrand, P. Telomeric repeat-containing RNA TERRA: a noncoding RNA connecting telomere biology to genome integrity. *Front. Genet*. **6,** 143 (2015).
59. Porro, A. *et al.* Functional characterization of the TERRA transcriptome at damaged telomeres. *Nat. Commun.* **5,** 5379 (2014).
60. Tutton, S. *et al.* Subtelomeric p53 binding prevents accumulation of DNA damage at human telomeres. *EMBO J.* **35,** 193–207 (2016).
61. Kuilman, T., Michaloglou, C., Mooi, W. J. & Peeper, D. S. The essence of senescence. *Genes Dev.* **24,** 2463–2479 (2010).
62. Velasco, G. *et al.* Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proc Natl Acad Sci USA* **107,** 9281–6 (2010).
63. Bodor, D. L., Rodríguez, M. G., Moreno, N. & Jansen, L. E. Analysis of protein turnover by quantitative SNAP-based pulse-chase imaging. *Curr. Protoc. Cell Biol.* Chapter **8**, Unit 8.8 (2012).

Acknowledgements

The authors would like to thank Sue Biggins for corrections of the manuscript and Oliver Bishof, Florent Hubé, Christian Jaulin, Patrick Lomonte and Claire Vourc'h for useful discussion about this work. We thank Dr. Kalitsis and Dr. Choo for the gift of the CENP-C antibody. Work in CF's lab is supported by grants from Agence Nationale pour la Recherche, Ligue Nationale Contre le Cancer and Gefluc. S.H. and G.G. are recipients of PhD fellowships from the French ministry of research and Fondation ARC.

Author Contributions

Conceptualization, C.F.; Methodology, C.F. and S.H.; Investigation, S.H.; Writing – Original Draft, S.H.; Writing – Review & Editing, C.F.; Funding Acquisition, C.F.; Resources, G.G., I.I. and G.V.; Supervision, C.F.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Hédouin, S. *et al.* CENP-A chromatin disassembly in stressed and senescent murine cells. *Sci. Rep.* **7**, 42520; doi: 10.1038/srep42520 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© The Author(s) 2017

CENP-A chromatin disassembly in stressed and senescent murine cells

Sabrine Hédouin, Giacomo Grillo, Ivana Ivkovic, Guillaume Velasco and Claire Francastel*

Université Paris Diderot, Sorbonne Paris Cité, Epigenetics and Cell Fate, CNRS UMR7216, 75205 Paris cedex, France

* Correspondance to:
Claire Francastel
UMR7216 - Epigénétique et Destin Cellulaire
Université Paris 7 Diderot - Bâtiment Lamarck - 4ème étage
Case Courrier 7042
35, rue Hélène Brion
75205 Paris Cedex 13
e-mail: claire.francastel@univ-paris-diderot.fr

FULL EXPERIMENTAL PROCEDURES

Cell lines and transfections

Mouse embryonic fibroblasts (MEFs) isolated from 12.5 days post conception (d.p.c.) embryos, NIH/3T3 cells and their CENP-A-SNAP derivatives were cultured in complete media (DMEM supplemented with 10% fetal bovine serum, 100 U/mL penicillin and 100µg/mL streptomycin, all from Life Technologies). CENP-A–SNAP was constructed by inserting a PCR-generated murine CENP-A cDNA flanked by NheI and EcoRI sites into the corresponding sites of pSNAPf vector (New England Biolabs) in frame with the SNAP tag. Mutation of the Ser-30 in alanine (S30A) was generated using the Q5® Site-Directed Mutagenesis Kit according to the manufacturer's recommendations (New England Biolabs) using the CENP-A-SNAP plasmid as a template. NIH/3T3 clones expressing CENP-A-SNAP WT and S30A were generated by stable transfection of plasmids using jetPEI (Polyplus transfection) and selection by G418 (1mg/ml; Gibco). Clones were expanded and examined by immunofluorescence and by Western blot using an anti-SNAP antibody (New England Biolabs) to identify lines expressing the CENP-A–SNAP fusion protein at levels smaller or comparable to that of the endogenous CENP-A, exhibiting the typical punctate pattern in the vicinity of chromocenters interphase cells.

The eukaryotic expression vector containing four minor satellite repeat units under the control of the CMV promoter in a pcDNA3 vector was generated as described previously (Bouzinba-Segard et al., 2006). As a control vector, we used the same vector in which the minor satellite sequences were replaced by an irrelevant sequence of the same size. Transient transfections of NIH/3T3 cells were performed using jetPEI (Polyplus transfection) according to manufacturer's recommendations, in conditions allowing a transfection efficiency of 60% as assessed by transfection with a pEGFP-N1 vector (Clontech).

Drug treatments

Cells were treated with the following conditions unless stated otherwise: 10µM Etoposide, 1mM Hydroxyurea, 10µg/ml Mitomycin C, 10µM Nocodazole, 5µg/ml Actinomycin D, 1% EtOH (all from Sigma), 200µg/ml Zeocin (Invitrogen), 50µM ATM inhibitor (KU-55933, Selleckchem) or 10µM ATR inhibitor (VE-821, Selleckchem). Duration of treatments is

indicated in figure legends. For global UVC exposure, adherent cells were washed in PBS and exposed to 20J/m² UVC (254 nm) using a low-pressure mercury lamp, then incubated in complete media and harvested 24h later. Conditions were set using a VLX-3W dosimeter (Vilbert-Lourmat). For heat shock, adherent cells were washed in PBS, overlaid with complete medium pre-warmed at 42°C and incubated at 42°C for 1 h. After heat shock, the medium was replaced with pre-warmed medium at 37°C. Cells were allowed to recover from that point on for 1 hr before harvesting for further experiments.

SNAP quench and pulse labelling

SNAP-tagged histones were pulse-labelled with 2 µM SNAP-cell TMR-star (New England Biolabs) for 15 min or quenched with 2 µM SNAP-cell Block (BTP, New England Biolabs) for 30 min. After quenching or pulse labelling, cells were washed twice with PBS and then incubated in complete medium for 30 min to allow excess compound to diffuse from cells. Then, cells were washed twice with PBS and incubated for 24h in complete medium containing or not 10µM Etoposide. Labelling of total CENP-A was then monitored by immunofluorescence as described below.

Immunofluorescence

Cells were grown on glass coverslips and treated as described in the drug treatment section followed by fixation with 4% paraformaldehyde in PBS (Electron Microscopy Sciences). After permeabilisation with 0.1% Triton X-100 in PBS, samples were blocked in 2% donkey serum (Jackson Immunoresearch) in PBS followed by incubation with primary antibodies (Table S3) in blocking buffer overnight at 4°C. After incubation with secondary antibodies conjugated to Alexa-Fluor 488 or 594 (Jackson ImmunoResearch), coverslips were mounted in Vectashield medium containing 2µg/ml DAPI (Vector Laboratories).

DNA & RNA FISH

Cells were directly grown on Superfrost Plus microscope slides (Menzel–Glaser, Braunschweig, Germany) then fixed and permeabilized as described above. For DNA-FISH, a denaturation step was added by incubating slides for 5 min in denaturation buffer (70%

formamide, 2X SCC) at 75°C. Slides for both DNA and RNA-FISH were dehydrated in successive ethanol baths (70, 90, 100%, 3 min each). Minor or major satellite oligonucleotide probes (Table S4) were labelled in 3' with fluorescent Cy3-dUTP or Alexa488-dUTP using Terminal Transferase (New England Biolabs) according to the manufacturer's recommendations. Slides were hybridized overnight at 37°C with 100 ng of labelled probe, 10 µg of yeast tRNA (Invitrogen) in Hybrisol VII buffer (MP Biomedicals). Slides were washed three times for 10 min in 2X SSC at 42°C and mounted in Vectashield medium with 2µg/ml DAPI (Vector Laboratories).

Of note, as permabilization of cells prior to fixation results in perturbations to nuclear architecture, in particular that of centromeric regions, control RNAse treatments for RNA-FISH experiments were not performed. Instead, we carefully verified that RNA-FISH signals for centromeric transcripts were undetectable prior to drug treatments. Hence, the increased signals that we detected matched PCR data and it was unlikely that they resulted from hybridization to DNA repeats. Likewise, RNA-FISH signals that spread throughout the nucleus cannot result from hybridization to centromeric DNA repeats.

Microscopy

Image acquisition was performed at room temperature on a fluorescence microscope (Axioplan 2; Zeiss) with a Plan-Neofluar 100X/1.3 NA oil immersion objective (Zeiss) using a digital cooled camera (CoolSNAP fx; Photometrics) and METAMORPH 7.04 software (Roper Scientific, Trenton, NJ). Images presented correspond to one focal plane. Analysis was performed by scoring at least 150 cells in each experiment.

siRNA and plasmid transfection

siRNA purchased from Sigma or Eurofins MWG Operon (Table S5) were transfected into cells using Interferin (Polyplus transfection) following manufacturer's instructions. Briefly, two rounds of transfection with synthetic siRNAs at a final concentration of 20 nM were performed: a first round on NIH/3T3 cells in suspension and the second when cells were allowed to adhere to the plate. Cells were collected 72 hr after the first transfection. Etoposide treatment was performed 48h post-transfection when this applies.

RNA extraction and RT-qPCR analysis

Total RNA from cell lines was isolated using TRI Reagent® (Sigma) according to manufacturer's instructions. Contaminant genomic DNA was eliminated with TURBO DNA-free kit (Ambion). Reverse transcription was carried out using 500ng DNA-free RNA and 50µM random hexamers (Roche), 20U of RNase inhibitor (New England Biolabs) and 20U of RevertAid Reverse Transcriptase (Fermentas). Complementary DNA reactions were used as templates for PCR reactions. Real-time PCR was performed using the LightCycler® DNA Master SYBR Green I mix (Roche) supplemented with 0.2µM specific primer pairs (Table S6). Real-time quantitative PCR were run on a light cycler rapid thermal system (LightCycler®480 2.0 Real time PCR system, Roche) with 20 sec of denaturation at 95°C, 20 sec of annealing at 60°C and 20 sec of extension at 72°C for all primers, and analyzed by the comparative CT($\Delta\Delta$CT) method using U6 RNA as an invariant RNA. As a control, complementary DNA reactions in which the reverse transcriptase was omitted were used to assess the absence of DNA contaminants in our templates. Primer efficiency was established by plotting a standard curve of Ct values from serial dilutions of cDNA and these were comparable for all DNA repeats. Therefore, the $\Delta\Delta$CT was applied to compare variations in the levels of a given type of repeat transcripts without correcting for primer efficiency.

Nuclear extracts and Western blot

Cells were washed and scrapped in PBS. After centrifugation, cell pellets were lysed in RIPA buffer (300 mM NaCl, 1% NP-40, 0.5% Na-deoxycholate, 0.1% SDS, 50 mM Tris pH 8.0, complete protease inhibitor cocktail) for total extracts. For nuclear extracts, cells were resuspended in buffer H (10 mM Tris pH 7.9, 10 mM KCl, 1.5 mM MgCl2, 0.5% NP40, complete protease inhibitor cocktail) 20 min on ice to lyse cellular membranes. Nuclei were pelleted by centrifugation and lysed in RIPA buffer (300 mM NaCl, 1% NP-40, 0.5% Na-deoxycholate, 0.1% SDS, 50 mM Tris pH 8.0, complete protease inhibitor cocktail).

For Western blot analysis, proteins were resolved by SDS–PAGE, transferred onto PVDF membrane (Thermo Fischer) and probed using the appropriate primary (Table S3) and secondary antibodies coupled to horse-radish peroxidase (Jackson Immunoresearch). Protein detection was performed with ECL reagents (Pierce).

Chromatin Immunoprecipitation

Cells were directly cross-linked with 1% formaldehyde for 10min at room temperature (RT) and the reaction was quenched by adding glycine at a final concentration of 0.125M for 5min at RT. Fixed cells were washed and harvested with PBS. Cell membranes were lysed in cell lysis buffer (5mM Pipes pH8.0, 85mM KCl, 0.5% NP40, Complete protease inhibitor cocktail (Roche)) for 20min, 4°C. Nuclei were pelleted by centrifugation and lysed in nuclei lysis buffer (50mM Tris HCl pH8.1, 150mM NaCl, 10mM EDTA, 1% SDS, Complete protease inhibitor cocktail (Roche)) for 20min, 4°C. Chromatin was subjected to sonication with Bioruptor Power-up (Diagenode) yielding genomic DNA fragments with a bulk size of 150-300 bp. Supernatant was diluted 10 times in IP dilution buffer (16.7 mM Tris pH8, 167 mM NaCl, 1.2 mM EDTA, 1.1% Triton X-100, 0.01% SDS, complete protease inhibitor cocktail (Roche)). IPs were performed overnight at 4°C using 20µg of chromatin and 2µg of specific antibodies against indicated proteins (Table S3). Immune complexes were recovered by adding pre-blocked protein G magnetic beads (Dynabeads, Life Technologies) and incubated for 2h at room temperature. Beads were washed once with Low salt buffer (0.1% SDS, 1% Triton, 2 mM EDTA, 20 mM Tris pH 8, 150 mM NaCl), once with High salt buffer (0.1% SDS, 1% Triton, 2 mM EDTA, 20 mM Tris pH 8; 500 mM NaCl), once with LiCl wash buffer (10 mM Tris pH 8.0, 1% Na-deoxycholate, 1% NP-40, 250 mM LiCl, 1 mM EDTA) and twice with TE (10mM Tris pH 8.0, 1mM EDTA pH 8.0, 50mM NaCl). Beads were eluted in Elution buffer (50mM Tris pH 8.0, 10mM EDTA pH 8.0, 1% SDS) at 65°C for 45min and cross-link was reversed O/N at 65°C after addition of NaCl to 0.2M final concentration. The eluted material was digested with 40µg of Proteinase K (Invitrogen) at 65°c for 2h then phenol/chloroform-extracted and ethanol-precipitated. DNA was resuspended in water and q-PCR performed using LightCycler® DNA Master SYBR Green I mix (Roche) and analyzed on a light cycler rapid thermal system (LightCycler®480 2.0 Real time PCR system, Roche). ChIP-qPCR results were normalized on input signal (% input) then normalized against the untreated condition. Sequences of primers are listed in Table S6. As a negative control, an additional mock ChIP using IgG was systematically added to each experiment to verify check that no more than 0.01% of the input was precipitated whatever the genomic region tested, including satellite repeats.

Statistical analysis

For scoring CENP-A subnuclear localization, centromere architecture or transcription, cells were classified into categories indicated in each figure. A total of 300 to 400 cells from 2 to 3 experiments were analyzed, the proportion of cells assigned to each category was calculated, and category distributions were compared through a Chi-square test.

For qPCR analysis, quantitative data are presented as means ± SEM. Values were compared with the unpaired Student's t test, with a P value <0.05 being considered statistically significant. Representative data are the result of at least three independent experiments performed on at least three independent RNA extractions or chromatin preparations.

LEGENDS TO SUPPLEMENTAL FIGURES

Supplemental Figure S1. Characterization of cells subjected to genotoxic stress.

(A) Immunoblots of p53 and γH2AX proteins from nuclear extracts of NIH/3T3 cells treated with the indicated drugs in conditions commonly used: Etoposide (ETOP; 10µM, 4 hr), Zeocin (ZEO; 200µg/ml, 4 hr), Hydroxyurea (HU; 1mM, 18 hr), Mitomycin C (MMC; 10µg/ml, 4 hr treatment plus 24 hr recovery), UVC rays (20J, 24 hr after exposure) and Nocodazole (NOC; 10µM, 16 hr). Ponceau red staining served as a loading control. (B) FACS analysis of NIH/3T3 cells repartition in the cell-cycle after treatment with the indicated drugs and at the indicated time points.

Supplemental Figure S2. CENP-A is specifically delocalized in response to various genotoxic stress.

(A) Immunofluorescence analysis of CENP-A delocalization in response to different DNA damaging agents in conditions indicated in S1A. Percentage of counted cells that displayed CENP-A delocalization are indicated for each condition (n>300 cells). (B-C) Immunofluorescence analysis of (B) CENP-B and (C) CENP-C after a 24 hr ETOP treatment. (D) Immunoblots of p53, phosphorylated H2A.X (γH2AX) and cleaved Caspase-3 from total extracts of untreated NIH/3T3 cells (CTRL) or treated with ETOP at the indicated time points or following a 24 hr recovery period (4+24H). Ponceau red staining served as a loading control.

Supplemental Figure S3. CENP-A delocalization correlates with persistence of stress signalling.

(A) Percentage of counted cells that displayed CENP-A delocalization as in S2A are shown for the indicated drugs, after a 24 hr treatment, or after a 4 hr treatment followed by a 24 hr recovery period. (n>300 cells). P-values (unpaired t-test) comparing control (CTRL) and drug treatments: *** $p<0.001$; ** $p<0.01$. (B) Immunoblots of γH2AX from total extracts of untreated NIH/3T3 cells (CTRL) or treated with the indicated drug at the indicated time points or following a 24 hr recovery period (4+24). Ponceau red staining served as a loading control. (C) FACS analysis of NIH/3T3 cells repartition in the cell-cycle following a 24 hr ETOP

treatment in the absence of or concomitant with inhibition of DDR signalling using inhibitors of ATM (ATMi) or ATR (ATRi).

Figure S4. Transcriptional analysis of centromeric and other DNA repeats in stress conditions.

(A) The histogram represents the results of RT-qPCR analysis of minor satellite repeats expression in NIH/3T3 cells treated with the indicated drugs in conditions used in S1A. Additional conditions: heat shock (HS; 1hr at 42°C followed by 1h at 37°C), and Ethanol (EtOH; 1%, 8 hr). Below is a western blot analysis of the induction of the DNA damage repair pathway in nuclear extracts from NIH/3T3 following the various drug treatments indicated in the upper graph, characterized by p53 and $p21^{CIP}$ stabilization and phosphorylation of H2AX (γH2AX) in stress conditions generating DNA strand breaks. Ponceau red served as a loading control (B) qRT-PCR analysis of expression of murine DNA repeats during the kinetics of ETOP treatment. Ribosomal DNA (rDNA); interspersed repeats: long (LINE) and short (SINE) interspersed nuclear elements; transposable elements: intracisternal A-particle (IAP); telomeric repeats from juxtacentromeric (TeloCen) or long arm of chromosome 11 (11q) positions. Histograms represent relative fold change values normalized to U6 RNA levels and compared to untreated cells (CTRL). Error bars represent S.E.M of at least three independent experiments. P-values (unpaired t-test) comparing control (CTRL) and drug treatments: *** $p<0.001$; ** $p<0.01$; *$p<0.05$.

Supplemental Figure S5. Transcription is required but centromeric transcripts are not sufficient to promote CENP-A delocalization.

(A) qRT-PCR analysis of minor satellites expression and (B) Immunofluorescence analysis of CENP-A delocalization in NIH/3T3 cells treated for 24 hr with Etoposide (ETOP, 10μM) in presence or not of the RNA polymerase inhibitor Actinomycin D (Act D; 5μg/ml). Error bars represent S.E.M of at least three independent experiments. (C) qRT-PCR analysis of minor satellites ectopic expression 24 hr after transfection of NIH/3T3 cells with a plasmid containing four minor satellite repeats under the control of the CMV promoter (MIN SAT) or a control plasmid transcribing a fragment of irrelevant RNA of the same size (CTRL). (D) Immunofluorescence analysis of CENP-A delocalization 24 hr after transfection of NIH/3T3

cells with the minor satellite (MIN SAT) or control (CTRL) expression vectors. P-values (unpaired t-test) comparing control (CTRL) and drug treatments: ** p<0.01; *p<0.05.

Supplemental Figure S6. The DDR effector p53, but not ATM/ATR, are required for transcriptional activation of minor satellite repeats in response to genotoxic stress.

(A) qRT-PCR analysis of minor satellites expression in NIH/3T3 cells with Etoposide (10μM) for 24 hr in presence or not of ATM (ATMi) or ATR (ATRi) inhibitors. P-values (unpaired t-test) comparing controls (CTRL) and drug treatments: ** p<0.01. (B) RT-qPCR analysis of $p21^{CIP}$ and minor satellite transcripts after treatment with Etoposide at the indicated concentration for 4 hr in $p53^{+/+}$ (NIH/3T3) or $p53^{-/-}$ fibroblasts, and immunoblots of p53 and γH2AX proteins from nuclear extracts of cells treated as above. Ponceau red staining served as a loading control. P-values (unpaired t-test) comparing controls (CTRL) and drug treatments: *** p<0.001; ** p<0.01. (C) Analysis of $p21^{CIP}$ and minor satellite transcripts after treatment of NIH/3T3 cells with Etoposide (E), Nutlin-3 (N3) or both (E+N3), for 4 hr, using RT-qPCR. Immunoblot of p53 and γH2AX from nuclear extracts of cells treated as above. Ponceau red staining served as a loading control. P-values (unpaired t-test) comparing controls (CTRL) and drug treatments: *** p<0.001; ** p<0.01. (D) FACS analysis of $p53^{-/-}$ cells repartition in the cell-cycle after ETOP treatment at the indicated time points compared to untreated cells (CTRL). Error bars represent S.E.M of at least three independent experiments.

Supplemental Figure S7. CENP-A delocalization is dependent on chromatin remodelers.

qRT-PCR analysis of siRNA-mediated knockdown efficiency for each chromatin remodelling factors in untreated (CTRL) and cells treated with etoposide for 24 hr (ETOP). Values represented are relative fold change compared to untreated (CTRL) control siRNA (siCTRL) and normalized against U6 RNA levels. Error bars represent S.E.M of at least three independent experiments.

Supplemental Figure S8. Increased transcription of minor satellite transcripts in stressed or senescent MEFS

(A) qRT-PCR analysis of minor satellites expression and the senescence marker $p16^{Ink4}$ in primary MEFs treated for the indicated time with Etoposide (ETOP, 2μM). Values represented are relative fold change compared to untreated (CTRL) cells and normalized against U6 RNA levels. Error bars represent S.E.M of at least three independent experiments. P-values (unpaired t-test) comparing controls (CTRL) and drug treatments: * $p<0.05$ (unpaired t-test). (B) Percentage of primary MEFs treated as in A that exhibit CENP-A delocalization. P-values (Chi-square test) comparing controls (CTRL) and drug treatments: *** $p<0.001$. (C) Senescence-associated β-Galactosidase staining in primary mouse embryonic fibroblasts (MEFs) at passage 2 and 6. Senescent cells are stained in blue. (D) Western blot analysis of markers of senescence: p53, $p21^{CIP}$ and $p16^{INK4}$ in primary MEFs at passage 2 (P2) and 6 (p6). Levels of γ-tubulin served as a loading control. (E) Western blot analysis of markers of senescence p53, $p21^{CIP}$ and $p16^{INK4}$ after treatment of primary MEFs at p2 with 10μM Nutlin-3 for the indicated time points. Levels of α-tubulin served as a loading control.

Supplemental Figure S9. Transcriptional, structural and epigenetic disorganization of centromeric repeats in senescent MEFs

(A) Analysis of minor satellites transcription in senescent MEF cells at passage 6 (p6), using RT-qPCR. Data are normalized to U6 RNA levels and compared to primary MEFs at passage 2 (p2). Error bars represent S.E.M of at least three independent experiments. P-values (unpaired t-test) comparing MEFp2 and p6: * $p<0.05$ (unpaired t-test). (B) Analysis of minor satellites transcription in senescent MEF after treatment with Nutlin-3 (10μM, 72hr), using RT-qPCR. Data are normalized to U6 RNA levels and compared to untreated primary MEFs at passage 2 (p2). Error bars represent S.E.M of at least three independent experiments. P-values (unpaired t-test) comparing untreated and treated MEFs: * $p<0.05$ (unpaired t-test). (C) Analysis of CENP-A localization in senescent MEFs at passage 6 compared to primary cells at passage 2 using anti-CENP-A antibodies and FITC-conjugated secondary antibodies. Insets are magnified an additional 3 fold. Scale bar, 10μm. The percentage of cells showing a fragmented pattern (as in C; middle panel) or both g-fragmented pattern and a CENP-A nucleolar staining (as in C; lower panel) are represented on the histogram ($n>200$ cells). P-

values (Chi-square test) comparing MEFp2 and p6: *** p<0.001. (D) Disorganization of minor and major satellites nuclear domains in senescent MEFs (p6) compared to primary MEFs (p2) as revealed by DNA-FISH using specific directly labelled probes. Scale bar, 10µm. The percentage of cells showing stretched signals for minor and major satellite repeats are represented on the histogram (n>200 cells). P-values (Chi-square test) comparing MEFp2 and p6: *** p<0.001.

SUPPLEMENTAL TABLES

Supplemental Table S1: List of drugs used in the study, the type of DNA lesions that they promote and their impact on cell-cycle

DRUG	DRUG abbreviation	TYPE OF DNA LESION	PHASE OF EXPECTED CELL CYCLE ARREST
Etoposide	**ETOP**	Double Strand Breaks	G2/M
Zeocin	**ZEO**	Single Strand Breaks Double Strand Breaks	G2/M
Hydroxyurea	**HU**	Replication fork stalling	G1/S
Mitomycin C	**MMC**	DNA crosslink	S, G2/M
Ultraviolet C	**UVC**	Pyrimidine dimers	G1/S
Nocodazole	**NOC**	Inhibition of spindle microtubule formation	G2/M

Supplemental Table S2: List of chromatin remodelers and histone chaperones tested in loss of function experiments

NAME	ASSOCIATED COMPLEX	(PERI)CENTROMERIC LOCALIZATION	REFERENCES
CHD3	NuRD (Mi-2)	Pericentromeres	Helbling Chadwick et al., 2009 Sims and Wade, 2011
CHD4			
BRG1	SWI/SNF	Pericentromeres, centromeres	Nielsen et al., 2002 Gkikopoulos et al., 2011 Bourgo et al., 2009
INO80	INO80	Pericentromeres, centromeres	Ogiwara et al., 2007 Chambers et al., 2012
ATRX	DAXX/ATRX	Pericentromeres, centromeres	Pluta et al., 1998 McDowell et al., 1999 Morozov et al., 2012 Lacoste et al., 2014
DAXX			
HELLS/LSH		Pericentromeres, centromeres	Dennis et al., 2001 Yan et al., 2003
SSRP1	FACT	Pericentromeres, centromeres	Foltz et al., 2006 Ishov et al., 2004 Lejeune et al., 2007 Okada et al., 2009

Supplemental Table S3. Primary Antibodies used in immunofluoscence and western blots.

TARGET	HOST	SUPPLIER/REFERENCE	APPLICATION
B23 (Nucleophosmin)	mouse	Santa Cruz Biotechnology (sc56622)	IF
CENP-A	rabbit	Abcam (ab33565)	ChIP, WB
CENP-A	rabbit	Cell Signaling Technology (2048)	IF
CENP-B	rabbit	Santa Cruz Biotechnology (sc22788)	IF
CENP-C	rabbit	Gift from P. Kalitsis	IF
γH2AX (pS139)	rabbit	Epitomics (2212-1)	WB
Histone H2A	rabbit	Cell Signaling Technology (12349)	ChIP
Histone H3	rabbit	Abcam (ab1791)	ChIP
Histone H4	rabbit	Abcam (ab10158)	ChIP
p53	rabbit	Santa Cruz Biotechnology (sc-6243)	WB
p21	mouse	Santa Cruz Biotechnology (sc-6246)	WB
p16	rabbit	Santa Cruz Biotechnology (sc-1207)	WB
γ-tubulin	mouse	Sigma (T6557)	WB
SSRP1	mouse	Biolegend (609701)	WB
Cleaved Caspase-3	rabbit	Cell Signaling Technology (9661)	WB, IF

Supplemental Table S4: List of oligonucleotide probes used in DNA or RNA-FISH assays.

NAME	SEQUENCE (5' to 3')	ASSAY
Minor satellites probe	ATGTTTCACATTGTAACTCATTGATATACACTGTTCTACAA ATCCCGTTTCCAACGAATGTGTTTT	RNA-FISH, DNA-FISH
Major satellites probe	AAATATGGCGAGGAAAACTGAAAAAGGTGGAAAATTTA GAAATGTCCACTGTAGGACGTG	DNA-FISH

Supplemental Table S5. Sequences of siRNA used in loss of function experiments.

siRNA	Target sequence (5’ -> 3’)
siATRX	GAUGCUAGAUCAUCAGUAA
siBRG1	CAAACUGGGCGUAUGAAUU
siCHD3	CCGCCGCAGGCGCCAAUUA
siCHD4	GUCUAAGUUCUUUCGAGUU
siCtrl	GCCGGUAUGCCGGUUAAGU
siDAXX	GUAACUCCGGUAGUAGGAA
siINO80	GAAACUUAUUUCUUACCAA
siLSH	GAGAAGUUGAUAAUCCAUA
siSSRP1	CAGUGUACCACAGGCAAGA

Supplemental Table S6. Sequences of primers used for RT-qPCR analysis.

(Primers specific for repeat families: # primers from Bouzinba-Segard et al. 2006; * primers from Martens et al. 2005; § primers from Deng et al. 2012)

PRIMER NAME	SEQUENCE 5'-> 3'	APPLICATION
Minor satellites F #	GAACATATTAGATGAGTGAGTTAC	RTqPCR, ChIP
Minor satellites R #	GTTCTACAAATCCCGTTTCCAAC	
Major satellites F *	GACGACTTGAAAAATGACGAAATC	RTqPCR, ChIP
Major satellites R *	CATATTCCAGGTCCTTCAGTGTGC	
rDNA F (intergenic spacer) *	CCTGTGAATTCTCTGAACTC	RTqPCR, ChIP
rDNA R (intergenic spacer) *	CCTAAACTGCTGACAGGGTG	
LINE-1 family F (ORF2) *	TTTGGGACACAATGAAAGCA	RTqPCR
LINE-1 family R (ORF2) *	CTGCCGTCTACTCCTCTTGG	
SINE B1 family F *	GTGGCGCACGCCTTTAATC	RTqPCR
SINE B1 family R *	GACAGGGTTTCTCTGTGTAG	
SINE B2 family F *	GAGATGGCTCAGTGGTTAAG	RTqPCR
SINE B2 family R*	CTGTCTTCAGACACTCCAG	
IAP family F (GAG) *	AGCAGGTGAAGCCACTG	RTqPCR
IAP family R (GAG) *	CTTGCCACACTTAGAGC	
TERRA TeloCen F §	CCAAAGTTTCTGCAAGGCAAA	RTqPCR
TERRA TeloCen R §	CCCAATCTGTTGGTGGTCTTTT	
TERRA 11q F §	TGCCATTGGAACACAGCAA	RTqPCR
TERRA 11q R §	CGTCTGCTGAGGTCCACAGA	
CHD3 F	AGCCCGCCTTCAGATGTC	RTqPCR
CHD3 R	GGAGGTGTGGCGTAAGGAC	
CHD4 F	GGCCATCTTGGGTTCTATTG	RTqPCR

CHD4 R	AACAGCTTGTGCTGGAGTGA	
BRG1 F	GACCAGCCCCTCCAAATTAC	RTqPCR
BRG1 R	CTTCTGTGGGGTGCTTGTG	
INO80 F	TTTCTTACCAATAGCCGAAAGC	RTqPCR
INO80 R	TCGAGGCAGAGTCTCCTTACA	
ATRX F	AGCCGTGACTCAGATGGAAT	RTqPCR
ATRX R	CGCAGGATGCATTTCTTACA	
LSH F	ACACGAGCTGGTGGTTTAGG	RTqPCR
LSH R	GCCTGGAGATCAGACTGAGG	
DAXX F	GTGGATGGAGGGAGCAGTAA	RTqPCR
DAXX R	AGTTTGTGGAGGAACGGAAC	
SSRP1 F	TGGGCTTAAACTGCTCACAA	RTqPCR
SSRP1 R	TCCATTAGCTCAAGGCGATA	
U6 F	CTCGCTTCGGCAGCACA	RTqPCR
U6 R	AACGCTTCACGAATTTGCGT	

SUPPLEMENTAL REFERENCES

Bourgo, R.J. et al. SWI/SNF deficiency results in aberrant chromatin organization, mitotic failure, and diminished proliferative capacity. *Mol. Biol.* Cell **20**, 3192–3199 (2009).

Bouzinba-Segard, H., Guais, A., & Francastel, C. Accumulation of small murine minor satellite transcripts leads to impaired centromeric architecture and function. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8709–8714 (2006).

Chambers, A.L. et al. The INO80 chromatin remodeling complex prevents polyploidy and maintains normal chromatin structure at centromeres. *Genes Dev.* **26**, 2590–2603 (2012).

Deng, Z. et al. Formation of telomeric repeat-containing RNA (TERRA) foci in highly proliferating mouse cerebellar neuronal progenitors and medulloblastoma. *J Cell Sci.* 125, 4383-4394 (2012).

Dennis, K., Fan, T., Geiman, T., Yan, Q., and Muegge, K. Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes Dev.* **15**, 2940–2944 (2001).

Foltz, D.R. et al. The human CENP-A centromeric nucleosome-associated complex. *Nat. Cell Biol.* **8**, 458–469 (2006).

Gkikopoulos, T. et al. The SWI/SNF complex acts to constrain distribution of the centromeric histone variant Cse4. *EMBO J.* **30**, 1919–1927 (2011).

Helbling Chadwick, L., Chadwick, B.P., Jaye, D.L., and Wade, P.A. The Mi-2/NuRD complex associates with pericentromeric heterochromatin during S phase in rapidly proliferating lymphoid cells. *Chromosoma* **118**, 445–457 (2009).

Ishov, A.M., Vladimirova, O.V., and Maul, G.G. Heterochromatin and ND10 are cell-cycle regulated and phosphorylation-dependent alternate nuclear sites of the transcription repressor Daxx and SWI/SNF protein ATRX. *J. Cell Sci.* **117**, 3807–3820 (2004).

Lacoste, N. et al. Mislocalization of the Centromeric Histone Variant CenH3/CENP-A in Human Cells Depends on the Chaperone DAXX. Mol. *Cell* **53**, 631–644 (2014).

Lejeune, E. et al. The Chromatin-Remodeling Factor FACT Contributes to Centromeric Heterochromatin Independently of RNAi. *Curr. Biol.* **17**, 1219–1224 (2007).

Martens, JH. et al. The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J.* **24**,800-812 (2005).

McDowell, T.L. et al. Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes. Proc. Natl. Acad. Sci. *96*, 13983–13988 (1999).

Morozov, V.M., Gavrilova, E.V., Ogryzko, V.V., and Ishov, A.M. Dualistic function of Daxx at centromeric and pericentromeric heterochromatin in normal and stress conditions. *Nucleus* **3**, 276–285 (2012).

Nielsen, A.L. et al. Selective interaction between the chromatin-remodeling factor BRG1 and the heterochromatin-associated protein HP1α. *EMBO J.* **21**, 5797–5806 (2002).

Ogiwara, H., Enomoto, T., and Seki, M. The INO80 Chromatin Remodeling Complex Functions in Sister Chromatid Cohesion. *Cell Cycle* **6**, 1090–1095. (2007).

Okada, M., Okawa, K., Isobe, T., and Fukagawa, T. CENP-H–containing complex facilitates centromere deposition of CENP-A in cooperation with FACT and CHD1. *Mol. Biol. Cell* **20**, 3986–3995 (2009).

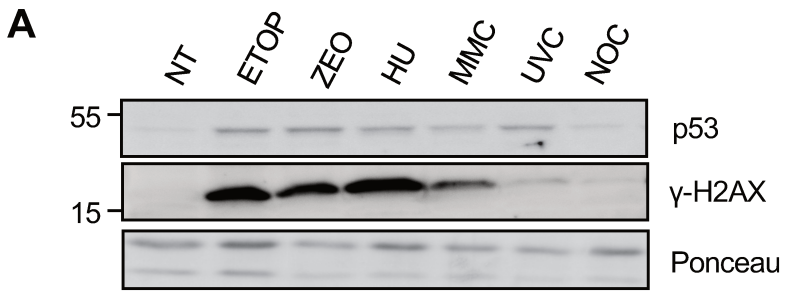
Pluta, A.F., Earnshaw, W.C., and Goldberg, I.G. Interphase-specific association of intrinsic centromere protein CENP-C with HDaxx, a death domain-binding protein implicated in Fas-mediated cell death. *J. Cell Sci.* **111**, 2029–2041 (1998).

Sims, J.K., and Wade, P.A. Mi-2/NuRD complex function is required for normal S phase progression and assembly of pericentric heterochromatin. *Mol. Biol. Cell* **22**, 3094–3102 (2011).

Yan, Q., Cho, E., Lockett, S., and Muegge, K. Association of Lsh, a Regulator of DNA Methylation, with Pericentromeric Heterochromatin Is Dependent on Intact Heterochromatin. *Mol. Cell. Biol.* **23**, 8416–8428 (2003).

Figure S1

Hedouin et al.



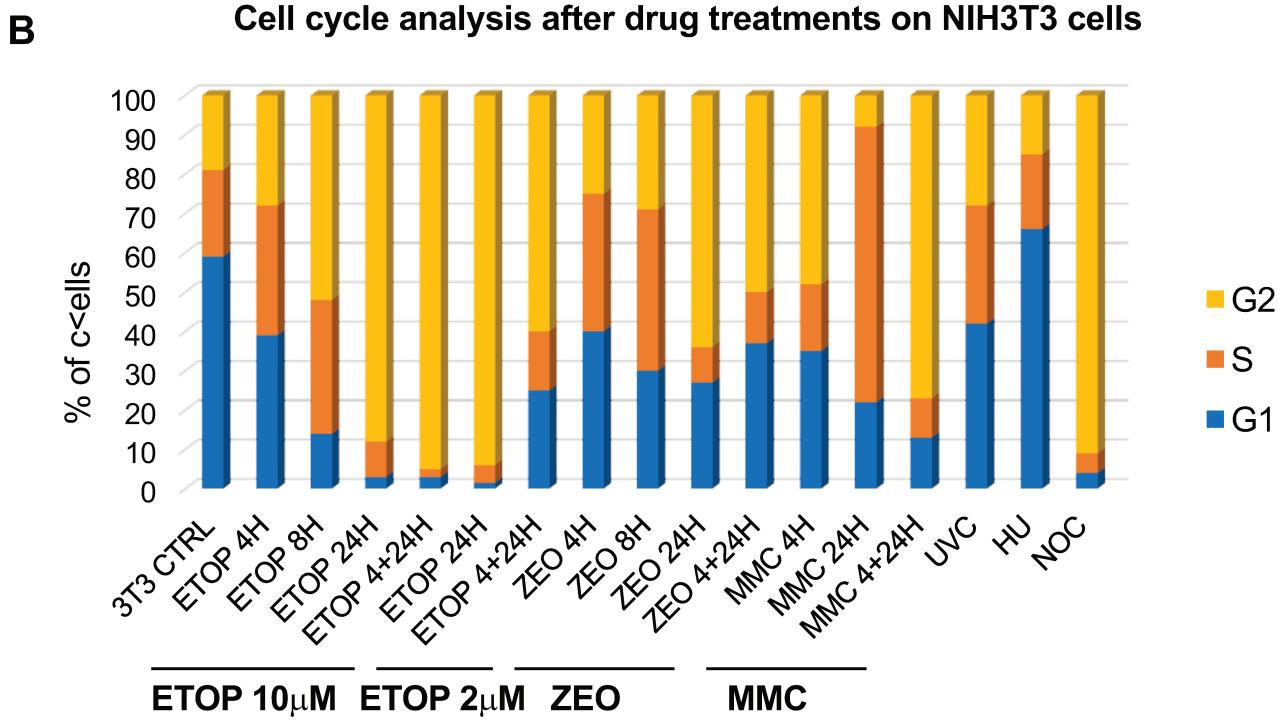
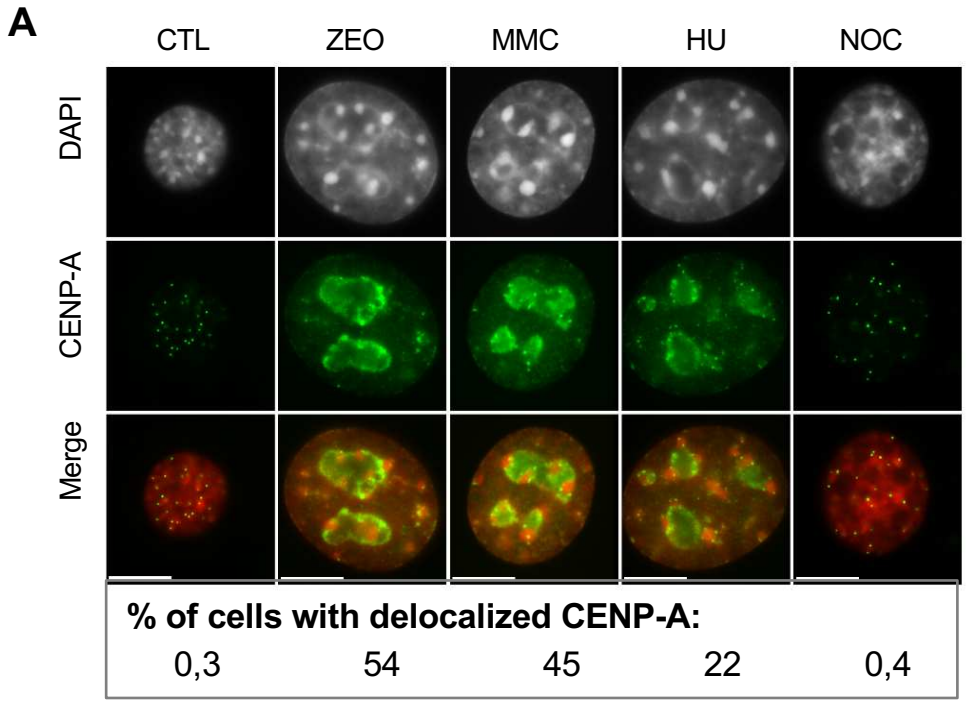
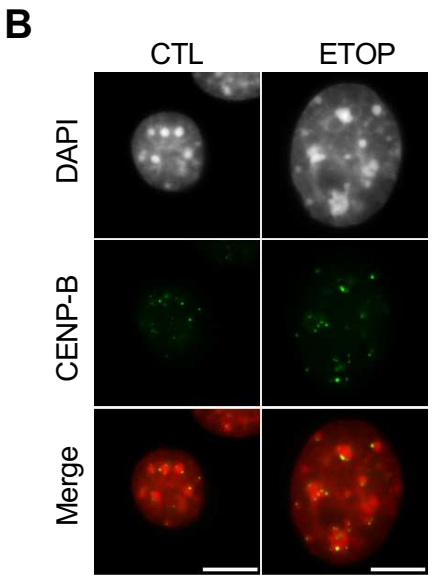
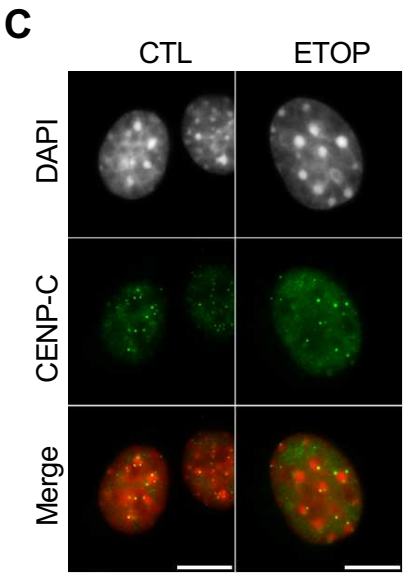


Figure S2

Hedouin et al.







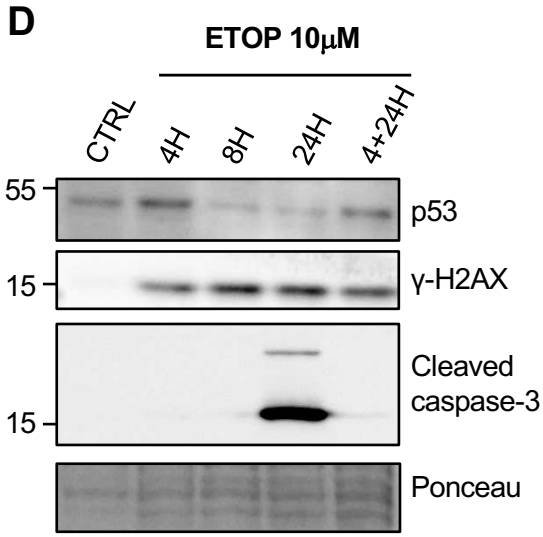


Figure S3

Hedouin et al.

