

AIX-MARSEILLE UNIVERSITE

FACULTE DE MEDECINE DE MARSEILLE

Ecole Doctorale des Sciences de la Vie et de la Santé (ED62)

THESE

Pour obtenir le grade de **Doctorat d'AIX-MARSEILLE UNIVERSITE**

Thèse de doctorat en Pathologie Humaine, **Spécialité Oncologie**

PLZF et les protéines du groupe Polycomb :
interaction et implication dans la différenciation
hématopoïétique normale et pathologique

Présentée et publiquement soutenue le 17 décembre 2015

par **Myriam KOUBI**

Membres du jury de thèse :

Pr Christian CHABANNON	Président du jury
Dr Philippe KASTNER	Rapporteur
Dr Thomas MERCHER	Rapporteur
Dr Fabien GUIDEZ	Examineur
Dr Estelle DUPREZ	Directrice de thèse

Travail de thèse réalisé au Centre de **Recherche en Cancérologie de Marseille**

REMERCIEMENTS

J'aimerais tout d'abord remercier les membres de mon jury, les Dr Philippe Kastner, Thomas Mercher et Fabien Guidez d'avoir accepté d'évaluer mes travaux de thèse.

Je remercie tout particulièrement le Professeur Christian Chabannon, qui a accepté de présider mon jury et qui a toujours été présent pour me conseiller et m'encourager malgré ses nombreuses responsabilités et obligations.

Je remercie enfin le Docteur Estelle Duprez d'avoir dirigé et suivi mes travaux tout au long de la thèse et de m'avoir fait confiance en me permettant de participer à des projets collaboratifs passionnants.

J'adresse mes sincères remerciements à nos collaborateurs. Tout d'abord, sur la plateforme bioinformatique du CRCM, je remercie le Dr Samad El Kaoutari qui a pris le relais de Mr Guillaume Tiberi et sur le site de l'IBDM à Luminy je remercie le Dr Andrew Saurin qui nous a aidés dans les analyses. Je remercie également le Docteur Bertrand Nadel et les membres de son équipe : Dominique, Jean-Marc et bien sûr Marie avec qui travailler a été un réel plaisir. Enfin, parmi mes collaborateurs, je compte bien sûr les membres de l'équipe « ED », actuels ou anciens : merci Christelle, Guillaume, Noémie, Nadine, Anne-Marie, Boris, Sarah, Christine, Amelle pour les nombreuses discussions, scientifiques ou non que nous avons eues et pour les moments passés ensemble. Merci à Frédérique Lembo pour ses nombreux conseils et pour son aide dans certaines manip.

Je tiens également à remercier deux personnes qui m'ont également aidée et toujours bien conseillée durant ma thèse, tout d'abord, le Dr Vincent Géli, mon parrain de thèse, et le Dr Patrice Dubreuil, conseiller bénévole, à qui je dois beaucoup.

Tous mes remerciements aux personnes du CRCM que j'ai souvent sollicitées pour des conseils sur les manip : Sébastien, Pascal, Danièle, Sylvie, Avais, Florence, Olivier, Audrey, Rémy....

Je n'oublie pas mes amis du labo ou non qui m'ont tous, sans exception, soutenue tout au long de cette thèse (remerciements personnalisés sur demande...) : Alexia, Sitraka, Anne, Christophe, Denise, Wilfried, Armelle, Mandy, Marion, Pascale, Besma, Khedidja, Cyndie, Kahina, Sophie, Michel, Amandine, Anne-Laure, Maria, Elodie, Luisa, Quentin, Jean-Hugues, Douglas, Rania, Emilie, Greg, Julie, Laetitia, Nadia, Kévin, Elodie, Vanessa, Aurélia, Licia, Audrey, Mike, Marie-Lou, Claude, Jacques...

J'ai également une pensée pour les Pr Pierre Cau et Patrice Roll qui m'ont permis de faire mes premiers pas dans la recherche et qui m'ont donné envie de faire une thèse.

Et le meilleur pour la fin, last but not least, je remercie toute ma famille qui m'a toujours soutenue durant mes (longues) études. Je remercie du fond de mon cœur mes parents, ma sœur et ma grand-mère pour leur soutien et leur patience.