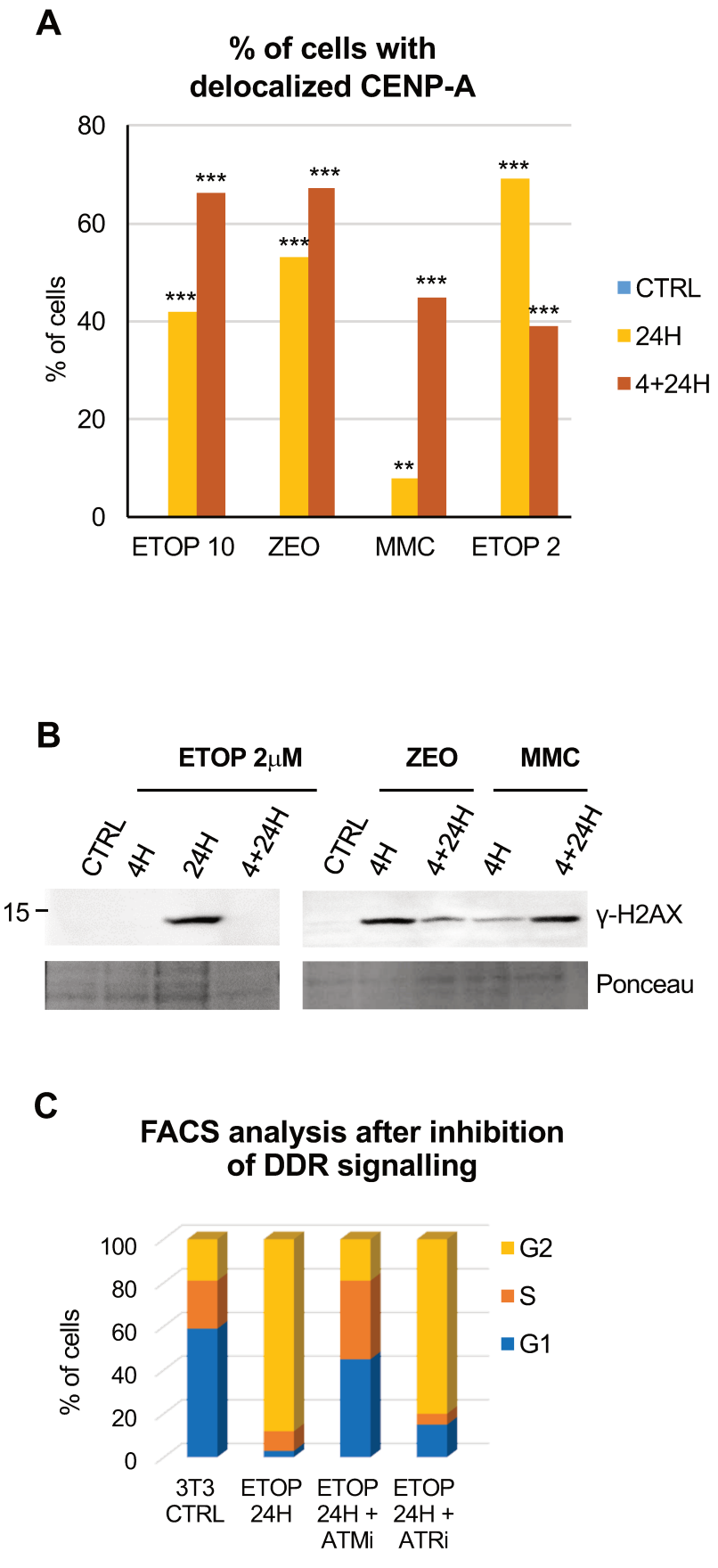
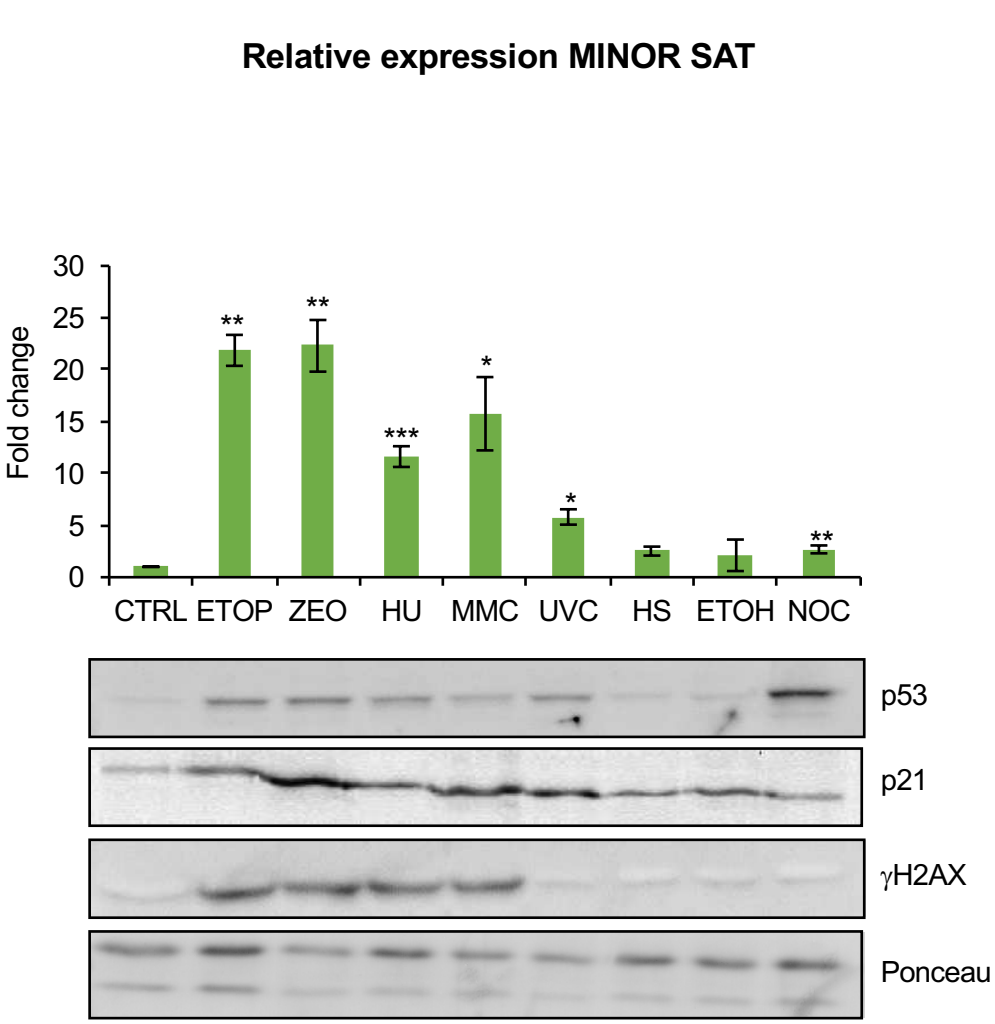


Figure S4

Hedouin et al.

A



B

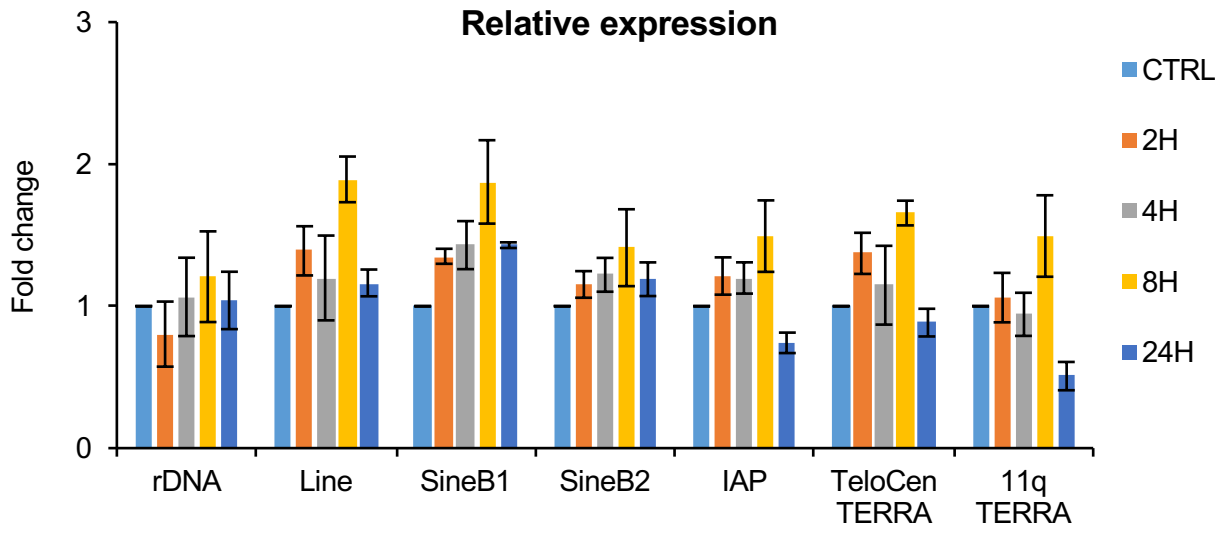
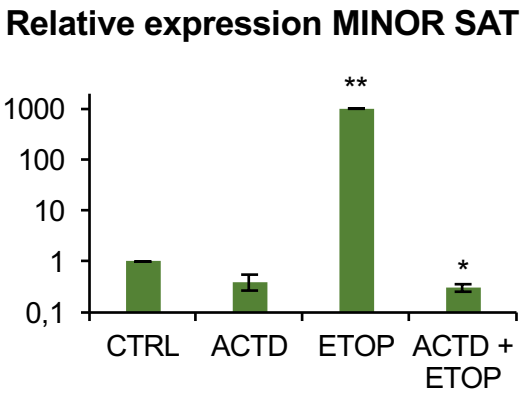


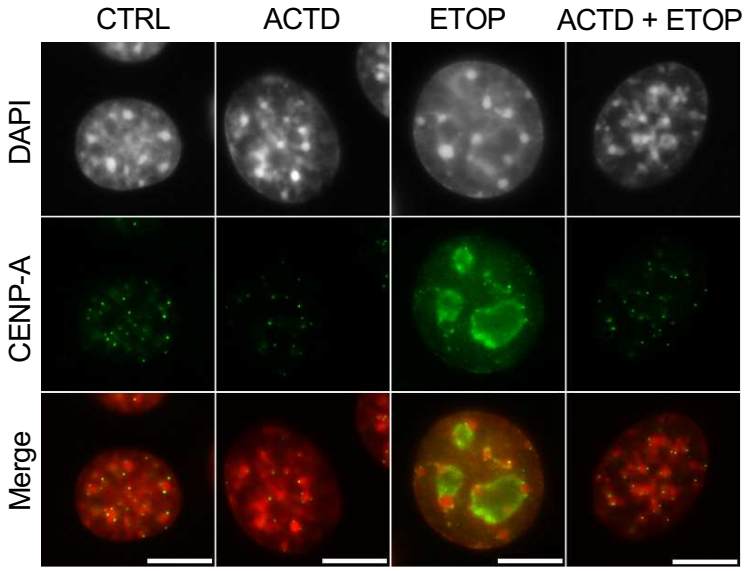
Figure S5

Hedouin et al.

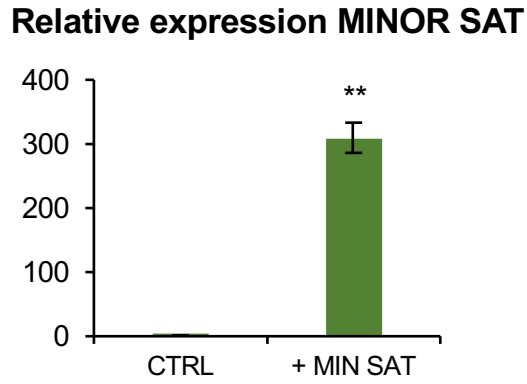
A



B



C



D

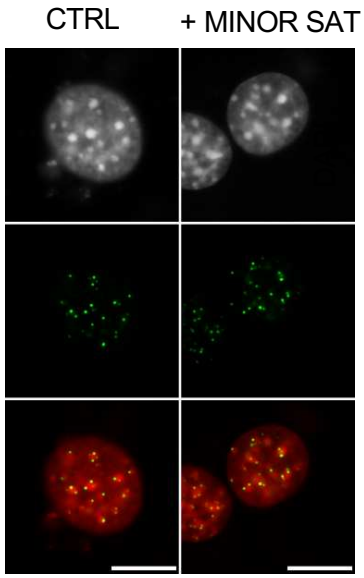


Figure S6

Hedouin et al.

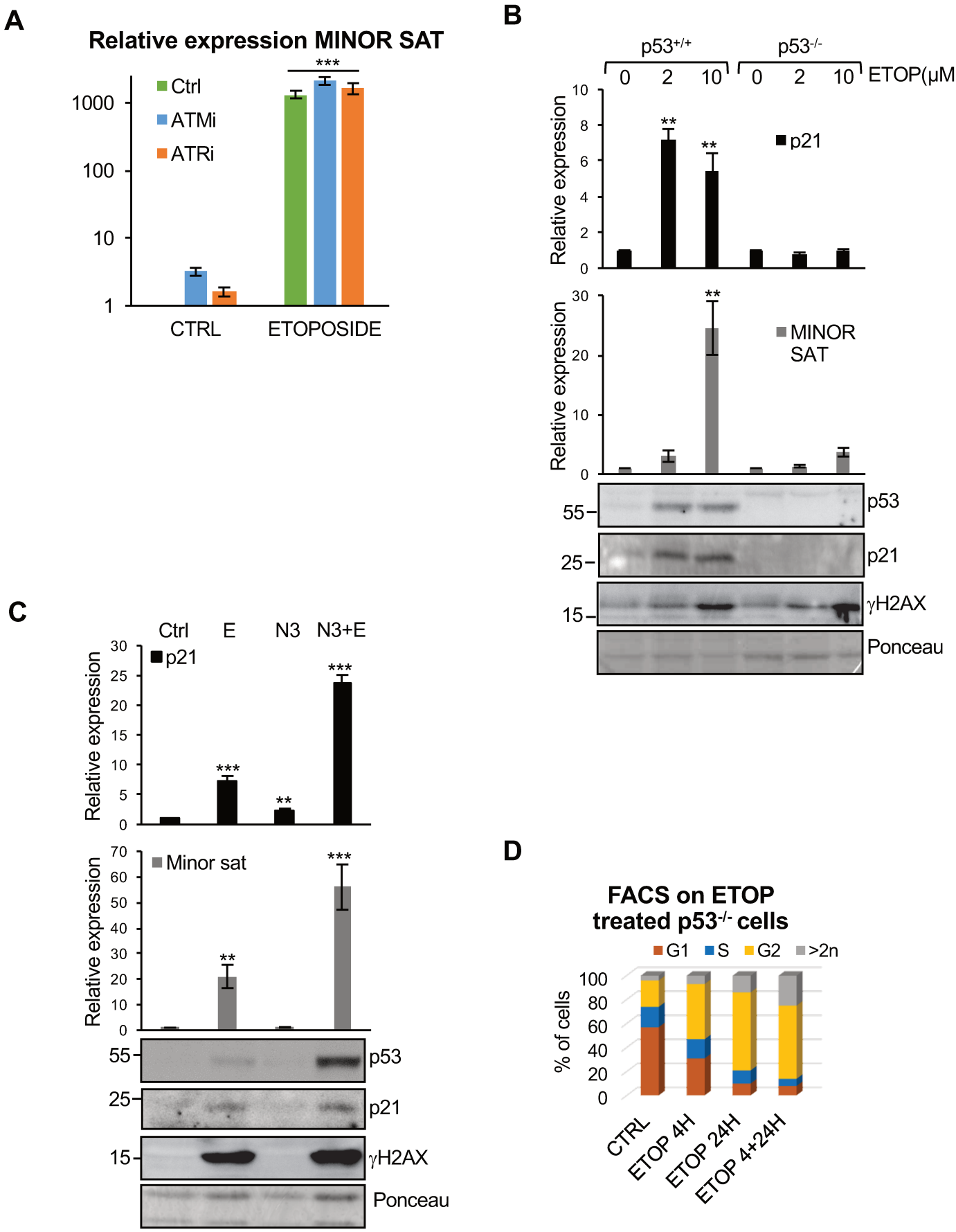


Figure S7

Hedouin et al.

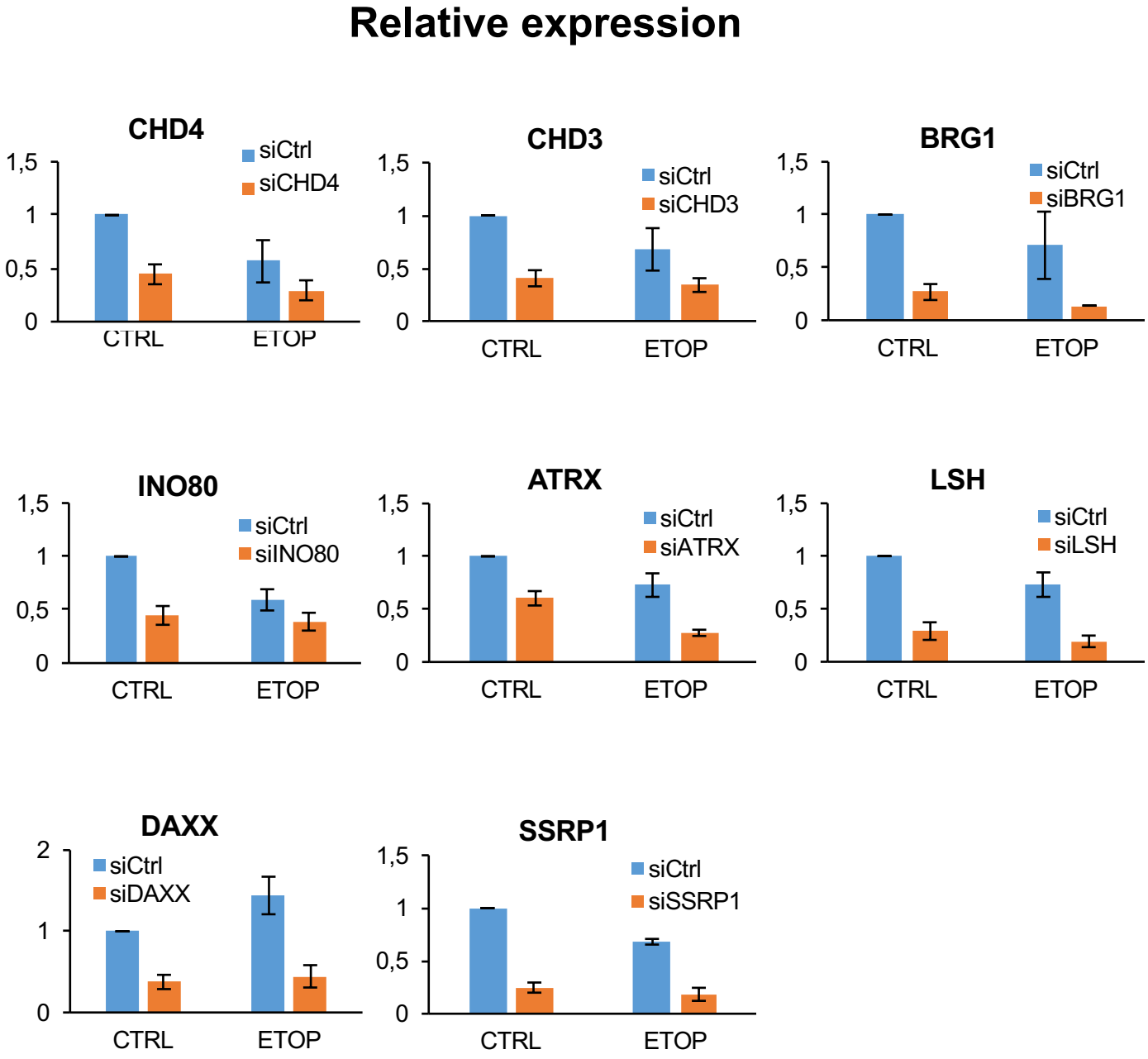
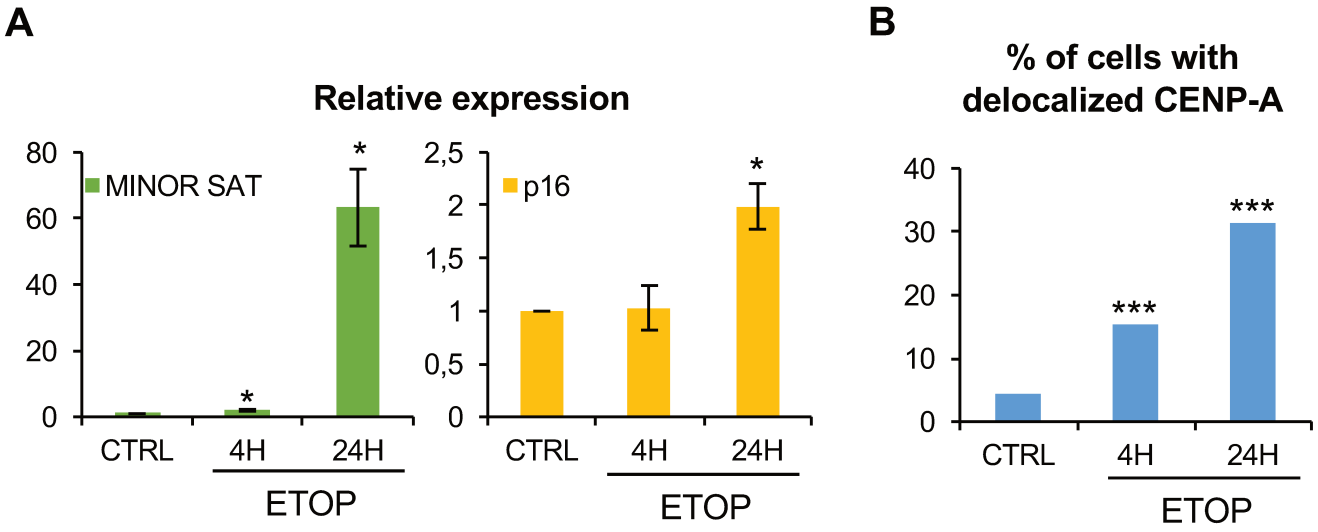
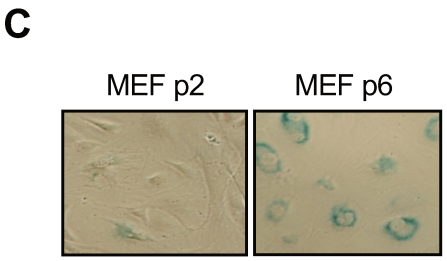
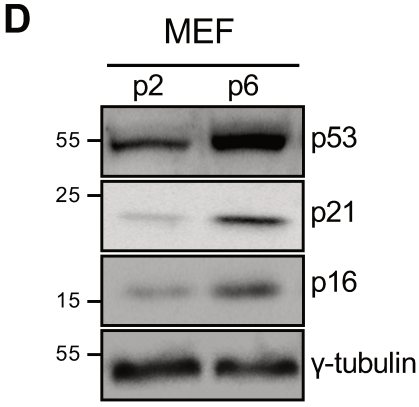


Figure S8


A
Relative expression
MINOR SAT
p16
CTRL
4H
24H
ETOP
B
% of cells with delocalized CENP-A
CTRL
4H
24H
ETOP


C
MEF p2
MEF p6


D
MEF
p2
p6
p53
p21
p16
γ-tubulin

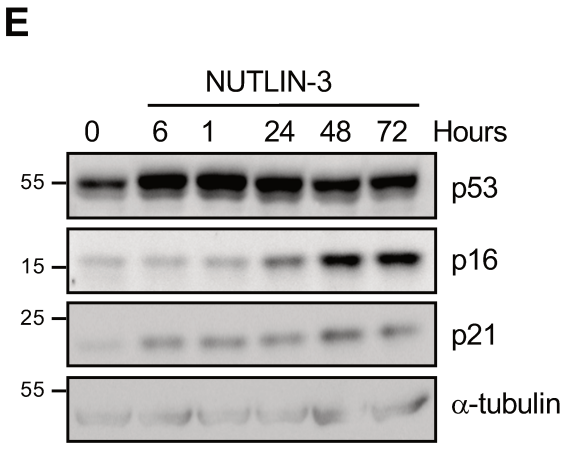
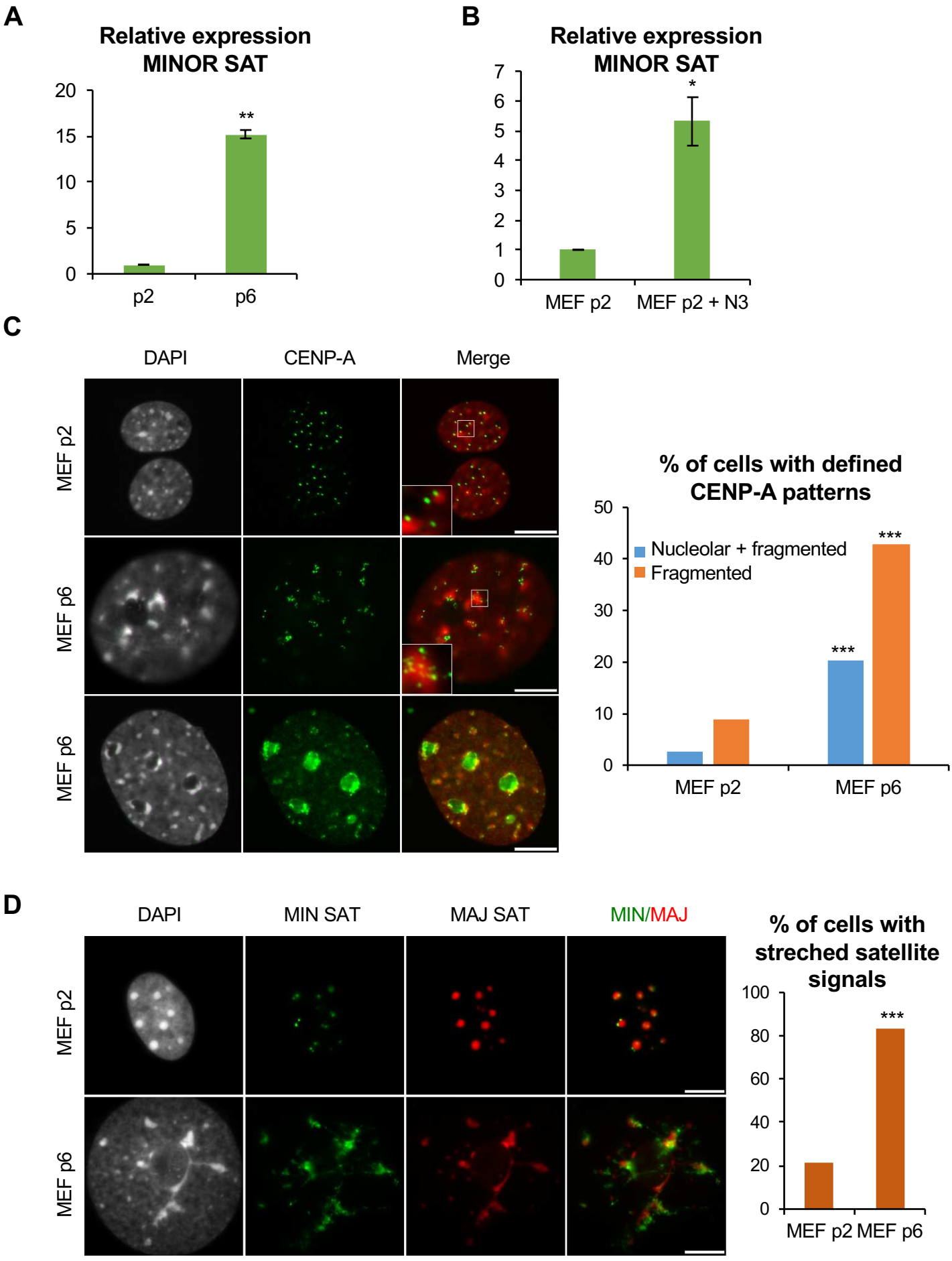

E
NUTLIN-3
0
6
1
24
48
72
Hours
p53
p16
p21
α-tubulin

Figure S9

Hedouin et al.



ANNEXE 2

Members of the DNA-binding protein with C2H2 zinc fingers and a BTB/POZ domain include ZBTB24, which has been the focus of my PhD and binds DNA independently of the methylation status of the underlying sequence. In contrast, ZBTB4, identified in the team of PA Defossez, has preferential affinity for methylated DNA and was shown to repress transcription through the recruitment of histone deacetylases (Filion et al., 2006). Ectopically expressed ZBTB4 localises at the chromocenters in interphase nuclei of wild-type cells but not in TKO, suggesting that ZBTB4 could be a reader of DNA methylation (Filion et al., 2006). ZBTB4 also binds certain non-methylated sequences but with a lower affinity compared to methylated DNA (Buck-Koehntop and Defossez, 2013).

In collaboration with the Defossez's team I participated in the characterisation of ZBTB4 function. First, they observed that a decreased expression of ZBTB4 correlated with increased genomic instability across many frequent human cancers. In addition, depletion of ZBTB4 in human cell lines led to increased formation of micronuclei and binucleated cells that was associated with aberrant expression of mitotic checkpoint proteins and an increased frequency of lagging chromosomes during mitosis. The derivation of the ZBTB4 knock-out (Z4KO) mice led to the observation that Z4KO mice were smaller compared to WT and heterozygous littermates. Primary cells, isolated from Z4KO mice, were characterised by genomic instability, increased mitotic defects and increased aneuploidy. Moreover, Z4KO mice were also more susceptible to 7,12-dimethylbenz(a)anthracene/12-O-tetradecanoylphorbol-13-acetate (DMBA/TPA)–induced skin carcinogenesis, suggesting that ZBTB4 could represent a tumour-suppressor gene and an important epigenetic regulator involved in maintenance of genome stability in mammals.

I contributed to this study performing experiments on ZBTB4 mutant fibroblasts. I also checked the methylation status of minor and major satellite repeats in in these cells and I could not see any difference compared to WT MEF. I also participated to the revision of the manuscript.

This work has been published in January 2017, in *Cancer Research*.

Published OnlineFirst November 4, 2016; DOI: 10.1158/0008-5472.CAN-16-1181

Molecular and Cellular Pathobiology

Cancer Research

Loss of the Methyl-CpG–Binding Protein ZBTB4 Alters Mitotic Checkpoint, Increases Aneuploidy, and Promotes Tumorigenesis

Audrey Roussel-Gervais[1], Ikrame Naciri[1], Olivier Kirsh[1], Laetitia Kasprzyk[1], Guillaume Velasco[1], Giacomo Grillo[1], Pierre Dubus[2], and Pierre-Antoine Defossez[1]

Abstract

Chromosome segregation during mitosis is monitored by the mitotic checkpoint and is dependent upon DNA methylation. ZBTB4 is a mammalian epigenetic regulator with high affinity for methylated CpGs that localizes at pericentromeric heterochromatin and is frequently downregulated in cancer. Here, we report that decreased ZBTB4 expression correlates with high genome instability across many frequent human cancers. In human cell lines, ZBTB4 depletion was sufficient to increase the prevalence of micronuclei and binucleated cells in parallel with aberrant mitotic checkpoint gene expression, a weakened mitotic checkpoint, and an increased frequency of lagging chromosomes during mitosis. To extend these findings, we generated Zbtb4-deficient mice. Zbtb4$^{-/-}$ mice were smaller than their wild-type littermates. Primary cells isolated from Zbtb4$^{-/-}$ mice exhibited diminished mitotic checkpoint activity, increased mitotic defects, aneuploid cells marked by a specific transcriptional signature, and increased genomic instability. Zbtb4$^{-/-}$ mice were also more susceptible to 7,12-dimethylbenz(*a*)anthracene/12-*O*-tetradecanoylphorbol-13-acetate (DMBA/TPA)–induced skin carcinogenesis. Our results establish the epigenetic regulator ZBTB4 as an essential component in maintaining genomic stability in mammals. *Cancer Res; 77(1); 62–73. ©2016 AACR.*

Introduction

More than a century ago, Theodor Boveri speculated that cancer is caused by chromosome gains or losses; this prediction has provided a very useful basis for further research (1).

Indeed, aneuploidy is observed in about 90% of solid tumors (2), and one of its causes is elevated levels of chromosome missegregation during mitosis (3). Cells have evolved different mechanisms to prevent missegregation, including the mitotic checkpoint [or spindle assembly checkpoint (SAC)], which promotes faithful sister chromatid segregation; it functions by delaying anaphase until all kinetochores have been correctly attached to the spindle microtubules (4). Among its key actors are the proteins MAD1, MAD2 (also called MAD2L1), BUB1, and BUB1B/BUBR1, which block the activity of the anaphase-promoting complex/cyclosome (APC/C) by sequestering its co-activator, CDC20. When CDC20 is released from this inhibition, it activates the ubiquitinating activity of APC/C, triggering the degradation of cyclin B and securin and therefore sister chromatid separation and anaphase (5).

Complete inactivation of the mitotic checkpoint is lethal in mammals (6), but a weakened mitotic checkpoint is commonly observed in cancer cells (7, 8). Studies have shown that the level of mitotic checkpoint proteins is critical for their function (9); accordingly, aberrant transcription and reduced expression of components of the SAC are often detected in human cancers (7, 10–12). However, genetic lesions of the core components of this checkpoint pathway are rare both in cancer cell lines and in primary tumors (12, 13), suggesting that epigenetic alterations or other events are responsible for weakening the mitotic checkpoint. Thus, the identification of new actors regulating the main proteins involved in this process is crucial.

Epigenetic marks, and especially DNA methylation, also play key roles in chromosome segregation (14): downregulating the DNA methyltransferase DMNT1 in cell lines or in animals is sufficient to cause aneuploidy and promote transformation (15, 16). The actors involved in translating DNA methylation into its functional genome-stabilizing and tumor-suppressing outcome are still unknown, even though they are predicted to play important roles in normal cells and in cancer.

ZBTB4 is a mammalian DNA-binding protein with C2H2 zinc fingers and a BTB/POZ domain. It was initially described as a paralog of Kaiso, which, like Kaiso, had preferential affinity for DNA containing methylated CpGs and repressed transcription through the recruitment of histone deacetylases (17). ZBTB4 also binds certain specific nonmethylated sequences yet with lower affinity than methylated sequences (18, 19). When expressed in mouse cells, ZBTB4 overlaps completely with chromocenters (17), which contain the highly methylated centromeric and pericentromeric repeats (20). This localization is lost in cells lacking DNA methylation (17), establishing that ZBTB4 is a reader of a critical epigenetic determinant of centromere and pericentromere identity.

[1]University Paris Diderot, Sorbonne Paris Cité, Epigenetics and Cell Fate, UMR 7216 CNRS, Paris, France. [2]University Bordeaux, UMR INSERM 1053, Bordeaux, France.

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Corresponding Author: Pierre-Antoine Defossez, University Paris Diderot, Box 7042, 75205 Paris Cedex 13, France. Phone: 33-1-57278916; Fax: 33-1-57278911; E-mail: pierre-antoine.defossez@univ-paris-diderot.fr

doi: 10.1158/0008-5472.CAN-16-1181

©2016 American Association for Cancer Research.

Downloaded from cancerres.aacrjournals.org on April 19, 2017. © 2017 American Association for Cancer Research.

Published OnlineFirst November 4, 2016; DOI: 10.1158/0008-5472.CAN-16-1181

ZBTB4 expression is reduced in several types of cancer (21–24), and low ZBTB4 expression is associated with decreased latency to relapse in breast and prostate tumors (21, 25). This suggests that ZBTB4 loss may confer a selective advantage to tumor cells. Potential reasons include ZBTB4 competing with Sp1 for binding to nonmethylated sites in the promoter of key protumoral genes, such as VEGF and VEGFR, and repressing their transcription (15). In addition, ZBTB4 is indirectly recruited to the p21 promoter by association with MIZ-1, and ZBTB4 loss promotes p21 induction, shifting cell fate toward survival (as opposed to apoptosis) upon p53 activation (23). Neither proposed role of ZBTB4 (Sp1 antagonism or p21 repression) rules out additional mechanisms by which ZBTB4 loss may benefit cancer cells, furthermore neither role involves the methyl-CpG–binding capacity of ZBTB4, nor its prominent localization at pericentric heterochromatin.

Here, we show that human tumors with low ZBTB4 expression have higher genome instability. Depletion of ZBTB4 in human cell lines causes mitotic abnormalities and widespread genome instability, accompanied by a weakening of the mitotic checkpoint. Cells isolated from *Zbtb4*$^{-/-}$ knockout mice have a weaker mitotic checkpoint, chromosome segregation defects, and increased genome instability. The mutant mice show decreased weight and increased susceptibility to chemically induced skin tumorigenesis.

Altogether these data show that the methyl-CpG reader ZBTB4 has a conserved role in the preservation of genomic stability, providing an example of an epigenetic reader participating in mitotic checkpoint regulation and in the control of chromosome segregation.

Materials and Methods

Cell culture and mouse embryonic fibroblast preparation

Mouse embryonic fibroblasts (MEF) were isolated from eviscerated E13.5 embryos. Tissues were rinsed with PBS, minced, and digested with trypsin-EDTA 30 minutes at 37°C. Trypsin was inactivated by addition of DMEM containing 10% FBS. For the growth curve, cells were plated at 3×10^5 in 6-well plates and counted each day in duplicate during 6 days using a hemocytometer. MEF populations from 5 different embryos were used for each genotype. HeLa, HCT116, U2OS, and MEFs were cultured in growth medium (DMEM), 10% FBS, 100 U/mL penicillin/streptomycin. HeLa, U2OS, and HCT116 were obtained from our own stocks and their identity was not formally verified for this study.

DNA paint

Samples were fixed in Carnoy fixative, and the FISH protocol was performed according to the manufacturer's instructions for DNA probes. Xcyting mouse chromosome painting probes XMP X green (Metasystems, D-1420-050-FI) and XMP 12 orange (Metasystems, D-1412-050-OR) were used.

Flow cytometric analysis

For determination of cell-cycle profiles, cells were plated in 60-mm plates and grown to 60% to 70% confluence. After trypsinization, the cells were collected and fixed in ice-cold 70% ethanol overnight, incubated with anti-BrdU-FITC (BD Pharmingen, 556028) and/or propidium iodide (PI) solution (50 μg/mL) in the presence of 0.2 mg/mL RNase A for 30 minutes. DNA content was measured using the BD FACSCalibur (BD Biosciences) and CellQuest Pro software. For determination of apoptotic cell population, In situ Direct DNA Fragmentation (TUNEL) Assay Kit (Abcam) was used according to the manufacturer's recommendations. For micronucleus quantification, 50 μL of retro-orbital blood from 1-year-old mice was collected in 300 μL heparin, fixed in cold methanol, and stained with anti-CD71/FITC antibody (Abcam) and PI solution as previously described (26). Ten thousand cells were acquired for all experiments.

Generation of *Zbtb4*$^{-/-}$ mice

To obtain a *Zbtb4*-null allele, we deleted exon 3, which contains the starting ATG and the coding sequence for the BTB domain and 2 of the 3 DNA-binding zinc fingers. The targeting construct containing the Neomycin selection marker and loxP sites on each side of exon 3 was electroporated into an embryonic stem cell line derived from the 129/Sv. Neomycin-resistant clones were selected and injected into the inner cell mass of a C57BL/6 blastocyst. Homozygous *Zbtb4* mice carrying the targeted allele were obtained. The Neomycin cassette and the exon 3 were then depleted by crossing allele-carrying mice with transgenic mice expressing the Cre recombinase under the control of the cytomegalovirus promoter.

The use of experimental animals was approved by the Buffon ethics committee and Ministry of Research and Higher Education (France). *Zbtb4*$^{-/-}$ generation and phenotyping were assessed by the Institut Clinique de la Souris (ICS, Strasbourg, France). The mice analyzed in this study were in a mixed genetic background C57BL6/J,129sv/Pas; and littermates were used for all experiments.

Immunofluorescence and confocal imaging

Cells were grown on glass coverslips, fixed with 4% paraformaldehyde (PFA) for 10 minutes, and then permeabilized with 0.5% Triton. For the incorporation of the DNA-labeling dye bromodeoxyuridine (BrdUrd), cells were treated with 10 μmol/L BrdUrd for 1 hour, collected by trypsination, fixed with 4% PFA, and incubated for 20 minutes in 2 mol/L HCl before permeabilization with 0.5% Triton. For mitotic checkpoint analysis, cells grown on glass coverslips were treated by DMSO (control) or by 0.3 μg/mL of nocodazole for 16 hours and then were fixed with 4% PFA for 10 minutes at room temperature followed by ice-cold MeOH for 10 minutes at −20°C.

Microarray and mRNA expression analysis

We used Affymetrix GeneChip Mouse Gene 2.0 ST arrays. We followed MIAME guidelines and the data were deposited on GEO (GSE77816). Venn diagrams were generated with Venny (http://bioinfogp.cnb.csic.es/tools/venny/). Data were analyzed with Partek tools and limma/EMA R packages. Gene Ontology (GO) term analysis was performed using Gprofiler (27) with a maximum *P* value of 10^{-3}. The was used to defined the enriched gene sets in *Zbtb4*$^{-/-}$ compared with trisomic cells of wild-type (WT) MEF cells. The data for trisomic MEFs (28) were reanalyzed with limma to generate a custom gene set, which was compared with the set of genes deregulated in *Zbtb4*$^{-/-}$ cells using the Gene Set Enrichment Analysis (GSEA) software.

Total RNA was extracted using Tri reagent (Sigma). For reverse transcription, first-strand cDNA was generated from 1 μg total RNA using Superscript III RT (LifeTechnologies) and random primers. qPCR was performed using Power SYBR Green (Applied Biosystems) on a ViiA 7 Real-Time PCR System (LifeTech).

Downloaded from cancerres.aacrjournals.org on April 19, 2017. © 2017 American Association for Cancer Research.

Published OnlineFirst November 4, 2016; DOI: 10.1158/0008-5472.CAN-16-1181

Expression values were normalized to the housekeeping genes *TBP*, *Rplp0*, and *B2M*.

Skin tumorigenesis protocol

Hair removal was performed at 7 weeks of age. One week after hair removal, 400 nmol 7,12-dimethylbenz(*a*)anthracene (DMBA) was painted on the shaved area and then 20 nmol 12-*O*-tetradecanoylphorbol-13-acetate (TPA) was applied twice per week for the following weeks. Mice were examined weekly for tumor formation and weighed. At 21 weeks, tumors were excised and fixed in 4% PFA, included in paraffin, and cut for histologic analysis. All mouse handling was overseen and approved by the Ethics Committee of our animal facility.

Time-lapse microscopy

Media containing vehicle or 0.125 μg/mL of nocodazole was replaced in wells immediately before the plate was placed on a time-lapse microscope (Leica DMI6000, with heated motorized stage, a mounted incubator unit and humidified CO_2). Brightfield (for MEFs) or fluorescence (for HeLa cells) images were recorded every 5 minutes for 24 hours at 5 regions per well on 4-well Lap-Tek. Image capture was controlled by MetaMorph Microscopy Automation & Image Analysis Software. Images from time-lapse capture were analyzed using Fiji (ImageJ). Measured parameters included (i) total cell numbers, (ii) number of mitotic events, (iii) intramitotic duration, (iv) the occurrence of polyploid nuclei, and (v) apoptotic/necrotic occurrences.

Virus production and infection

293T were transfected with the vector of interest, psPAX2 and pCMV VSVg plasmids (Addgene), using Lipofectamine (Life Biotechnologies) according to the manufacturer's recommendations. Cell supernatants containing infectious retroviral pseudotypes were harvested 24 to 36 hours posttransfection and filtered on a 0.45-μm membrane. Exponentially growing HeLa or U2OS cells were infected with virus-containing culture supernatant containing 10 mg/mL polybrene (Sigma). After 72 hours, cells were selected with puromycin (Sigma).

Western blotting

HeLa cells were synchronized in prometaphase by a 16-hour treatment with nocodazole (0.3 mg/mL). Cells were collected by mitotic shake off, washed with PBS, and replated in normal medium. Cells were then harvested at the indicated time post-release and lysed in RIPA buffer, treated with benzonase, sonicated (series of 5 minutes ON, 5 minutes OFF during 5–10 minutes; Bioruptor, Diagenode), and supplemented with NuPage Sample Reducing Agent and LDS Sample Buffer (LifeTech). Forty micrograms of protein was loaded for each samples.

Results

Human tumors with low ZBTB4 expression have highly altered genomes

We previously observed that ZBTB4 is downregulated in several cancers. To gain more precise information, we used cBioportal (29) to mine data from The Cancer Genome Atlas (TCGA). In particular, we surveyed genome instability, as estimated by the percentage of the genome affected by copy number alterations (CNA). We ranked each tumor cohort by ZBTB4 expression level, then compared CNA prevalence in tumors with low ZBTB4 expression (first decile, "ZBTB4 low") and in the remaining 90% of tumors ("Others"). We assessed 8 cancer types from tissues that normally express ZBTB4: breast-invasive carcinoma, prostate adenocarcinoma, uterine endometrioid carcinoma, kidney renal cell carcinoma, cervical squamous cell carcinoma, lung adenocarcinoma, stomach adenocarcinoma, and head and neck squamous cell cancer. In all of these, low ZBTB4 expression correlated with a higher percentage of the genome being affected by CNAs (Fig. 1A). As a control, we performed the same analysis for ZBTB33/Kaiso (30), but CNA frequency was never increased in ZBTB33-low tumors (Supplementary Fig. S1A). These data show that, in many cancer types, lower ZBTB4 expression correlates with higher genome instability.

Widespread genomic abnormalities are found in human cells with ZBTB4 knockdown

To determine whether ZBTB4 downregulation plays a causal role in genome instability, we examined aneuploidy in ZBTB4-depleted human cell lines. For this, we used lentivirus-delivered shRNAs against ZBTB4, which we validated in 2 distinct cell lines: HeLa (p53-negative) and U2OS (p53-positive; Supplementary Fig. S1B and S1C). By DAPI staining of fixed cells, we found that ZBTB4-depleted HeLa cells had 2-fold more micronuclei than control cells (Fig. 1B). We also observed a 50- to 100-percent increase in bi-nucleated cells in the *shZBTB4* population compared with the control (Fig. 1C). Chromosomes that lag behind at anaphase can cause both micronuclei (by formation of a nuclear envelope around them) and binucleated cells (by getting trapped in the cytokinesis furrow, causing its regression), so we analyzed the prevalence of lagging chromosomes. We observed a frequency of lagging chromosomes 1.5 to 2.5 times higher in *shZBTB4* than in control cells (Fig. 1D and Supplementary Fig. S1D). Finally, we tested whether some of our observations held true in the quasi-diploid cell line HCT116. In this context as well, knocking down ZBTB4 significantly increased polynucleation (Supplementary Fig. S1E).

We conclude that ZBTB4 loss in human cells negatively affects mitotic chromosome segregation and genome integrity.

ZBTB4-depleted cells show a defective mitotic checkpoint

Mitosis is the key phase in which chromosome segregation is monitored and executed (12); we therefore next examined the unfolding of mitosis. We used HeLa cells stably expressing H2B-mCherry (to visualize chromosomes), EB3-GFP (to visualize microtubules), and sh*ZBTB4* or the sh*CTRL*. We followed cells undergoing mitosis by live microscopy and tracked the duration of each phase (Fig. 2A and Supplementary Fig. S2A). While sh*ZBTB4* cells had a duration of prometaphase, anaphase, and telophase similar to controls, they had a metaphase duration significantly shorter than control cells (Fig. 2A). This is reminiscent of cells with a weakened mitotic checkpoint, which fails to maintain the cells in metaphase long enough to resolve attachment problems (31). This led us to examine the mRNA expression level of key mitotic checkpoint regulators and we found, both in HeLa and U2OS cells, that ZBTB4 depletion led to a downregulation of *BUB1B* and *MAD2L1* (Fig. 2B); other genes such as *BUB3* and *CCNB1* (cyclin B1 coding gene) were not affected. Therefore, we conclude that cells lacking ZBTB4 have deregulated mitotic checkpoint gene expression, as well as a shorter metaphase, compatible with a reduced checkpoint activity. We also determined the mitotic

Downloaded from cancerres.aacrjournals.org on April 19, 2017. © 2017 American Association for Cancer Research.

Published OnlineFirst November 4, 2016; DOI: 10.1158/0008-5472.CAN-16-1181

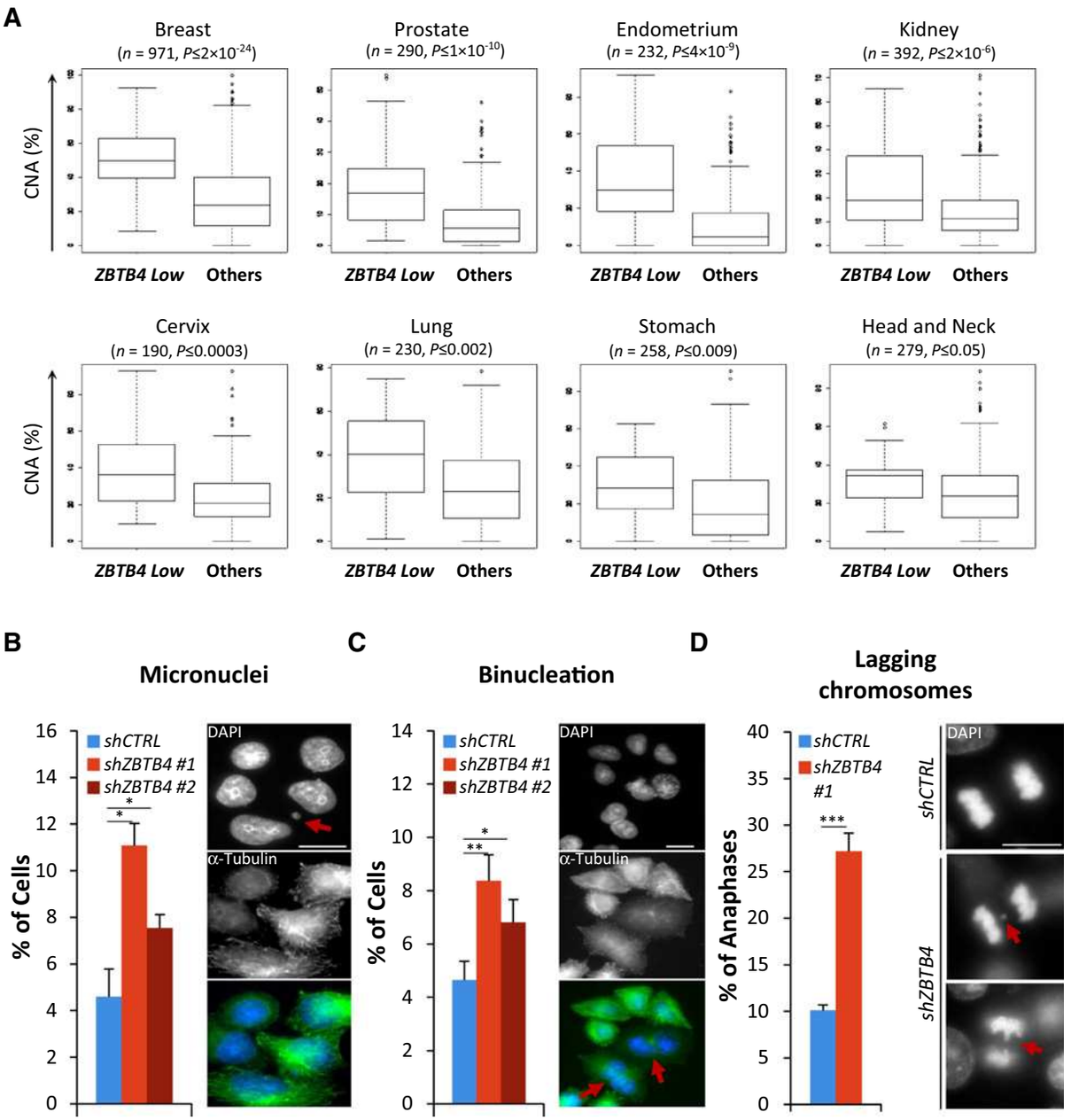


Figure 1.

Loss of ZBTB4 leads to genome instability. **A,** Percentage of the genome affected by CNA in the lowest decile for ZBTB4 expression ("ZBTB4-low") and the remaining tumors ("Others"); *P* value calculated by *t* test. **B,** Left, quantification of HeLa cells containing micronuclei after ZBTB4 knockdown. Right, representative image. Top, DAPI staining; middle, α-tubulin staining; bottom, merge. Arrow, micronucleus. Scale bar, 5 μm. **C,** Left, quantification of HeLa cells containing two nuclei after ZBTB4 knockdown. Right, as in **B**, arrow shows a binucleated cell. **D,** Left, quantification of anaphase HeLa cells showing lagging chromosomes after ZBTB4 knockdown. Right, illustrations, DAPI staining, arrow shows lagging chromosomes. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.005$. n.s., no significant difference.

index of U2OS cells before, during and after release from nocodazole arrest. We observed that few control cells had exited the block after 1 hour, whereas about half of the shZBTB4 cells had (Fig. 2C), supporting the conclusion that ZBTB4 depletion makes cells less efficient at sustaining a mitotic checkpoint.

To directly evaluate the activation of the mitotic checkpoint, we examined the amount and localization of key SAC components, localized at kinetochores when the checkpoint is activated. We evaluated their dynamics under conditions of checkpoint activation by nocodazole and subsequent release in anaphase. Both *shCTRL* and *shZBTB4* were efficiently arrested in mitosis by the

Downloaded from cancerres.aacrjournals.org on April 19, 2017. © 2017 American Association for Cancer Research.

Published OnlineFirst November 4, 2016; DOI: 10.1158/0008-5472.CAN-16-1181

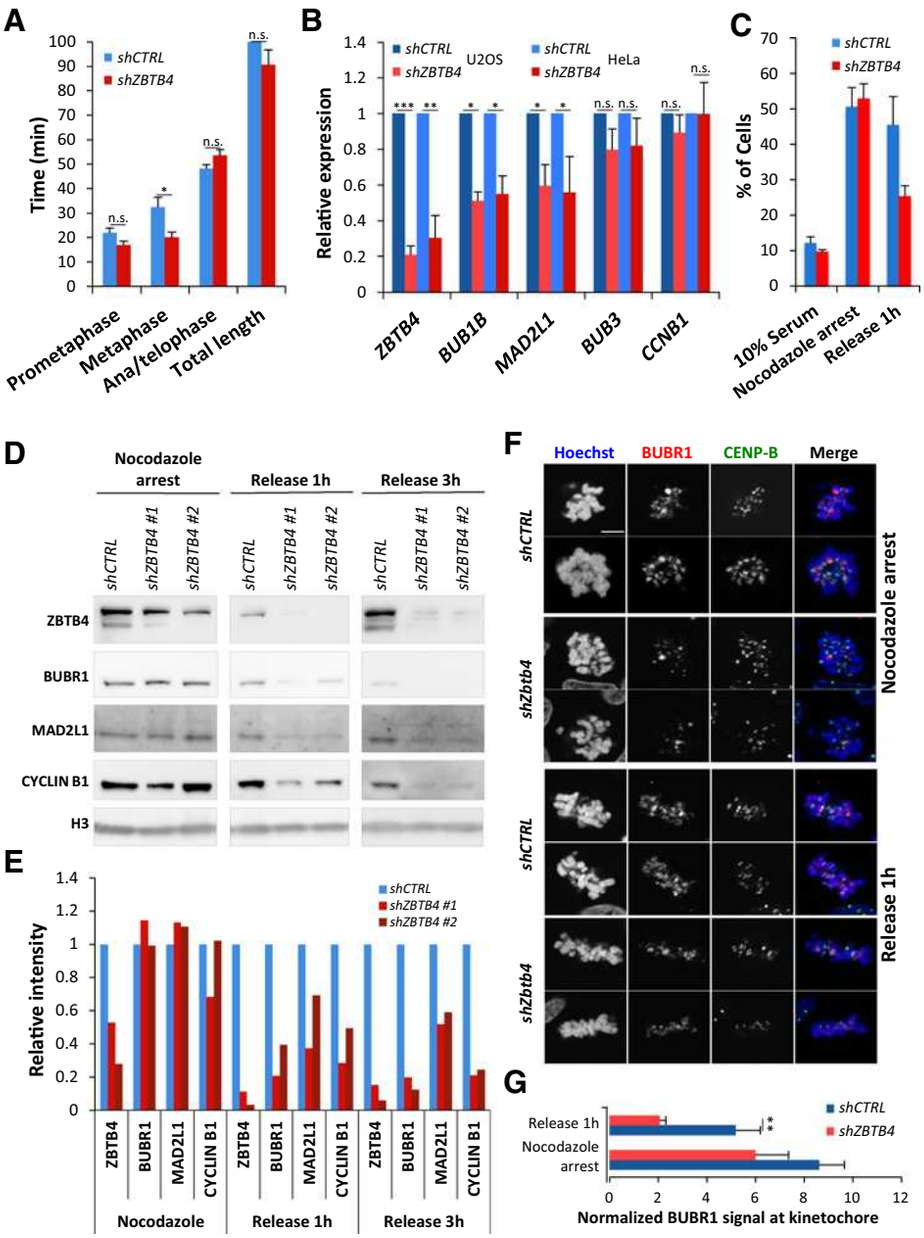


Figure 2.
Cells depleted of ZBTB4 have a weaker mitotic checkpoint. **A,** Videomicroscopy measurement of mitotic phase durations in HeLa cells. **B,** qRT-PCR analysis of mitotic checkpoint genes. **C,** Mitotic index of U2OS cells in the indicated conditions. **D,** Western blotting on whole-cell lysates from HeLa cells blocked with nocodazole or released for the indicated times. Histone H3, loading control. **E,** Quantification of **D**. **F,** Representative *z*-stack reconstruction of BUBR1 and CENP-B localization in U2OS-CENP-B-GFP cells expressing sh*CTRL* or sh*ZBTB4*. Cells were blocked in mitosis with nocodazole (top) and released into mitosis for 1 hour (bottom). Scale bar, 5 μm. **G,** Quantification of BUBR1 intensity signal at kinetochore, normalized to CENP-B intensity. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.005$. n.s., no significant difference.

nocodazole treatment, as determined by FACS (Supplementary Fig. S2B and S2C). No difference was observed in the abundance of BUBR1, MAD2L1, and cyclin B1, in cells depleted for ZBTB4 versus control cells under nocodazole arrest (Fig. 2D, left). However, after nocodazole washout, BUBR1 and MAD2L1 were less abundant in *shZBTB4* cells than in controls. In parallel, the downregulation of cyclin B1 protein, which reflects checkpoint inactivation, APC/C activation and progression through anaphase, was accelerated in sh*ZBTB4* cells (Fig. 2D, middle and right, and Fig. 2E). These data are, again, consistent with a mitotic checkpoint defect. We further confirmed these observations by immunofluorescence, using U2OS cells stably expressing CenpB-GFP (Fig. 2F and G). As expected, BUBR1 was recruited to kinetochores in control cells treated with nocodazole, and it was still present, albeit at a reduced level, 1 hour after release into nocodazole-free medium. However, in cells depleted of ZBTB4, BUBR1 accumulated at the kinetochores to lower levels in the presence of nocodazole. The difference with control cells was more pronounced 1 hour after nocodazole release, with about 2.5-fold less BUBR1 at kinetochores in *shZBTB4* cells relative to control cells (Fig. 2F and G). Altogether, these data indicate that ZBTB4-depleted cells are less efficient than control cells at sustaining an activated mitotic checkpoint, which may contribute to the heightened aneuploidy and genome instability we observed in these cells.

Zbtb4$^{-/-}$ mice are smaller than their littermates

We next generated *Zbtb4*-mutant mice (B6N.129-*Zbtb4*^{tm1defo}): using homologous recombination, we constitutively deleted exon 3 of the *Zbtb4* gene, which contains the translation-initiating ATG,

Downloaded from cancerres.aacrjournals.org on April 19, 2017. © 2017 American Association for Cancer Research.

Published OnlineFirst November 4, 2016; DOI: 10.1158/0008-5472.CAN-16-1181

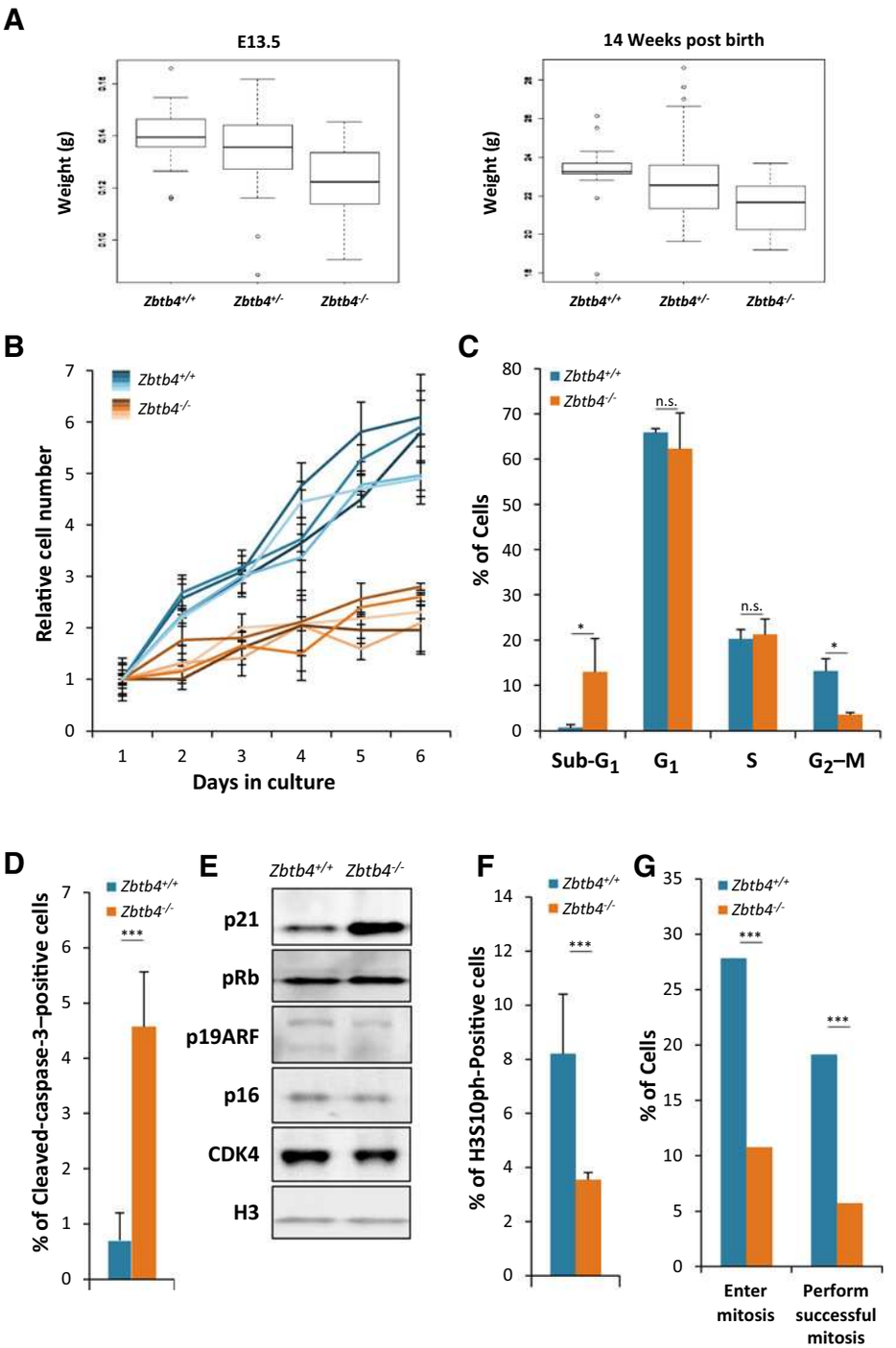


Figure 3.
Zbtb4$^{-/-}$ mice have a lower weight than controls, and their primary MEFs show increased apoptosis and G_2–M defects. **A,** Weight of animals at embryonic day 13.5 (E13.5; $n = 32$, $P \leq 0.2$ WT vs. KO) and at 14 weeks post-birth ($n = 36$, $P \leq 0.03$ WT vs. KO). **B,** Growth curves for 5 independent *Zbtb4*$^{+/+}$ and *Zbtb4*$^{-/-}$ primary MEF cultures. **C,** Flow cytometric analysis of cell-cycle phases (PI staining) for 4 independent *Zbtb4*$^{+/+}$ and *Zbtb4*$^{-/-}$ passage 3 MEFs. **D,** Immunofluorescent quantification of cleaved caspase-3–positive in cells from **C**. **E,** Western blotting on whole-cell protein lysates from cells in **C**. **F,** Quantification of H3S10ph staining. **G,** Analysis of 24-hour time-lapse microscopy of primary *Zbtb4*$^{+/+}$ and *Zbtb4*$^{-/-}$ MEFs. *, $P < 0.05$; ***, $P < 0.005$. n.s., no significant difference.

and encoded the BTB domain as well as 2 of the 3 zinc fingers required for DNA binding (Supplementary Fig. S3A). Both genotyping (Supplementary Fig. S3B) and expression analysis (Supplementary Fig. S3C) confirmed exon 3 deletion in mice. We will refer to the allele we generated as *Zbtb4*$^{-}$.

Zbtb4$^{-/-}$ animals were recovered at the expected frequency; males and females were fertile and had a lifespan comparable to control littermates. Nevertheless, the body weight of *Zbtb4*$^{-/-}$ animals was significantly lower than that of control littermates; this difference was already perceptible *in utero* (E13.5) and persisted throughout adulthood (Fig. 3A), showing that ZBTB4 function is required for normal organismal size.

Zbtb4$^{-/-}$ primary fibroblasts display decreased proliferation, a deficit in the G_2–M compartment, and increased apoptosis

Aneuploidy often causes a proliferative disadvantage at the cellular level and can decrease organ size (32), so we examined the proliferation of primary MEFs. *Zbtb4*$^{-/-}$ MEFs proliferated at half

Downloaded from cancerres.aacrjournals.org on April 19, 2017. © 2017 American Association for Cancer Research.

Published OnlineFirst November 4, 2016; DOI: 10.1158/0008-5472.CAN-16-1181

the rate of control cells (Fig. 3B), possibly resulting from changes in senescence, cell-cycle distribution, apoptosis, or a combination thereof. Staining with the senescence marker SA-β-gal gave similar results in WT and knockout cells (Supplementary Fig. S3D), ruling out a major role for senescence. In contrast, cell-cycle analysis by flow cytometry revealed that the *Zbtb4*$^{-/-}$ MEFs had G_1 and S populations comparable to WT (Fig. 3C) but significantly more sub-G_1 cells and significantly fewer G_2–M cells. Staining with BrdUrd confirmed that the S-phase fraction was similar in WT and knockout cells (Supplementary Fig. S3E). Staining of cleaved caspase-3 revealed about 8-fold more apoptosis in knockout cells relative than WT (Fig. 3D), likely accounting for the increased sub-G_1 fraction. TUNEL staining independently confirmed increased apoptosis in knockout cells (Supplementary Fig. S3F). The proliferation markers pRb and CDK4, and the senescence markers p16 and p19ARF, were not differentially expressed between WT and knockout cells (Fig. 3E). In contrast, p21 was markedly more abundant in *Zbtb4*$^{-/-}$ MEFs (Fig. 3E), agreeing with earlier results (23). The p53 protein was not detectably upregulated in *Zbtb4*$^{-/-}$ MEFs relative to controls (Supplementary Fig. S3G).

Besides increased apoptosis, the other marked phenotype of knockout cells was a deficit in the G_2–M subpopulation (3-fold smaller than WT, $P < 0.05$; Fig. 3C). This deficit was corroborated by immunofluorescence with an anti-H3S10ph antibody (Fig. 3F). We then used time-lapse microscopy to directly monitor mitosis in growing cells and found that, in a 24-hour period, about 3-fold less cells entered mitosis (as judged by nuclear envelope breakdown), and about 4-fold less successfully completed mitosis in knockout than in WT cells (Fig. 3G); of note, mitotic catastrophes were not increased in the mutant cells. From these data, we conclude that mouse cells without ZBTB4 have defects affecting mitosis, as human cells do.

Transcriptome analysis reveals G_2–M checkpoint deregulation and a signature of aneuploidy in *Zbtb4*$^{-/-}$ primary fibroblasts

We then turned to transcriptomic analysis (data available on GEO under reference GSE77816), using 5 independent primary MEF isolates from WT and *Zbtb4*$^{-/-}$ embryos at E13.5; 4 of each genotype isolates correctly clustered by genotype in an unsupervised clustering analysis and were used for subsequent analysis. We found that 222 genes were upregulated and 313 genes downregulated in absence of ZBTB4 (Supplementary Fig. S4A).

GSEA revealed that several gene sets were differentially expressed between knockout and WT cells. One of the gene sets most different between samples was "G_2–M checkpoint," which mostly includes genes involved in mitosis, such as *Bub1B* and *Cdc25c*; these genes were less expressed in knockout than in WT cells (Fig. 4A). We verified the validity of the microarray results by standard qRT-PCR (Fig. 4B), which confirmed, among others, the downregulation of *Bub1B* and *Cdc25c* in knockout cells. Therefore, MEFs lacking ZBTB4 downregulate important genes necessary for mitotic checkpoint function, paralleling our observations in human cells.

Studies in humans and mice have shown that aneuploidy causes a transcriptional response indicative of stress and slow cell growth (28, 33, 34). Thus, we asked whether the "aneuploidy signature" was identifiable in the transcriptome of *Zbtb4*$^{-/-}$ MEFs. We first compared the GO terms common to trisomic (TS) and to *Zbtb4*$^{-/-}$ MEFs. These terms included cell-cycle progression ($P < 10^{-14}$), mitosis ($P < 10^{-14}$), DNA replication ($P < 10^{-2}$), chromosome segregation ($P < 10^{-13}$), DNA repair ($P < 10^{-3}$), organelle fission ($P < 10^{-13}$), and DNA metabolism ($P < 10^{-3}$; Fig. 4C). Altogether, we observed a 34% overlap with GO terms enriched in aneuploid trisomic MEFs and human cells (Fig. 4D and Supplementary Table S1; ref. 35).

We then performed a second, more precise analysis. We reanalyzed available data (28, 33) to extract the differentially expressed genes and generate a custom gene set associated with aneuploidy in mouse trisomic cells. We then used GSEA to ask whether this gene set was statistically misregulated in *Zbtb4*$^{-/-}$ cells and the answer was positive (Fig. 4E). Therefore, cells lacking ZBTB4 have a transcriptional signature indicative of heightened aneuploidy.

We performed additional GSEA analyses using publically available genesets in MSigDB. The genes significantly altered in *Zbtb4*$^{-/-}$ cells correspond to genes associated with the gene sets: "G_2–M checkpoint," "E2F targets," "Myc targets v1," "Mitotic spindle," "mTORC signaling," and "DNA repair" (Supplementary Fig. S4B).

Increased aneuploidy in MEFs lacking ZBTB4

The transcriptome results prompted us to test for signs of chromosome missegregation and aneuploidy in *Zbtb4*$^{-/-}$ cells. We started with missegregation, and the 2 types of abnormal events we examined—presence of micronuclei (Fig. 5A) and binucleation (Fig. 5B)—were significantly more prevalent in *Zbtb4*$^{-/-}$ than WT MEFs.

We also used DNA probes for chromosomes 12 and X to evaluate the incidence of aneuploidy in female cells. We observed prevalent aneuploidy in iMEFs *Zbtb4*$^{-/-}$ with a broad distribution of chromosome numbers (range 1–13 X chromosomes per cell in knockout cells, for example; Fig. 5C and D). These data demonstrate that high chromosomal instability occurs when ZBTB4 is absent. Of note, we ruled out centrosome abnormalities as a possible cause (Supplementary Fig. S5).

Zbtb4$^{-/-}$ mice exhibit more aneuploidy and are more sensitive to tumor induction

We finally asked whether our conclusions also held true in a living animal. ZBTB4 is expressed in many mouse tissues (36), including erythroid progenitors. Reticulocytes and erythrocytes retain existing micronuclei when extruding their nucleus and aneuploidy *in vivo* can be precisely assessed in these cells by measuring the abundance of micronuclei with FACS (26). We found that both reticulocytes and erythrocytes from *Zbtb4*$^{-/-}$ mice had significantly more micronuclei than their WT counterpart (Fig. 6A). This key observation shows that, *in vivo*, ZBTB4 loss increases genome instability.

ZBTB4 is expressed in skin, and increased aneuploidy promotes DMBA/TPA-induced skin carcinogenesis (37), so we tested whether *Zbtb4*$^{-/-}$ mice might more susceptible than their littermates in this system. Indeed, we found that the number of tumors per animal was higher in knockout than in WT, with about 2-fold difference from 9 to 15 weeks posttreatment (Fig. 6B). The latency of tumor development was also reduced in the *Zbtb4*$^{-/-}$ mice: they started exhibiting tumors by 8 weeks of treatment, and by 10, weeks all animals (8 of 8) were affected (Fig. 6C). In contrast, the WT started developing tumors by 9 weeks, and only by 19 weeks was penetrance total (Fig. 6C). The heterozygous *Zbtb4*$^{+/-}$ animals had a tumor burden and latency comparable with WT. Histologic analysis showed that most of the occurring tumors were papillomas. In addition, 2 of 8 knockout animals developed large tumors, which dissection and staining revealed to be

Downloaded from cancerres.aacrjournals.org on April 19, 2017. © 2017 American Association for Cancer Research.

Published OnlineFirst November 4, 2016; DOI: 10.1158/0008-5472.CAN-16-1181

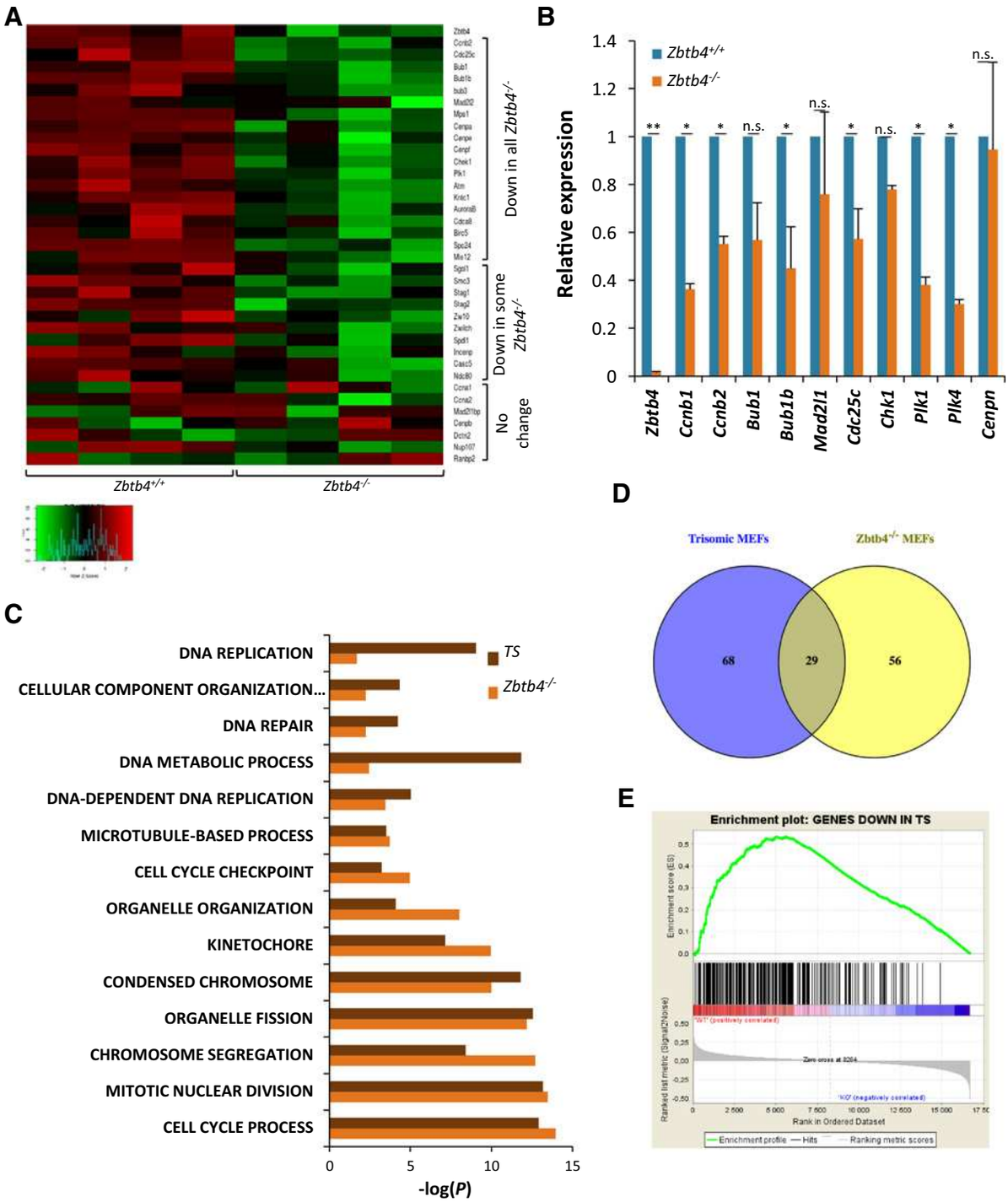


Figure 4.

Loss of ZBTB4 induces transcriptional alterations indicative of aneuploidy and mitotic checkpoint deregulation. **A,** Mitotic checkpoint genes are frequently deregulated in *Zbtb4*$^{-/-}$ MEFs. Microarrays performed on passage 3 *Zbtb4*$^{+/+}$ and *Zbtb4*$^{-/-}$ MEFs from four different mice. **B,** qRT-PCR analysis of mitotic checkpoint genes in passage 3 MEFs (two clones per genotype). *, $P < 0.05$; **, $P < 0.01$. n.s., no significant difference. **C,** Comparative GO terms analysis using gprofiler in *Zbtb4*$^{-/-}$ and trisomic MEFs. **D,** Venn diagram of GO terms found in trisomic and *Zbtb4*$^{-/-}$ MEFs. **E,** GSEA revealed an enrichment of genes related to the aneuploid signature of trisomic MEFs in *Zbtb4*$^{-/-}$ MEFs.

Downloaded from cancerres.aacrjournals.org on April 19, 2017. © 2017 American Association for Cancer Research.

Published OnlineFirst November 4, 2016; DOI: 10.1158/0008-5472.CAN-16-1181

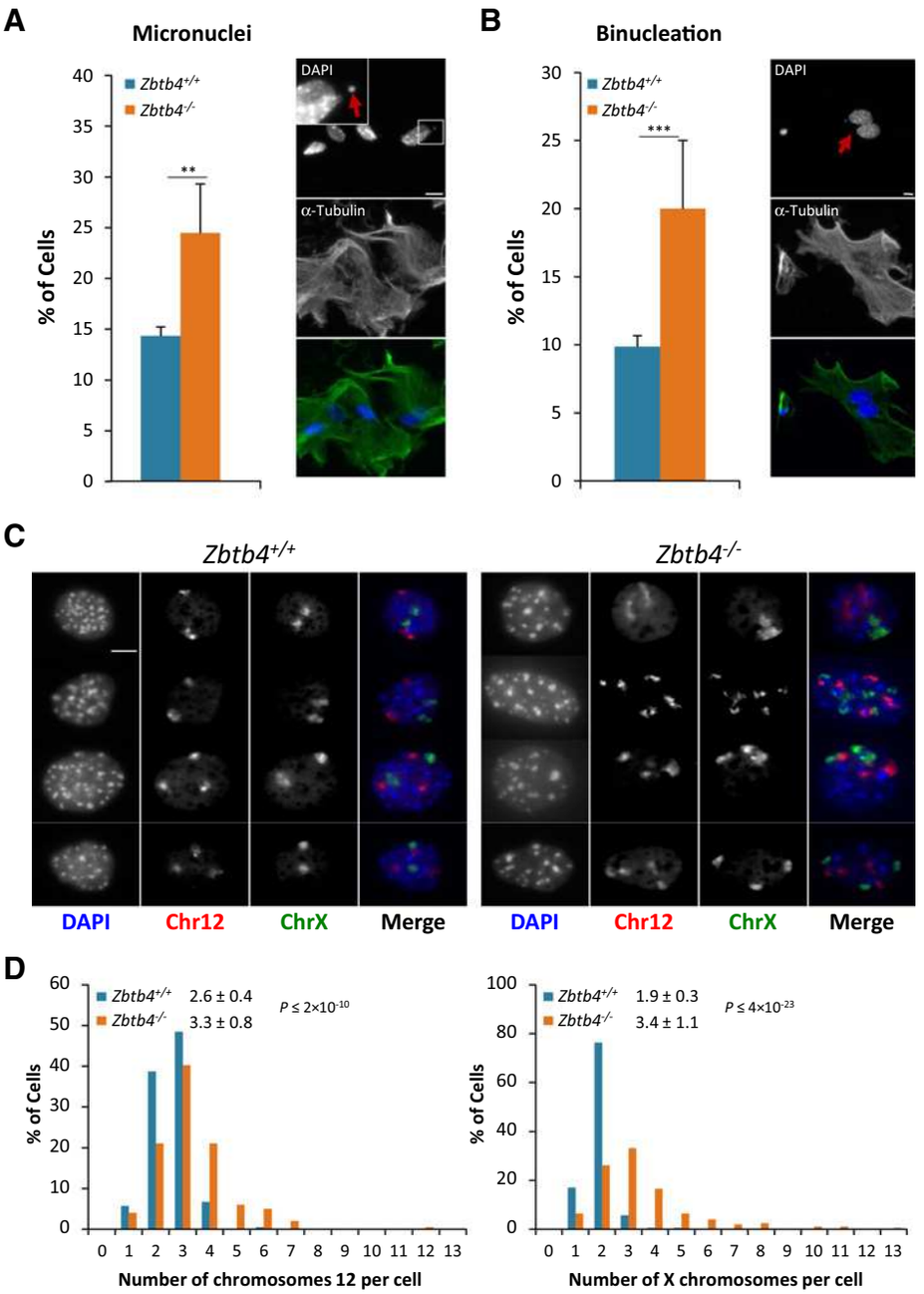


Figure 5.
Chromosome segregation defects in *Zbtb4*$^{-/-}$ mouse fibroblasts. **A,** Prevalence of micronuclei in primary MEFs (Hoechst staining). Left, quantitation; right, representative image. Top, DAPI staining; middle, α-tubulin staining; bottom, merge. Arrow, micronucleus. Scale bar, 5 μm. **B,** Prevalence of binucleated cells in primary MEFs Left, quantitation; right, as in A, arrow shows a binucleated cell. **C,** Representative results of chromosome quantification in interphase cells by DNA paint for chromosomes 12 (red) and chromosome X (green) in iMEFs. Scale bar, 5 μm. **D,** Quantification of **C** (two lines per genotype). **, $P < 0.01$; ***, $P < 0.005$.

keratoacanthomas (Fig. 6D and E). One heterozygote of 12 developed a keratoacanthoma; upon dissection and analysis of this tumor, we found that it lacked mRNA expression of *Zbtb4* (Supplementary Fig. S6A and S6B), possibly indicative of LOH. None of 8 WT mice developed keratoacanthoma.

To summarize, ZBTB4 inactivation renders mice more susceptible to a DMBA/TPA challenge: knockouts develop papillomas in higher numbers and with higher speed than control animals. In addition, ZBTB4 loss facilitates the appearance of keratoacanthoma, which we never observed in WTs. From these results we conclude that loss of ZBTB4 is sufficient to accelerate tumorigenesis in a mouse model.

Discussion

In this report, we establish a role of ZBTB4 in chromosome stability. In both mice and human cell lines, the depletion of ZBTB4 weakens the mitotic checkpoint, increases the frequency of lagging chromosomes, of micronuclei, and of binucleated cells. In a ZBTB4 knockout model, there is more aneuploidy measured *in vivo*, and the animals are more sensitive to induced skin tumorigenesis. Cancer genomics data also show that lower ZBTB4 expression correlates with higher genome instability across a broad range of tumor types. Altogether our data support a conserved role for ZBTB4 in preserving genome integrity.

Downloaded from cancerres.aacrjournals.org on April 19, 2017. © 2017 American Association for Cancer Research.

Published OnlineFirst November 4, 2016; DOI: 10.1158/0008-5472.CAN-16-1181

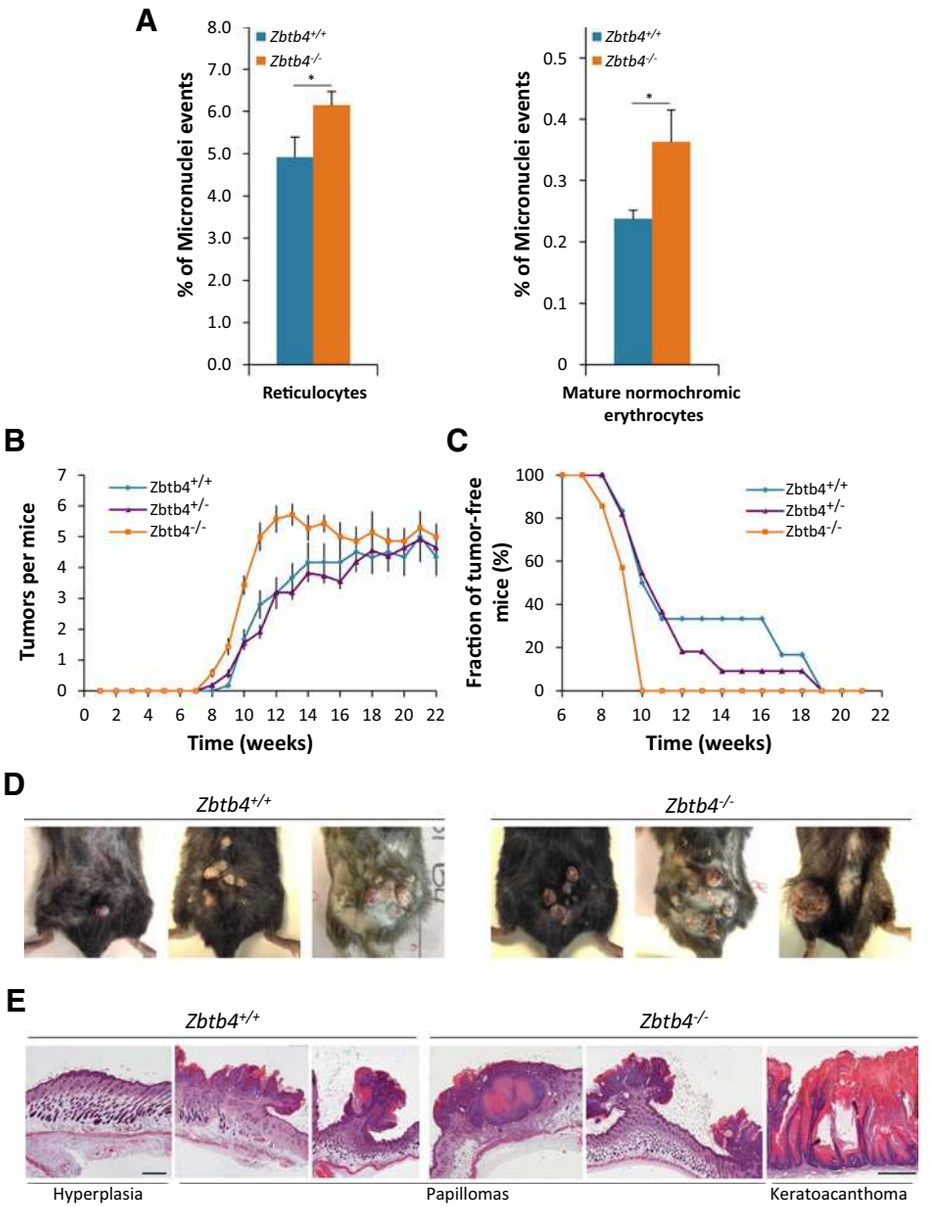


Figure 6.
Increased aneuploidy and skin tumorigenesis in *Zbtb4*$^{-/-}$ mice. **A,** FACS quantification of micronuclei in circulating reticulocytes and erythrocytes [$n = 8$ for each genotype, P (Student) $= 0.03$ and 0.02]. *, $P < 0.05$. **B,** Average tumor number per mouse [P (Student) $< 2 \times 10^{-4}$ *Zbtb4*$^{-/-}$ vs. *Zbtb4*$^{+/+}$). $n = 8$ *Zbtb4*$^{-/-}$,12 *Zbtb4*$^{+/-}$, and 8 *Zbtb4*$^{+/+}$ littermates of mixed genetic background (C57BL6/J,129sv/Pas). **C,** Fraction of papilloma-free mice over time after treatment [P (log-rank) $< 2 \times 10^{-3}$ for *Zbtb4*$^{-/-}$ vs. *Zbtb4*$^{+/+}$). **D,** Dorsal view of representative mice after 21 weeks of DMBA/TPA treatment. **E,** Hematoxylin and eosin staining of representative tumor sections after 21 weeks of DMBA/TPA treatment. Scale bar, 100 μm.

Lack of ZBTB4 leads to cell-cycle disturbances and increased genome instability

We have generated knockout mice lacking the critical exon 3 of ZBTB4. Removal of ZBTB4 in human cell lines by shRNA reproduces the phenotypes seen in ZBTB4-knockout MEFs, supporting the idea that the deletion allele is null.

The *Zbtb4*$^{-/-}$ MEFs have several abnormalities, including increased apoptosis and disturbed cell-cycle distribution, with fewer G_2–M cells. This is somewhat surprising, as previous data indicated that p21 was induced in response to ZBTB4 depletion—a finding we also observe.

While p21 has roles in G_2 arrest, its best-characterized function is in triggering G_1 arrest, but we observe no significant G_1 arrest in knockout cells. Using human cells, we have found that metaphase proceeds faster in cells lacking ZBTB4, but metaphase is a short fraction of mitosis (less than 10 minutes in HeLa cells), so the shortening we observe cannot account for the large decrease of the G_2–M compartment in knockout cells. Instead, it seems more likely, and consistent with our videomicroscopy data, that some knockout cells die in mitosis, thus enriching the sub-G_1 fraction, and depleting the G_2–M compartment. This does not rule out the possibility that some cells die in G_1 or in G_2 before getting to M phase—further experiments in the future could address this possibility.

Many cells, both in the mouse and human system, show chromosome abnormalities in the absence of ZBTB4: lagging chromosomes, micronuclei, and binucleated cells all occur at elevated frequencies in the knockout cells. These may be causally linked as lagging chromosomes nucleate micronuclei. In addition, when they get caught in the cytokinetic furrow, lagging

Downloaded from cancerres.aacrjournals.org on April 19, 2017. © 2017 American Association for Cancer Research.

Published OnlineFirst November 4, 2016; DOI: 10.1158/0008-5472.CAN-16-1181

chromosomes cause the furrow to regress, leading to the formation of a binucleated cell. Interestingly, these cells then induce p21 and enter G_1 arrest (38). It is possible that these cells are numerous enough to account for the p21 increase that is observed but not so numerous that they would measurably affect the G_1 compartment. We also note that p53 is not detectably upregulated in MEFs lacking ZBTB4 relative to controls. This could suggest, again, that only a small proportion of cells upregulate p53 or, alternatively, that the robust p21 induction we detect is p53-independent. The Myc/Miz1 axis is a tantalizing candidate for this p53-independent regulation, as ZBTB4 interacts with Miz1 (21, 23).

Mice with artificially induced aneuploidy in the brain have smaller brains, because aneuploid cells are cleared out by apoptosis and not completely replaced (32). Mice lacking ZBTB4 are lighter than their WT siblings, and we speculate that this is also due to the removal, *in vivo*, of aneuploid cells.

ZBTB4 loss weakens the mitotic checkpoint: possible causes

Human HeLa cells lacking ZBTB4 progress faster through mitosis and, more precisely, they have a faster metaphase-to-anaphase transition. This is compatible with a weakened mitotic checkpoint, as it is the activity of this checkpoint that prolongs metaphase until all kinetochores are correctly attached and under tension. Further support for a weaker checkpoint in cells lacking ZBTB4 comes from our observation that these cells re-enter mitosis faster than control cells after nocodazole removal. At present, we cannot determine whether the mitotic checkpoint in knockout cells is harder to activate or easier to inactivate, or both, and further work will be required to solve this question.

The removal of ZBTB4 leads to decreased transcription of mitotic checkpoint components, both in mouse and in human cells. Our previous experiments using reporter systems suggest that ZBTB4 is a transcriptional repressor (17), and all the published evidence from other groups is also consistent with ZBTB4 being a transcriptional repressor (21, 23, 39). Therefore, we do not favor the hypothesis that ZBTB4 controls mitotic gene expression directly, by binding the promoter of these genes and activating their transcription. We also doubt that this downregulation of checkpoint gene expression is an adaptation, helping cells cope with the absence of ZBTB4, because we observe the same phenomenon during acute depletion of ZBTB4 by shRNA. As ZBTB4 colocalizes with pericentric heterochromatin in mouse cells, we speculate that its absence could have an effect in cis, such as disturbing the attachment and/or function of checkpoint proteins at the kinetochore, eventually leading to reduced expression of mitotic checkpoint genes by a possible feedback mechanism.

Loss of ZBTB4 accelerates tumorigenesis

Mice lacking ZBTB4 are more susceptible to chemically induced skin tumorigenesis. This supports the idea that ZBTB4 loss is not merely a consequence of transformation, but actually a contributing event. It has been recently shown that increased aneuploidy in the skin is sufficient to increase tumorigenesis (37), so we hypothesize that ZBTB4 loss renders the mice more sensitive to skin cancer by increasing aneuploidy; this does not rule out the possibility that ZBTB4 has other functions that protect the organism against cancer. We also note that 0 of 8 *Zbtb4*$^{+/+}$ animals developed keratoacanthoma whereas 2 of 8 *Zbtb4*$^{-/-}$ mice did. One *Zbtb4*$^{+/-}$ heterozygote developed a keratoacanthoma and, upon examination, this tumor had lost ZBTB4 expression. These numbers are still too small to hold statistical power, but they suggest that ZBTB4 might be particularly important to preserve the cells of origin of keratoacanthoma, which reside in the hair follicle (40).