Abréviations

ADN : Acide DésoxyriboNucléique
AEBP2 : Adipocyte Enhancer-Binding Protein 2
AML1 : Acute Myeloid Leukemia 1
AR : récepteur aux androgènes
ASXL1: Additional Sex Combs like 1
ATRA : All-Trans Retinoic Acid
BTB/POZ : Bric à Brac, Tramtrack and Broad-complex/ Pox virus and Zinc finger
CBP : CREB Binding Protein
Cbx: ChromoBoX
CDKs : Cyclin-Dependant Kinases
CGI : CpG Islands
ChIP(-seq) : Chromatin Immunoprecipitation (sequencing)
Co-IP : co-immunoprécipitation
CMH-I : complexe majeur d'histocompatibilité de type I
CpG : Cytosine-phosphate-Guanine
CRPC : castration-resistant prostate cancer
CSH : cellules souches hématopoïétiques
Dnmt3a et b : DNA méthyltransférases 3a et b
EED : embryonic ectoderm development
ER : récepteur aux œstrogènes
EZH 1/2 : Enhancer of Zeste Homolog 1/2
FDR : False Discovery Rate
HAT : Histones Acétyl-Transférases
HDAC : Histone Deacetylase
HMT : Histone Methyltransferase
ILC : Innate Lymphoid Cells
IFN : Interféron
Jarid2: Jumonji, AT Rich Interactive Domain 2
KD : knock-down
KDM : Lysines Déméthylases
KDM2B : lysine (K)-specific demethylase 2B
KO : knockout
LAM : leucémies aiguës myéloblastiques
LAL-T ou LAL-B : leucémies aiguës lymphoblastiques T ou B
L3MBTL2 : L(3)Mbt-Like 2 (Drosophila)
Lnc-RNA : longs ARN non-codants
LT-HSC : Long term Hematopoietic Stem Cells
MDS : syndrome myélodysplasique
N-Cor : Nuclear Co-repressor
NK: cellules Natural Killer
NKT : cellules Natural Killer T
NuRD : Nucleosome Remodeling Deacetylase
Oct4 : Octamer-binding transcription factor 4
PARP: poly(ADP-ribose) polymérase
Pb : paires de bases
P/CAF : p300/CBP-associated factor

PcG : Polycomb Group
 Pcgf : Polycomb Group ring Finger
 PCL : PolyComb Like
 Pho : Pleiohomeotic
 PhoRC : Pho Repressive Complex
 PLZF : Promyelocytic Leukemia Zinc Finger
 PML : Pro-Myelocytic Leukemia
 PRC1/2 : Polycomb Repressive Complex 1/2
 PREs : Polycomb Response Elements
 RAREs : éléments de réponse à l'acide rétinoïque
 RAR α : récepteur α à l'acide rétinoïque
 RbAp48/46 : Retinoblastoma-associated proteins 46 and 48
 RD2 : Repression Domain 2
 ROR α : Retinoic acid-related orphan nuclear receptor α
 ROS : Reactive Oxygen Species
 Rybp : RING1 And YY1 Binding Protein
 Sall1/4 : Sal-Like protein 1/4
 shRNA : short hairpin RNA
 SMRT : Silencing Mediator for Retinoid and Thyroid receptors
 SUMO : Small Ubiquitin-like Modifier
 SUZ12 : SUppressor of Zeste 12
 TAL1 : T-cell Acute Lymphocytic Leukemia 1
 TERT : Telomerase Reverse transcriptase
 TFs : Transcription factors
 TrxG : Trithorax Group
 TSS: Transcription Start Site
 Xist : X-inactive specific transcript
 YY1 : Yin Yang 1
 Zbtb16 : Zinc Finger and BTB domain containing 16

Glossaire

Hétérochromatine facultative : elle correspond à des régions transcriptionnellement inactives mais contenant des gènes qui peuvent être activés en fonction du développement cellulaire, de l'état de différenciation, des stimuli,...

Région enhancer : région d'ADN qui correspond à une séquence régulatrice de nucléotide (<10 paires de bases) qui peut fixer des protéines nommées facteurs de transcription pour stimuler la transcription d'un gène. Ce dernier peut posséder plusieurs régions enhancer qui sont généralement situées assez loin du gène (jusqu'à 100 000 nucléotides). Le gène et sa région enhancer peuvent aussi ne pas être proches l'un de l'autre et même être sur deux chromosomes différents. Cependant le repliement de l'ADN dans le noyau leur permet une proximité physique.

Paralogue : Protéine homologues qui ont des fonctions différentes mais liées, au sein d'un même organisme.