ZBTB4 is frequently downregulated in human cancers, both at the RNA and at the protein levels. Several mechanisms seem to be involved: one is downregulation by oncogenic miRNAs (21, 41) and another genomic deletions (22, 42). A recent report has shown that deletions affecting TP53 in cancer are typically large (up to 4 Mb), remove up to 100 genes, and that some of these losses synergize with TP53 loss to drive transformation (43). ZBTB4 is located on human chromosome 17p13.1, about 180 kb away from the TP53 gene, and well within the interval that is frequently codeleted with TP53. We speculate that ZBTB4 loss could synergize with TP53 loss because it increases genome instability, which permits faster tumor evolution.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: A. Roussel-Gervais, O. Kirsh, P.-A. Defossez
Development of methodology: A. Roussel-Gervais, I. Naciri, L. Kasprzyk, G. Velasco, G. Grillo
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): A. Roussel-Gervais, I. Naciri, L. Kasprzyk, G. Velasco, G. Grillo, P. Dubus
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A. Roussel-Gervais, O. Kirsh, P. Dubus
Writing, review, and/or revision of the manuscript: A. Roussel-Gervais, O. Kirsh, G. Velasco, G. Grillo, P.-A. Defossez
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): A. Roussel-Gervais, O. Kirsh
Study supervision: A. Roussel-Gervais, P.-A. Defossez

Acknowledgments

For helpful discussions, we are grateful to: Allison Bardin, Renata Basto, Katia Boyarchuk, Gael Cristofari, Christine Dozier, Arnaud Echard, Claire Francastel, Olivier Gavet, Patrick Heun, Roger Karess, and Vincent Vanoosthuyse. We thank for the gift of useful reagents: David Bernard, Valérie Doye, Jan Ellenberg, Kevin Sullivan, and Irina Stancheva. We thank Institut Clinique de la Souris (Strasbourg, France) for generating and phenotyping the *Zbtb4*$^{-/-}$ animals and the animal facility and Institut Jacques Monod (Paris, France) for housing the animals. We acknowledge the ImagoSeine facility, member of the France BioImaging infrastructure supported by the French National Research Agency (ANR-10-INSB-04, « Investments of the future ») for imaging. We also thank the Cochin HistIM Facility for histology preparation and the Cochin Genom'IC Facility for microarray experiments. We are indebted to Niambi Brewer who performed some experiments on this project. We warmly thank Laure Ferry for her help.

Grant Support

A.-R. Roussel-Gervais was supported by doctoral fellowships from Ligue Nationale contre le Cancer and Fonds de recherche en santé du Québec (FRSQ), Canada. Work in the laboratory of P.-A. Defossez is supported by Association pour la Recherche contre le Cancer (ARC), Ligue Nationale contre le Cancer, Groupement des Entreprises Françaises dans la Lutte contre le Cancer (GEFLUC), and by ANR-11-LABX-0071 under program ANR-11-IDEX-0005-01.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received May 4, 2016; revised October 1, 2016; accepted October 26, 2016; published OnlineFirst November 4, 2016.

Downloaded from cancerres.aacrjournals.org on April 19, 2017. © 2017 American Association for Cancer Research.

Published OnlineFirst November 4, 2016; DOI: 10.1158/0008-5472.CAN-16-1181

References

1. Holland AJ, Cleveland DW. Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. Nat Rev Mol Cell Biol 2009;10:478–87.
2. Albertson DG, Collins C, McCormick F, Gray JW. Chromosome aberrations in solid tumors. Nat Genet 2003;34:369–76.
3. Gordon DJ, Resio B, Pellman D. Causes and consequences of aneuploidy in cancer. Nat Rev Genet 2012;13:189–203.
4. London N, Biggins S. Signalling dynamics in the spindle checkpoint response. Nat Rev Mol Cell Biol 2014;15:736–47.
5. Zich J, Hardwick KG. Getting down to the phosphorylated "nuts and bolts" of spindle checkpoint signalling. Trends Biochem Sci 2010;35:18–27.
6. Dobles M, Liberal V, Scott ML, Benezra R, Sorger PK. Chromosome missegregation and apoptosis in mice lacking the mitotic checkpoint protein Mad2. Cell 2000;101:635–45.
7. Cahill DP, Lengauer C, Yu J, Riggins GJ, Willson JK, Markowitz SD, et al. Mutations of mitotic checkpoint genes in human cancers. Nature 1998;392:300–3.
8. Lara-Gonzalez P, Westhorpe FG, Taylor SS. The spindle assembly checkpoint. Curr Biol CB 2012;22:R966–80.
9. Musacchio A, Hardwick KG. The spindle checkpoint: structural insights into dynamic signalling. Nat Rev Mol Cell Biol 2002;3:731–41.
10. Li Y, Benezra R. Identification of a human mitotic checkpoint gene: hsMAD2. Science 1996;274:246–8.
11. Shichiri M, Yoshinaga K, Hisatomi H, Sugihara K, Hirata Y. Genetic and epigenetic inactivation of mitotic checkpoint genes hBUB1 and hBUBR1 and their relationship to survival. Cancer Res 2002;62:13–7.
12. Kops GJPL, Weaver BAA, Cleveland DW. On the road to cancer: aneuploidy and the mitotic checkpoint. Nat Rev Cancer 2005;5:773–85.
13. Schvartzman J-M, Sotillo R, Benezra R. Mitotic chromosomal instability and cancer: mouse modelling of the human disease. Nat Rev Cancer 2010;10:102–15.
14. Herrera LA, Prada D, Andonegui MA, Dueñas-González A. The epigenetic origin of aneuploidy. Curr Genomics 2008;9:43–50.
15. Gaudet F, Hodgson JG, Eden A, Jackson-Grusby L, Dausman J, Gray JW, et al. Induction of tumors in mice by genomic hypomethylation. Science 2003;300:489–92.
16. Barra V, Schillaci T, Lentini L, Costa G, Di Leonardo A. Bypass of cell cycle arrest induced by transient DNMT1 post-transcriptional silencing triggers aneuploidy in human cells. Cell Div 2012;7:2.
17. Filion GJP, Zhenilo S, Salozhin S, Yamada D, Prokhortchouk E, Defossez P-A. A family of human zinc finger proteins that bind methylated DNA and repress transcription. Mol Cell Biol 2006;26:169–81.
18. Sasai N, Nakao M, Defossez P-A. Sequence-specific recognition of methylated DNA by human zinc-finger proteins. Nucleic Acids Res 2010;38:5015–22.
19. Buck-Koehntop BA, Defossez P-A. On how mammalian transcription factors recognize methylated DNA. Epigenetics 2013;8:131–7.
20. Almouzni G, Probst AV. Heterochromatin maintenance and establishment: lessons from the mouse pericentromere. Nucl Austin Tex 2011;2:332–8.
21. Kim K, Chadalapaka G, Lee S-O, Yamada D, Sastre-Garau X, Defossez P-A, et al. Identification of oncogenic microRNA-17-92/ZBTB4/specificity protein axis in breast cancer. Oncogene 2012;31:1034–44.
22. Ross-Adams H, Lamb AD, Dunning MJ, Halim S, Lindberg J, Massie CM, et al. Integration of copy number and transcriptomics provides risk stratification in prostate cancer: a discovery and validation cohort study. EBioMedicine 2015;2:1133–44.
23. Weber A, Marquardt J, Elzi D, Forster N, Starke S, Glaum A, et al. Zbtb4 represses transcription of P21CIP1 and controls the cellular response to p53 activation. EMBO J 2008;27:1563–74.
24. Yamada D, Pérez-Torrado R, Filion G, Caly M, Jammart B, Devignot V, et al. The human protein kinase HIPK2 phosphorylates and downregulates the methyl-binding transcription factor ZBTB4. Oncogene 2009;28:2535–44.
25. Kim K, Chadalapaka G, Pathi SS, Jin U-H, Lee J-S, Park Y-Y, et al. Induction of the transcriptional repressor ZBTB4 in prostate cancer cells by drug-induced targeting of microRNA-17-92/106b-25 clusters. Mol Cancer Ther 2012;11:1852–62.
26. Balmus G, Karp NA, Ng BL, Jackson SP, Adams DJ, McIntyre RE. A high-throughput *in vivo* micronucleus assay for genome instability screening in mice. Nat Protoc 2015;10:205–15.
27. Reimand J, Kull M, Peterson H, Hansen J, Vilo J. g:Profiler–a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res 2007;35:W193–200.
28. Sheltzer JM, Torres EM, Dunham MJ, Amon A. Transcriptional consequences of aneuploidy. Proc Natl Acad Sci 2012;109:12644–9.
29. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 2013;6:pl1.
30. Defossez P-A, Stancheva I. Biological functions of methyl-CpG-binding proteins. Prog Mol Biol Transl Sci 2011;101:377–98.
31. Meraldi P, Draviam VM, Sorger PK. Timing and checkpoints in the regulation of mitotic progression. Dev Cell 2004;7:45–60.
32. Marthiens V, Rujano MA, Pennetier C, Tessier S, Paul-Gilloteaux P, Basto R. Centrosome amplification causes microcephaly. Nat Cell Biol 2013;15:731–40.
33. Sheltzer JM. A transcriptional and metabolic signature of primary aneuploidy is present in chromosomally unstable cancer cells and informs clinical prognosis. Cancer Res 2013;73:6401–12.
34. Dürrbaum M, Kuznetsova AY, Passerini V, Stingele S, Stoehr G, Storchová Z. Unique features of the transcriptional response to model aneuploidy in human cells. BMC Genomics 2014;15:139.
35. Williams BR, Prabhu VR, Hunter KE, Glazier CM, Whittaker CA, Housman DE, et al. Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells. Science 2008;322:703–9.
36. Wu C, Jin X, Tsueng G, Afrasiabi C, Su AI. BioGPS: building your own mash-up of gene annotations and expression profiles. Nucleic Acids Res 2016;44:D313–316.
37. Serçin Ö, Larsimont J-C, Karambelas AE, Marthiens V, Moers V, Boeckx B, et al. Transient PLK4 overexpression accelerates tumorigenesis in p53-deficient epidermis. Nat Cell Biol 2016;18:100–10.
38. Andreassen PR, Lohez OD, Lacroix FB, Margolis RL. Tetraploid state induces p53-dependent arrest of nontransformed mammalian cells in G1. Mol Biol Cell 2001;12:1315–28.
39. Yang WS, Chadalapaka G, Cho S-G, Lee S-O, Jin U-H, Jutooru I, et al. The transcriptional repressor ZBTB4 regulates EZH2 through a MicroRNA-ZBTB4-specificity protein signaling axis. Neoplasia 2014;16:1059–69.
40. Zito G, Saotome I, Liu Z, Ferro EG, Sun TY, Nguyen DX, et al. Spontaneous tumour regression in keratoacanthomas is driven by Wnt/retinoic acid signalling cross-talk. Nat Commun 2014;5:3543.
41. Bomben R, Gobessi S, Dal Bo M, Volinia S, Marconi D, Tissino E, et al. The miR-17∼92 family regulates the response to Toll-like receptor 9 triggering of CLL cells with unmutated IGHV genes. Leukemia 2012;26:1584–93.
42. Grasso CS, Wu Y-M, Robinson DR, Cao X, Dhanasekaran SM, Khan AP, et al. The mutational landscape of lethal castration-resistant prostate cancer. Nature 2012;487:239–43.
43. Liu Y, Chen C, Xu Z, Scuoppo C, Rillahan CD, Gao J, et al. Deletions linked to TP53 loss drive cancer through p53-independent mechanisms. Nature 2016;531:471–5.

Downloaded from cancerres.aacrjournals.org on April 19, 2017. © 2017 American Association for Cancer Research.

Published OnlineFirst November 4, 2016; DOI: 10.1158/0008-5472.CAN-16-1181

Cancer Research
The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

AACR American Association for Cancer Research

Loss of the Methyl-CpG–Binding Protein ZBTB4 Alters Mitotic Checkpoint, Increases Aneuploidy, and Promotes Tumorigenesis

Audrey Roussel-Gervais, Ikrame Naciri, Olivier Kirsh, et al.

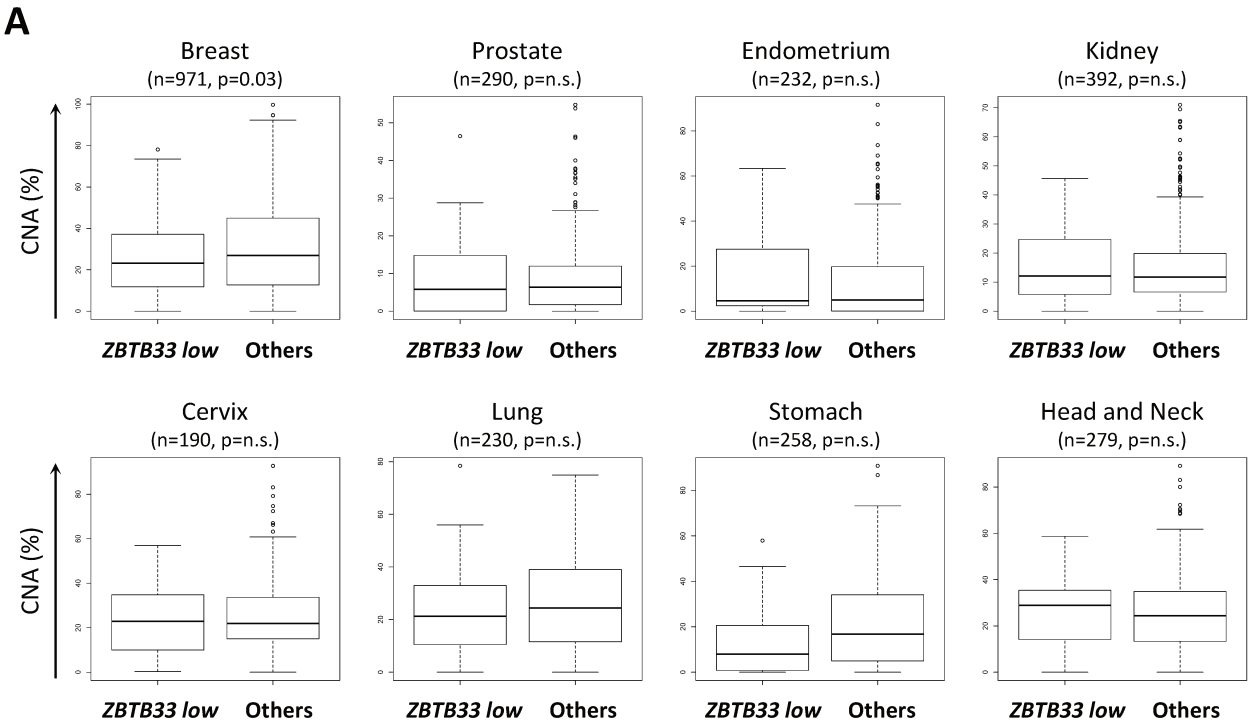
Cancer Res 2017;77:62-73. Published OnlineFirst November 4, 2016.

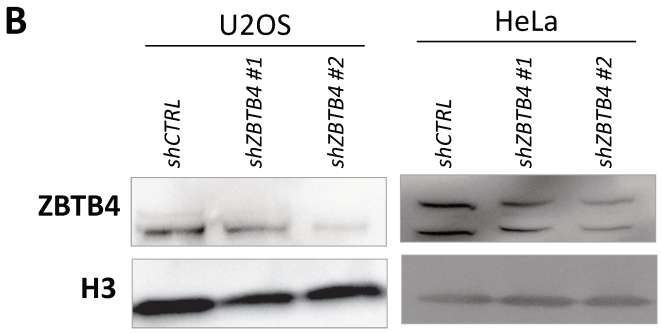
Updated version	Access the most recent version of this article at: doi:10.1158/0008-5472.CAN-16-1181
Supplementary Material	Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2016/11/04/0008-5472.CAN-16-1181.DC1

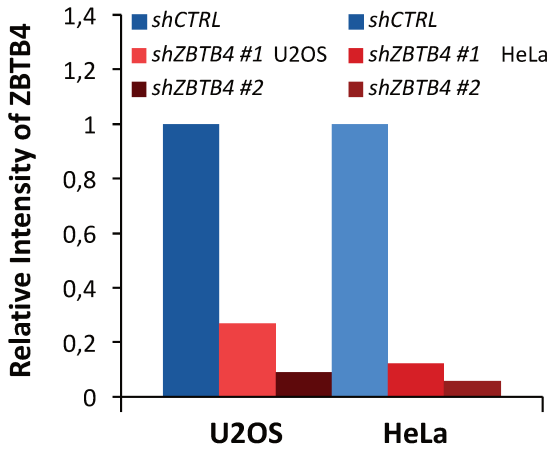
Cited articles	This article cites 43 articles, 14 of which you can access for free at: http://cancerres.aacrjournals.org/content/77/1/62.full.html#ref-list-1

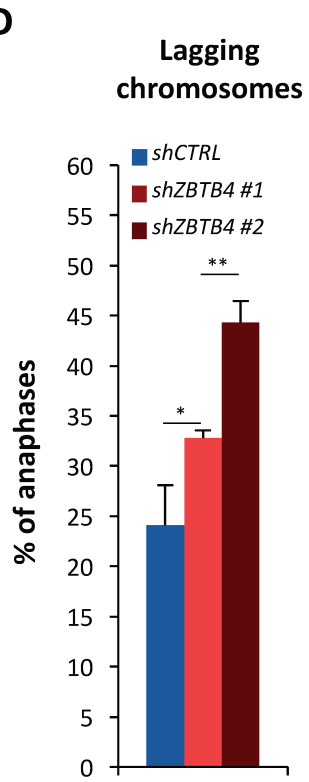
E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.

Downloaded from cancerres.aacrjournals.org on April 19, 2017. © 2017 American Association for Cancer Research.









E

Sample	Total cells observed	Bi- and polynucleated cells	% bi- and polynucleated cells	Chi-square proportion test
HCT116 shControl	n=266	n=19	7.1%	p=0.029
HCT116 shZBTB4	n=457	n=62	13.6%	

Figure S1.

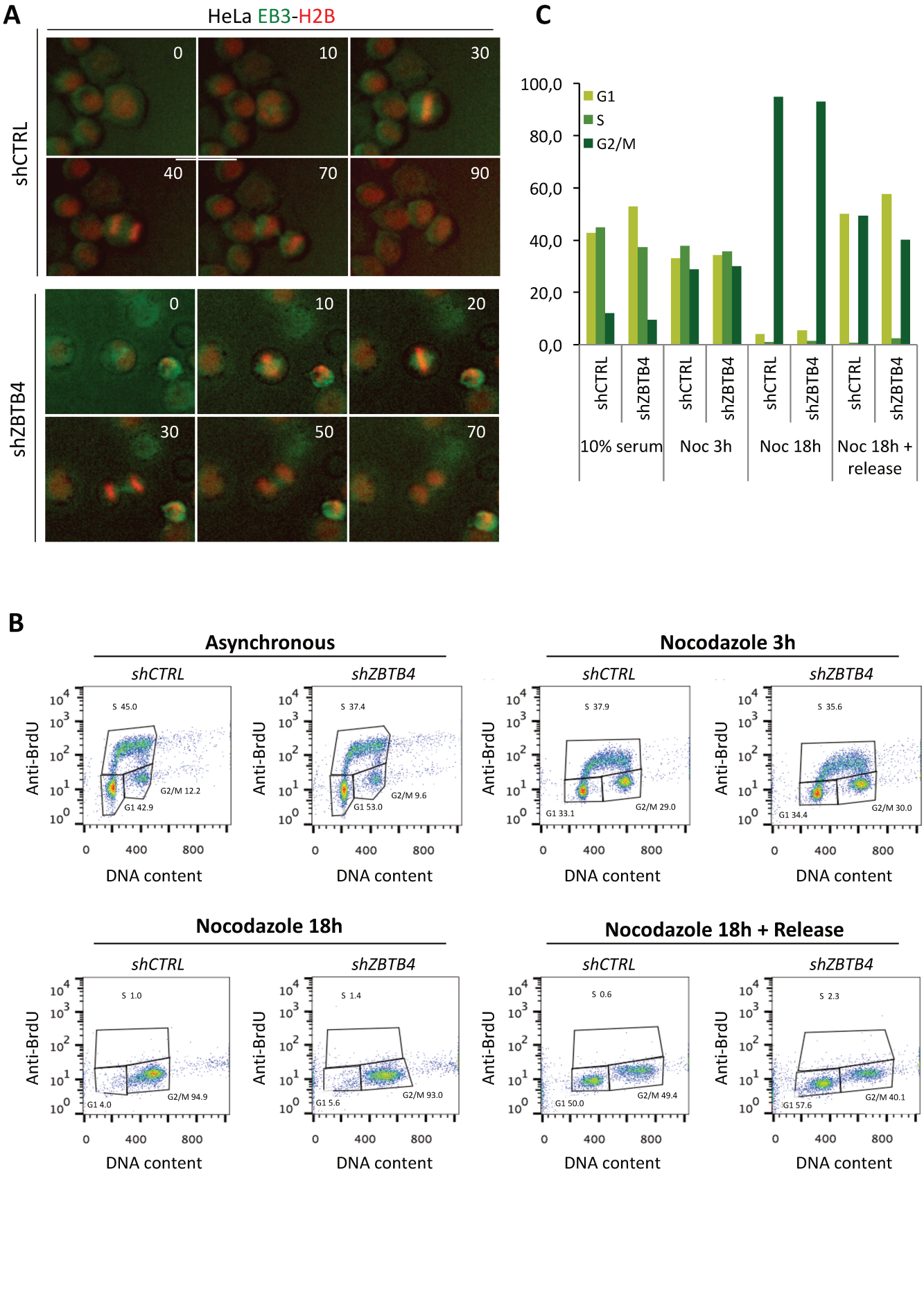

A
HeLa EB3-H2B
shCTRL
0
10
30
40
70
90
shZBTB4
0
10
20
30
50
70
C
100,0
80,0
60,0
40,0
20,0
0,0
G1
S
G2/M
shCTRL
shZBTB4
10% serum
Noc 3h
Noc 18h
Noc 18h + release
B
Asynchronous
Nocodazole 3h
Nocodazole 18h
Nocodazole 18h + Release
shCTRL
shZBTB4
Anti-BrdU
DNA content
S 45.0
G2/M 12.2
G1 42.9
S 37.4
G2/M 9.6
G1 53.0
S 37.9
G1 33.1
G2/M 29.0
S 35.6
G1 34.4
G2/M 30.0
S 1.0
G1 4.0
G2/M 94.9
S 1.4
G1 5.6
G2/M 93.0
S 0.6
G1 50.0
G2/M 49.4
S 2.3
G1 57.6
G2/M 40.1

Figure S2.

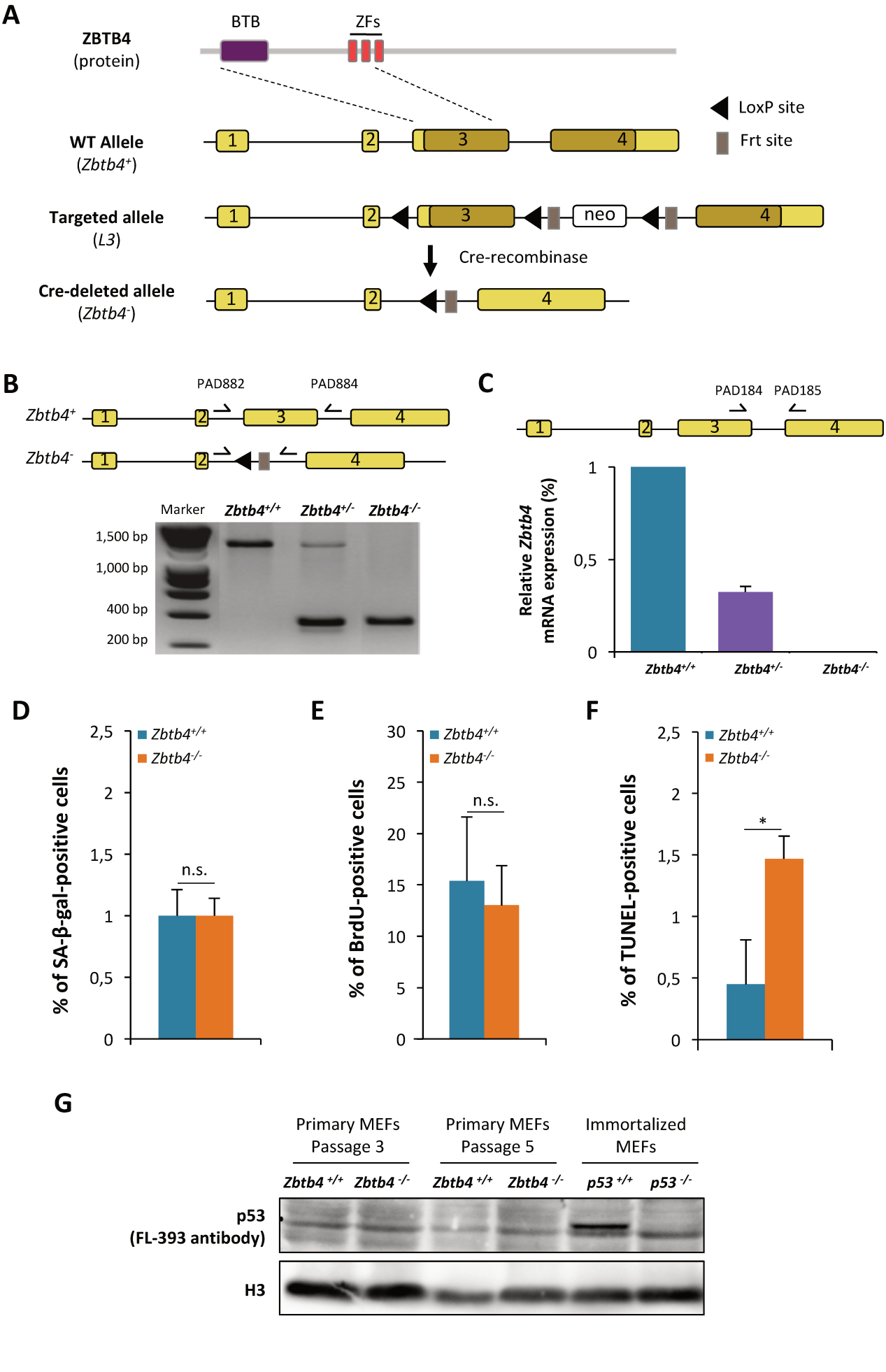
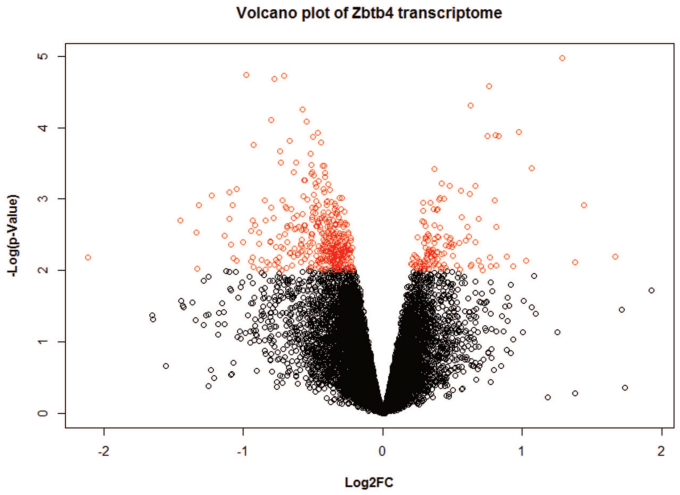

A
ZBTB4
(protein)
BTB
ZFs
LoxP site
Frt site
WT Allele
(Zbtb4+)
Targeted allele
(L3)
neo
Cre-recombinase
Cre-deleted allele
(Zbtb4-)
B
PAD882
PAD884
Zbtb4+
Zbtb4-
Marker
Zbtb4+/+
Zbtb4+/-
Zbtb4-/-
1,500 bp
1,000 bp
400 bp
200 bp
C
PAD184
PAD185
Relative Zbtb4 mRNA expression (%)
D
% of SA-β-gal-positive cells
n.s.
E
% of BrdU-positive cells
n.s.
F
% of TUNEL-positive cells
*
G
Primary MEFs Passage 3
Primary MEFs Passage 5
Immortalized MEFs
Zbtb4 +/+
Zbtb4 -/-
p53 +/+
p53 -/-
p53
(FL-393 antibody)
H3

Figure S3.

A



B **Down in *Zbtb4*$^{-/-}$ MEFs**

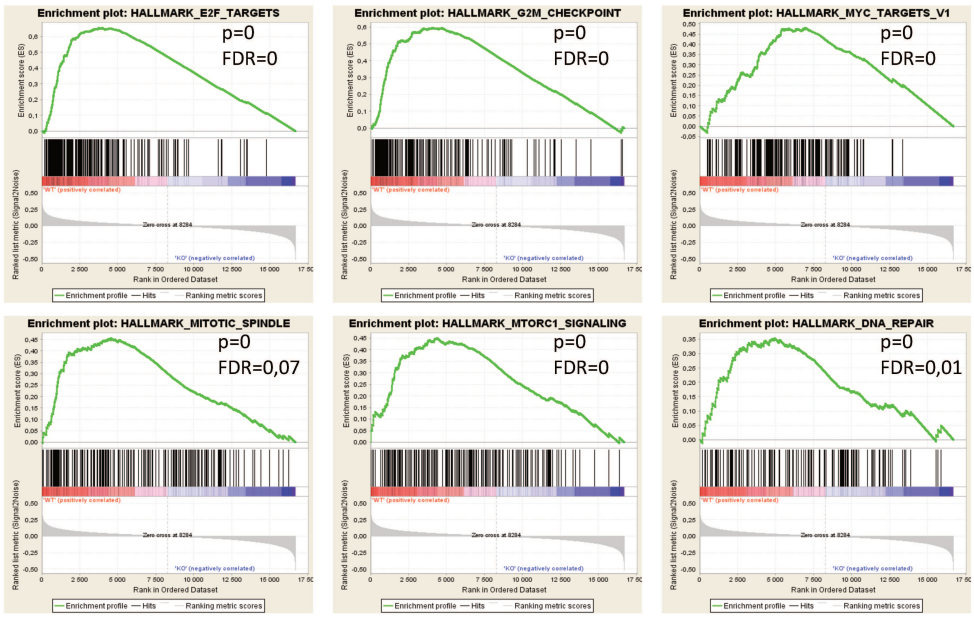


Figure S4.

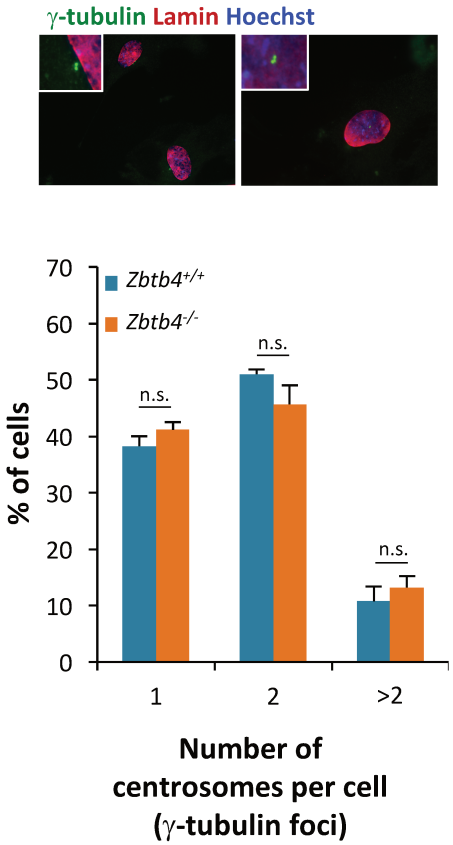

γ-tubulin Lamin Hoechst
70
60
50
40
30
20
10
0
% of cells
Zbtb4+/+
Zbtb4-/-
n.s.
n.s.
n.s.
1
2
>2
Number of centrosomes per cell (γ-tubulin foci)

Figure S5.

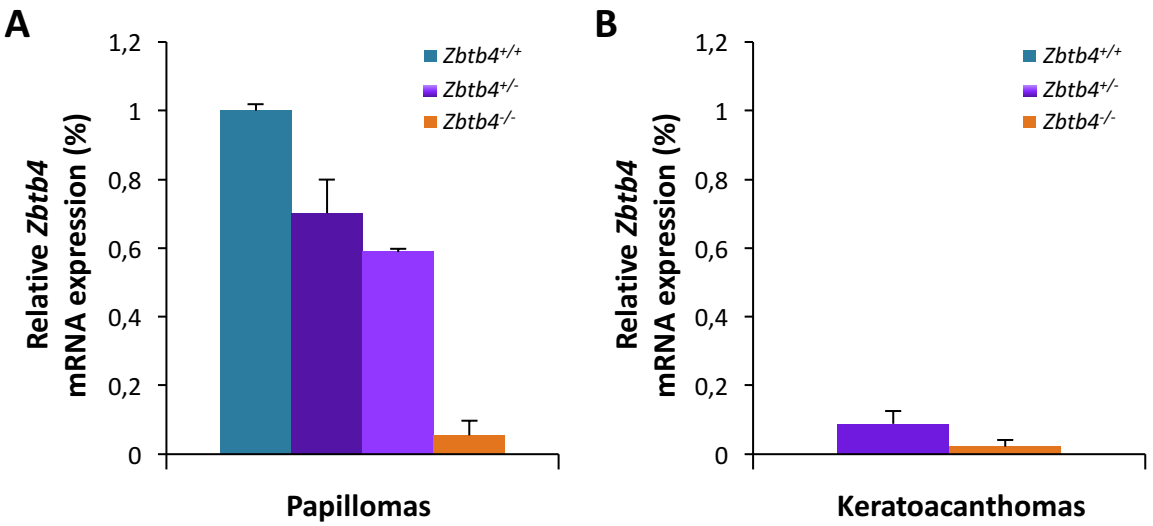

A
1,2
1
0,8
0,6
0,4
0,2
0
Relative Zbtb4 mRNA expression (%)
Zbtb4+/+
Zbtb4+/-
Zbtb4-/-
Papillomas
B
1,2
1
0,8
0,6
0,4
0,2
0
Relative Zbtb4 mRNA expression (%)
Zbtb4+/+
Zbtb4+/-
Zbtb4-/-
Keratoacanthomas

Figure S6.

Supplementary methods

Genotyping

PCR was used to genotype the *Zbtb4*$^{-/-}$ mice using genomic DNA extracted from the tail or the brain. The genomic DNA was extracted by Proteinase K (Roche) digestion followed by an isopropanol precipitation. Genomic DNA was used for PCR with the following primers using GoTaq (Promega) according to the manufacturer's instructions; PAD882: 5'-TGCCCTTTGCTGTGGCCTGA-3 and PAD884: 5'-GCACCCTGGCTCCTTCCAGC-3'.

Immunofluorescence and confocal imaging

All coverslips were incubated 1 hour at room temperature with 0.1% bovine serum albumin (BSA) and next incubated 1 hour at RT with the antibodies directed against the following antigens: mouse anti-phospho-histone H3-Ser10 (Cell Signaling, 9706, 1:100), rat anti-BrdU (Abcam, ab6326, 1:200), rabbit anti-cleaved caspase 3 (Cell Signalling, 9661, 1:500), mouse anti-alpha-Tubulin (Abcam, ab7291, 1:2000), mouse anti-gamma-Tubulin (Abcam, ab11316, 1:500), rabbit anti-LAMIN B1 (Abcam, ab16048, 1:500), mouse anti-BUBR1 (Thermo Scientific, MA1-16577, 1:200). DNA was visualized with Hoechst 33342 (Sigma, 0.5 µg/ml) or DAPI in Vectashield (Vector Laboratories).

Images were collected with either a LEICA DMI6000 B or a ZEISS LSM 710 Laser Scanning Confocal microscope with a 63×-immersion oil objective. Image capture was controlled by MetaMorph® Microscopy Automation & Image Analysis Software (LEICA) or Zen Software (ZEISS). For all samples, an optimal setting of the laser power and PMT voltage was chosen to avoid pixel saturation and minimize photobleaching. The settings were kept constant so that valid

comparisons could be made between measurements from different samples. FiJi (ImageJ) software was used for the 3D quantification analysis.

RT-qPCR primers

ZBTB4 : Forward (F) 5'-AGGAAGTACCCCTGCCGCTA-3', Reverse (R) 5'-TTGTAGCCTCCATTGGGTGT-3'. *BUB1B* : (F) 5'-TAGGGCGTTTATGCAATGAGC-3', (R) 5'-TCCTGAAATATCGCATCTGCTTT-3'. *MAD2L1* : (F) 5'-GTTCTTCTCATTCGGCATCAACA-3', (R) 5'-GAGTCCGTATTTCTGCACTCG-3'. *BUB3* : (F) 5'-CTGCGCCTTCTACGATCCAA-3', (R) 5'-GCATCATGGGTCCCAACAAGAT-3'. *CCNB1* : (F) 5'-AATAAGGCGAAGATCAACATGGC-3', (R) 5'-TTTGTTACCAATGTCCCCAAGAG-3'. *B2M* : (F) 5'-GGCTATCCACGTACTCCAAA-3', (R) 5'-CGGCAGGCATACTCATCTTTTT-3'. *RPLP0 :* (F) 5'-TCCAGAGGCACCATTGAAATT-3', (R) 5'-TCGCTGGCTCCCACCTT-3'. *TBP:* (F) 5'-CCACTCACAGACTCTCACAAC-3', (R) 5'-CTGCGGTACAATCCCAGAACT-3'.

Ccnb1: (F) 5'-GCGTGTGCCTGTGACAGTTA-3', (R) 5'-CCTAGCGTTTTTGCTTCCCTT-3'. *Ccnb2* : (F) 5'-GCCAAGAGCCATGTGACTATC-3', (R) 5'-CAGAGCTGGTACTTTGGTGTTC-3'. *Bub1 :* (F) 5'-AGAATGCTCTGTCAGCTCATCT-3', (R) 5'-TGTCTTCACTAACCCACTGCT-3'. *Bub1b* : (F) 5'-GAGGCGAGTGAAGCCATGT-3', (R) 5'-TCCAGAGTAAAAGCGGATTTCAG-3'. *Mad2L1*: (F) 5'-GTGTTCTCCGTTCGATCTAGTG-3', (R) 5'-CCACGCTGATACAAAATGCTG-3'. *Cdc25c :* (F) 5'-GGCAAACCTAAGCATTCTGTCG-3', (R) 5'-CCAGAGGTCCAGATGAATCCA-3'. *Chk1* : (F) 5'-ATCGTGAACGCTTACTGAACAA-3', (R) 5'-CTGACAGCTATCACTGGGCT-3'. *Plk1* : (F) 5'-TGTAGTTTTGGAGCTCTGTCG-3', (R) 5'-

TCCCTGTGAATGACCTGATTG-3'. *Plk4* : (F) 5'-GGATCGAGGACTTTAAGGTTGG-3', (R) 5'-TCTCTGTACCATTCCAGCTTTG-3'. *Cenpn* : (F) 5'-CACAGTATTCACAGCCAAACC-3', (R) 5'-CAATCTGATGGTGTTTGCTGG-3'.

Vector Virus Production and Infection

ZBTB4 shRNA plasmid vectors (Sigma) contain the following hairpin sequences:

sh*ZBTB4*#1 (clone NM_020899.2-1293s1c1)

CCGGCGCTATTGTGAGAAAGTGTTTCTCGAGAAACACTTTCTCACAATAGCGTTTTT

sh*ZBTB4* #2 (clone NM_020899.2-1389s1c1)

CCGGGAGACCTTTGTCACTTACTATCTCGAGATAGTAAGTGACAAAGGTCTCTTTTT

Western blotting

Western blots were performed using the following antibodies: rabbit anti-ZBTB4 (homemade #120), mouse anti-BUB1B (Thermo Scientific, MA1-16577, 1:2000), rabbit anti-MAD2L1 (Abcam, ab97777, 1:500), mouse anti-Cyclin B1 (Thermo Scientific, MA5-14327, 1:500), rabbit anti-Hitone H3 (Abcam, ab1791, 1:10000), rabbit anti-p53 (Santa Cruz, sc-6243, 1:500), mouse anti-p21Cip1 (Santa Cruz, sc-6246, 1:1000), rabbit anti-p19ARF (Novus Biologicals, NB200-106 1:500), rabbit anti-Phospho Rb (Cell Signaling, 9308, 1 :1000), rabbit anti-p16Ink4a (Cell Signaling, 4824, 1:1000), mouse anti-CDK4 (Cell Signaling, 2906, 1:2000). Secondary HRP donkey anti-mouse IgG (Jackson ImmunoResearch, 715-035-150, 1 :10000) and HRP donkey anti-rabbit IgG (Jackson ImmunoResearch , 711-035-152, 1 :10000) Bands were visualized by horseradish peroxidase labeled antibodies and SuperSignal West Dura Extended Duration Substrate (Thermo Scientific). FiJi (ImageJ) software was used for the intensity quantification analysis.

Supplementary Figure Legends

Figure S1. Controls related to Figure 1.

A, Boxplots showing the percentage of the genome affected by Copy Number Alterations (CNA) in the indicated tumor types, comparing the lowest decile for ZBTB33 expression ("ZBTB33 low") to the remaining tumors ("Others"); p-value calculated by T-test, n.s: no statistically significant. **B,** Western blotting on whole-cell protein lysates from U2OS and HeLa expressing sh*CTRL*, sh*ZBTB4* #1 or sh*ZBTB4* #2. H3: Histone H3. **C,** Quantification of panel B **D,** Quantification of anaphase U2OS cells showing lagging chromosomes after ZBTB4 knockdown with *shZBTB4* #1 and #2. **E,** ZBTB4 knockdown in HCT116 cells leads to increased polynucleation.

Figure S2. Illustrations and controls related to Figure 2.

A, Representative mitosis of HeLa cells stably expressing EB3-GFP, H2B-mCherry and sh*CTRL* or sh*ZBTB4* captured by time-lapse videomicroscopy at 40X. Scale bar = 10 μm. **B,** Representative gating of flow cytometry cell cycle analysis by PI and Br-dUTP staining, for cells of the indicated genotypes, before, during, and after nocodazole block. **C,** Summary of cell cycle analysis by PI and Br-dUTP staining showing percentage of cells in G1, S and G2/M (non gating cells excluded).

Figure S3. Controls related to Figure 3.

A, Diagram of the ZBTB4 protein indicating the BTB/POZ domain and the DNA binding domain, composed of 3 zinc fingers (ZF). The exons of the *Zbtb4* gene are indicated in yellow and the coding exons appear in dark yellow. The portion of ZBTB4 protein encoded by exon 3 is

indicated. Exon 3 was replaced by a cassette introducing LoxP sites and the *neo* gene to generate the targeted allele. The knock-out allele of *Zbtb4* lacks the entire exon 3. **B,** Primers used in PCR for genotyping are indicated on top. The figure shows a representative PCR on genomic DNA purified from brain of wild-type, *Zbtb4*$^{+/-}$ and *Zbtb4*$^{-/-}$ mice. The top band of the PCR is derived from the wild-type allele, whereas the lower band is derived from the targeted allele. **C,** Primers used in PCR for mRNA expression analysis are indicated on top. The relative expression level of *Zbtb4* mRNA in the brain of wild-type, *Zbtb4*$^{+/-}$ and *Zbtb4*$^{-/-}$ adult mice is shown below. Expression values are normalized to a housekeeping gene and expression level in wild-type set up to 1 (n=3). **D,** *Zbtb4*$^{+/+}$ and *Zbtb4*$^{-/-}$ MEF cultures at passages 5 were stained for senescence-associated-beta-galactosidase activity (SA-β-gal) and quantified. **E,** Quantification of BrdU-positive cells in *Zbtb4*$^{+/+}$ and *Zbtb4*$^{-/-}$ MEF cultures at passage 3. **F,** Flow cytometry analysis of the TUNEL-positive population in *Zbtb4*$^{+/+}$ and *Zbtb4*$^{-/-}$ MEFs. **G,** Western blotting for p53 on total cellular extracts from *Zbtb4*$^{+/+}$ and *Zbtb4*$^{-/-}$ MEFs and the indicated controls.

Figure S4. Additional data related to Figure 4.

A, Volcano plot, Log2FC against pvalue, of *Zbtb4*$^{-/-}$ vs WT MEFs, red dots indicates differentially expressed genes (pvalue ≤ 0.01). **B,** Gene sets significantly deregulated in *Zbtb4*$^{-/-}$ MEFs, as determined using available genesets and GSEA analysis.

Figure S5. Additional data related to Figure 5.

Representative figures and quantification of γ-tubulin foci per cells.

Figure S6. Additional data related to Figure 6.

A, The relative expression level of *Zbtb4* mRNA in papillomas from *Zbtb4*$^{+/+}$, *Zbtb4*$^{+/-}$ and *Zbtb4*$^{-/-}$ mice. Expression values are normalized to a housekeeping gene and expression level in wild-type set up to 1. **B,** As in A., but for keratoacanthomas.

ANNEXE 3

Mutations in DNMT3B affecting its catalytic activity characterize the ICF1 patients (Xu et al., 1999; Hagleitner et al., 2007; Weemaes et al., 2013). By unknown mechanisms, mutant-DNMT3B interferes with lymphoid-specific pathways leading to immunodeficiency, which is more severe in ICF1 patients compared to the other subtypes (Weemaes et al., 2013; Sterlin et al., 2016). Although recent studies showed that DNMT3B methylates gene bodies of actively transcribed genes (Yang et al., 2014; Baubec et al., 2015; Neri et al., 2017), the impact of mutations in DNMT3B in this process has not been investigated yet.

In collaboration with Matarazzo's team, we investigated the impact of mutations in DNMT3B through transcriptomic and epigenomic approaches in lymphoblastoid cell lines derived from ICF1 patients. We report that mutations in DNMT3B altered DNA methylation of intragenic regions impairing multiple aspects of transcriptional regulation, including alternative TSS usage, antisense transcription and splicing. These alterations associated preferentially with changes in intragenic H3K4me3 levels and, to a lesser extend, to H3K27me3 and H3K36me3 levels. Altered transcription was observed for genes relevant for the disease, including the memory-B cell marker CD27 and PTPRC genes, providing new insights into the aetiology of ICF1 syndrome. Our study also provides new mechanistic insights into how the recruitment of mutant-DNMT3B at gene bodies can modulate sense- and antisense-transcription as well as exon splicing, through interaction with RNA molecules and RNA-binding proteins. This study contributes to shed light on the role of DNMT3B in the DNA methylation of gene bodies and its impact on transcriptional output of the human genome.

I contributed to this study by comparing the methylation results we obtained in ICF1 blood samples by HM450K (Article 2) with the methylation results obtained by Gatto and colleagues on LCL derived from ICF1 patients by RRBS. This comparison showed that most of the hypomethylated regions identified by Gatto and colleagues (more that 75%) were found hypomethylated also in our analysis.

This work has been published in February 2017, in *Nucleic Acids Research*.

Nucleic Acids Research, 2017 **1**
doi: 10.1093/nar/gkx163

ICF-specific DNMT3B dysfunction interferes with intragenic regulation of mRNA transcription and alternative splicing

Sole Gatto[1,2,†], Miriam Gagliardi[1,3,*,†], Monica Franzese[3], Sylwia Leppert[1], Mariarosaria Papa[1], Marco Cammisa[1,4], Giacomo Grillo[5], Guillame Velasco[5], Claire Francastel[5], Shir Toubiana[6], Maurizio D'Esposito[1,7], Claudia Angelini[3] and Maria R. Matarazzo[1,*]

[1]Institute of Genetics and Biophysics 'Adriano Buzzati-Traverso', CNR, Naples 80131, Italy, [2]Sanford Burnham Prebys Medical Discovery Research Institute, La Jolla, CA, USA, [3]Institute for Applied Mathematics 'Mauro Picone', CNR, Naples 80131, Italy, [4]Department of Environmental, Biological and Pharmaceutical Sciences and Technologies, Second University of Naples, Caserta 81100, Italy, [5]CNRS UMR7216, Epigenetics and Cell Fate, Université Paris Diderot, Sorbonne Paris Cité, Paris 75205, France, [6]Molecular Medicine Laboratory, Rambam Health Care Campus and Rappaport Faculty of Medicine, Technion, Haifa, Israel and [7]IRCCS Neuromed, Pozzilli, Italy

Received July 01, 2016; Revised February 07, 2017; Editorial Decision February 24, 2017; Accepted February 28, 2017

ABSTRACT

Hypomorphic mutations in DNA-methyltransferase DNMT3B cause majority of the rare disorder Immunodeficiency, Centromere instability and Facial anomalies syndrome cases (ICF1). By unspecified mechanisms, mutant-DNMT3B interferes with lymphoid-specific pathways resulting in immune response defects. Interestingly, recent findings report that DNMT3B shapes intragenic CpG-methylation of highly-transcribed genes. However, how the DNMT3B-dependent epigenetic network modulates transcription and whether ICF1-specific mutations impair this process remains unknown. We performed a transcriptomic and epigenomic study in patient-derived B-cell lines to investigate the genome-scale effects of DNMT3B dysfunction. We highlighted that altered intragenic CpG-methylation impairs multiple aspects of transcriptional regulation, like alternative TSS usage, antisense transcription and exon splicing. These defects preferentially associate with changes of intragenic H3K4me3 and at lesser extent of H3K27me3 and H3K36me3. In addition, we highlighted a novel DNMT3B activity in modulating the self-regulatory circuit of sense-antisense pairs and the exon skipping during alternative splicing, through interacting with RNA molecules. Strikingly, altered transcription affects disease relevant genes, as for instance the memory-B cell marker *CD27* and *PTPRC* genes, providing us with biological insights into the ICF1-syndrome pathogenesis. Our genome-scale approach sheds light on the mechanisms still poorly understood of the intragenic function of DNMT3B and DNA methylation in gene expression regulation.

INTRODUCTION

DNA methylation plays an important role in epigenetic signaling, having an impact on gene regulation, chromatin structure, development and disease. Generally, most mammalian genomes are largely methylated except at active or 'poised' promoters, enhancers and CpG islands, where it has a repressive effect. Nevertheless, gene body DNA methylation has been associated with high expression levels (1).

DNA methylation is established and maintained by the combined function of three active DNA methyltransferases DNMT3A, DNMT3B and DNMT1 (2). Although it has been largely studied, much remains unknown regarding how genomic DNA methylation patterns are determined in human cells, and which are the mechanisms that guide recruitment and activity of DNMTs *in vivo* (1). Mouse models suggest that although exhibiting overlapping functions, DNMT3A and DNMT3B have unique expression patterns and genomic targets during development (3–5). In line with

*To whom correspondence should be addressed. Tel: +39 0816132426/424; Fax: +39 0816132606; Email: maria.matarazzo@igb.cnr.it
Correspondence may also be addressed to Dr. Miriam Gagliardi. Tel: +39 0816132424; Fax: +39 0816132606; Email: miriam.gagliardi@igb.cnr.it
†These authors contributed equally to this work as first authors.

© The Author(s) 2017. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

this view, latest genome-wide studies showed that gene body DNA methylation patterns of highly expressed genes is dependent on DNMT3B activity (6–8). DNA methylation in transcribed regions might potentially silence alternative promoters, antisense transcripts, transcription factor binding sites, retrotransposon elements (LINEs, SINEs, LTRs and other retroviruses) and other functional elements to ensure the efficiency of transcription (9–11).

In addition, it is increasingly clear that DNA methylation plays a role in the processing of mRNAs during transcription modulating the elongation or splicing (12–14). Accordingly, DNA methylation level has been recently found associated with inclusion rate of alternative exons (15,16). Modulation of alternative splicing might represent a specific and evolutionary conserved activity of *de novo* DNA methyltransferases, as suggested by knockdown studies of DNA methyltransferase 3 (dnmt3) in honeybee (17). A kinetic model, in which epigenetic modifications affect the rate of transcriptional elongation, and/or a recruitment model, in which adaptor proteins bind to epigenetic modifications recruiting splicing factors, have been proposed (18). However, how exactly chromatin factors influence the DNMT3B activity and methylation profiles at transcribed regions remains to be elucidated.

In humans, hypomorphic *DNMT3B* mutations are sufficient to cause majority of the rare autosomal recessive disorder Immunodeficiency, Centromere instability and Facial anomalies (ICF) syndrome (MIM 242860) cases, reported as ICF type1 (19–21). Patients are characterized by DNA hypomethylation and decondensation of specific heterochromatic and euchromatic regions, and show alterations in tissue-specific gene silencing (22,23). ICF1-specific DNA methylation defects give rise to severe chromosomal rearrangements only in lymphocytes, probably acting in the onset of immunological phenotype. Defective steps of B-cell terminal differentiation might contribute to the agammaglobulinemia in ICF syndrome, given that ICF peripheral blood only contain naive B cells, while memory and gut plasma cells are absent (24). Most ICF1 patients carry missense mutations in or near the catalytic domain of DNMT3B (21). Nonsense mutations always occur as compound heterozygous, highlighting that the DNMT3B protein is essential for life, according to mouse models (3,5,25). Mutations severely perturb the DNA methylation profile at satellite 2 and 3 of juxtacentromeric heterochromatin and at telomeric/subtelomeric repeats, where it associates with chromosomal instability and abnormal shortening of telomeres, respectively (26–28). Treatment of human lymphocytes with the DNA methylation inhibitor 5-azacytidine (5-AzaC) induces ICF-like pericentromeric chromosomal abnormalities and missegregation of chromosomes 1, 16 and 9, suggesting that it is the DNA hypomethylation to trigger these anomalies (29,30).

Defects in lymphoid-specific pathways directly associated with the impaired DNMT3B activity have not been identified yet. The main reason is that alterations of DNA methylation pattern were previously pursued through candidate gene approaches based on promoter region analysis, which mostly failed to clarify the molecular pathogenesis of ICF syndrome (22,31,32). In this view, we carried out an integrated genome-wide study by combining maps of DNMT3B binding sites, differentially methylated CpGs, H3K36me3, H3K4me3 and H3K27me3 histone marks and differentially expressed genes (DE-genes) in ICF1 patients derived B lymphoblast cell lines (B-LCLs). Examining the genome-wide effect of DNMT3B dysfunction in this disease-relevant model, we identified its target genes and gained a wider understanding of the protein function in modulating the intragenic regulation of transcription.

All together, our observations demonstrate that DNMT3B controls the proper mRNA transcription through the regulation of transcript-isoform TSS usage, cryptic-TSS expression, sense–antisense transcription and alternative exon splicing, by affecting the intragenic epigenetic signature and/or by directly interacting with pre-mRNA molecules. Since the in-depth epigenomic and transcriptomic defects in ICF1 syndrome have not been previously described, we provided novel insights into the molecular mechanisms associated to ICF1-specific DNMT3B mutations.

MATERIALS AND METHODS

Cell culture

Epstein Barr Virus immortalized lymphoblast cell lines (B-LCL) were derived from peripheral blood mononuclear cells (PBMCs) of two unrelated ICF patients with missense mutations in DNMT3B, ICF1p1 (female, compound heterozygous A603T and intron 22 G to A mutation resulting in insertion of three amino acids (STP) in DNMT3B, Coriell Cell Repository) and ICF1p2 (male, compound heterozygous V699G and R54X mutation in DNMT3B, provided by Dr R.S. Hansen). Control EBV-immortalized B-LCLs were from unaffected unrelated individuals (Ctrl1 and Ctrl4) and phenotypically normal parents of ICF1p1 patient (Ctrl2 and Ctrl3, female and male respectively, Coriell Cell Repository). The B-LCLs from the ICF patients' parents had a similar passage history as the patients' B-LCLs. We confirmed the genome-wide results in additional B-LCLs (kindly provided by Dr Francastel, INSERM) deriving from ICF1 patients (ICF1pT and ICF1pY, with homozygous mutations D817G/D817G and T775I/T775I, respectively).

For drug treatment, 1×10^6 control cells were plated and treated with 1 μM 5-AzaC (Sigma) for 24 h. Then, the medium containing 5-AzaC was removed and replaced with RPMI + FBS 10% and cells were harvested at different time points (48, 72 and 120 h) for further analyses.

For siRNA knockdown, endogenous DNMT3B was knocked down using the pre-designed siRNA Smart Pool (Dharmacon), while siGLO RNAi control was used as negative control (Dharmacon). The siRNA transfection was performed with Nucleofector transfection (Amaxa, Nucleofector, Kit V) according to the manufacturer's instructions. SiRNA treated cells were processed at 24, 48, 72 and 120 h post-transfection.

DNA extraction was performed using Wizard Genomic DNA (Promega). Total RNA was extracted using the TRIzol reagent (Life Technologies) and purified from residual DNA contaminant using the TURBO DNA-*free*TM Kit

(Ambion, Austin, TX, USA) according to the manufacturer's instructions. RNA quality was checked on the Agilent 2100 Bioanalyzer and quantity was measured using the Qubit instrument (Life Technologies). Whole cell lysates were extracted with ice-cold lysis buffer (100 mM Tris–HCl, pH 8, 1% NP-40, 0.2% SDS, 140 mM NaCl, 20 mM EDTA pH 8) supplemented with protease inhibitor cocktail (Roche Applied Sciences).

Reduced representation bisulfite sequencing (RRBS) and data processing

RRBS libraries were prepared and sequenced according to Illumina's instructions. High quality reads were aligned to the reference genome using the Bismark aligner (33) and methylation call was performed with methylation extractor script. We quantified the methylation level of 1 291 564 CpGs spanning a number of different functional regions in the genome and belonging to different CpG contexts, and with sequencing depth of at least 10 reads in each condition (the median depth was >23 reads among these CpGs in each condition). A detailed description of tools used for data processing is reported in Supplementary Experimental Procedures. Gene-specific DNA methylation level was evaluated by MethylMiner Methylated DNA Enrichment Kit (Invitrogen) and bisulfite sequencing using EpiTech Bisulfite Kit (Qiagen) according to the manufacturer's instructions. Gene-specific RT-qPCR were performed and data were expressed and normalized as $\log_2$(methylated/unmethylated), representing the enrichment of the target region in the methylated compared with the unmethylated fraction. Primer sequences are reported in Supplementary Table S1.

RNA sequencing and data processing

RNA isolation and library construction were performed according to Illumina's instructions. After mapping the reads to the reference genome using TopHat2 (34), differential gene expression between two different cell conditions was calculated using DESeq2 implemented in R (35). The relative isoform abundance estimation was performed using BitSeq (36) and the measurement of exon usage was carried out using DEXSeq (37), according to the script published in (38). Gene ontology analysis was performed using DAVID Bioinformatics Resource (39), while enrichment of specific pathways was analyzed using Ingenuity pathway analysis (IPA; http://www.ingenuity.com). A detailed description of tools used for data processing is reported in Supplementary Experimental Procedures.

Quantitative real time PCR

Total RNA from B-LCLs was reverse-transcribed using iScript cDNA Synthesis kit (Bio-Rad San Diego, CA, USA). Quantitative real-time PCR (qPCR) was performed using SsoAdvanced™ universal SYBR® Green supermix (Bio-Rad) on Bio-Rad iCycler according to the manufacturer's protocols. The ΔΔCt method was used to determine relative quantitative levels. *GAPDH* or *ACTB* genes were used to normalize the data, considering their unaltered expression in ICF1 samples (Supplementary Figure S1). Primer sequences for gene expression analysis are shown in Supplementary Table S1. All qPCR assays were performed according to the MIQE guidelines (40), and relative informations are provided in the supporting material Supplementary Table S2.

ChIP sequencing and data processing

Chromatin immunoprecipitation was performed as previously described (41). Suitable amount of chromatin was incubated with 5 μg of the indicated antibodies against H3K27me3, H3K4me3, H3K36me3, DNMT3A, CTCF (Abcam) and anti-DNMT3B (Diagenode). Immunoprecipitated complexes were recovered with protein A sepharose (Pharmacia), washed with low and high salt buffers, reverse-crosslinked, and purified. Immunoprecipitated samples were used for qPCR and results were expressed as site occupancy (normalization relative to a negative region amplified in the same sample) to exclude technical variations between samples. We first normalize to the input amount and then, the obtained %input is used to calculate the enrichment over %input of a negative region (*GAPDH* promoter, *MB* exon2 or *ACTB* promoter; Supplementary Figure S1). Mean of results deriving from independent ChIP replicates were reported. Primers sequences are reported in Supplementary Table S1. Immunoprecipitated samples for H3K4me3, H3K27me3, H3K36me3 and DNMT3B were used for preparing libraries and sequencing. 50bp reads were first analyzed for quality filtering and aligned to the reference genome. All the reported results were obtained by normalizing for the total library size. A detailed description of tools used for mapping, peak calling and identification of differentially enriched peaks is reported in Supplementary Experimental Procedures.

RNA immunoprecipitation (RIP) sequencing and data processing

Native and cross-linked RNA immunoprecipitation experiments were performed using the Magna RIP™ RNA-Binding Protein Immunoprecipitation Kit (Millipore) and the Magna Nuclear RIP (Cross-Linked) RNA-Binding Protein Immunoprecipitation Kit (Millipore), respectively, according to the manufacturer's instructions. Antibodies for RIP assays were anti-DNMT3B (Diagenode), anti-DNMT3A (Abcam) and anti-hnRNP-LL (Aviva). Immunoprecipitated fractions were retro-transcribed and cDNAs were used for qPCR with gene-specific primers and/or for preparing libraries. Sequences (50 bp reads) were analyzed for quality filtering and aligned to the reference genome using TopHat2 (34). FPKMs were calculated using Cufflinks (42) on Galaxy (43) after selecting default parameters. A detailed description of tools used for data analysis is reported in Supplementary Experimental Procedures.

Flow cytometry analysis

Determination of cell surface expression of CD45RABC and CD45RO molecules was carried out by cytofluorimetric analysis using the FACS ARIA cell-sorting system and

DIVA software (BD Biosciences). Direct immunofluorescence was performed using PerCP and FITC mouse anti-human CD45RABC and CD45RO antibodies respectively, along with the appropriate mouse IgG isotype controls (BioLegend). Staining, washing and analysis were performed following the manufacturer's recommendations.

Co-immunoprecipitation

Co-immunoprecipitation experiments were performed using Nuclear Complex Co-IP kit (Active Motif) and following the manufacturer's instructions. The antibodies used were the following: DNMT3B (Diagenode), and hnRNP-LL (Aviva). Input samples were also used as nuclear extract in protein expression assay.

Western blot

Whole cell and nuclear protein lysates were quantified by Bradford protein assay (Bio-Rad). Twenty microgram of total protein lysates were separated by SDS polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to PVDF membranes (Millipore), blocked with 2% BSA in TBS buffer with 0.05% Tween20 and incubated with primary antibodies for 1 h at RT, followed by incubation with an appropriate HRP-conjugated secondary antibody for 1 h at RT. Densitometric analysis was performed using Typhoon Scan and ImageQuant software.

Statistical analysis

Gene expression, gene specific ChIP, RIP and methylation analysis were presented as means ± standard deviations (SD) from at least three independent experiments. Statistical analyses were performed using Student's *t*-test (two-tails). *P*-values were adjusted with BH method and we generally considered the following values as statistically significant: **P*-adj <0.05; ***P*-adj <0.005; ****P*-adj <0.0005. The significance of the overlapping between two gene lists was calculated using the hypergeometric test available in R. Statistics relative to DNA methylation comparison between samples was analyzed using Kolmogorov–Smirnov test and values were corrected with Bonferroni method. CpG methylation clustering diagram between samples and principle component analysis (PCA) of mean methylation levels were performed in R using the Methylkit package (44).

RESULTS

Altered CpG methylation affects intragenic regions in mutant-DNMT3B human cells

Previous candidate gene approaches were unsuccessful in identifying ICF1-specific DNA hypomethylation at CpG island promoters of differentially expressed genes (DE-genes), suggesting either that only few of them are DNMT3B direct targets or that differences in DNA methylation level are outside promoters (22,31). Methylome studies carried out in ICF1 B-LCLs suggested that DNA methylation defects might be more extensive than what supposed until now (45,46). However, these studies did not explore which DNA methylation defects are dependent on mutant-DNMT3B, how they functionally impact on gene expression regulation, and whether this effect occurs influencing other epigenetic marks.

In this light, we combined data from Whole Genome Bisulfite Sequencing (WGBS) (45), Reduced Representation Bisulfite Sequencing [RRBS; (47)], RNA-Seq and ChIP-Seq to map DNMT3B binding sites and histone marks (H3K4me3, H3K36me3 and H3K27me3) in B-LCLs derived from peripheral blood of two unrelated ICF patients with DNMT3B missense mutations, ICF1 patient 1 (ICF1p1) and ICF1 patient 2 (ICF1p2), compared to control B-LCLs. These cells represent a suitable model for ICF studies because of the central role of B cells in abnormal immunoglobulin production in ICF cases and the scarcity of primary B cells from patients, of which only few reach adulthood. Despite cell culturing, B-LCLs maintain highly significant ICF-specific differences in mRNA levels of immunoglobulin genes [(31,32) and the present study] and high frequencies of karyotype anomalies described in mitogen-stimulated ICF lymphocytes (27).

We identified 79 521 and 174 030 hypo-methylated CpGs in ICF1p1 and ICF1p2 respectively (Figure 1A and Supplementary Figure S2A–C). When we compared DNA methylation values from RRBS and the whole-genome bisulfite sequencing (WGBS), previously reported for one ICF1 sample (45), we found that they were significantly concordant (Supplementary Figure S2D). Moreover, because DNA methylation profile may be heavily influenced by immortalization and culture condition, we validated our results by comparing the RRBS results with DNA methylation assessment by Infinium HumanMethylation 450K in PBMCs of ICF1 patients (Dr. Francastel, personal communication; Supplementary Figure S2E).

Besides the expected hypomethylation, ICF1 samples also showed gain of methylation with 75 573 and 35 872 hypermethylated CpGs (ICFp1 and ICFp2 respectively; Supplementary Figure S2A). In contrast to observations in ICF1 iPSCs (48) and similar to a recent report in cancer (49), our results did not show substantial levels of DNA methylation at non-CpG sites in ICF1 or control B-LCLs (data not shown). We identified 2381 in ICF1p1 (1833 hypo- and 548 hyper-DMRs) and 3520 DMRs in ICF1p2 (3053 hypo- and 467 hyper-DMRs), which affected all the chromosomes (Supplementary Figure S3A). Of note, 74% of hypomethylated DMRs in ICF1p1 were shared with ICF1p2, while 13% of hypermethylated DMRs were shared between the two ICF1 samples. As previously reported, pericentromeric heterochromatin (PCH) and repetitive sequences (satellite, LINE, SINE, LTRs, etc.) were considerably hypomethylated in ICF1 samples (Supplementary Figure S3B and data not shown).

To assess the effect of inactivating mutations on the genome-wide binding pattern of DNMT3B protein, we performed ChIP-Seq using an antibody against the endogenous protein. We observed that the ability of mutant-DNMT3B to bind DNA was not impaired in terms of total number of peaks both at genic and intergenic regions (Supplementary Figure S3C), but rather in terms of their extension depending on the DNMT3B variant. Indeed, DNMT3B peaks larger than 100kb were increased in

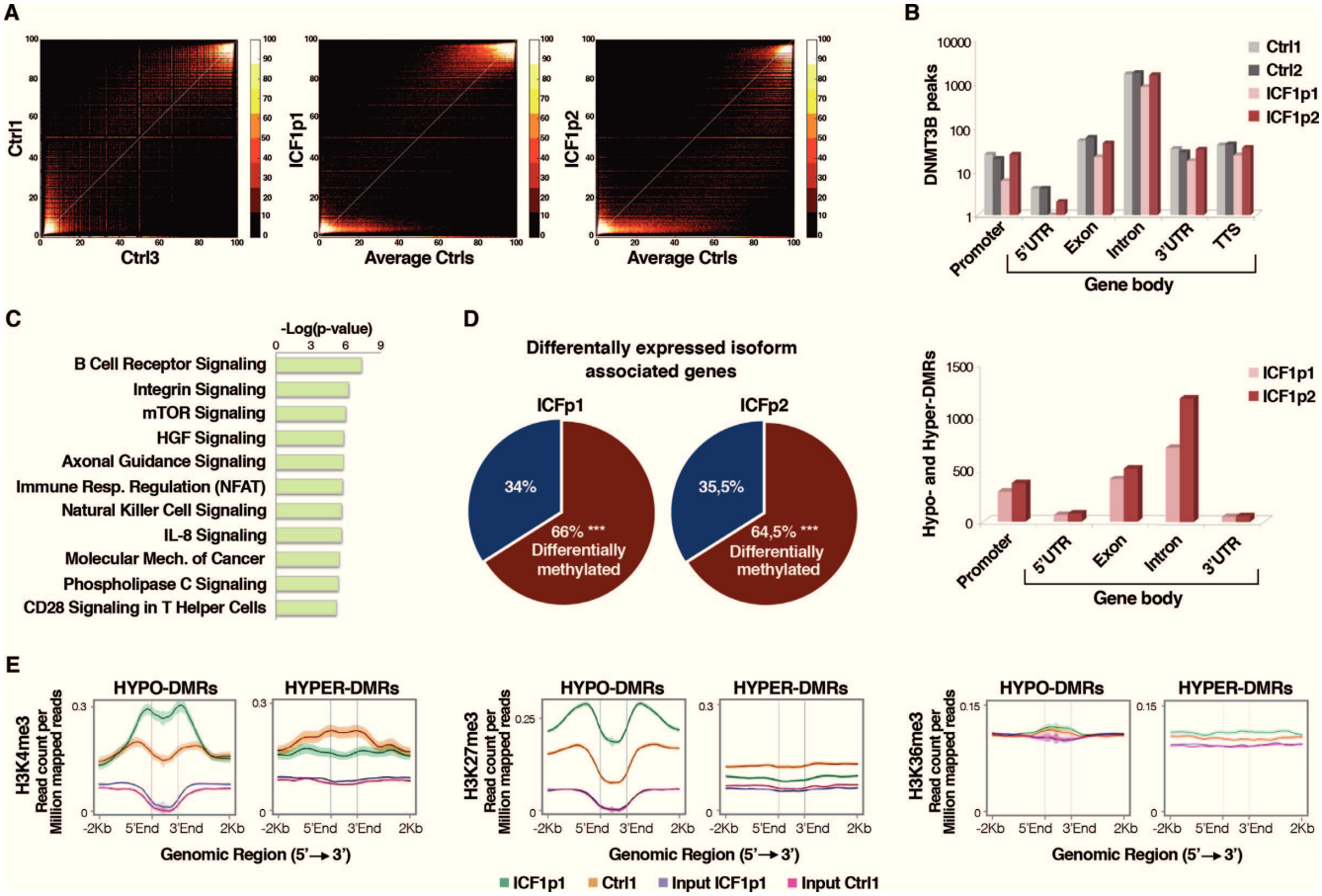


Figure 1. Genome-wide analysis of CpG methylation, gene expression, DNMT3B and histone marks (H3K4me3, H3K27me3 and H3K36me3) profiles in ICF1-specific DNMT3B mutant cells. (**A**) Scatter plots and density color codes for DNA methylation data (RRBS) of all autosomes. Pairwise comparisons of methylation percentage between controls (Ctrls) and between ICF1p1 or ICF1p2 and average Ctrls are shown; (**B**) Distribution of DNMT3B peaks (upper panel) and hypo and hyper-DMRs ($\Delta \geq |25\%|$; lower panel) along the genic features (promoter, 5′UTR, exon, intron, 3′UTR and TTS) calculated using Homer tool (85); (**C**) Enriched gene pathways (IPA) among the differential expressed isoform associated genes (DE-isoforms) (*P*-values corrected by BH method were considered); (**D**) Pie chart showing the percentage of DE-isoforms associated genes affected by DNA methylation changes (*P*-values $<10^{-16}$ were calculated using one-tail hypergeometric test); (**E**) Density plots of H3K4me3, H3K27me3 and H3K36me3 (read count per million mapped reads) at the hypo- and hyper-DMRs (−2 kb upstream, +2 kb downstream).

ICF1p1 sample, compared to ICF1p2 sample and two controls (data not shown). Given that the extensive loss of DNA methylation at pericentromeric heterochromatin represents a hallmark of ICF molecular phenotype (Supplementary Figure S3B), we examined the DNMT3B binding at these regions, finding that it was mostly unaltered in ICF1 samples compared to controls, as we observed at genes (Supplementary Figure S3D and E).

Because human DNMT3B protein has recently been proposed as responsible for intragenic DNA methylation, we examined the distribution of DNMT3B binding sites along the genic features. Remarkably, both wild-type and mutant-DNMT3B preferentially targeted gene bodies particularly at exons and introns rather than at promoters (Figure 1B, upper panel). Accordingly, most hypo and hyper-DMRs matched with CpG-rich regions at gene-bodies (Figure 1B, lower panel), indicating that most DNA methylation defects occurred at intragenic regions. CpG methylation level at mutant-DNMT3B genomic targets clearly decreased compared to controls, as demonstrated in both RRBS and WGBS, indicating that ICF1-specific mutations prevalently cause hypomethylation (Supplementary Figure S3F). Interestingly, the ICF1p1 sample exhibited a higher level of DNA hypermethylation (Supplementary Figure S2A) associated with a lower number of DNMT3B peaks at gene body regions (Figure 1B, upper panel) compared to ICF1p2 sample. This suggests that DNMT3A might be responsible for the aberrant DNA hypermethylation.

Taken together, these results indicate that ICF1-specific mutations in DNMT3B severely affect gene body CpG methylation.

Intragenic CpG methylation defects associate with altered expression of transcript isoforms

The Gene Ontology analysis of DMR-associated genes and DNMT3B target genes interestingly showed significant enrichment of the alternatively splicing functional category in both control and disease samples [~45% of DNMT3B targets and DMR-genes, *P*-value: 3.53×10^{-35} and 2.56×10^{-5}, respectively]. This observation together with the finding that DNMT3B-mediated CpG hypomethylation defects

occur at intragenic regions suggested that DNMT3B might play a role in the regulation of alternative isoform expression.

To address this possibility and to investigate whether this process is perturbed by ICF1-specific DNMT3B mutations, we analyzed the RNA-Seq datasets using the BitSeq tool, which evaluates the exon–exon junction usage to measure the transcript isoform abundance (36; Supplementary Table S3).

Nearly 55% of differentially expressed isoform (DE-isoform) associated genes were shared by the two ICF1 samples when compared to controls. The newly identified genes belonged to molecular pathways relevant for ICF immunological phenotype (Figure 1C). The most significant pathway affected in both ICF1 samples was the B cell receptor signaling (36% of genes within the pathway showed differential isoform expression; Benjamini–Hochberg (BH) adjusted *P*-value: 1×10^{-3} and 1×10^{-7} for ICF1p1 and ICF1p2, respectively), presumably accounting for the agammaglobulinemia commonly reported in ICF patients (Supplementary Figure S4).

We found that roughly 65% of DE-isoform associated genes were affected by differential DNA methylation (TSS ± 2 kb) in both ICF1 samples (Figure 1D), supporting the hypothesis that ICF1-specific DNMT3B mutations might perturb the transcription from canonical TSSs by shifting it toward alternative isoform-specific TSSs. Moreover, it has been reported that histone modifications, in particular the H3K4me3, H3K36me3 and H3K27me3 marks, define the transcriptional status at intragenic isoform-specific TSS (50). Therefore, we wondered whether these histone marks had altered profiles in the context of DNA methylation changes in ICF condition. We carried out ChIP-Seq experiments and integrated the results with the RRBS, WGBS and RNA-Seq datasets.

First, we found that H3K4me3 level inversely changed at hypo- and hyper-DMRs, suggesting that CpG methylation defects associate with H3K4me3 changes throughout the genome (Figure 1E, left panel). Conversely, the repressive H3K27me3 mark level significantly increased only at hypo-DMRs indicating that a compensatory mechanism might occur to overcome specific events of loss of gene silencing (Figure 1E; middle panel). At the same time, H3K36me3 was not influenced by increase or decrease of DNA methylation at genome-wide level (Figure 1E; right panel) consistently with the other observations that H3K36me3 mark is not affected in the more severely demethylated cell line HCT116 DKO, which has lost 90% of DNA methylation (6,51).

We then decided to examine whether and how H3K4me3, H3K27me3 and H3K36me3 were involved in the TSS differential expression observed, dependently or not on DNA methylation changes. For this reason, we focused the analysis on differentially expressed TSS (DE-TSS) of annotated isoforms in ICF1 and control samples (Figure 2A and Supplementary Figure S5A; TSS ± 500 bp). We found CpG hypo- and/or hyper-methylation at a majority of them, indicating that ICF-specific DNMT3B mutations affect the isoform transcription by altering CpG methylation at their TSS. Next, we measured the histone marks level at the affected TSS finding that H3K4me3 mark showed the most significant alteration associated with DNA methylation changes, while H3K27me3 and H3K36me3 remained unaffected at most DE-TSS (Figure 2A). Similarly, no significant enrichment of these histone marks was observed at DE-TSS that were not differentially methylated.

Alteration of isoform transcription was found at *TCEA2* gene, which resulted from the analysis of DE-isoform associated genes and further validated by qPCR. The long isoform transcribed from the most upstream TSS was abnormally silenced, while the transcription from the internal TSS of the short isoform was upregulated in both ICF1 samples (Figure 2B and C). Silencing of long isoform was linked to significant DNA methylation increase and concomitant reduction of H3K4me3 mark at TSS, while short isoform-TSS showed H3K4me3 increase but no CpG hypomethylation (Figure 2B and D and Supplementary Figure S5B). A gain of mutant-DNMT3B and DNMT3A binding was observed at the hypermethylated CpG island, supporting the idea that a methylating activity of both proteins is aberrantly recruited in ICF1 samples (Supplementary Figure S5C). To investigate whether DNA hypermethylation was directly responsible for the long isoform silencing we treated ICF1 cells with 5-AzaC and monitored the expression of the two *TCEA2* isoforms upon DNA hypomethylation induction. We found reactivation of the long isoform, while the short isoform was simultaneously downregulated, confirming that the DNA methylation change alone is sufficient to disturb the relative level of isoform expression (Supplementary Figure S5D and E). Moreover, to examine whether mutant-DNMT3B was directly involved in long isoform downregulation, we reduced its expression using short interfering RNAs (siRNAs) in ICF1 cells. Indeed, repression of *TCEA2* long isoform was attenuated upon DNMT3B knockdown (Supplementary Figure S5F and G), indicating that mutant DNMT3B protein was necessary for DNMT3A/3B complex to aberrantly methylate its TSS. Collectively, our findings support an essential role for DNMT3B activity in regulating the transcription of alternative transcript isoforms by differential TSS methylation, which correlates with H3K4me3 mark changes.

Activation of intragenic cryptic-TSS associates with altered CpG and histone methylation

Recent studies have proposed that cryptic or alternative promoters, marked by H3K4me3 and detected by the capped analysis of gene expression (CAGE) sequencing, may be characterized by promoter-like methylation in gene bodies, providing a mechanism to ensure proper transcriptional elongation (10,52–54). To deeper investigate the status of the aberrant transcription initiations in ICF1 samples, we predicted the position of intragenic cryptic-TSS by integrating several datasets from ENCODE Project. We overlapped CAGE tags, Polymerase II binding sites and H3K4me3-enriched regions filtering them for intragenic position and after excluding the annotated TSS.

We then measured the differential expression of the predicted intragenic cryptic-TSS between ICF1 samples and controls using DESeq tool (55) and subsequently searched for differentially methylated CpGs at these regions. We found that about 80% of DE cryptic-TSS were differentially

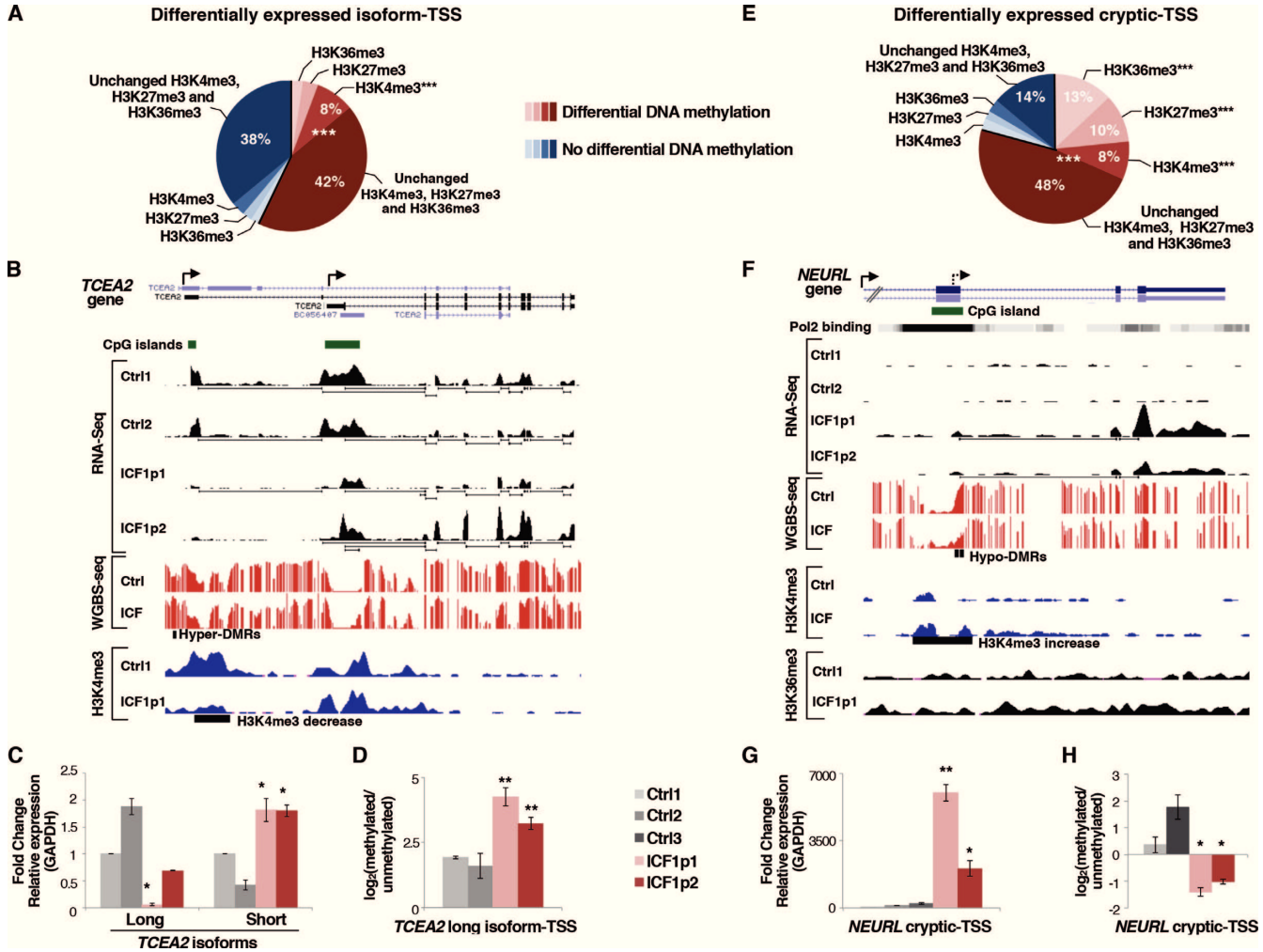


Figure 2. DNMT3B dysfunction affects isoform-TSS usage and transcription from cryptic-TSS in ICF1 cells. (**A**) Pie chart showing the percentage of differentially expressed isoform-TSS (±500 bp) with at least five CpGs differentially methylated and/or H3K4me3, H3K27me3 and H3K36me3 marks changes, obtained by integrating the RNA-Seq, WGBS and ChIP-Seq datasets (*P*-values $< 10^{-16}$ were calculated using one-tail hypergeometric test); (**B**) UCSC genome browser screenshot of *TCEA2* gene showing differentially expressed long and short transcript isoforms, CpG methylation status and H3K4me3 enrichment at corresponding TSSs. The setting of vertical viewing range is the same for ICF1 and control samples in each experiment; (**C**) mRNA level (qPCR) of long and short isoforms at *TCEA2* locus and, (**D**) CpG methylation level [$\log_2$ (methylated/unmethylated)] of *TCEA2* long isoform-TSS in ICF1 samples compared to controls; (**E**) Pie chart showing the percentage of differentially expressed cryptic-TSS (CAGE-tag, Pol II binding sites, H3K4me3 peaks ± 500 bp) with at least five CpGs differentially methylated and/or histones marks changes (*P*-values $< 10^{-16}$ were calculated using one-tail hypergeometric test); (**F**) UCSC genome browser screenshot representing a cryptic-TSS (Pol II binding site; UCSC Accession: wgEncodeEH001426) at *NEURL* gene associated with CpG hypomethylation, H3K4me3 increase and transcriptional upregulation, as shown by WGBS, RNA-Seq, H3K4me3 and H3K36me3 ChIP-Seq; (**G**) Expression level of *NEURL* cryptic-TSS and, (**H**) CpG methylation [log2 (methylated/unmethylated)] in ICF1 samples compared to controls. The enrichment of histone modification changes in differentially methylated category compared to the unchanged methylation category was evaluated using χ^2-test. **P*-adj <0.05; ***P*-adj <0.005; ****P*-adj <0.0005.

methylated (Figure 2E and Supplementary Figure S6A; *P*-value $< 10^{-16}$ in both comparisons).

Remarkably, H3K4me3, H3K27me3 and H3K36me3 marks were significantly affected only when associated with differential CpG methylation at the DE cryptic-TSS, suggesting that these histone marks tightly cooperate with DNA methylation to regulate the intragenic spurious transcription (Figure 2E).

Next, we investigated whether cryptic sites mapped to known repetitive elements, and we found the enrichment of two specific classes, LTRs and DNA repeats, at differentially methylated cryptic-TSS (Supplementary Figure S6B). Moreover, we examined the transcription directionality at cryptic-TSS calculating the directionality score that measures the transcriptional bias to either plus or minus strand of each cryptic-TSS. We found that the large majority of them were transcribed in the sense direction (Supplementary Figure S6C).

Among the compromised genes with a cryptic-TSS, we report the case of *NEURL* gene, where CpG island hypomethylation at the exon 5, corresponding to an intragenic Polymerase II binding site, resulted in the activation of a cryptic H3K4me3-enriched TSS in ICF1 samples (Figure 2F–H and Supplementary Figure S6D). The activation of this spurious intragenic TSS was linked to the expression of an aberrant short isoform including the last three exons of

NEURL gene (Figure 2G). Either upon induction of CpG hypomethylation in control cells using 5-AzaC treatment or DNMT3B gene knockdown, we observed increased expression of the cryptic-TSS at *NEURL* gene, indicating that this altered regulatory mechanism is directly controlled by DNMT3B-mediated DNA methylation level (Figure S6E–G). Collectively, these results provide evidence that mutant-DNMT3B affects the proper regulation of gene expression also by inducing intragenic changes of CpG methylation and histone marks level at cryptic-TSSs.

Natural antisense transcription is impaired in mutant-DNMT3B cells and affects sense transcription by altering the epigenetic-based self-regulatory circuit

Intragenic DNA methylation and chromatin context have been involved in the regulation of the sense-antisense gene pair transcription (56). Therefore, we hypothesized that DNMT3B mutations might affect this epigenetic-based regulatory mechanism and that it might involve specific histone mark alteration. As a first step, we examined the RNA-Seq dataset and found 153 altered natural antisense transcripts (NATs). Then, by analyzing the CpG methylation and the H3K4me3, H3K27me3 and H3K36me3 marks status at TSS (±500 bp) of NATs, we highlighted that 68% showed significant CpG methylation changes, and altered histone marks were detected at a significant fraction of them (Figure 3A and Supplementary Figure S7A). Interestingly, 66 out of 153 were associated with altered expression of their sense gene and significant changes of CpG methylation, H3K4me3, H3K27me3 and H3K36me3 were observed at the TSS of these genes (Figure 3B). These results would suggest that DNMT3B activity is essential for the sense-antisense pairs transcriptional regulation by ensuring a proper epigenetic context.

In an attempt to identify functionally relevant pairs of deregulated sense-antisense transcripts, we focused our attention on the memory B-cell marker *CD27* gene, known to be aberrantly silenced in peripheral blood of ICF1 patients and ICF1 cell lines. By performing gene-specific qPCR, and CpG methylation level analysis, we highlighted that *CD27* gene repression in ICF1p1 and p2 samples was linked to TSS CpG hypermethylation (Figure 3C–E). Intriguingly, the overlapping antisense long non-coding transcript *CD27-AS1* showed upregulation in ICF1 samples, as revealed by isoform-specific analysis of RNA-Seq dataset and qPCR (Figure 3C and F).

To identify the mechanism responsible for the inverse transcriptional correlation between sense and antisense expression at *CD27* locus and investigate how this mechanism is impaired in ICF1 condition, we first analyzed H3K4me3 and H3K36me3 marks at TSS and gene body of sense and antisense transcripts by ChIP assay. Interestingly, in control cells we observed high H3K4me3 level at the sense TSS, while it was poorly enriched at *CD27-AS1* TSS (Figure 3G). Conversely, an opposite H3K4me3 profile was observed in ICF1 samples (lower level at *CD27* TSS, higher level at *CD27-AS1* TSS) according with *CD27* sense silencing and higher antisense transcription in disease cells (Figure 3G).

Consistently with the active *CD27* gene expression in control samples, H3K36me3 mark was enriched downstream the *CD27* TSS along the gene body (Figure 3C). Because accumulation of H3K36me3 has been considered a general mechanism for reversible gene silencing when it occurs at promoters (57), we analyzed the enrichment of this histone mark at *CD27-AS1* TSS. Notably, we detected a high H3K36me3 level overlapping the antisense promoter in control cells, which is presumably responsible for suppressing the antisense transcription (Figure 3C and Supplementary Figure S7B). Accordingly, H3K36me3 level was significantly reduced in ICF1 samples compared to control samples at *CD27-AS1 TSS*, being correlated with *CD27-AS1* antisense upregulation in disease cells (Figure 3C and Supplementary Figure S7B).

We next examined the interaction of DNMT3A and DNMT3B proteins at the aberrantly hypermethylated *CD27* promoter region by ChIP assay. The active DNMT3A and the catalytically inactive DNMT3B both interacted with the *CD27* sense promoter only in ICF1 samples, suggesting that these proteins were aberrantly recruited and responsible for the hypermethylation and consequent gene silencing (Figure 3H). We also examined DNMT3A/B binding at *CD27-AS1* TSS, without detecting any enrichment (Supplementary Figure S7C).

To prove that DNA hypermethylation was able to silence the *CD27* gene in ICF1 condition, we induced CpG hypomethylation by treating the ICF1p2 cells with 5-AzaC and examined the expression of the sense transcripts at different time points. We found that *CD27* gene repression was released, thus confirming that DNA hypermethylation was sufficient to aberrantly silence *CD27* transcription in ICF1 samples (Supplementary Figure S7D and E). Interestingly, we observed a concomitant reduction of the *CD27-AS1* transcription in the treated ICF1p2 cells, indicating that a self-regulatory circuit takes place at *CD27* locus, which implies a transcriptional perturbation between sense and antisense transcripts (Supplementary Figure S7D). Similar effects on the *CD27*/*CD27-AS1* reciprocal transcription were obtained when we ectopically silenced DNMT3B in *CD27*-hypermethylated ICF1p2 cells, suggesting that DNMT3B plays a role in this regulatory network, either by methylating itself with its residual catalytic activity and/or by tethering DNMT3A at *CD27* TSS (Supplementary Figure S7F). Remarkably, in 5-AzaC-treated control cells *CD27-AS1* resulted overexpressed and its TSS was hypomethylated, suggesting the *CD27-AS1* transcription can be also directly modulated by CpG methylation itself (Supplementary Figure S7G and H).