Avant propos

L'épigénétique est un domaine de la biologie relativement récent qui a bouleversé les dogmes établis par la découverte de la double hélice d'ADN par Watson et Crick en 1953. Selon ces dogmes, nous sommes conditionnés par un programme génétique hérité de nos parents et dont la dérégulation entraîne de nombreuses pathologies, plus ou moins graves. La découverte d'une hérédité épigénétique est cependant venue bousculer la toute puissance du gène en mettant en relief de nouveaux mécanismes de régulation au sein des cellules. En effet, l'épigénétique correspond aux modifications réversibles et héréditaires qui touchent la chromatine et qui influencent l'expression des gènes sans toutefois impliquer de changement de la séquence d'ADN.

Concrètement, les modifications épigénétiques sont matérialisées par des marques apposées par des enzymes spécialisées sur l'ADN ou sur les protéines histones qui le structurent. Les marques les mieux caractérisées sont les groupements méthyle apposés sur l'ADN, ainsi que diverses modifications chimiques des histones (méthylation – tout comme l'ADN mais via des protéines différentes -, acétylation...). Si la méthylation de l'ADN est le plus souvent associée à une répression des gènes, les marques histones ne déterminent pas, par leur seule présence, un état permissif ou non de la chromatine. En effet, si certaines marques sont plutôt associées à une chromatine compactée ou relâchée, c'est l'ensemble des marques histones sur un gène donné qu'il faut considérer pour savoir si le gène est actif ou non. Il existe donc de nombreuses protéines, organisées le plus souvent en complexes, capables de reconnaître, d'ajouter ou d'enlever les marques histones.

Les processus épigénétiques interviennent en effet dans la régulation de nombreux événements tels que la division, la différenciation, la survie et la mobilité cellulaires. Il est désormais largement admis que des anomalies épigénétiques contribuent au développement et à la progression de maladies, en particulier de cancers. Des anomalies épigénétiques activant des oncogènes ou inhibant des gènes suppresseurs de tumeurs ont pu être mises en évidence. Si de nombreuses études sont menées afin de comprendre quels sont les mécanismes épigénétiques impliqués dans le développement et la progression des cancers, il reste de nombreuses questions sans réponses.

Le travail présenté ici s'inscrit donc dans cette vaste problématique. L'objectif principal étant d'apporter, dans la mesure du possible, des éléments de compréhension aux mécanismes de régulation épigénétiques mis en jeu dans des conditions normales et pathologiques. Nous avons pris ici pour modèle l'hématopoïèse.

SOMMAIRE

REMERCIEMENTS

ABREVIATIONS

AVANT PROPOS

INTRODUCTION

I. Généralités sur l'épigénétique	p 1
1.1 Définitions	p 1
1.2 Lecture et analyse combinatoire des modifications post-traductionnelles des histones	p 3
1.3 Conclusion	p 4
II. Les protéines du groupe Polycomb	p 5
2.1 Présentation des protéines du groupe Polycomb	p 5
2.1.1 PhoRC	p 5
2.1.2 PRC1	p 5
2.1.3 PRC2	p 6
2.2 Mode d'action des protéines PcG	p 7
2.2.1 Généralités	p 7
2.2.2 Modèle de recrutement hiérarchique des protéines PcG	p 8
2.3 Recrutement des protéines PcG à la chromatine	p 10
2.3.1 Les éléments de réponse aux Polycomb chez la Drosophile	p 10
2.3.2 Les éléments de réponse aux Polycomb chez les vertébrés	p 11
2.3.3 Les îlots CpG	p 11
2.3.4 Recrutement du complexe PRC2 par des facteurs de transcription	p 12
2.3.5 Rôle des modifications d'histones dans le recrutement des PRC2	p 13
a) Densité du nucléosome	p 13
b) Rôle activateur des modifications d'histones	p 14
c) Rôle inhibiteur des modifications d'histones	p 14
d) Interaction des PRC2 avec les ARNs	p 15
2.4 Fonctions des protéines et complexes PcG	p 16
2.4.1 Rôle des protéines PcG dans le développement embryonnaire	p 16
2.4.2 Rôle des protéines PcG dans la différenciation et l'auto-renouvellement cellulaires	p 17