Antisense expression has been shown to regulate transcription initiation by affecting DNA methylation at sense gene CpG islands/promoter in several contexts (58–61). While the process of antisense transcription may itself regulate sense gene transcritpion, in other instances it has been shown that NATs can directly regulate promoter DNA methylation via recruitment of DNMTs (62). Therefore, to explore the involvement of this specific mechanism we performed native RNA immunoprecipitation (RIP) assay followed by qPCR using antibodies against DNMT3B and DNMT3A. Then, we evaluated the interaction of the antisense transcript *CD27-AS1* with the two DNA methyltransferases. Remarkably, in ICF1 cells *CD27-AS1* was selectively pulled-down by both the antibodies (after normal-

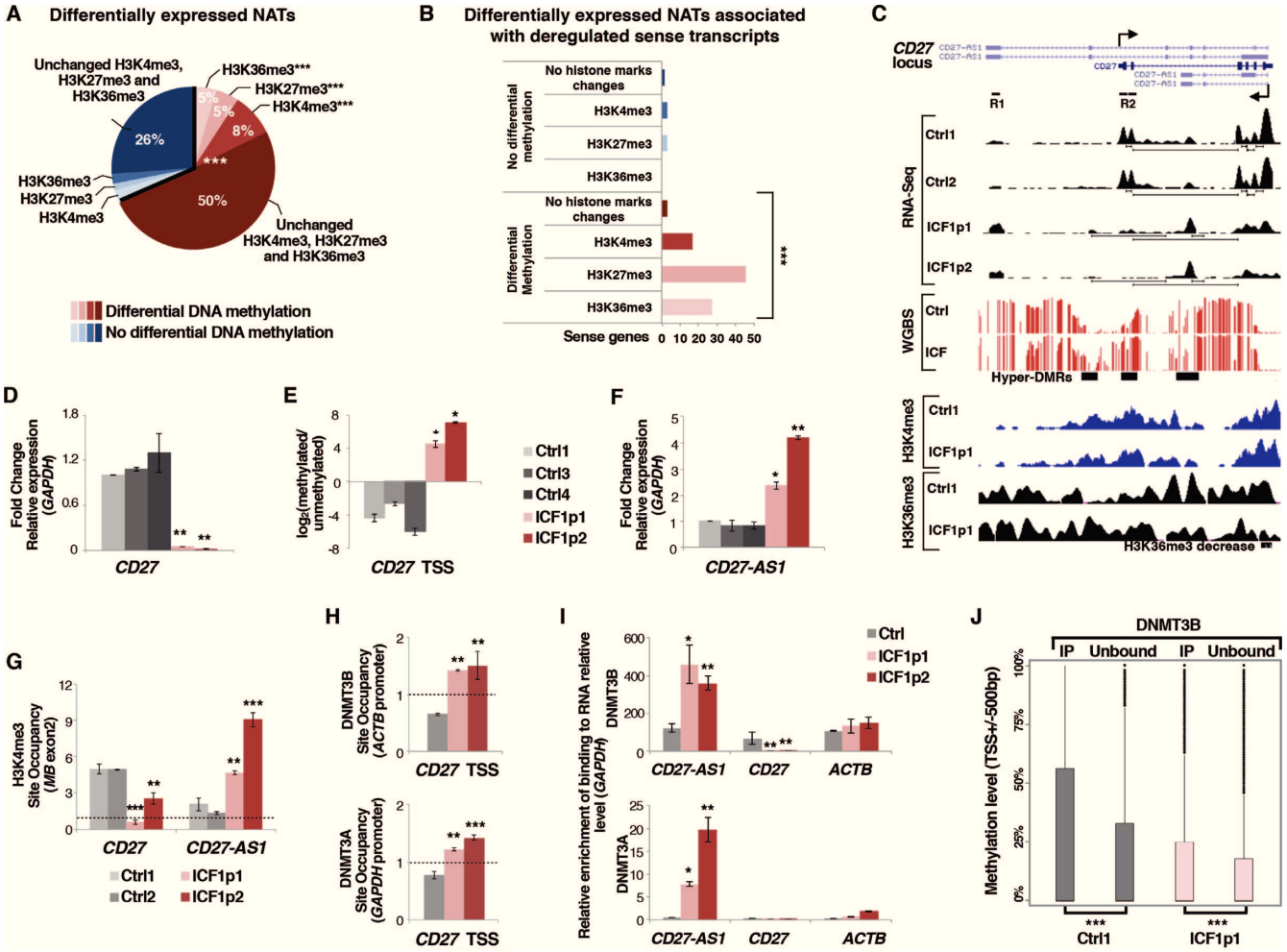


Figure 3. DNMT3B dysfunction affects sense-antisense pair genes transcription in ICF1 cells. (**A**) Pie chart showing the percentage of differentially expressed natural antisense transcripts (±500 bp) with at least 5 CpGs differentially methylated and/or H3K4me3, H3K27me3 and H3K36me3 marks changes, obtained by integrating the RNA-Seq, WGBS and ChIP-Seq datasets (*P*-values $<< 10^{-16}$ were calculated using one-tail hypergeometric test); (**B**) Differentially expressed sense genes associated with deregulated NATs are significantly enriched for DNA methylation and H3K4me3, H3K27me3 and H3K36me3 changes (*P*-values $<< 10^{-16}$ were calculated using one-tail hypergeometric test). The enrichment of histone modification changes in differentially methylated category compared to the unchanged methylation category was evaluated using χ^2-test; (**C**) UCSC genome browser screenshots of *CD27* gene showing the aberrant silencing of the sense transcript and the concomitant overexpression of the antisense transcript *CD27-AS1* (RNA-Seq), the CpG methylation status (WGBS), and H3K4me3 and H3K36me3 enrichments (ChIP-Seq); (**D**) mRNA, and (**E**) CpG methylation level [$\log_2$ (methylated/unmethylated)] of *CD27* sense gene; (**F**) mRNA level of *CD27-AS1* antisense in ICF1 samples compared to controls (gene specific primers were used to detect exclusively the antisense transcript; Supplementary Table S1); (**G**) H3K4me3 enrichment (ChIP-qPCR) at *CD27* and *CD27-AS1* TSS; (**H**) DNMT3B and DNMT3A enrichment to *CD27* TSS sense region (ChIP-qPCR); (**I**) DNMT3B and DNMT3A interaction with *CD27-AS1* antisense transcript and with *CD27* sense and *ACTB* transcripts as negative controls (RIP-qPCR); (**J**) Boxplot showing the methylation level at TSS of sense genes, which expression is reduced compared to the antisense partner (antisense/sense ratio ≥ 1.5 FPKM) and which interacts with DNMT3B at transcript level comparing the immunoprecipitated (DNMT3B-interacting transcripts) and unbound RNA fractions in each cell line. **P*-adj <0.05; ***P*-adj <0.005; ****P*-adj <0.0005.

ization to the transcript amount), indicating that *CD27-AS1* forms a complex with DNMT3A and DNMT3B proteins (Figure 3I). Given that DNMT3A and DNMT3B were also both enriched at *CD27* sense promoter (Figure 3H), the recruitment of DNMTs at this region, mediated by the antisense non-coding transcript, may eventually leads to the aberrant CpG hypermethylation. All together, our findings demonstrate that DNMT3B dysfunction affects the antisense-mediated gene regulatory networks through altering the epigenetic marks and the ncRNA-dependent guide of DNMT3B.

Next, we wondered whether this NAT-mediated DNMT3s tethering at sense promoters might be more generally involved in the self-regulatory circuit of sense-antisense gene pairs and whether DNMT3B mutations affect this mechanism at wider scale in disease context.

Thus, we first selected the overexpressed antisense transcripts compared to their sense partner in ICF samples and controls. Next, we identified the overexpressed antisense transcripts interacting with DNMT3B, which we obtained sequencing the immunoprecipitated fraction of the RIP assay (RIP-Seq; Supplementary Figure S8A–D). Interestingly, we found that the DNMT3B-bound fraction of

overexpressed antisense transcripts significantly correlated with the highest CpG methylation level at TSS of the sense partner, compared to the unbound fraction both in control and disease cells (Figure 3J). This suggests that the interaction between DNMT3B and non-coding antisense RNAs to the sense promoter might be a general mechanism to recruit the methyltransferase activity.

DNMT3B protein is actively involved in modulating exon exclusion through its RNA-binding property

Intragenic DNA methylation has been proposed as mechanism able to modulate alternative splicing events (63). Therefore, we hypothesized that DNMT3B mutations could affect the alternative splicing process and examined changes in exon skipping and inclusion between ICF1 and control samples using DEXSeq tool (37). We identified 3821 and 2330 differentially skipped and included exons, respectively, in ICF1p1 and 6436 and 3753 differentially skipped and included, respectively, in ICF1p2, and 42%-46% of them were differentially methylated (Figure 4A and Supplementary Figure S9A and B). However, this subtype was not significantly enriched among the differentially included exons, when compared with the total number of exons affected by altered DNA methylation in ICF1. This suggests that most of these events are a secondary effect of DNMT3B mutations or are caused by a regulatory DNMT3B function independent on methyltransferase activity.

To discriminate between these two possibilities we analyzed RNA expression levels of all spliceosome components in control and ICF1 samples and found no significant alterations (Supplementary Table S3), implying that most of the changes in alternative splicing in ICF1 samples are a direct outcome of DNMT3B dysfunction.

Therefore, considering that DNMT3B is capable of interacting with components of the transcriptional machinery (64), we asked whether the RNA-binding ability of DNMT3B might play a role in the alternative exon inclusion process. First, we compared the list of genes showing differentially spliced exons with the list of DNMT3B-interacting genes (RIP-Seq dataset) and found a significant overlap (Figure 4B and Supplementary Figure S9C). Next, we selected genes with differentially spliced exons showing an increased binding of DNMT3B at their transcripts in ICF samples and measured the number of exon skipping or inclusion events. From this analysis, we found significant enrichment for either skipping or inclusion events, indicating that acquisition of DNMT3B binding to mRNAs similarly impacts on both categories of alternative splicing (Figure 4C). All together, our studies suggest that in human cells ICF1-specific mutations of DNMT3B may affect the alternative splicing of mRNAs through a novel function of the protein, implying its ability to interact with mRNA molecules.

Among the aberrantly spliced genes, we identified the Protein Tyrosine Phosphatase, Receptor type, C gene (*PTPRC*; also known as *CD45*), which is a trans-membrane protein tyrosine phosphatase essential for antigen receptor-mediated signaling in lymphocytes (65). Differences in exon skipping and glycosylation of the protein produce various isoforms of CD45 that are present in a cell-specific manner. CD45RO, the smallest isoform, distinguishes effector-memory T cells, whereas primary B cells are known to express predominantly the largest CD45 protein isoform (encoded by CD45RABC, including exons 4–5–6). In line with the *in vivo* findings, the cultured B cells and B-LCLs express almost uniquely the RABC and RBC isoforms (66).

By comparing control and ICF1p2 RNA-seq tracks we found that exons 4–5–6 were differentially spliced (Figure 4D) and this intriguingly correlated with an increased binding of DNMT3B protein to *CD45* transcript in ICF1 sample (Figure 4D). To validate these observations, we performed isoform-specific PCR and found that ICF1 cells showed significantly higher CD45RO isoform expression than controls, while the CD45RABC isoform decreased, suggesting that DNMT3B activity is necessary to ensure a proper inclusion of the exons 4–5–6 (Figure 4E and F). The altered isoform transcription led to a change from $CD45RABC^{high}/CD45RO^{low}$ population in control cells to $CD45RABC^{low}/CD45RO^{high}$ population in ICF1 cells, which is more or less pronounced depending on DNMT3B mutations (Supplementary Figure S10A).

Given the fact that splicing occurs co-transcriptionally, epigenetic marks might affect the splicing decisions both by influencing the kinetics of transcriptional elongation (kinetic model), or through adaptor proteins that bind to epigenetic modifications and recruit splicing factors [recruitment model; (18)]. Both processes have been described for the splicing regulation mediated by DNA methylation (15,16,67). Therefore, we decided to examine thoroughly the involvement of these two mechanisms.

In line with WGBS results, we found that the alternative exons 4–5–6 were highly CpG methylated, either in control or ICF1 samples, and showed significant increase between the two conditions only in ICF1p2 cells (Figure 4D and Supplementary Figure S10B). Also, we analyzed the methylation status at single CpG level of exon 5, which has been reported alternatively spliced in naïve and activated CD4+ T cells according to its methylation level (15). We did not detect any changes at CpG methylation and consistently exon 5 did not show differential CTCF binding (Supplementary Figure S10C and D) (15). Similarly, we measured the enrichment of H3K36me3 marks at the spliced exons without finding any significant change (Supplementary Figure S10E).

Because we did not detect epigenetic modifications affecting the elongation rate of *CD45* gene, we explored in detail the mechanism suggested by our genome-wide studies (RIP-Seq), implying the direct binding of DNMT3B to the spliced exons of the transcript. Indeed, by carrying out native RNA immunoprecipitation with anti-DNMT3B we highlighted the interaction of the protein with *CD45* pre-mRNA in ICF1 samples (Figure 5A), which was further confirmed by cross-linked RNA Immunoprecipitation (Supplementary Figure S10F). This interaction was absent when analyzing different genes, which represented negative controls (e.g. *HOXC4*, *GAPDH, ACTB*).

As suggested by the recruitment model, chromatin-binding proteins, which bind to specific epigenetic modifications, can gather RNA-binding proteins. These RNA-binding proteins are then transferred to the mRNA molecule as it is being transcribed and change its alterna-

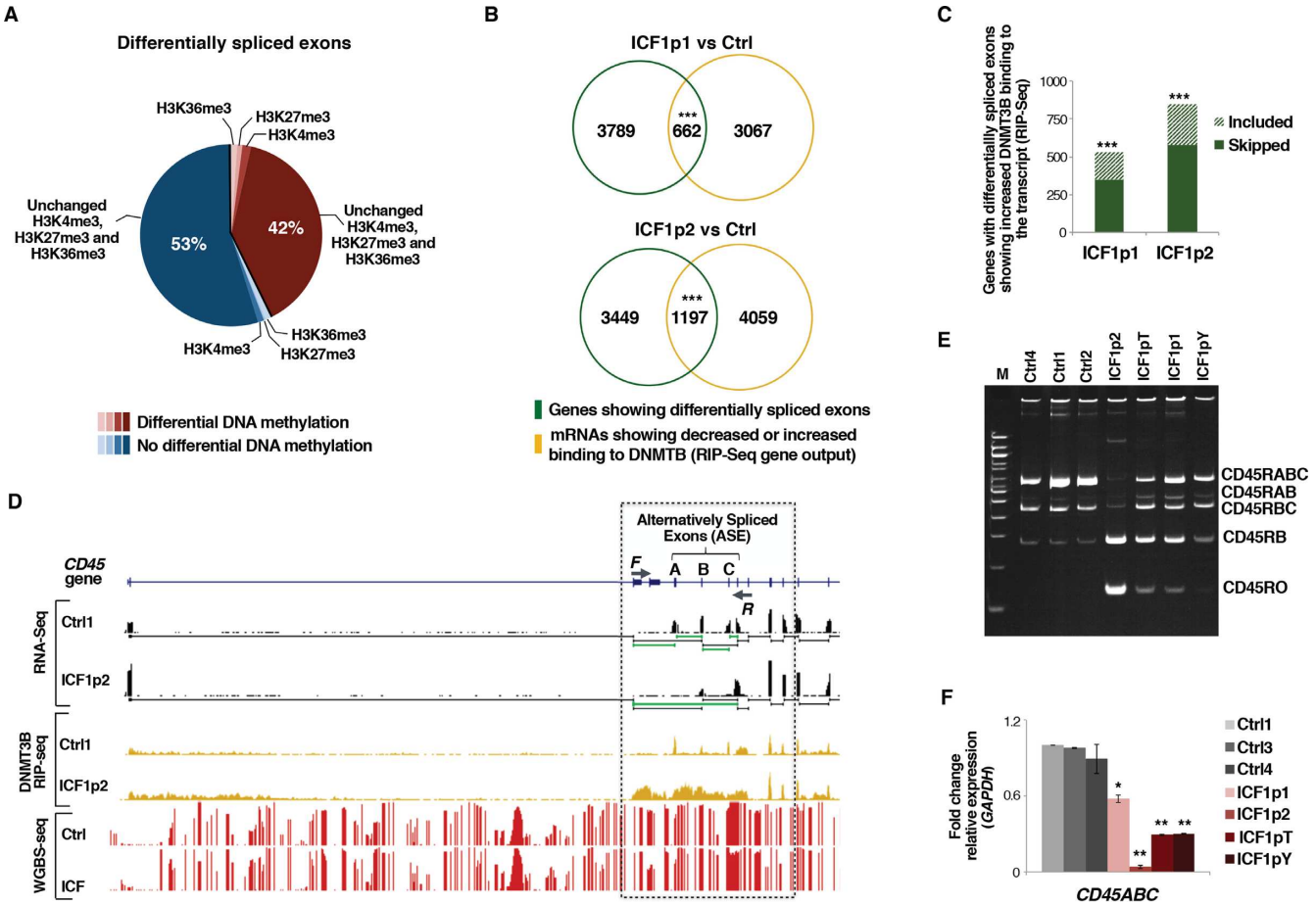


Figure 4. DNMT3B dysfunction affects alternative exon splicing in ICF1 cells. (**A**) Pie chart showing the percentage of differentially spliced exonic parts with differentially methylated CpGs and/or H3K4me3, H3K27me3 and H3K36me3 marks changes (statistical analysis was performed using one-tail hypergeometric test, which shows no significant enrichment for these epigenetic modification at differentially included exons); (**B**) Venn diagram showing the overlap between genes with differentially spliced exons and genes with decreased or increased DNMT3B binding to their transcripts (RIP-Seq) (*P*-values $<< 10^{-16}$ were calculated using one-tail hypergeometric test); (**C**) Histogram showing the number of genes with differentially spliced exons and increased DNMT3B binding to the transcript (normalized to IgG fraction; RIP-Seq). Gain of DNMT3B binding significantly associates with skipping or inclusion of exons (*P*-values $<< 10^{-16}$ were calculated using one-tail hypergeometric test); (**D**) UCSC genome browser screenshots of *PTPRC* (*CD45*) gene showing exon 4–5–6 skipping (RNA-Seq) and CpG methylation status (WGBS); (**E**) Semi-quantitative PCR amplification of *CD45* splicing isoforms in controls and ICF1 samples. Primers used are F and R as shown in D; (**F**) expression level (qPCR) of *CD45RABC* using isoform-specific oligonucleotides. **P*-adj <0.05; ***P*-adj <0.005; ****P*-adj <0.0005.

tive splicing pattern. We thus, verified the hypothesis that DNMT3B may function according to this model. First, we analyzed the DNMT3B binding to DNA by ChIP assay and we found a significant increase of DNMT3B interaction to the *CD45* exons 4–5–6 in ICF1 samples (Figure 5B). Interestingly, DNMT3A was enriched at these exons, indicating that high CpG methylation level at the exons was presumably ascribed to DNMT3A binding (Supplementary Figure S10G).

Then, we searched for molecules potentially participating to this aberrant exon skipping. *CD45* gene alternative splicing is tightly controlled by the tissue-specific ribonucleoprotein heterogeneous nuclear RNA-binding protein L-Like (hnRNP-LL). *Hn-RNP-LL* expression is specifically induced in terminally differentiated lymphocytes, including effector T cells and plasma cells, where it mediates the transition from CD45RA or CD45RABC to CD45RO, respectively (68,69). Current models of combinatorial alternative splicing propose that hnRNP-LL cooperates with the heterogeneous ribonucleoprotein hnRNP-L on *CD45* pre-mRNA, bridging exons 4 and 6 and looping out exon 5, thereby achieving full repression of the three variable exons (70).

Therefore, we first examined the expression of *hnRNP-LL* in ICF1 samples. Remarkably, *hnRNP-LL* was overexpressed in ICF1 cells compared to controls, and this correlated with CpG hypomethylation of its CpG island-associated promoter (Figure 5C–E and Supplementary Figure S11A). Conversely, the transcription of the protein interactor *hnRNP-L* was unaffected (data not shown). To confirm that overexpression of *hnRNP-LL* was able to modulate exons 4–5–6 skipping in an independent context, we treated control cells with 5-AzaC and verified that CpG hypomethylation led to the up-regulation of *hnRNP-LL* transcription (Supplementary Figure S11B and C) and resulted

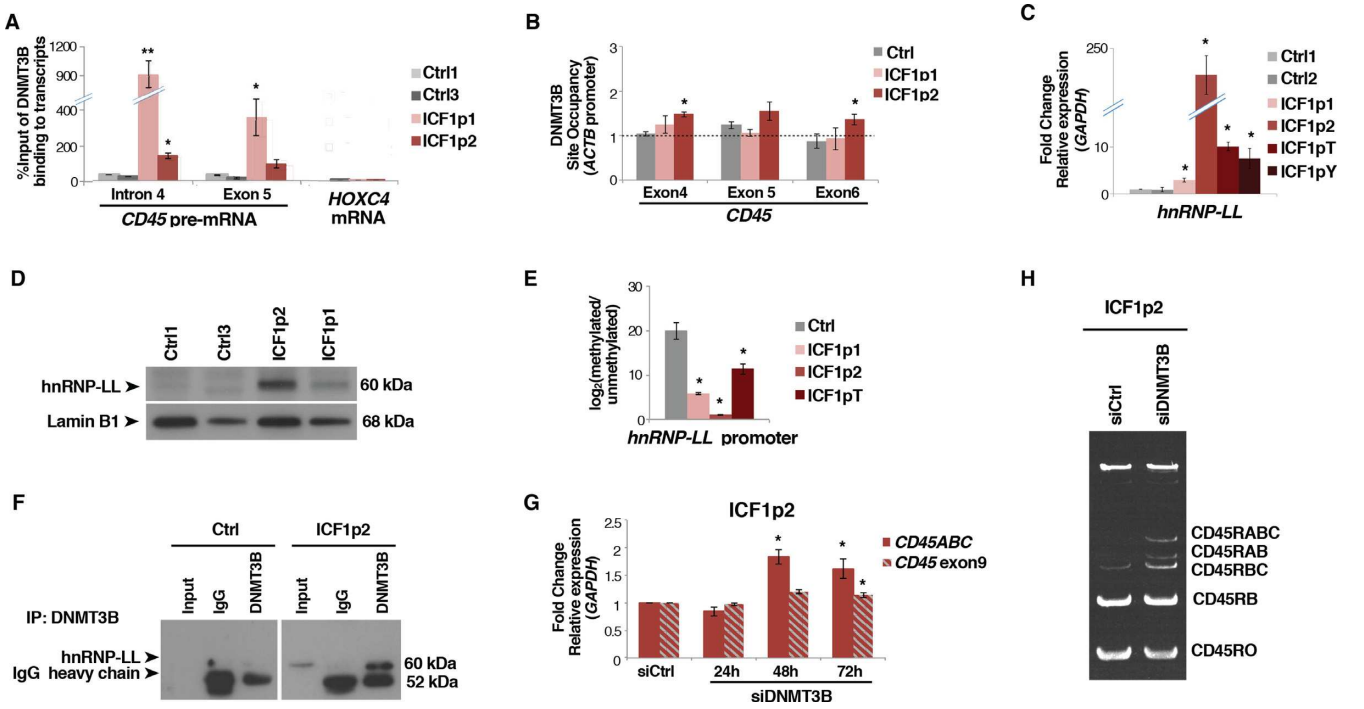


Figure 5. DNMT3B modulates exon 4–5–6 skipping by physically interacting with hnRNP-LL and *CD45* pre-mRNA. (**A**) Native RNA immunoprecipitation assay (RIP-qPCR) shows DNMT3B interaction with *CD45* pre-mRNA in control and ICF1 samples. Binding to *HOXC4* mRNA is reported as negative control; (**B**) DNMT3B enrichment at DNA (exons 4, 5 and 6) of *CD45* gene by ChIP assay; (**C** and **D**) Expression level of *hnRNP-LL* in ICF1 samples and controls by qPCR and western blot, respectively; (**E**) CpG methylation level [log2 (methylated/unmethylated)] at *hnRNP-LL* promoter in ICF1 samples; (**F**) DNMT3B and hnRNP-LL physically interact in ICF1p2 sample as shown in co-immunoprecipitation experiments; (**G**) Expression level (qPCR) of *CD45RABC* and the constitutive exon9 using isoform specific oligonucleotides upon siDNMT3B and control siRNAs transfection; (**H**) Semi-quantitative PCR amplification of *CD45* splicing isoforms in overexpressing-hnRNP-LL ICF1p2 sample upon siDNMT3B and control siRNAs transfection. **P*-adj <0.05; ***P*-adj <0.005; ****P*-adj <0.0005.

in the increased exons 4–5–6 exclusion (Supplementary Figure S11D).

Interestingly, we observed that the overexpressed *hnRNP-LL* interacted with *CD45* pre-mRNA in ICF1 cells, by performing RNA immunoprecipitation using anti-hnRNP-LL antibody (Supplementary Figure S11E). Collectively, these results suggest that DNMT3B might be recruited to pre-mRNA by the increased expression of *hnRNP-LL*. To address this hypothesis, we performed a co-immunoprecipitation assay, showing that hnRNP-LL and DNMT3B proteins physically interacted in ICF1 cells, which suggests the formation of a protein complex perturbing the correct inclusion of exons 4–5–6 (Figure 5F). Absence of the interaction was observed in control cells where *hnRNP-LL* is not expressed, as expected (Figure 5F).

DNMT3B might exert an active role in promoting the aberrant exons 4–5–6 skipping in ICF1 cells by interacting with the unspliced exons or being a secondary effect of the increased hnRNP-LL activity. To discriminate between these two scenarios and demonstrate that the DNMT3B activity is functionally relevant to influence the exons 4–5–6 skipping, we knocked down DNMT3B in ICF1p2 cells, where *hnRNP-LL* shows the greatest overexpression, by transfecting DNMT3B siRNAs. The DNMT3B knockdown did not affect the hnRNP-LL protein level, as shown by western blot analysis (Supplementary Figure S11F). Interestingly, when we measured the inclusion level of exons 4–5–6, we found that DNMT3B knockdown partially reverted the exon skipping, increasing the abundance of the longest isoform *CD45ABC* (Figure 5G,H). This indicates that despite the presence of hnRNP-LL in ICF1p2 cells DNMT3B function directly influences this process. In order to verify if hnRNP-LL and DNMT3B are reciprocally regulated we knocked down the heteronucleoriboprotein in ICF1p2 cells. The western blot analysis showed an unaltered level of DNMT3B protein (Supplementary Figure S11G), indicating that there is not a regulation feedback between these two proteins.

Overall, these results demonstrate that ICF1-specific DNMT3B mutations lead to CpG hypomethylation and overexpression of hnRNP-LL, which is engaged to *CD45* pre-mRNA in a protein complex through interacting with the mutant-DNMT3B. Here, DNMT3B activity specifically promotes aberrant exon skipping.

DISCUSSION

In this manuscript we tackled the important but still unanswered question of how hypomorphic mutations of DNMT3B affect the intragenic regulation of transcription in ICF1 condition. The integrative genome-wide study that we present shows that ICF1-specific mutations in DNMT3B catalytic domain slightly alter the ability of the protein to target DNA genomic regions, while the methyltransferase activity is rather impaired. The fact that mutant-

DNMT3B is able to bind DNA is expected, considering that ICF1-specific mutations mainly affect the catalytic domain, while the DNA binding capability depends on the PWWP domain within the N-terminal region (71). By combining genomic profiles of DNMT3B binding and CpG methylation defects, we observed that the DNA hypomethylation is prevalently explained by the catalytically inactive mutant-DNMT3B, which may function as negative dominant. Conversely, given that mutations inhibit the DNMT3B catalytic activity, the gain of CpG methylation is presumably ascribed to DNMT3A, as suggested by the findings that DNMT3A and mutant-DNMT3B are both mistargeted at hypermethylated sites, e.g. *TCEA2* long isoform TSS, *CD27* sense TSS and CD45 exon 4-5-6. In support of this conclusion, it has recently been confirmed that DNMT3B may play an accessory role by inducing (recruiting and stimulating) DNMT3A methylation activity (8).

Our results demonstrate that most DNMT3B binding sites occur at intragenic positions of genes, where DNA methylation defects largely take place in ICF1 cells. This is a novel finding in the context of ICF1 syndrome studies and explains why previous attempts to identify profound differences in DNA methylation profile of deregulated genes failed using candidate gene approaches focused at CpG island-associated promoters (22,31,32). These results paved the way to the identification of unprecedented aspects related to DNMT3B deficient activity in ICF1 syndrome, such as the alteration of isoform transcription by modulating the epigenetic signature (DNA methylation and H3K4me3, H3K27me3 and H3K36me3) at intragenic TSS. Notably, we found deregulation of alternative isoforms at genes relevant for ICF1-specific phenotype.

The large majority of affected genes were direct targets of DNMT3B, suggesting that proper binding and methylating activity of the protein is required for the appropriate relative abundance of transcript isoform. For instance, we found CpG-island hypermethylation overlapping the *TCEA2* long isoform TSS dependent on the aberrant mutant-DNMT3B/3A methylating activity in ICF1 cells, which silences its transcription, thus promoting higher level of the short isoform transcript (Figure 6). This was associated with a significant change of H3K4me3 at the affected TSS, but not at H3K27me3 and H3K36me3 marks, indicating that alteration of CpG methylation is sufficient to impact on TSS selection and tightly cooperate with H3K4me3 levels. That CpG methylation changes alone are sufficient to determine the TSS usage is further confirmed upon induction of long-isoform TSS hypomethylation in 5-AzaC treated ICF1p2 cells, where we observed the rescue of the *TCEA2* long isoform expression. Rescue of the long isoform level upon DNMT3B knockdown in ICF1 cells demonstrates that the binding activity of mutant DNMT3B might be important for DNMT3A recruitment to CpG island. Overall, together with the evidence that cryptic-TSS are illegitimately expressed in ICF1 samples this finding suggests that a proper intragenic DNMT3B-mediated CpG methylation contributes to achieve an efficient mRNA transcriptional initiation and elongation (Figure 6).

Antisense transcription is increasingly being recognized as an important regulator of gene expression. The locus specificity of natural antisense RNAs (NATs) opposite sense genes indicates that they might be part of self-regulatory circuits that allow genes to regulate their own expression. Studies have shown that antisense expression can induce a threshold-dependent on-off switch on sense-gene expression (56).

Noncoding antisense loss or gain of expression in a sense-antisense gene pair has often been implicated in human disease (59,72). The antisense-mediated repression may involve transcriptional perturbation of the sense partner but may also include engagement of chromatin remodeling complexes, with local accumulation of histone modifications and DNA methylation. Alternatively, chromatin modulation based on the process of transcription *per se* might be involved (73). This model is supported by the association of Pol II with chromatin remodelers like Set2, which catalyzes the deposition of H3K36me3 in transcribed regions (74). H3K36me3 in turn may be recognized by chromatin factors correlating with transcriptional repression (75–77).

Despite the cases where the requirement of the RNA molecule is clearly necessary, it is presumable that several complementary mechanisms might act cooperatively to ensure the faithful regulation of overlapping gene pairs. In this light, a remarkable example is the memory B-cell marker *CD27* gene, which transcription aberrantly switches from the sense to the overlapping antisense in ICF1 condition (Figure 6). In control cells, the high-level expression of sense *CD27* gene associates with the spreading of a transcription-induced accumulation of H3K36me3. In turn, H3K36me3 enrichment over the *CD27-AS1* TSS determines a repressive chromatin environment disfavoring the antisense transcription (Figure 6). We did not detect recruitment of DNMT3A/3B, which is known to bind H3K36me3, suggesting that different chromatin factors, like H3K4me3 demethylases or histone deacetylases, may be implicated.

Conversely, the epigenetic silencing of *CD27-AS1* promoter (CpG methylation and H3K4me3 level) is altered in ICF1 cells, fostering its transcription. The overexpressed antisense transcript interacts with DNMT3B/3A proteins inducing their recruitment to *CD27* sense TSS, which in turn acquires CpG hypermethylation and loses H3K4me3 mark. Strikingly, the epigenetic marks promote aberrant *CD27* sense gene repression, which then correlates with a reduced H3K36me3 level at *CD27-AS1,* thereby sustaining the antisense transcription.

CD27 protein plays a key role during the lymphocyte differentiation to memory B-cells and *CD27* gene expression is known to be severely affected in ICF condition. Indeed, complete lack of the CD19-CD27+ memory B-cell compartment in ICF patients in contrast to controls was detected from biochemical studies of ICF patients' peripheral blood (24). Peripheral B-cell maturation blockage was thought to contribute to agammaglobulinemia and to immune response dysfunction, which are the main clinical signs of the disease. Strikingly, our findings provided a molecular explanation to *CD27* aberrant silencing in ICF cells, contributing to decipher the molecular mechanisms underlying the disease pathogenesis.

There is a growing evidence for co-transcriptional regulation of splicing decisions and the current view is that CpG methylation may modulate the alternative splicing

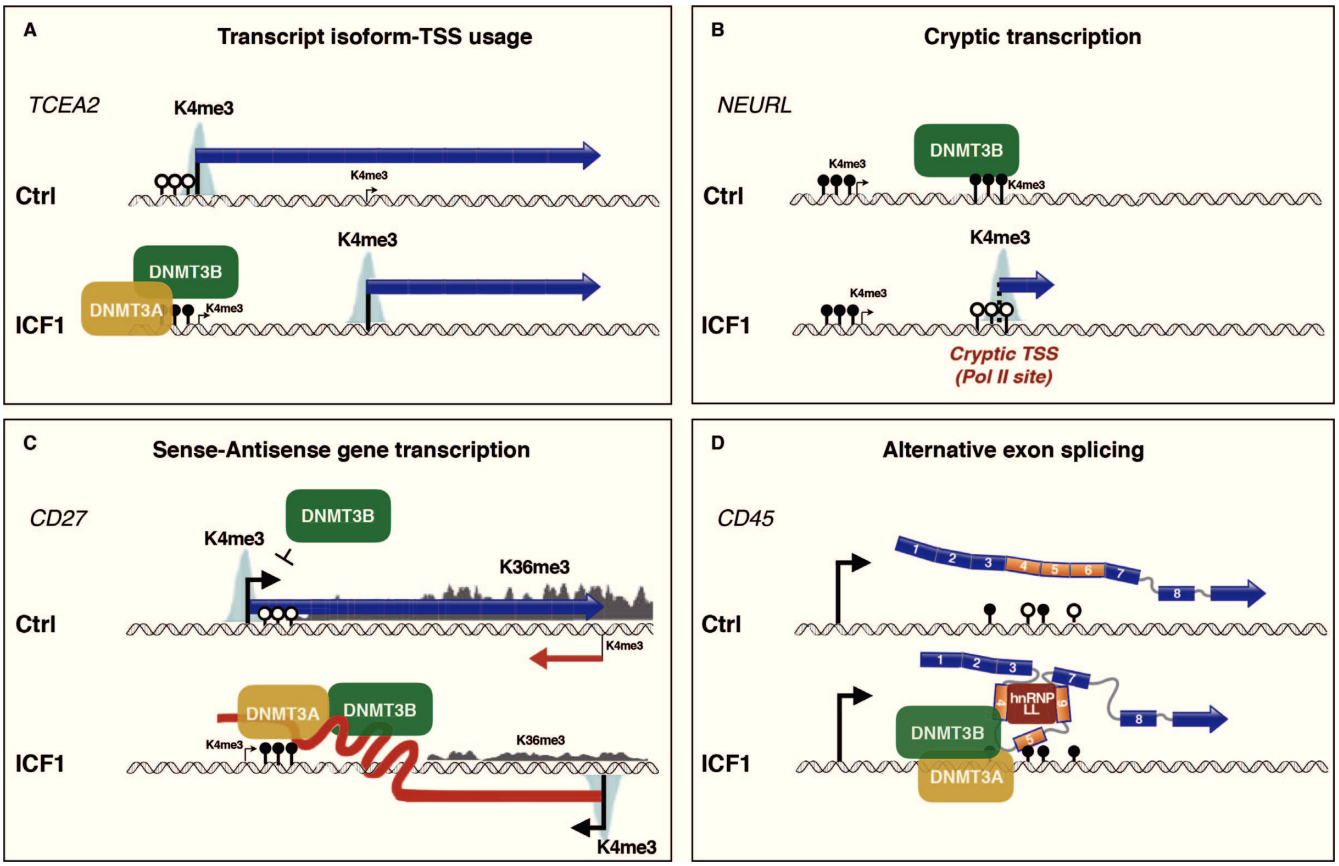


Figure 6. Summarizing model showing the mechanisms affected by ICF1-specific DNMT3B dysfunction in the regulation of (**A**) alternate isoform-TSS usage, (**B**) cryptic transcription, (**C**) sense–antisense gene self-regulatory circuit and (**D**) alternative exon splicing.

process, at least at a fraction of transcripts (63). Our findings point to a crucial role of DNMT3B in this process. According to our observations, DNMT3B might play a structural role acting as an adaptor protein, able to bind methylated DNA together with DNMT3A, and to recruit hnRNPs, which in turn tether splicing factors to the pre-mRNA. We proved that DNMT3B modulates the exon splicing through a mechanism implying the physical interaction of endogenous protein with mRNA molecules. The RNA binding capability of DNMTs was predicted based on several evidences, particularly involving non-coding RNAs (78,79). However, here we report the first example of interaction between DNMT3B and pre-mRNA molecules, which has the function to modulate the alternative exon splicing. At a genome-scale level we demonstrate that DNMT3B/mRNA interaction occurs in physiological conditions, suggesting a more general involvement of DNMT3B in mediating the alternative splicing. Further studies will be necessary to dissect the contribution of this regulatory mechanism to gene expression program in normal cells.

In disease context, this DNMT3B/mRNA binding alters the proper sequence of events leading to the inclusion of alternative exons 4–5–6 at *CD45* gene. We deciphered the molecular mechanism underlying the illegitimate exon exclusion in ICF1 samples, finding that DNMT3B is recruited to these exons and physically interacts with hnRNP-LL, which is ectopically expressed. To exclude that exons 4–5–6 skipping in ICF1 cells was exclusively dependent on hnRNP-LL overexpression we induced the DNMT3B knockdown in highly expressing hnRNP-LL ICF1 cells, and we were able to revert the exons 4–5–6 skipping. This provides direct evidence that DNMT3B does have an active role in modulating exon splicing at *CD45* gene. Here we describe a mechanism, which does not involve CTCF mediated regulation, which are instead observed at *CD45* exon5 in naïve and activated CD4+ T cells (15,80), indicating that specific epigenetic-based regulation of alternative splicing occurs in different cell contexts.

Notably, one study in embryonic stem cells suggests that DNMT3B would be able to interact with proteins involved in various aspects of Pol II-mediated transcription, as for instance PAF1 complex components, and heterogeneous ribonucleoproteins (64). We report a specific interaction between DNMT3B and hnRNP-LL in somatic cells, which would confirm this observation.

Within the complex with hnRNP-LL, the interaction between DNMT3B and pre-mRNA is expected to be direct as suggested by the native RNA immunoprecipitation and it may occur through the catalytic domain, which has been reported able to bind RNA molecules in case of DNMT1 protein (78). Mechanistically, the regulation of alternative exon inclusion/exclusion at DNMT3B-interacting pre-mRNAs, might involve a still unknown role of the protein in stimulating the m6A or the m5C RNA methylation process. In-

deed, the m6A RNA methylation has been reported as a modulator of mRNA splicing (81). Whatever the machinery involved, it is remarkable that it might represent a general DNMT3B-mediated mechanism of mRNA alternative exon splicing, as suggested by the RIP-Seq results in control cells. Strikingly, we identified a significant correlation between DNMT3B binding to mRNA transcripts and their alternative exon splicing. It will be interesting to assess the biological consequence of the altered transcriptome at global protein level.

Collectively, our results support the concept that gene body CpG methylation impacts on multiple regulatory mechanisms at intragenic level, including the repression of cryptic transcription from alternative promoters, selection of the transcribed isoform, antisense transcription and alternative exon splicing (9,10,82). Our findings contribute to decipher how the epigenetic modifications, which occur co-transcriptionally, interfere with a proper resultant mRNA transcript (in size and exon-intron content). Here, we provide evidence that DNMT3B plays a key role in regulating each of these events highlighting that dysfunction of DNA methylating activity or its aberrant ncRNA-mediated recruitment causes altered control of mRNA transcription and processing. These novel aspects concerning DNMT3B function deserve to be deeply studied, because they represent a conceptual proof that DNA methylation and mRNA processing functionally converge, pointing to a central role for DNMT3B protein.

Besides providing us with novel insights into DNMT3B-dependent transcriptional regulatory network, our integrative multi-omic approach contributes to clarify the disease-associated molecular defects. A terminal B-cell differentiation blockage at the transitional B-cell stage was reported in ICF1 patients (24). Consistently, we found that ICF samples exhibited transcriptional defects at genes involved in the B cell receptor-signaling pathway. Deregulation of this pathway typically causes the absence of circulating mature B cells and of all immunoglobulin isotypes, accompanied by the accumulation of pre-B cells in the bone marrow (83). Of note, among the newly identified differentially expressed genes we found either receptor-type PTPs, (e.g. *PTPRC*, *PTPRJ* and *PTPRB*) and nonreceptor-type PTPs (e.g. *PTPN13*). These are known to positively or negatively regulate lymphocyte activation and development, thus providing novel candidate genes contributing to the immune response defects in ICF1 patients.

The most striking example was *CD45*, which main role in B and T lymphocytes promotes cells activation. Remarkably, CD45 activity is negatively modulated by dimerization (84). Larger CD45RABC isoforms exist predominantly as monomeric active phosphatase, while the smaller size of the CD45RO extracellular domain facilitates dimerization, rendering it less active and increasing the signal transduction threshold. In T or B cells this would contribute to cessation of the primary immune response (66). In this light, the aberrant exon skipping detected in ICF cells might perturb the CD45 protein activity by increasing the abundance of the CD45RO isoform.

In summary, we have performed an integrated epigenomic and transcriptomic analysis in ICF1-derived B-LCL. Compared to previous classical transcriptional studies, we identified a new level of DNMT3B-mediated transcriptional regulation, which highlighted defects in gene pathways and functions predicted and long sought, thus representing a step forward towards a more careful characterization of the ICF1 immune phenotype. Furthermore, our study highlights the power of integrative analyses to clarify the DNMT3B-mediated intricate regulatory epigenetic network modulating the proper tissue-specific transcriptome.

ACCESSION NUMBER

Data for the RNA-seq, ChIP-seq, RRBS and RIP-seq in ICF1 and Ctrl cell lines have been deposited in the GEO database with accession number GSE95747.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Tom Roberts, Maria Strazzullo, Giovanna Del Pozzo and Andrea Riccio for helpful discussion, careful reading and critical comments of the manuscript. We also thank Sarah Teichmann, Henk Stunnenberg for their valuable help in the initial phases of the study and Maria Frola, Marianna Mastroianni for their help in the validation and organization of the data. Holger Heyn, Juan Sandoval and Manel Esteller are acknowledged for providing us with the WGBS data. Luigi Leone and Vincenzo Mercadante (NGS Core Facility) are acknowledged for technical assistance during histone marks ChIP-Sequencing experiments. We thank the IGA Technology Service (Udine) for RRBS, RNA-Seq and ChIP-Seq experiments and Laura Pisapia and Pasquale Barba (IGB FACS Facility) for surface phenotypic cell analysis. We are grateful to Giulio Pavesi, Francesco Russo, Dario Righelli and Marta Milo for bioinformatic support.

FUNDING

'Epigenomics Flagship Project (EPIGEN) Italian Ministry University and Research—CNR' [M.R.M., C.A. and M.D.E.]; UE Initial Training Network Project 'DISCHROM' [238242 to M.R.M. and M.D.E]; Telethon [GGP15209 to M.R.M.]. Funding for open access charge: Telethon GGP15209.
Conflict of interest statement. None declared.

REFERENCES

1. Jeltsch,A. and Jurkowska,R.Z. (2014) New concepts in DNA methylation. *Trends Biochem. Sci.*, **39**, 310–318.
2. Jurkowska,R.Z., Jurkowski,T.P. and Jeltsch,A. (2011) Structure and function of mammalian DNA methyltransferases. *ChemBioChem*, **12**, 206–222.
3. Okano,M., Bell,D.W., Haber,D.A. and Li,E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, **99**, 247–257.
4. Chen,T., Ueda,Y., Dodge,J.E., Wang,Z. and Li,E. (2003) Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol. Cell. Biol.*, **23**, 5594–5605.