2.5 Les protéines PcG et leur implication dans les cancers	p 19
2.5.1 Rôle général des protéines PcG dans les cancers	p 19
a) Implication des protéines du complexe PRC2 dans les cancers	p 19
b) Implication des protéines du complexe PRC1 dans les cancers	p 19
2.6 Cas de la protéine EZH2	p 20
2.6.1 Rôle d'EZH2 dans les hémopathies et les cancers solides	p 20
a) Anomalies d'EZH2 dans les hémopathies : EZH2, oncogène ou suppresseur de tumeur ?	p 20
b) Anomalies d'EZH2 dans les cancers solides : EZH2, une protéine à double facette	p 21
2.6.2 Nouvelles fonctions récemment identifiées d'EZH2	p 22
a) EZH2 : un activateur transcriptionnel dans les cancers	p 22
b) EZH2 : une enzyme qui ne modifie pas que les histones	p 23
2.7 CONCLUSION	p 23
III. La protéine PLZF : Promyelocytic Leukemia Zinc Finger	p 23
3.1 Caractéristiques de PLZF	p 23
3.2 Rôle général et expression de PLZF	p 24
3.3 Modèles animaux de PLZF	p 24
3.4 Fonctions biologiques de PLZF	p 25
3.4.1 Auto-renouvellement versus différenciation	p 25
3.4.2 Cycle cellulaire et apoptose	p 26
3.4.3 Remodelage de la chromatine (via les HDAC)	p 27
3.4.4 PLZF et immunité	p 27
3.4.5 PLZF et hématopoïèse	p 28
3.5 Rôle activateur de PLZF	p 29
3.6 Modifications post-traductionnelles de PLZF	p 29
3.6.1 Acétylation	p 29
3.6.2 Phosphorylation	p 29
3.6.3 Sumoylation	p 30
3.6.4 Ubiquitination	p 30
3.7 PLZF et cancers	p 30
3.7.1 Translocation PLZF- RAR α	p 30
3.7.2 Implication de PLZF dans les cancers (forme non fusionnée)	p 31
3.8 Conclusion	p 32

ARTICLE 1 : A non-conventional genomic association between PLZF and EZH2 revealed through genome-wide analyses	p 37
ARTICLE 2 : PLZF mutation results in cell cycle alteration and accelerated mouse hematopoietic stem cell aging	p 77
ARTICLE 3 : Site- and allele-specific polycomb dysregulation in T-cell leukaemia	p 109
 DISCUSSION ET PERSPECTIVES	
I. Interaction entre PLZF et EZH2	p 123
II. Expériences de ChIP-seq : Pertinence des données et artéfacts éventuels liés à l'analyse	p 124
III. EZH2 et activation des gènes	p 125
IV. Quel rôle de la protéine EZH1 ?	p 127
V. Double rôle de PLZF sur la transcription de ses gènes cibles	p 127
VI. Exemple d'une dérégulation d'enhancer par insertion : cas du gene TAL1	p 129
1. Activation mono-allélique de TAL1	p 129
2. Création d'un super enhancer en amont du gène TAL1	p 129
 BIBLIOGRAPHIE	p 133
 ANNEXE	p 151

Introduction

I. Généralités sur l'Epigénétique

1.1.Définitions

Le terme « épigénétique » a été introduit pour la première fois par C. Waddington en 1942 (Waddington, 2012) et a d'abord été utilisé pour décrire l'apparition de changements héréditaires dans l'activité des gènes sans modification du code génétique. La définition s'est ensuite élargie pour inclure les mécanismes de régulation de l'ADN qui n'induisent pas de changement dans la séquence nucléotidique, en dépit du caractère héréditaire (Berger et al., 2009).

Le matériel génétique des cellules eucaryotes est condensé au sein du noyau. La compaction de l'ADN est assurée par les nucléosomes qui sont constitués d'octamères de protéines histones (H2A, H2B, H3 et H4). L'ensemble ADN/nucléosomes forme la chromatine dont l'organisation structurale permet de : i) prévenir les dommages à l'ADN, ii) renforcer l'ADN lors des mitoses (stress physique), iii) contrôler l'expression des gènes et la réplication de l'ADN.

Les nucléosomes font l'objet de réarrangements structuraux en fonction du degré d'enroulement de l'ADN et sont sujets à des modifications covalentes. Ces modifications, qui constituent les « marques épigénétiques », concernent aussi bien les histones que l'ADN. Tandis que la modification principale de l'ADN est la méthylation, les histones sont susceptibles de porter de nombreuses modifications post-traductionnelles. En effet, lorsque les histones sont associées pour former un nucléosome, leurs extrémités N-terminales dépassent du nucléosome (« queues histones ») et sont ainsi accessibles aux différentes enzymes responsables de l'ajout ou du retrait des modifications post-traductionnelles.