5. Ueda,Y., Okano,M., Williams,C., Chen,T., Georgopoulos,K. and Li,E. (2006) Roles for Dnmt3b in mammalian development: a mouse model for the ICF syndrome. *Development*, **133**, 1183–1192.
6. Yang,X., Han,H., De Carvalho,D.D., Lay,F.D., Jones,P.A. and Liang,G. (2014) Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell*, **26**, 577–590.
7. Baubec,T., Colombo,D.F., Wirbelauer,C., Schmidt,J., Burger,L., Krebs,A.R., Akalin,A. and Schubeler,D. (2015) Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature*, **520**, 243–247.
8. Duymich,C.E., Charlet,J., Yang,X., Jones,P.A. and Liang,G. (2016) DNMT3B isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells. *Nat. Commun.*, **7**, 11453.
9. Kulis,M., Heath,S., Bibikova,M., Queiros,A.C., Navarro,A., Clot,G., Martinez-Trillos,A., Castellano,G., Brun-Heath,I., Pinyol,M. *et al.* (2012) Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat. Genet.*, **44**, 1236–1242.
10. Maunakea,A.K., Nagarajan,R.P., Bilenky,M., Ballinger,T.J., D'Souza,C., Fouse,S.D., Johnson,B.E., Hong,C., Nielsen,C., Zhao,Y. *et al.* (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*, **466**, 253–257.
11. Wolff,E.M., Byun,H.M., Han,H.F., Sharma,S., Nichols,P.W., Siegmund,K.D., Yang,A.S., Jones,P.A. and Liang,G. (2010) Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer. *PLoS Genet.*, **6**, e1000917.
12. Anastasiadou,C., Malousi,A., Maglaveras,N. and Kouidou,S. (2011) Human epigenome data reveal increased CpG methylation in alternatively spliced sites and putative exonic splicing enhancers. *DNA Cell Biol.*, **30**, 267–275.
13. Chodavarapu,R.K., Feng,S., Bernatavichute,Y.V., Chen,P.Y., Stroud,H., Yu,Y., Hetzel,J.A., Kuo,F., Kim,J., Cokus,S.J. *et al.* (2010) Relationship between nucleosome positioning and DNA methylation. *Nature*, **466**, 388–392.
14. Gelfman,S., Cohen,N., Yearim,A. and Ast,G. (2013) DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure. *Genome Res.*, **23**, 789–799.
15. Shukla,S., Kavak,E., Gregory,M., Imashimizu,M., Shutinoski,B., Kashlev,M., Oberdoerffer,P., Sandberg,R. and Oberdoerffer,S. (2011) CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature*, **479**, 74–79.
16. Yearim,A., Gelfman,S., Shayevitch,R., Melcer,S., Glaich,O., Mallm,J.P., Nissim-Rafinia,M., Cohen,A.H., Rippe,K., Meshorer,E. *et al.* (2015) HP1 is involved in regulating the global impact of DNA methylation on alternative splicing. *Cell Rep.*, **10**, 1122–1134.
17. Li-Byarlay,H., Li,Y., Stroud,H., Feng,S., Newman,T.C., Kaneda,M., Hou,K.K., Worley,K.C., Elsik,C.G., Wickline,S.A. *et al.* (2013) RNA interference knockdown of DNA methyl-transferase 3 affects gene alternative splicing in the honey bee. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 12750–12755.
18. Iannone,C. and Valcarcel,J. (2013) Chromatin's thread to alternative splicing regulation. *Chromosoma*, **122**, 465–474.
19. Hagleitner,M.M., Lankester,A., Maraschio,P., Hulten,M., Fryns,J.P., Schuetz,C., Gimelli,G., Davies,E.G., Gennery,A., Belohradsky,B.H. *et al.* (2008) Clinical spectrum of immunodeficiency, centromeric instability and facial dysmorphism (ICF syndrome). *J. Med. Genet.*, **45**, 93–99.
20. Matarazzo,M.R., De Bonis,M.L., Vacca,M., Della Ragione,F. and D'Esposito,M. (2009) Lessons from two human chromatin diseases, ICF syndrome and Rett syndrome. *Int. J. Biochem. Cell Biol.*, **41**, 117–126.
21. Weemaes,C.M., van Tol,M.J., Wang,J., van Ostaijen-ten Dam,M.M., van Eggermond,M.C., Thijssen,P.E., Aytekin,C., Brunetti-Pierri,N., van der Burg,M., Graham Davies,E. *et al.* (2013) Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects. *Eur. J. Hum. Genet.*, **21**, 1219–1225.
22. Jin,B., Tao,Q., Peng,J., Soo,H.M., Wu,W., Ying,J., Fields,C.R., Delmas,A.L., Liu,X., Qiu,J. *et al.* (2007) DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis, and immune function. *Hum. Mol. Genet.*, **17**, 690–709.
23. Matarazzo,M.R., Boyle,S., D'Esposito,M. and Bickmore,W.A. (2007) Chromosome territory reorganization in a human disease with altered DNA methylation. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 16546–16551.
24. Blanco-Betancourt,C.E., Moncla,A., Milili,M., Jiang,Y.L., Viegas-Pequignot,E.M., Roquelaure,B., Thuret,I. and Schiff,C. (2004) Defective B-cell-negative selection and terminal differentiation in the ICF syndrome. *Blood*, **103**, 2683–2690.
25. Velasco,G., Hube,F., Rollin,J., Neuillet,D., Philippe,C., Bouzinba-Segard,H., Galvani,A., Viegas-Pequignot,E. and Francastel,C. (2010) Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 9281–9286.
26. Jeanpierre,M., Turleau,C., Aurias,A., Prieur,M., Ledeist,F., Fischer,A. and Viegas-Pequignot,E. (1993) An embryonic-like methylation pattern of classical satellite DNA is observed in ICF syndrome. *Hum. Mol. Genet.*, **2**, 731–735.
27. Gisselsson,D., Shao,C., Tuck-Muller,C.M., Sogorovic,S., Palsson,E., Smeets,D. and Ehrlich,M. (2005) Interphase chromosomal abnormalities and mitotic missegregation of hypomethylated sequences in ICF syndrome cells. *Chromosoma*, **114**, 118–126.
28. Yehezkel,S., Segev,Y., Viegas-Pequignot,E., Skorecki,K. and Selig,S. (2008) Hypomethylation of subtelomeric regions in ICF syndrome is associated with abnormally short telomeres and enhanced transcription from telomeric regions. *Hum. Mol. Genet.*, **17**, 2776–2789.
29. Ji,W., Hernandez,R., Zhang,X.Y., Qu,G.Z., Frady,A., Varela,M. and Ehrlich,M. (1997) DNA demethylation and pericentromeric rearrangements of chromosome 1. *Mutat. Res.*, **379**, 33–41.
30. Prada,D., Gonzalez,R., Sanchez,L., Castro,C., Fabian,E. and Herrera,L.A. (2012) Satellite 2 demethylation induced by 5-azacytidine is associated with missegregation of chromosomes 1 and 16 in human somatic cells. *Mutat. Res.*, **729**, 100–105.
31. Ehrlich,M., Buchanan,K.L., Tsien,F., Jiang,G., Sun,B., Uicker,W., Weemaes,C.M., Smeets,D., Sperling,K., Belohradsky,B.H. *et al.* (2001) DNA methyltransferase 3B mutations linked to the ICF syndrome cause dysregulation of lymphogenesis genes. *Hum. Mol. Genet.*, **10**, 2917–2931.
32. Gatto,S., Della Ragione,F., Cimmino,A., Strazzullo,M., Fabbri,M., Mutarelli,M., Ferraro,L., Weisz,A., D'Esposito,M. and Matarazzo,M.R. (2010) Epigenetic alteration of microRNAs in DNMT3B-mutated patients of ICF syndrome. *Epigenetics*, **5**, 427–443.
33. Krueger,F. and Andrews,S.R. (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, **27**, 1571–1572.
34. Kim,D., Pertea,G., Trapnell,C., Pimentel,H., Kelley,R. and Salzberg,S.L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.*, **14**, R36.
35. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
36. Glaus,P., Honkela,A. and Rattray,M. (2012) Identifying differentially expressed transcripts from RNA-seq data with biological variation. *Bioinformatics*, **28**, 1721–1728.
37. Anders,S., Reyes,A. and Huber,W. (2012) Detecting differential usage of exons from RNA-seq data. *Genome Res.*, **22**, 2008–2017.
38. Cho,V., Mei,Y., Sanny,A., Chan,S., Enders,A., Bertram,E.M., Tan,A., Goodnow,C.C. and Andrews,T.D. (2014) The RNA-binding protein hnRNPLL induces a T cell alternative splicing program delineated by differential intron retention in polyadenylated RNA. *Genome Biol.*, **15**, R26.
39. da Huang,W., Sherman,B.T. and Lempicki,R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.
40. Bustin,S.A., Benes,V., Garson,J.A., Hellemans,J., Huggett,J., Kubista,M., Mueller,R., Nolan,T., Pfaffl,M.W., Shipley,G.L. *et al.* (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.*, **55**, 611–622.
41. Matarazzo,M.R., Lembo,F., Angrisano,T., Ballestar,E., Ferraro,M., Pero,R., De Bonis,M.L., Bruni,C.B., Esteller,M., D'Esposito,M. *et al.* (2004) In vivo analysis of DNA methylation patterns recognized

by specific proteins: coupling CHIP and bisulfite analysis. *Biotechniques*, **37**, 666–673.
42. Trapnell,C., Williams,B.A., Pertea,G., Mortazavi,A., Kwan,G., van Baren,M.J., Salzberg,S.L., Wold,B.J. and Pachter,L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.*, **28**, 511–515.
43. Afgan,E., Baker,D., van den Beek,M., Blankenberg,D., Bouvier,D., Cech,M., Chilton,J., Clements,D., Coraor,N., Eberhard,C. *et al.* (2016) Tathe Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.*, **44**, W3–W10.
44. Akalin,A., Kormaksson,M., Li,S., Garrett-Bakelman,F.E., Figueroa,M.E., Melnick,A. and Mason,C.E. (2012) methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.*, **13**, R87.
45. Heyn,H., Vidal,E., Sayols,S., Sanchez-Mut,J.V., Moran,S., Medina,I., Sandoval,J., Simo-Riudalbas,L., Szczesna,K., Huertas,D. *et al.* (2012) Whole-genome bisulfite DNA sequencing of a DNMT3B mutant patient. *Epigenetics*, **7**, 542–550.
46. Simo-Riudalbas,L., Diaz-Lagares,A., Gatto,S., Gagliardi,M., Crujeiras,A.B., Matarazzo,M.R., Esteller,M. and Sandoval,J. (2015) Genome-wide DNA methylation analysis identifies novel hypomethylated non-pericentromeric genes with potential clinical implications in ICF syndrome. *PLoS One*, **10**, e0132517.
47. Smith,Z.D., Gu,H., Bock,C., Gnirke,A. and Meissner,A. (2009) High-throughput bisulfite sequencing in mammalian genomes. *Methods*, **48**, 226–232.
48. Huang,K., Wu,Z., Liu,Z., Hu,G., Yu,J., Chang,K.H., Kim,K.P., Le,T., Faull,K.F., Rao,N. *et al.* (2014) Selective demethylation and altered gene expression are associated with ICF syndrome in human-induced pluripotent stem cells and mesenchymal stem cells. *Hum. Mol. Genet.*, **23**, 6448–6457.
49. Berman,B.P., Weisenberger,D.J., Aman,J.F., Hinoue,T., Ramjan,Z., Liu,Y., Noushmehr,H., Lange,C.P., van Dijk,C.M., Tollenaar,R.A. *et al.* (2012) Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat. Genet.*, **44**, 40–46.
50. Hahn,M.A., Wu,X., Li,A.X., Hahn,T. and Pfeifer,G.P. (2011) Relationship between gene body DNA methylation and intragenic H3K9me3 and H3K36me3 chromatin marks. *PLoS One*, **6**, e18844.
51. Baubec,T., Colombo,D.F., Wirbelauer,C., Schmidt,J., Burger,L., Krebs,A.R., Akalin,A. and Schubeler,D. (2015) Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature*, **520**, 243–247.
52. Illingworth,R.S., Gruenewald-Schneider,U., Webb,S., Kerr,A.R., James,K.D., Turner,D.J., Smith,C., Harrison,D.J., Andrews,R. and Bird,A.P. (2010) Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet.*, **6**, e1001134.
53. Deaton,A.M., Webb,S., Kerr,A.R., Illingworth,R.S., Guy,J., Andrews,R. and Bird,A. (2011) Cell type-specific DNA methylation at intragenic CpG islands in the immune system. *Genome Res.*, **21**, 1074–1086.
54. Kulis,M., Queiros,A.C., Beekman,R. and Martin-Subero,J.I. (2013) Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer. *Biochim. Biophys. Acta*, **1829**, 1161–1174.
55. Anders,S. and Huber,W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, **11**, R106.
56. Pelechano,V. and Steinmetz,L.M. (2013) Gene regulation by antisense transcription. *Nat. Rev. Genet.*, **14**, 880–893.
57. Wagner,E.J. and Carpenter,P.B. (2012) Understanding the language of Lys36 methylation at histone H3. *Nat. Rev. Mol. Cell Biol.*, **13**, 115–126.
58. Lister,R., Pelizzola,M., Dowen,R.H., Hawkins,R.D., Hon,G., Tonti-Filippini,J., Nery,J.R., Lee,L., Ye,Z., Ngo,Q.M. *et al.* (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, **462**, 315–322.
59. Tufarelli,C., Stanley,J.A., Garrick,D., Sharpe,J.A., Ayyub,H., Wood,W.G. and Higgs,D.R. (2003) Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat. Genet.*, **34**, 157–165.
60. Lyle,R., Watanabe,D., te Vruchte,D., Lerchner,W., Smrzka,O.W., Wutz,A., Schageman,J., Hahner,L., Davies,C. and Barlow,D.P. (2000) The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1. *Nat. Genet.*, **25**, 19–21.
61. Latos,P.A., Pauler,F.M., Koerner,M.V., Senergin,H.B., Hudson,Q.J., Stocsits,R.R., Allhoff,W., Stricker,S.H., Klement,R.M., Warczok,K.E. *et al.* (2012) Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science*, **338**, 1469–1472.
62. Li,Q., Su,Z., Xu,X., Liu,G., Song,X., Wang,R., Sui,X., Liu,T., Chang,X. and Huang,D. (2012) AS1DHRS4, a head-to-head natural antisense transcript, silences the DHRS4 gene cluster in cis and trans. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 14110–14115.
63. Lev Maor,G., Yearim,A. and Ast,G. (2015) The alternative role of DNA methylation in splicing regulation. *Trends Genet.*, **31**, 274–280.
64. Rigbolt,K.T., Prokhorova,T.A., Akimov,V., Henningsen,J., Johansen,P.T., Kratchmarova,I., Kassem,M., Mann,M., Olsen,J.V. and Blagoev,B. (2011) System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. *Sci. Signal.*, **4**, rs3.
65. Rhee,I. and Veillette,A. (2012) Protein tyrosine phosphatases in lymphocyte activation and autoimmunity. *Nat. Immunol.*, **13**, 439–447.
66. Hermiston,M.L., Xu,Z. and Weiss,A. (2003) CD45: a critical regulator of signaling thresholds in immune cells. *Annu. Rev. Immunol.*, **21**, 107–137.
67. Maunakea,A.K., Chepelev,I., Cui,K. and Zhao,K. (2013) Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res.*, **23**, 1256–1269.
68. Chang,X., Li,B. and Rao,A. (2015) RNA-binding protein hnRNPLL regulates mRNA splicing and stability during B-cell to plasma-cell differentiation. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E1888–E1897.
69. Oberdoerffer,S., Moita,L.F., Neems,D., Freitas,R.P., Hacohen,N. and Rao,A. (2008) Regulation of CD45 alternative splicing by heterogeneous ribonucleoprotein, hnRNPLL. *Science*, **321**, 686–691.
70. Preussner,M., Schreiner,S., Hung,L.H., Porstner,M., Jack,H.M., Benes,V., Ratsch,G. and Bindereif,A. (2012) HnRNP L and L-like cooperate in multiple-exon regulation of CD45 alternative splicing. *Nucleic Acids Res.*, **40**, 5666–5678.
71. Qiu,C., Sawada,K., Zhang,X. and Cheng,X. (2002) The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. *Nat. Struct. Biol.*, **9**, 217–224.
72. Ladd,P.D., Smith,L.E., Rabaia,N.A., Moore,J.M., Georges,S.A., Hansen,R.S., Hagerman,R.J., Tassone,F., Tapscott,S.J. and Filippova,G.N. (2007) An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals. *Hum. Mol. Genet.*, **16**, 3174–3187.
73. Loos,F., Loda,A., van Wijk,L., Grootegoed,J.A. and Gribnau,J. (2015) Chromatin-mediated reversible silencing of sense-antisense gene pairs in embryonic stem cells is consolidated upon differentiation. *Mol. Cell. Biol.*, **35**, 2436–2447.
74. Li,J., Moazed,D. and Gygi,S.P. (2002) Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation. *J. Biol. Chem.*, **277**, 49383–49388.
75. Carrozza,M.J., Li,B., Florens,L., Suganuma,T., Swanson,S.K., Lee,K.K., Shia,W.J., Anderson,S., Yates,J., Washburn,M.P. *et al.* (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell*, **123**, 581–592.
76. Xie,L., Pelz,C., Wang,W., Bashar,A., Varlamova,O., Shadle,S. and Impey,S. (2011) KDM5B regulates embryonic stem cell self-renewal and represses cryptic intragenic transcription. *EMBO J.*, **30**, 1473–1484.
77. Dhayalan,A., Rajavelu,A., Rathert,P., Tamas,R., Jurkowska,R.Z., Ragozin,S. and Jeltsch,A. (2010) The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J. Biol. Chem.*, **285**, 26114–26120.
78. Di Ruscio,A., Ebralidze,A.K., Benoukraf,T., Amabile,G., Goff,L.A., Terragni,J., Figueroa,M.E., De Figueiredo Pontes,L.L., Alberich-Jorda,M., Zhang,P. *et al.* (2013) DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature*, **503**, 371–376.
79. Schmitz,K.M., Mayer,C., Postepska,A. and Grummt,I. (2010) Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev.*, **24**, 2264–2269.

80. Marina,R.J., Sturgill,D., Bailly,M.A., Thenoz,M., Varma,G., Prigge,M.F., Nanan,K.K., Shukla,S., Haque,N. and Oberdoerffer,S. (2016) TET-catalyzed oxidation of intragenic 5-methylcytosine regulates CTCF-dependent alternative splicing. *EMBO J.*, **35**, 335–355.
81. Fu,Y., Dominissini,D., Rechavi,G. and He,C. (2014) Gene expression regulation mediated through reversible m(6)A RNA methylation. *Nat. Rev. Genet.*, **15**, 293–306.
82. Varley,K.E., Gertz,J., Bowling,K.M., Parker,S.L., Reddy,T.E., Pauli-Behn,F., Cross,M.K., Williams,B.A., Stamatoyannopoulos,J.A., Crawford,G.E. *et al.* (2013) Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res.*, **23**, 555–567.
83. Durandy,A., Kracker,S. and Fischer,A. (2013) Primary antibody deficiencies. *Nat. Rev. Immunol.*, **13**, 519–533.
84. Xu,Z. and Weiss,A. (2002) Negative regulation of CD45 by differential homodimerization of the alternatively spliced isoforms. *Nat. Immunol.*, **3**, 764–771.
85. Heinz,S., Benner,C., Spann,N., Bertolino,E., Lin,Y.C., Laslo,P., Cheng,J.X., Murre,C., Singh,H. and Glass,C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.

Supplementary Experimental Procedures

A multi-step workflow summarizing the multi-dimensional integrated analysis of genome-wide datasets resulting from RRBS, ChIP-Seq, RNA-Seq and RIP-Seq is shown in FigS12.

Reference genome and transcriptome

For sequencing alignment we used the human reference genome assembly GRCh37/hg19 (http://ftp.ensembl.org/Homo_sapiens.GRCh37.75.dna_sm.primary_assembly.fa.gz), while for transcriptome annotation the version 82 of the GRCh37 (http://ftp.ensembl.org/pub/grch37/release-82/gtf/homo_sapiens/Homo_sapiens.GRCh37.82.chr.gtf.gz) was used.

Reduced representation bisulfite sequencing (RRBS) and data processing

Two technical replicates for each DNA sample with independent bisulfite conversion and library preparation were produced using 1-10μg of genomic DNA RRBS libraries were generated and sequenced at IGA Technology Services (Italy), by using the NuGEN Ovation Ultralow Methyl-Seq Library System and 50bp single-end sequencing on the Illumina HiSeq2500 platform. Sequence reads were processed by adaptor trimming (Illumina Pipeline Casava), subjected to quality control (FastQC) and low quality reads were discarded (Trim Galore). CpG clustering and PCA analysis showed that samples from patients and controls are distinguishable by their methylome (FigS2B,C). A summary of mapped reads is presented in TableS4. Single CpG methylation percentage was calculated on pooled replicates by dividing the number of methylated Cs by the total coverage on that base using the R library package methylKit. Differentially methylated CpGs were obtained comparing CpG methylation level in each ICF1 sample to the mean CpG methylation percentage of control samples.

Differentially methylated regions (DMRs) between samples were defined as sequences with minimum 10 (less than 2Kb apart) common CpGs with concordant methylation difference (≥ l25%l). The whole genome bisulfite sequencing (WGBS) data were downloaded from the Gene Expression Omnibus (GSE37578, www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37578) (1). DMR annotation to genes was performed using the R library package ChIPpeakAnno v.2.16.4 (2). Distances between peaks and genes were calculated with a home-made script and TSS-2Kb/TTS+2Kb was considered as distance.

The validation of the RRBS results using the methylation pattern in ICF1 PBMCs was performed examining RRBS-covered CpGs located into the hypomethylated probes of 450K at all the genic features (promoter, TSS, gene body and 3'UTR). These CpGs were significantly hypomethylated also in both our ICF1 samples**.** As a whole, we observed that PBMCs and B-LCLs (ICF1p1 and ICF1p2) significantly shared 52% and 57% of hypomethylation sites respectively, confirming the

validity of B-LCLs to study the ICF1 DNA methylation defects (FigS2E).

Repetitive sequence coordinates were downloaded from RepBase repository, while pericentromeres were defined as the region between the nearest gene at the centromere, after removing the 10 kb proximal to the gene.

Analyses of DMR distribution among the genic features were performed using HOMER suite.

RNA sequencing, RIP sequencing and data processing

Two independent RNA-seq experiments were carried out. Reads were mapped to the reference genome using TopHat2 v.2.0.13 (3). In the first step we aligned the reads against the transcriptome. We used the following non-default TopHat2 parameters: -r 250, -m 2, --min-coverage-intron 50, --max-coverage-intron 100000, —mate-std-dev 50, —segment-length 17. The total tag number obtained for each sample is listed in TableS4. The counts of sequenced reads per annotated gene were derived with the use of htseq-count script distributed with HTSeq (4). We used the R library package DESeq2 v.1.8.2 (5) for measuring differential gene expression between two different cell conditions, considering the two RNA-Seq experiments as biological replicates. DESeq2 treats gene expression data as count data modeled under a negative binomial distribution. We picked out genes with p-value <0.05.

The relative isoform abundance estimation was performed using BitSeq (6), which is a tool for inferring transcript expression with a probabilistic model of the read generation process based on a Markov chain Monte Carlo (MCMC) algorithm for Bayesian inference over the model. We used the default settings (-p) and we extract as output the RPKM (-outType). From BitSeq output we selected only the genes shared by two independent replicates, which were associated to isoforms with: $(\sigma^2_{ICF} + \sigma^2_{wt}) <$ 95th percentile, RPKM ≥ 0.1 (at least in one condition), $\log_2(RPKM_{ICF}/RPKM_{wt}) \geq |1.5|$. We first identified the differentially expressed isoforms (DE-isoform), we next extrapolated the genes associated with DE-isoform and finally we considered only the genes resulting shared between the two independent RNA-seq experiments. Consistently with previous findings, we identified some germline genes as targets of DNMT3B, hypomethylated at CpG islands and derepressed (7). Moreover, down-regulation of previously reported ICF1-specific genes was confirmed in our RNA-Seq datasets, such as the memory B cell marker CD27 and immunoglobulin genes [(8,9); TableS3].

We found significant differences among the isoforms expressed in disease and control cell lines revealing major alterations at a larger number of genes than after performing classical RNA-Seq analysis with DESeq2 tool (35)(FigS3G and Table S3).

Intragenic cryptic-TSS coordinates were obtained by integrating several datasets from ENCODE Project (from many different cell types). We overlapped CAGE tags, Polymerase II binding sites

and H3K4me3-enriched regions filtering them for intragenic position and after excluding the annotated TSS. We assayed the differential expression of the predicted intragenic cryptic-TSS between ICF1 samples and controls using DESeq tool (10) finding about 1,056 and 2,278 DE cryptic-TSS for ICF1p1 and ICF1p2, respectively. We subsequently selected cryptic-intragenic TSS associated with at least 5CpGs with D> |25%|.

The analysis of differential exon inclusion was performed using as input dataset a list of non-overlapping exons created from the transcriptome annotation by the script dexseq_prepare_annotation.py available in DEXSeq package (version 1.14.2)(11). The read count on these regions was performed using dexseq_count.py script, also part of the DEXSeq package. Identification of differential exon usage between control and ICF1 samples was analysed using the DEXSeq package according to the script published in (12). Genes associated at least to one differential used exon (pval<0.05) were considered. Considering the small size of exonic parts, we set the threshold of one differentially methylated CpG in the integrative analysis with WGBS.

Transcription directionality at cryptic-TSS was assessed calculating the directionality score (D) that measures the transcriptional bias to either plus (P) or minus (M) strand of each cryptic-TSS D = P/(P+M) (13). Based on these measurements, if $0.25 \leq D \leq 0.75$, cryptic TSSs were considered "bidirectional", while $D < 0.25$ and $D > 0.75$ represented unidirectional transcription of negative strand and positive strand, respectively.

For RIP-Sequencing, the sequenced reads were mapped using TopHat2, following the same parameter as above described for RNA-Seq dataset. The total tag number obtained for each sample is listed in TableS4. Cufflinks settings were the following: max intron length 300000, min isoform fraction 0.1, pre-mRNA fraction 0.15, using effective length correction. For each cell line we considered only genes with input FPKM ≥ 1 and with enrichment values in DNMT3B immunoprecipitated (IP) sample over background ($\log2(FPKM_{DNMT3B}/FPKM_{IgG})$) ≥ 1.5. The enrichment of exons in DNMT3B immunoprecipitated RNA fraction over background (IgG fraction) was calculated according to the following steps. Using HTSeq (4) we counted the mapped reads on non-overlapping exon list as described above. Then, we measured the differential exon abundance in DNMT3B and IgG immunoprecipitated transcripts using DESeq2 (5), considering exons with p-value ≤ 0.05 and log2 (DNMT3B/IgG) ≥1.5.

For the analysis of the sense expression regulation mediated by the interaction between DNMT3B and antisense RNAs, we first focused on the overexpressed antisense transcripts in ICF samples and controls compared to their sense partner, by selecting the sense-antisense gene pairs showing antisense-to-sense FPKM ratio ≥1.5 within the RNA-Seq datasets. Next, we analyzed the CpG methylation level at TSS of this specific sense genes subset (with a coverage of at least 5 CpGs) and compared the list of their corresponding antisense with the list of transcripts interacting with DNMT3B, which we obtained after sequencing the immunoprecipitated fraction of the RIP

assay (RIP-Seq; FigS8A-D) carried out using anti-DNMT3B antibody (corresponding to ~2000 transcribed genes in control samples and 3289 and 4162 transcribed genes in ICF1p1 and ICF1p2 samples, respectively; Fig3J)).

ChIP sequencing and data processing

DNMT3B and H3K36me3 immunoprecipitated samples were sequenced at IGA Technology Services (Udine, Italy). Reads alignment on the reference genome was performed using Bowtie v. 1.1.1 (14). The parameters used for Bowtie were –a, –m3, –-best, and –-strata.

H3K4me3 and H3K27me3 immunoprecipitated samples were sequenced at IGB NGS Core facility (Naples, Italy) with SOLiD System 4.0 (Applied Biosystems). Reads were first analyzed by the Applied's pipeline software for quality filtering and aligned to the reference genome by using Bowtie; we selected as parameters -C, -k1, -m3, --best, and --strata.

The total tag number obtained for each sample is listed in TableS4. We selected for further analysis only uniquely mapped reads and removed the PCR amplification artefacts and the technical replicates were pooled together. All the reported results were obtained by normalizing for the total library size. DNMT3B peak calling was performed using MACS v.1.4.2 (15). The sample enrichments were normalized for cell line specific input. MACS parameter settings were: band width of 200bp (--bw), p-value 1e-4 (-p) and the range of high-confidence enrichment ratio against background including setting between 8 and 30 (-m). Moreover, we detected large peaks using EDD (16), setting a bin size of 2Kb (log_ratio_bin_size), a gap penalty of 5 (-g) and 20000 Monte Carlo trials (-n). Confidence intervals were calculated using the normal approximation method for binomial proportions. The SICER v.1.1 (17) peak-finding algorithm was used to identify the H3K4me3-, H3K27me3- and H3K36me3-enriched sites throughout the genome. For all these histone marks we selected peaks with a false discovery rate (FDR) of 1e-5. We used as window size and gap different values accordingly with the profile of analyzed histone marks. In particular, for H3K4me3 we used window size 200 bp and gap size 200 bp, for H3K27me3 500 bp and 1000 bp, while for H3K36me3 800bo and 4000bp respectively. The gap size selection was performed as described (Zang et al., 2009). We counted the reads on each detected histone mark peak in all conditions using Bedtools (18) and we detected differentially enriched domain using DESeq (10). We picked out differentially enriched regions with p-value <0.01. Density plots were obtained using NGSplot tool (19). We calculated the read count per million mapped reads using fragment length (-FL) 300, as reference region (-R) the position of interesting regions extended of (-L) 2000 and smoothing window (-MW) 10 were used.

DNMT3B peak annotation to genes was performed using the R library package ChIPpeakAnno v.2.16.4 (2). The distances between peaks and genes were calculated with a home-made script

and TSS-2Kb/TTS+2Kb was considered. The pericentromeric domains were defined as the regions between the nearest gene to the centromere and the centromere, and filtering out the 10 kb proximal to the gene. Bedtools and Bedops (20) toolkits were used for the above described data analysis. Analysis of DNMT3B-peaks distribution among the genic features were performed using HOMER suite.

Gene-specific ChIP assays were carried out by quantitative Real Time PCR, by using SYBR Green quantitative PCR (SsoAdvanced Universal SYBR Green Supermix, Biorad) according to CFX96TM Real Time PCR Detection Systems. The enrichment of DNA was calculated in terms of % input = $2^{-\Delta Ct \times 100}$, where ΔCt (threshold cycle) is determined by $Ct_{IP\ sample}$ - Ct_{Input} and 100 refers to the input, which is 1% of the starting chromatin or Site Occupancy= $2^{-(\Delta Ct_{Target\ region} - \Delta Ct_{Negative\ control})}$.

Supplementary Figures

FigS1. UCSC genome browser screenshots showing expression profile of *GAPDH, ACTB* and *MB* genes used as controls.

FigS2. A, Histogram showing total number of hypo- and hypermethylated CpGs in ICF1p1 and ICF1p2 compared to controls; B, CpG methylation clustering diagram between samples. Distance method: "correlation", linkage method: "Ward". Samples analyzed, ICF1p1, ICF1p2, Ctrl1 and Ctrl3. C, Principle component analysis (PCA) of mean methylation levels from ICF1 samples and controls. The PCA revealed a clear separation between ICF1 samples and controls by the first two components. As expected the controls are homogeneous in terms of methylome, while ICF1 samples are heterogeneous because of the diverse DNMT3B mutations. Conversely, technical replicates among the samples (a,b) are highly concordant. D, Overlap between differentially methylated regions (DMRs) detected by whole-genome bisulfite sequencing [WGBS; (1)] and RRBS (the present study); E, DNA methylation level measured by RRBS in ICF1 and control B-LCLs at CpGs hypomethylated in patients' PBMCs (Infinium 450K Illumina microarray), occurring at promoters, TSS, gene bodies and 3'UTR (p-value was calculated with Kolmogorov-Smirnov method and adjusted with Bonferroni method).

FigS3. A, Distributions of hypo- and hyper-DMRs at all chromosomes in ICF1p1 and ICF1p2 compared to controls; B, CpG methylation at pericentromeric heterochromatin regions (PCH) in ICF1 and control samples (p-value was calculated with Kolmogorov-Smirnov method and adjusted with Bonferroni method); C; Distribution of DNMT3B peaks at genic and intergenic regions, expressed as Log of DNMT3B covered bases; D, DNMT3B binding at pericentromeric heterochromatin regions (PCH) in ICF1 and control samples reported as average of fold change

(IP/INPUT) calculated in 200 bins; E, UCSC genome browser screenshot showing DNMT3B, WGBS and RRBS profiles at pericentromeric and centromeric regions of Chromosome 16; F, Venn diagrams showing the overlap between DMRs and DNMT3B peaks in ICF1 samples; G, Venn Diagrams showing the overlap between differentially expressed genes (DE-genes) obtained using DESeq2 and DE-isoform associated genes obtained using BitSeq tool.

FigS4. B-cell receptor-signaling pathway detected using IPA Ingenuity Pathway analysis. Red highlights indicate DE-isoform associated genes.

FigS5. A, Pie chart showing the percentage of differentially expressed isoform-TSS (+/-500bp), with at least 5 CpGs differentially methylated in ICF1p2 sample compared to controls, obtained by integrating the RNA-Seq and WGBS datasets; B, H3K4me3 mark enrichment (ChIP-qPCR) at *TCEA2* long and short isoforms TSS; C, DNMT3A and DNMT3B enrichment at *TCEA2* hypermethylated long isoform-TSS (ChIP-qPCR); D, *TCEA2* long and short isoform mRNAs level (qPCR) upon CpG hypomethylation induction in ICF1p2 cells using the demethylating agent 5-AzaC; E, CpG methylation level [log2 (methylated/unmethylated)] at *TCEA2* long isoform TSS upon 5AzaC treatment; F, DNMT3B protein level after transfecting ICF1p2 cells with siRNA Smart Pool against DNMT3B (siDNMT3B) and with control siRNAs; G, *TCEA2* long and short isoform mRNA level (qPCR) upon *DNMT3B* knockdown in ICF1p2 cells. *p-adj<0.05; **p-adj<0.005; ***p-adj<0.0005

FigS6. A, Pie chart showing the percentage of differentially expressed cryptic-TSS (+/-500bp), with at least at 5 CpGs differentially methylated in ICF1p2 sample compared to controls, obtained by integrating the RNA-Seq and WGBS datasets; B, H3K4me3 mark enrichment (ChIP-qPCR) at *NEURL* cryptic-TSS in ICF1 and control samples; C; transcription level (qPCR) of *NEURL* cryptic-TSS upon CpG hypomethylation induction in control cells using the demethylating agent 5-AzaC; D, CpG methylation level [log2 (methylated/unmethylated)] at *NEURL* cryptic-TSS upon 5-AzaC treatment of control cells; E, transcription level (qPCR) of *NEURL* cryptic-TSS upon *DNMT3B* knockdown in control cells. *p-adj<0.05; **p-adj<0.005; ***p-adj<0.0005

FigS7. A, Pie chart showing the percentage of differentially expressed natural antisense transcript (NATs) (+/-500bp), with at least 5 CpGs differentially methylated in ICF1p2 sample compared to controls, obtained by integrating the RNA-Seq and WGBS datasets; B, H3K36me3 mark enrichment (ChIP-qPCR) at *CD27-AS1* antisense TSS in ICF1 and control samples; C, DNMT3B and DNMT3A enrichment at *CD27-AS1* antisense TSS in ICF1 and control samples (qPCR-ChIP); D, *CD27* sense and *CD27-AS1* antisense mRNAs level (qPCR) upon CpG hypomethylation

induction in ICF1p2 cells using 5-AzaC; E, CpG methylation level [log2 (methylated/unmethylated)] at *CD27* sense in 5-AzaC-treated ICF1p2 cells; F, *CD27* sense and *CD27-AS1* antisense mRNAs level (qPCR) upon *DNMT3B* knockdown in ICF1p2 cells; G, *CD27-AS1* antisense mRNAs level (qPCR) upon CpG hypomethylation induction in control cells using 5-AzaC; H, CpG methylation level [log2 (methylated/unmethylated)] at *CD27-AS1* antisense TSS in 5-AzaC-treated control cells. *p-adj<0.05; **p-adj<0.005; ***p-adj<0.0005.

FigS8. DNMT3B-associated RNAs identified by RIP-Sequencing (RIP-Seq). A, Outline of the experimental strategy utilized to identify DNMT3B-associated RNAs; B, Histograms showing the total number of RNAs interacting with DNMT3B in control and ICF1 samples. Protein coding mRNAs, pseudogenes, antisense transcripts and lincRNAs are the most represented RNA molecules identified (% of expressed transcripts detected in the Input (FPKM>=1) are reported for each bar); C, Venn Diagram showing the number of RNAs associated with DNMT3B in control and ICF1 samples; D, Experimental validation of RIP-Seq results. Transcript level (RIP-qPCR) of *TDRD12*, *SYCP2 and HOXC4 genes* in RNA immunoprecipitated fraction of each cell lines normalized by input.

FigS9. A, Pie chart showing the percentage of differentially spliced exons, which are differentially methylated at CpGs in ICF1p2 sample compared to controls, obtained by integrating the RNA-Seq and WGBS datasets; B, UCSC genome browser screenshots showing differentially spliced exons (RNA-Seq) associated to differential CpG methylation (WGBS); C, UCSC genome browser screenshots showing differentially spliced exons (RNA-Seq) associated to differential DNMT3B-binding to transcripts (RIP-Seq).

FigS10. A, FACS analysis confirming that ICF1 samples express higher level of CD45RO than CD45RABC compared to controls; B, CpG methylation level [log2 (methylated/unmethylated)] at exons 4, 5 and 6 of *CD45* gene in ICF1 and control samples; C, Bisulfite sequencing showing single CpG methylation at exon 5 in ICF1 samples and controls; D, CTCF enrichment at *CD45* exon 4, exon 5 and intron 6 (ChIP-qPCR); E, H3K36me3 mark enrichment (ChIP-qPCR) at exons 4, 5 and 6 of *CD45* gene in ICF1 and control samples; F, Crosslinked-RNA Immunoprecipitation confirming the interaction between DNMT3B and *CD45* exon 4; G, DNMT3A enrichment (ChIP-qPCR) at exons 4, 5 and 6 of *CD45* gene in ICF1 and control samples. *p-adj<0.05; **p-adj<0.005; ***p-adj<0.0005.

FigS11. A, UCSC genome browser screenshots showing CpG methylation (WGBS) and expression profile (RNA-Seq) at *hnRNP-LL* gene; B, Native RNA immunoprecipitation showing hnRNP-LL binding to *CD45* exons 4 and 6 exclusively in ICF1 samples (qPCR-RIP); C, CpG methylation level [log2 (methylated/unmethylated)] at *hnRNP-LL* CpG-island upon CpG hypomethylation induction in control cells using 5-AzaC; D, *hnRNP-LL* mRNA level (qPCR) upon CpG hypomethylation in 5-AzaC-treated control cells; E, *CD45ABC* and *CD45ex9-10* transcription level (qPCR) in 5-AzaC-treated control cells; F, hnRNP-LL protein level upon *DNMT3B* knockdown in ICF1p2 cells; G, DNMT3B protein level upon *hnRNP-LL* knock-down in ICF1p2 cells. *p-adj<0.05; **p-adj<0.005; ***p-adj<0.0005

FigS12. A multi-step workflow summarizing the multi-dimensional integrated analysis of genome-wide datasets derived from RRBS, ChIP-Seq, RNA-Seq and RIP-Seq experiments. The tools selected for data analysis are reported for each "omic" experiment.

Table S1. Primer sequences used for gene-specific analysis.

Table S2. MIQE checklist

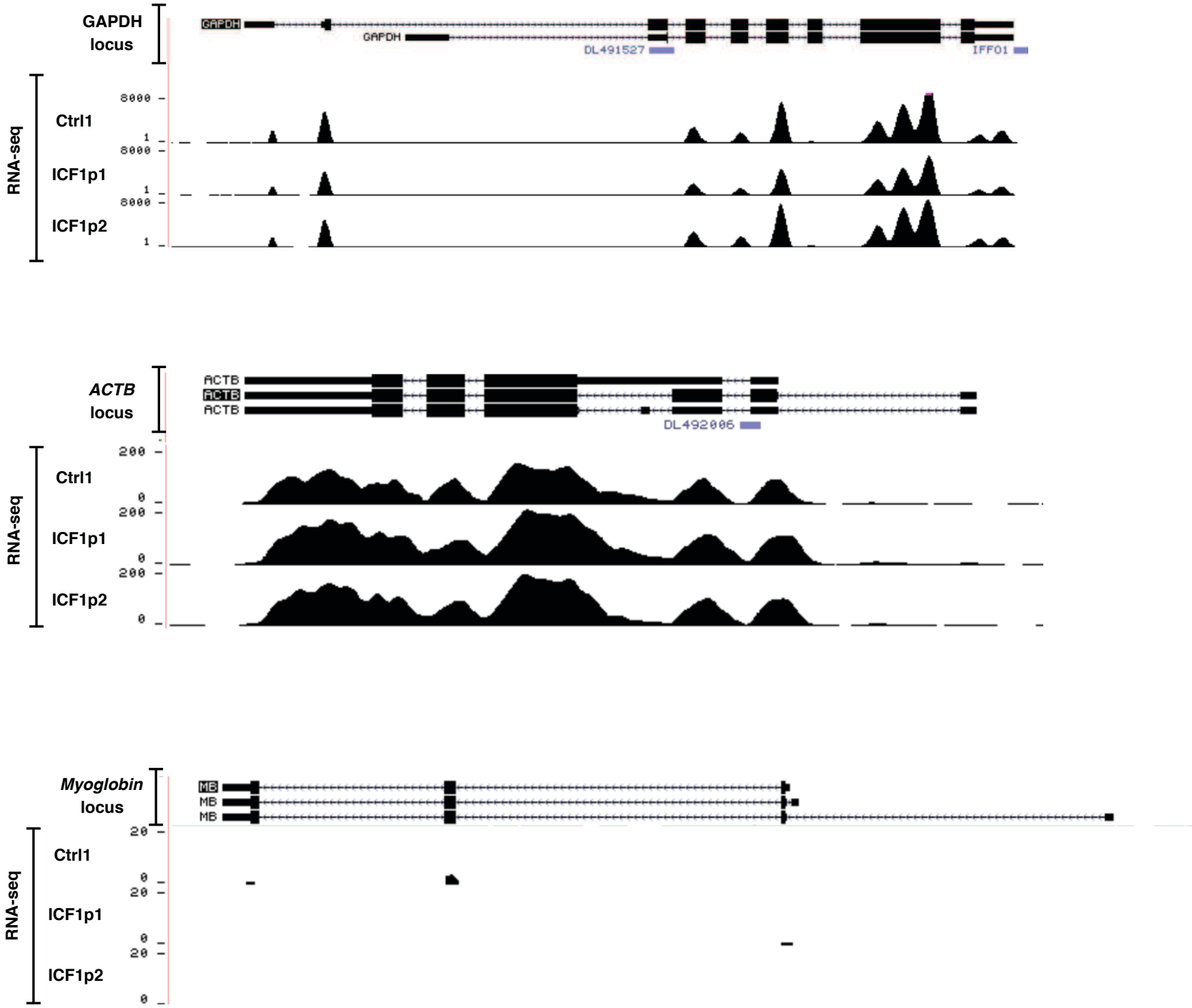
Table S3. List of differentially expressed genes identified using DESeq2 or BitSeq and genes associated to DMRs or DNMT3B binding, genes interacting with DNMT3B identified from RIP-Seq.

Table S4. Summary of total and mapped tag numbers for each sequencing experiment.

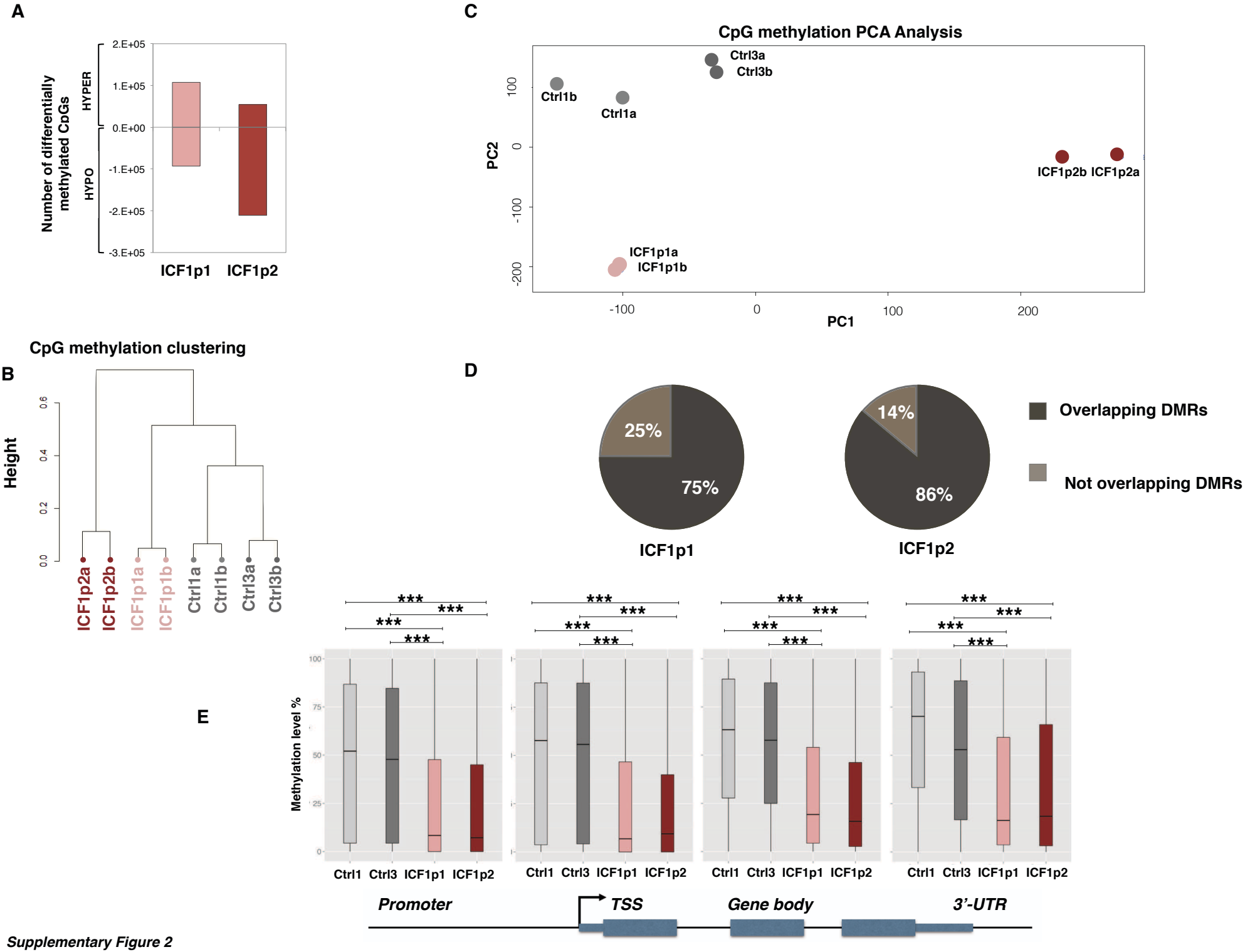
Supplementary references:Uncategorized References

1. **Heyn, H., Vidal, E., Sayols, S., Sanchez-Mut, J.V., Moran, S., Medina, I., Sandoval, J., Simo-Riudalbas, L., Szczesna, K., Huertas, D. *et al.* (2012) Whole-genome bisulfite DNA sequencing of a DNMT3B mutant patient. *Epigenetics : official journal of the DNA Methylation Society*, 7, 542-550.**
2. **Zhu, L.J., Gazin, C., Lawson, N.D., Pages, H., Lin, S.M., Lapointe, D.S. and Green, M.R. (2010) ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC bioinformatics*, 11, 237.**
3. **Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology*, 14, R36.**
4. **Anders, S., Pyl, P.T. and Huber, W. (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31, 166-169.**
5. **Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15, 550.**

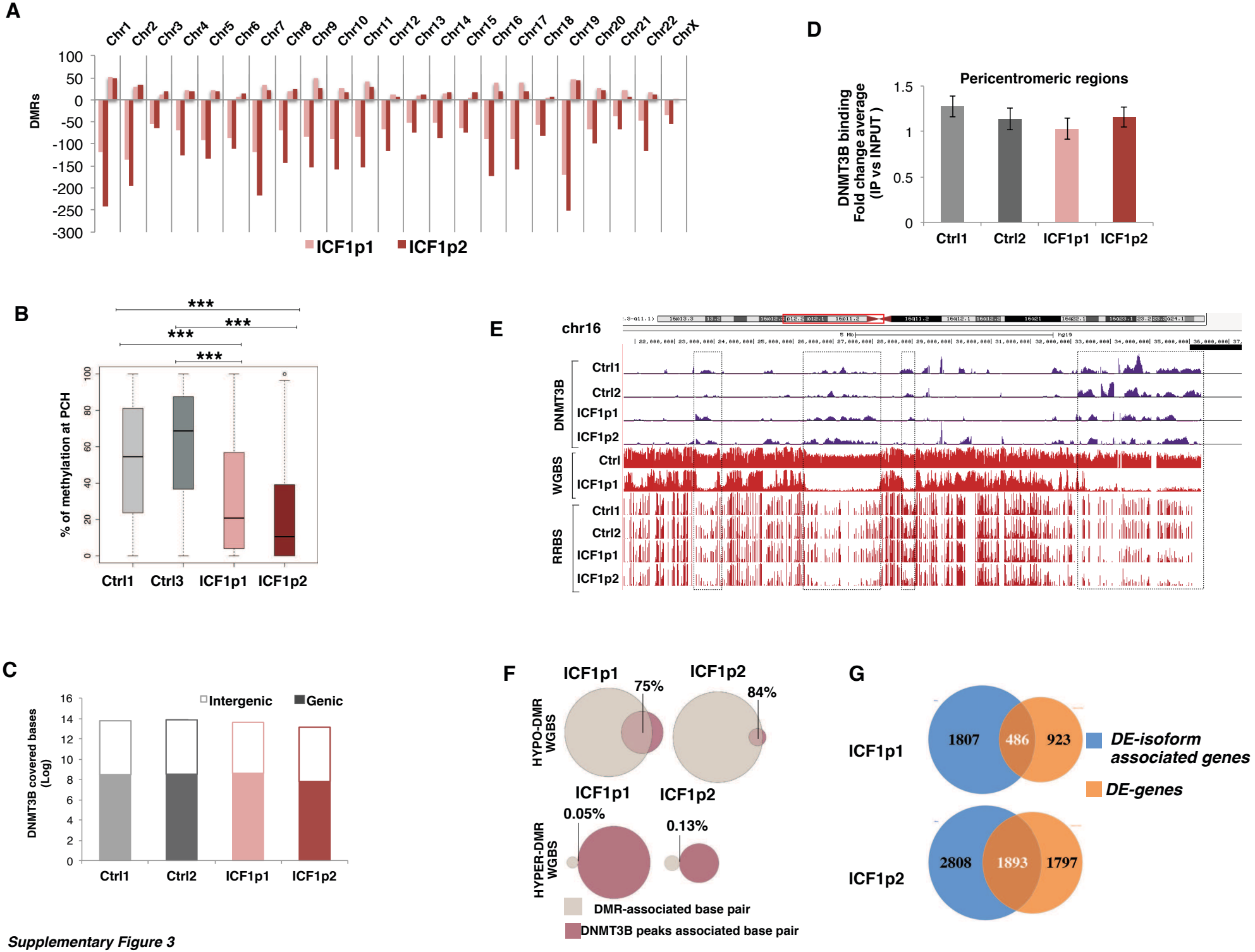
6. Glaus, P., Honkela, A. and Rattray, M. (2012) Identifying differentially expressed transcripts from RNA-seq data with biological variation. *Bioinformatics*, 28, 1721-1728.
7. Velasco, G., Walton, E.L., Sterlin, D., Hedouin, S., Nitta, H., Ito, Y., Fouyssac, F., Megarbane, A., Sasaki, H., Picard, C. *et al.* (2014) Germline genes hypomethylation and expression define a molecular signature in peripheral blood of ICF patients: implications for diagnosis and etiology. *Orphanet J Rare Dis*, 9, 56.
8. Ehrlich, M., Buchanan, K.L., Tsien, F., Jiang, G., Sun, B., Uicker, W., Weemaes, C.M., Smeets, D., Sperling, K., Belohradsky, B.H. *et al.* (2001) DNA methyltransferase 3B mutations linked to the ICF syndrome cause dysregulation of lymphogenesis genes. *Hum Mol Genet*, 10, 2917-2931.
9. Jin, B., Tao, Q., Peng, J., Soo, H.M., Wu, W., Ying, J., Fields, C.R., Delmas, A.L., Liu, X., Qiu, J. *et al.* (2007) DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis, and immune function. *Hum Mol Genet*.
10. Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome biology*, 11, R106.
11. Anders, S., Reyes, A. and Huber, W. (2012) Detecting differential usage of exons from RNA-seq data. *Genome research*, 22, 2008-2017.
12. Cho, V., Mei, Y., Sanny, A., Chan, S., Enders, A., Bertram, E.M., Tan, A., Goodnow, C.C. and Andrews, T.D. (2014) The RNA-binding protein hnRNPLL induces a T cell alternative splicing program delineated by differential intron retention in polyadenylated RNA. *Genome biology*, 15, R26.
13. Andersson, R., Refsing Andersen, P., Valen, E., Core, L.J., Bornholdt, J., Boyd, M., Heick Jensen, T. and Sandelin, A. (2014) Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *Nat Commun*, 5, 5336.
14. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25, 1754-1760.
15. Feng, J., Liu, T., Qin, B., Zhang, Y. and Liu, X.S. (2012) Identifying ChIP-seq enrichment using MACS. *Nature protocols*, 7, 1728-1740.
16. Lund, E., Oldenburg, A.R. and Collas, P. (2014) Enriched domain detector: a program for detection of wide genomic enrichment domains robust against local variations. *Nucleic acids research*, 42, e92.
17. Zang, C., Schones, D.E., Zeng, C., Cui, K., Zhao, K. and Peng, W. (2009) A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics*, 25, 1952-1958.
18. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841-842.
19. Shen, L., Shao, N., Liu, X. and Nestler, E. (2014) ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC genomics*, 15, 284.
20. Neph, S., Kuehn, M.S., Reynolds, A.P., Haugen, E., Thurman, R.E., Johnson, A.K., Rynes, E., Maurano, M.T., Vierstra, J., Thomas, S. *et al.* (2012) BEDOPS: high-performance genomic feature operations. *Bioinformatics*, 28, 1919-1920.



Supplementary Figure 1

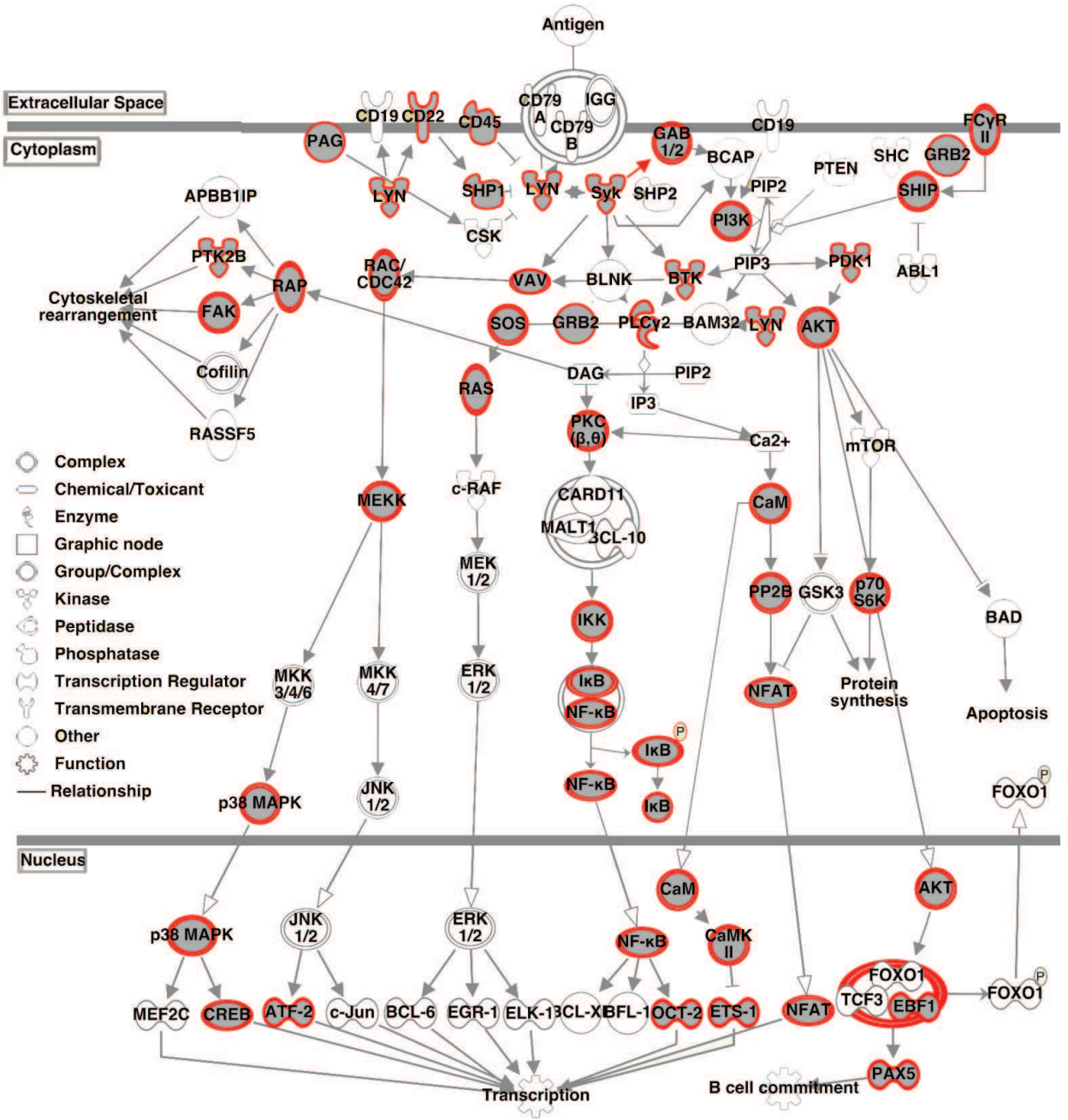


Supplementary Figure 2

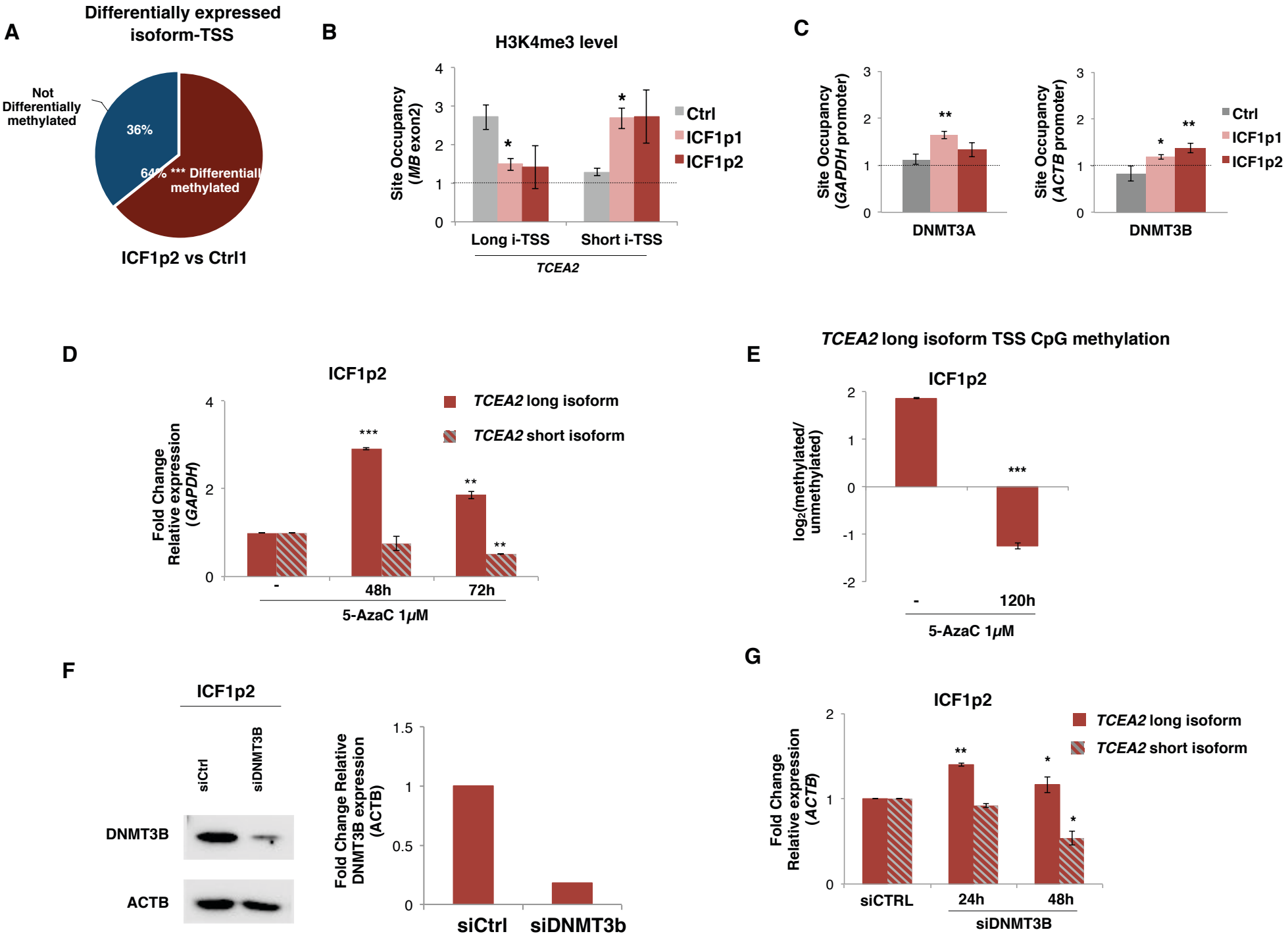


Supplementary Figure 3

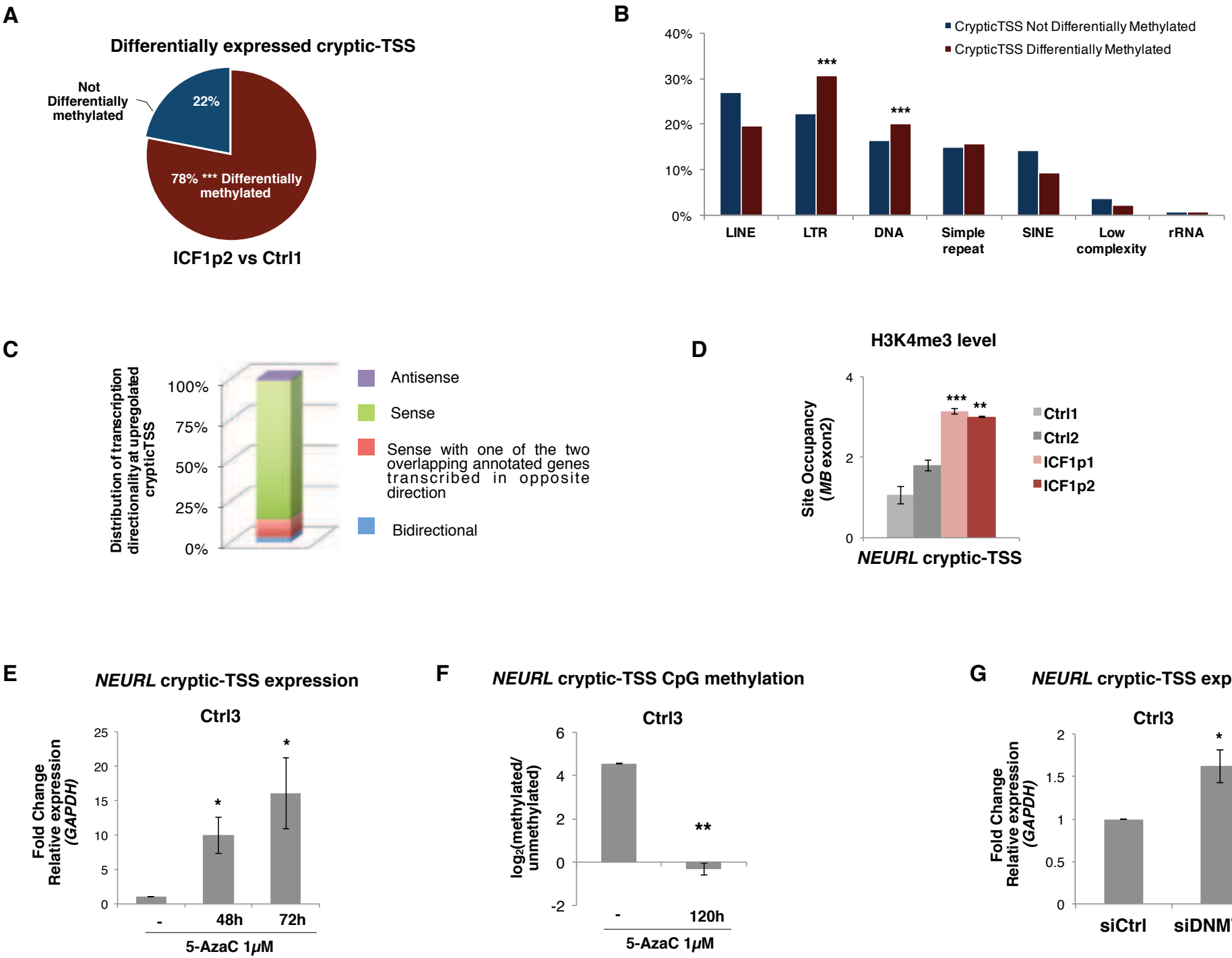
A



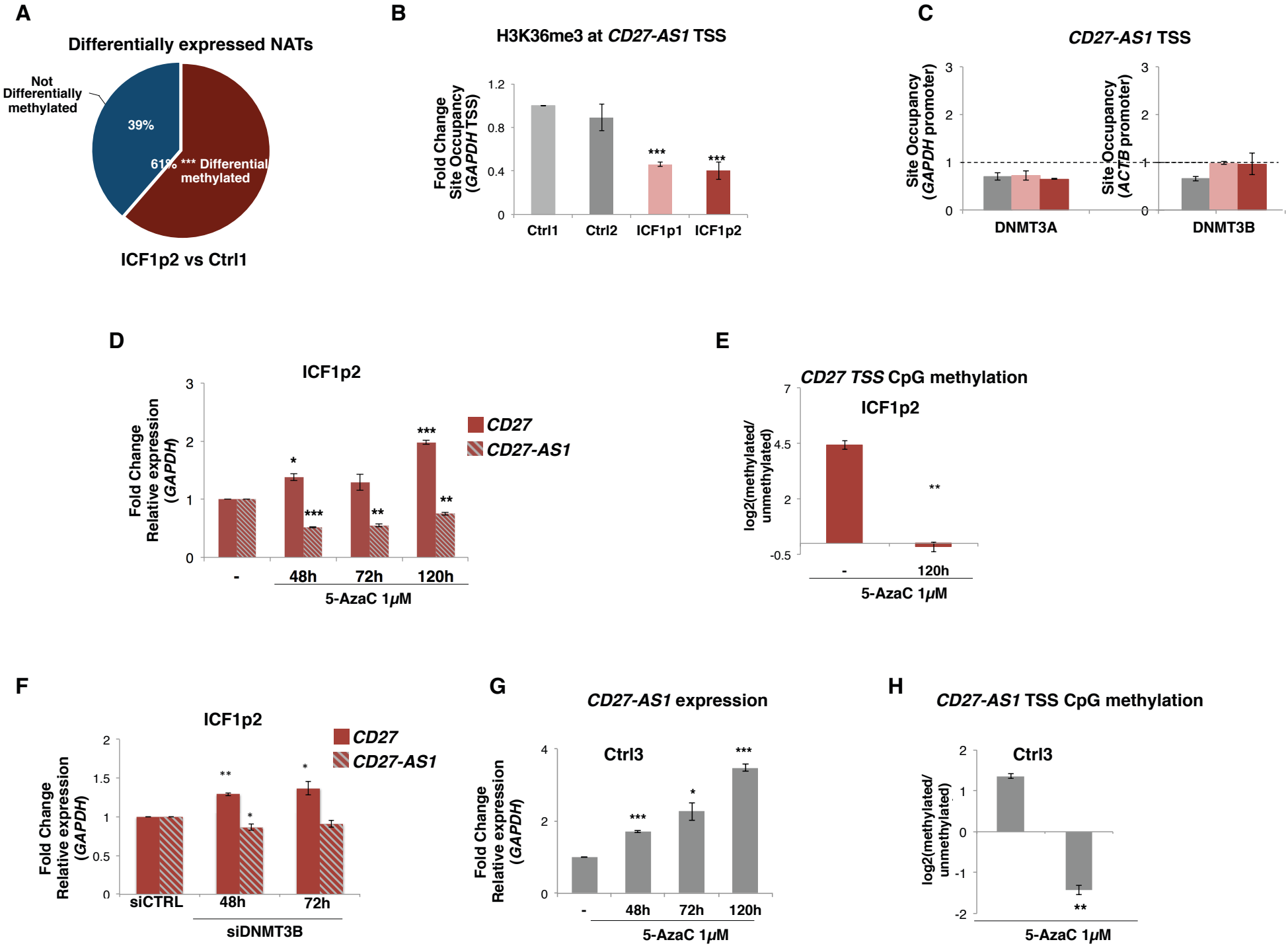
Supplementary Figure 4



Supplementary Figure 5

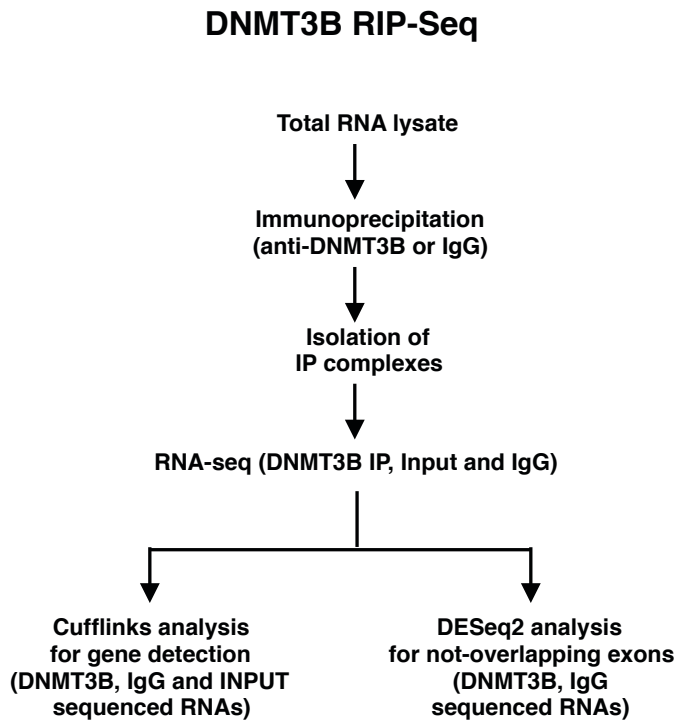


Supplementary Figure 6

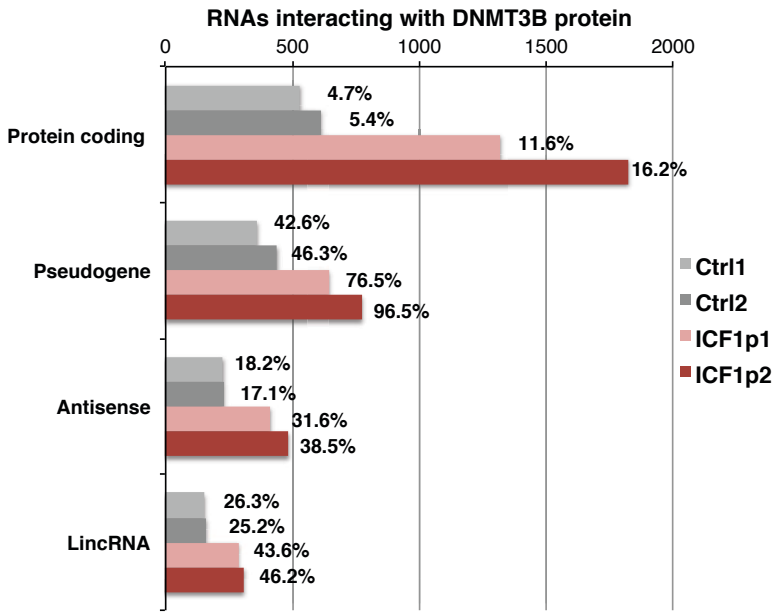


Supplementary Figure 7

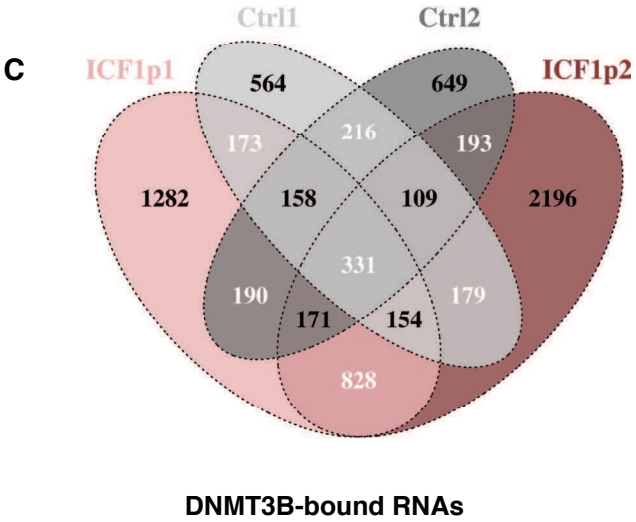
A



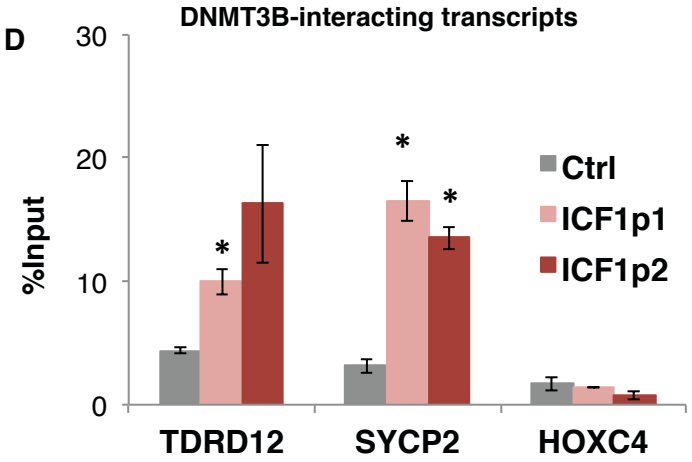
B



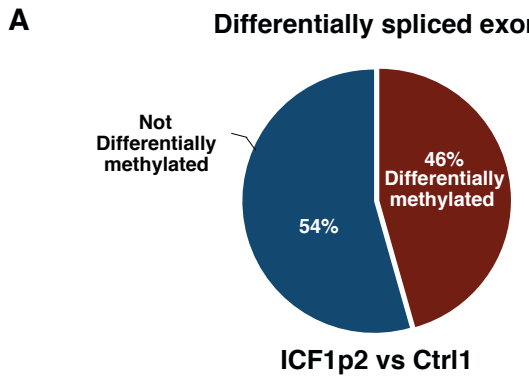
C

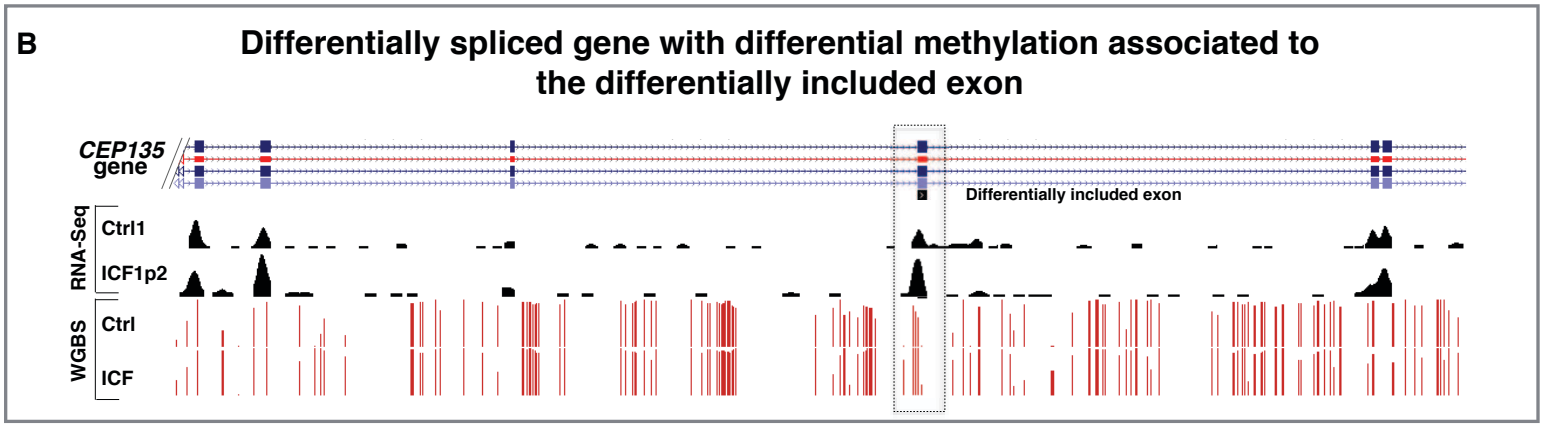


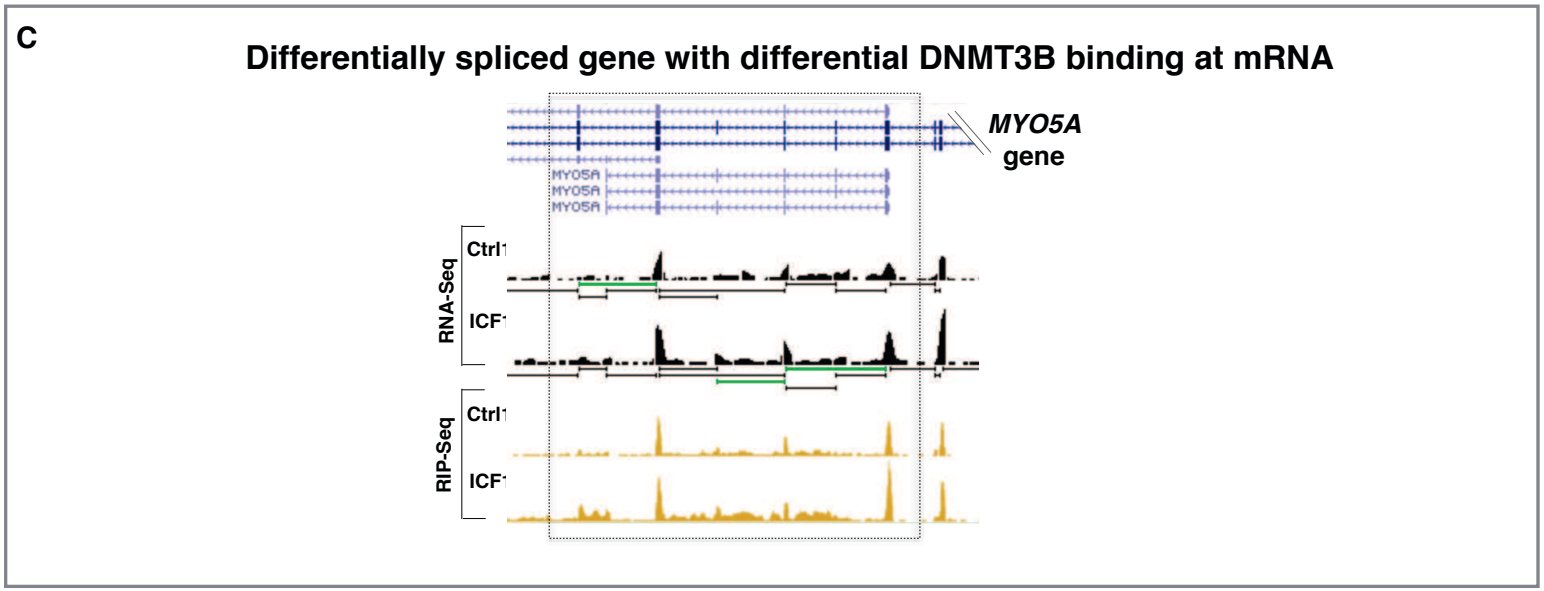
D



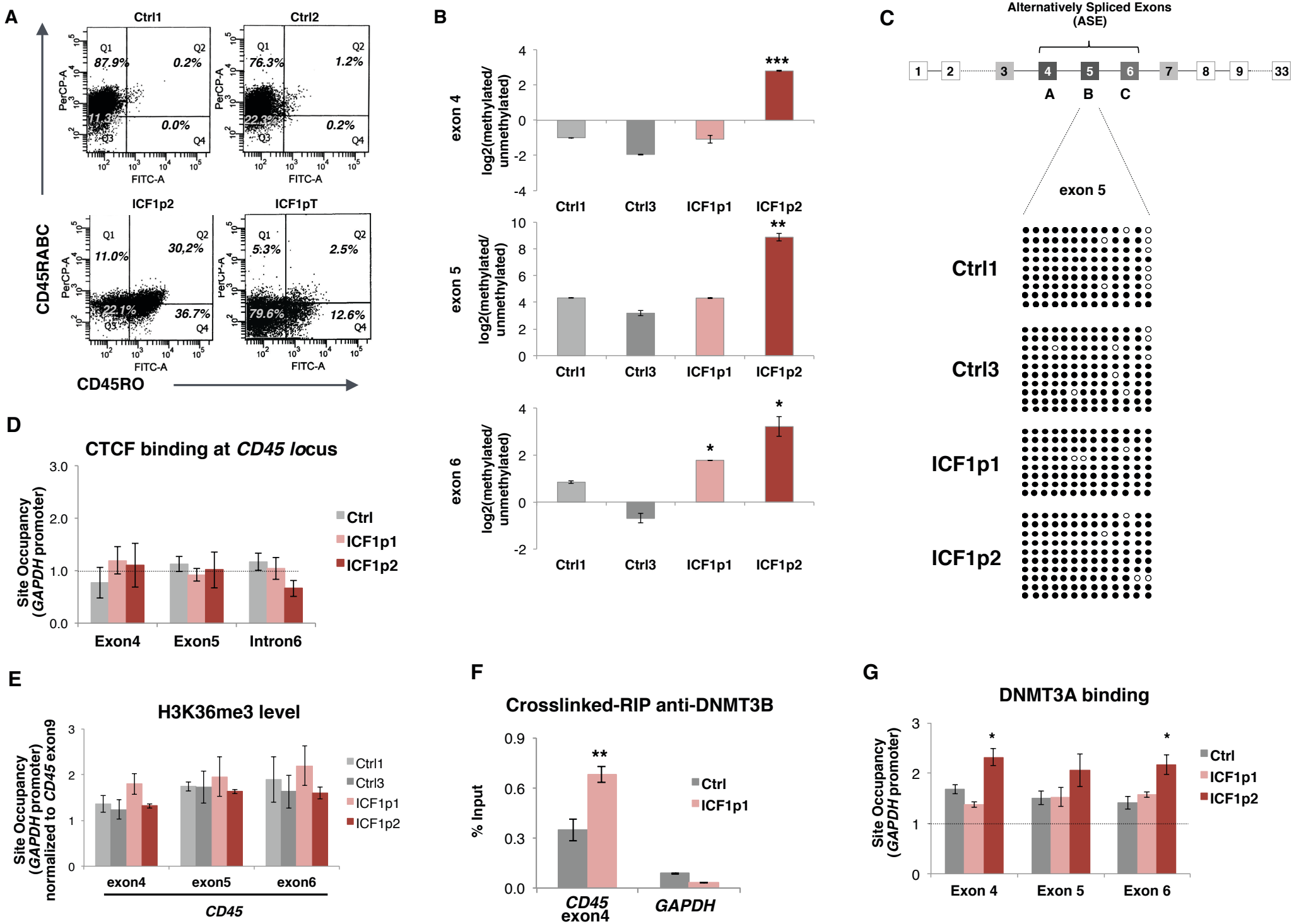
Supplementary Figure 8



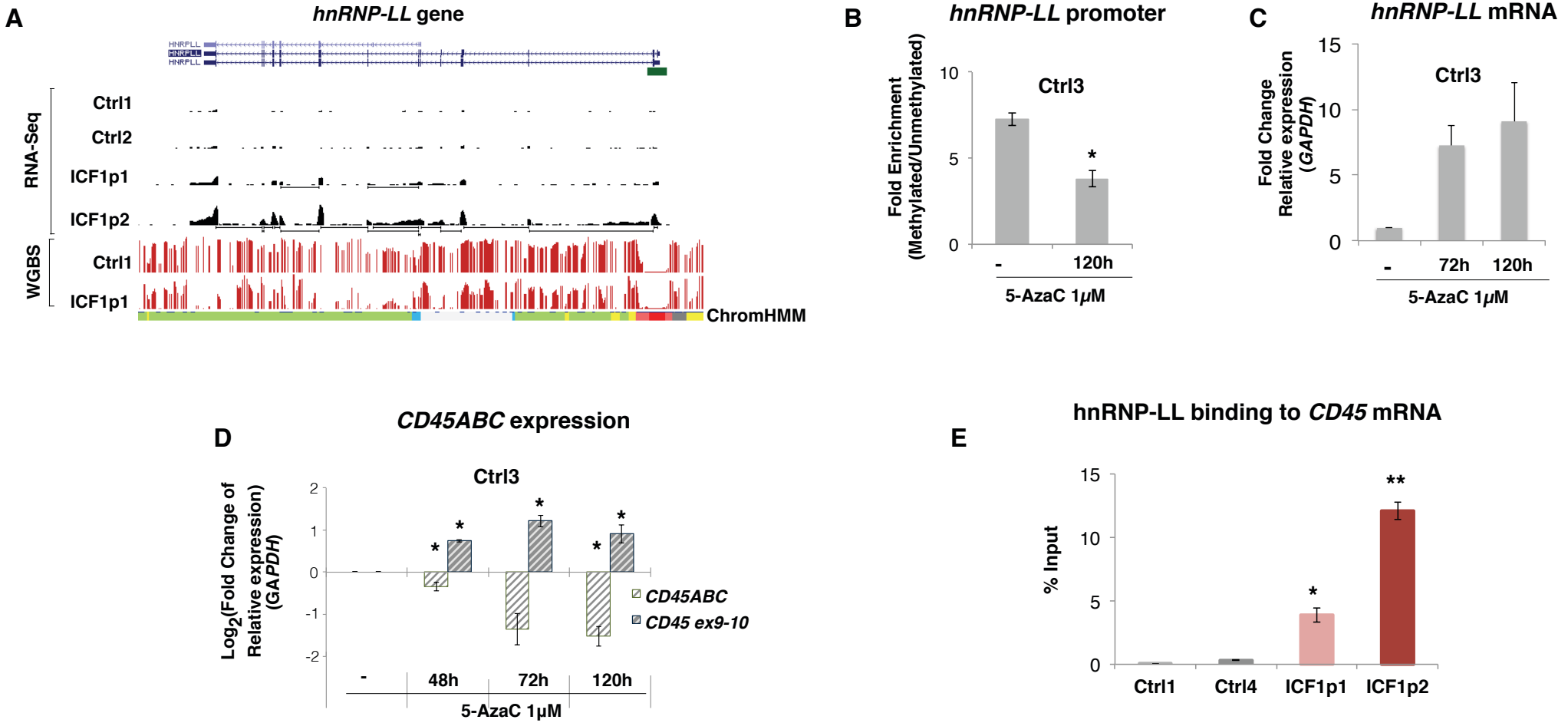




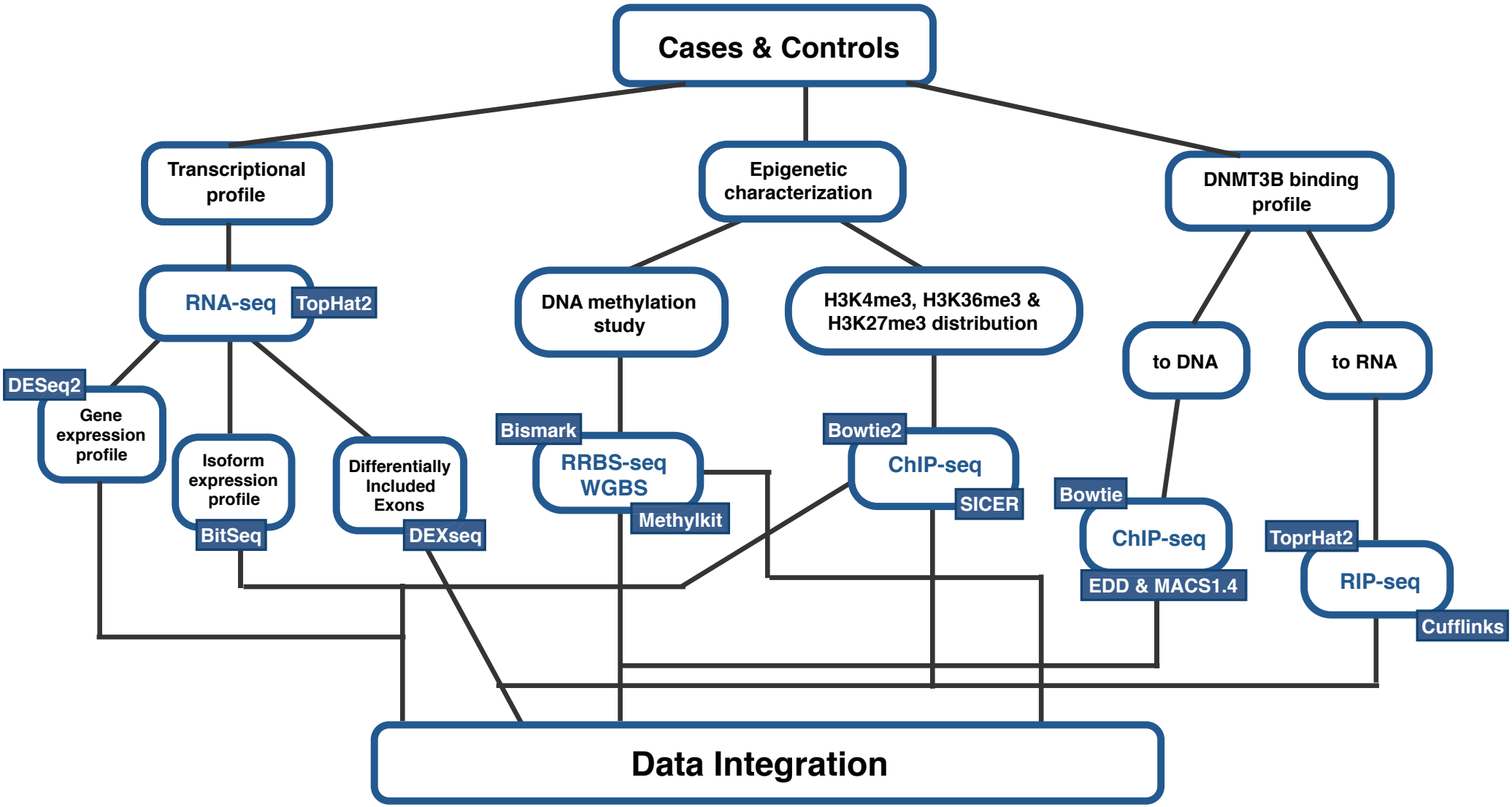
Supplementary Figure 9



Supplementary Figure 10



Supplementary Figure 11



Supplementary Figure 12

THE ICF SYNDROME AND EMERGENT PLAYERS IN DNA METHYLATION AND DEVELOPMENT: WHEN STUDYING A RARE GENETIC DISEASE SHEDS NEW LIGHT ON AN "OLD" FIELD

DNA methylation is an essential process for the development of mammals. Its abnormal distribution, particularly at the level of the repeated regions of the genome, is a pathological signature. The discovery of hereditary diseases affecting DNA methylation and the stability of the genome allowed a considerable progress in the identification of their actors and mechanisms. We chose to study the ICF (Immunodeficiency, Centromeric Instability and Facial Abnormalities) syndrome, the first genetic disorder identified with defects in the distribution of DNA methylation, linked to chromosomal instability. When I started my PhD, mutations in two genes had been described to cause the ICF syndrome: DNMT3B and ZBTB24. However, the genetic origin of a subset of ICF patients remained unknown. We identified mutations in CDCA7 and HELLS as causative of the ICF syndrome. I showed that their loss of function in somatic cells results in the loss of DNA methylation at centromeric repeats, strongly suggestive of a role DNA methylation maintenance. Hence, the study of the aetiology of a genetic disease provided new candidate "guardians" of DNA repeats and genome stability, with virtually unknown functions but with exciting potential roles in the DNA methylation machinery and in development. During my PhD, I established methylation maps in ICF patients cells to identify common and distinct targets of these factors, as well as their genomic and epigenomic characteristics. In contrast to DNMT3B mutations, those in ZBTB24, CDCA7 and HELLS affect methylation at CpG-poor regions in intergenic genomic locations and at interspersed DNA repeats, and more generally, at genomic locations with heterochromatic features. Their integrity is required for the methylated status of coding and non-coding clusters of genes, some of which are expressed in a monoallelic manner. To better characterize the role of ZBTB24 in development and DNA methylation pathways, we generated a mouse model carrying mutations in ZBTB24. We showed that ZBTB24 is essential for early development, while it seemed to be dispensable for *in vitro* differentiation of murine ES cells. We implicated ZBTB24 in the establishment of DNA methylation at DNA repeats, both in tandem or interspersed, in differentiating ES cells. Interestingly, ZBTB24 seems to be also implicated in the establishment of the repressive mark H3K9me3 suggesting that ZBTB24 may indirectly control DNA methylation through an interplay with histone marks. As a whole, our work sheds light on how DNA methylation and heterochromatin marks are established and maintained at unique genes and DNA repeats, and provides new actors and mechanisms to consider in studies of the maintenance of genome stability.

SYNDROME ICF ET ACTEURS EMERGENTS DANS LA METHYLATION DE L'ADN ET LE DEVELOPPEMENT : L'ETUDE D'UNE MALADIE GENETIQUE RARE APPORTE UN REGARD NOUVEAU SUR UN « ANCIEN » DOMAINE

La méthylation de l'ADN est un processus vital pour le développement des mammifères. Sa distribution anormale, notamment au niveau des régions répétées du génome, est une signature pathologique. La découverte de maladies héréditaires touchant la stabilité du génome a permis des avancées considérables dans l'identification des acteurs et des mécanismes. Nous avons choisi d'étudier le syndrome ICF (Immunodéficience, instabilité Centromérique et anomalies Faciales), première maladie génétique identifiée avec des défauts de la méthylation de l'ADN, liés à une instabilité chromosomique. Lorsque j'ai commencé ma thèse, des mutations dans les gènes DNMT3B et ZBTB24 avaient été décrites comme causes génétiques du syndrome. Cependant, d'autres causes génétiques restaient inconnues. Nos travaux ont permis d'identifier deux nouveaux gènes, CDCA7 et HELLS, dont les mutations sont responsables du syndrome. J'ai montré que leur perte de fonction dans les cellules somatiques entraîne un défaut de méthylation des répétitions centromériques, suggérant leur rôle dans le maintien de la méthylation de l'ADN. Par conséquent, l'étude de l'étiologie d'une maladie génétique rare a permis d'identifier de nouveaux « gardiens » de la stabilité du génome, avec des fonctions jusqu'alors insoupçonnées dans les processus de méthylation de l'ADN et dans le développement. Au cours de mon doctorat, j'ai établi des cartes de méthylation des cellules de patients ICF afin d'identifier les cibles communes et distinctes de ces facteurs, ainsi que leurs caractéristiques génomiques et épigénomiques. Contrairement aux mutations de DNMT3B, celles de ZBTB24, CDCA7 et HELLS affectent la méthylation dans des régions pauvres en CpG, dans des régions intergéniques et dans des répétitions d'ADN intercalées. Plus généralement, ce sont les régions d'hétérochromatine qui sont les plus touchées et en particulier des clusters des gènes codants et non codants, dont certains sont exprimés de manière monoallélique. Pour mieux caractériser le rôle de ZBTB24 dans le développement et la méthylation de l'ADN, nous avons généré un modèle murin mutant qui nous a permis de monter que ZBTB24 était essentielle pour le développement embryonnaire précoce. De plus, ZBTB24 jouerait un rôle dans l'établissement de la méthylation des séquences répétées de l'ADN, à la fois en tandem ou intercalé. Fait intéressant, ZBTB24 semble être également impliqué dans l'établissement de la marque répressive H3K9me3, suggérant un rôle de la protéine dans le "dialogue" entre la méthylation de l'ADN et celle des histones. Dans l'ensemble, mon travail met l'accent sur la façon dont la méthylation de l'ADN et les marques d'hétérochromatine sont établies et maintenues à des gènes uniques et des répétitions de l'ADN, et fournit de nouveaux acteurs et mécanismes à considérer dans les études sur le maintien de la stabilité du génome.