Ces modifications influencent les interactions entre les histones et l'ADN, les histones entre elles ou sont les cibles de protéines effectrices.

De très nombreuses modifications post-traductionnelles des histones ont été répertoriées parmi lesquelles on compte :

- Sur les résidus lysine : l'acétylation, la méthylation, l'ubiquitination, la SUMOylation, la crotonylation, la butyrylation et la propionylation
- Sur les résidus arginine : la méthylation, la citrullination et l'ADP-ribosylation
- Sur les résidus sérine et thréonine : la phosphorylation et la glycosylation

La conjugaison de ces modifications dicte l'état transcriptionnel de la chromatine et dans les années 2000, Strahl et Allis ont évoqué l'existence d'un « code histone » (figure 1) dont les différentes combinaisons conduiraient à différentes fonctions en aval (Strahl and Allis, 2000).

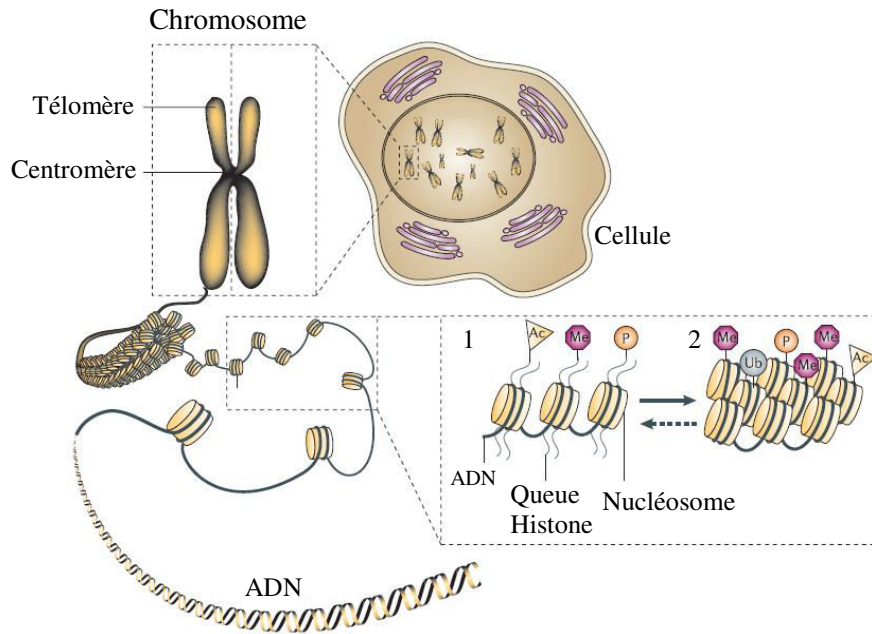


Figure 1. Le code histone. L'information génétique dans les cellules eucaryotes est compactée en chromosomes au sein desquels l'ADN est enroulé autour des histones pour former la chromatine. Les changements dynamiques de structure de la chromatine sont directement influencés par des modifications post-traductionnelles des histones. Les différentes combinaisons formées par ces modifications constituent le code histone qui permet d'établir un état de répression (2) ou d'activation (1) de l'activité transcriptionnelle des gènes. Adapté à partir de (Sparmann and van Lohuizen, 2006).

Suite à la découverte du code histone, sont apparus les termes de “writer”, “eraser” et “reader”, formulés afin de décrire respectivement les protéines ou complexes protéiques en charge de déposer, de retirer ou encore de reconnaître les différentes modifications post-traductionnelles des histones. Schématiquement, les « writers » ont pour rôle d'apposer les diverses marques histones au niveau des acides aminés constituant les queues histones. Ce sont des enzymes telles que les Histones AcétylTransférases (HATs), les Histones MéthylTransférases (HMTs) ou encore les kinases. Les « erasers » sont des enzymes qui ont pour rôle de retirer les marques catalysées par les « writers ». On trouve plusieurs types « d'erasers » : les Histones DésACétylases (HDACs), les lysines DéMéthylases (KDMs) et les phosphatases. Les « readers » sont des protéines possédant des domaines spécialisés dans la reconnaissance des différentes marques histones.

En plus des modifications des histones, d'autres facteurs tels que la méthylation de l'ADN, les variants d'histones, le positionnement des histones... sont nécessaires à l'ajustement de la structure et de la fonction de la chromatine. L'ensemble de ces facteurs constitue la « machinerie épigénétique ».

1.2. Lecture et analyse combinatoire des modifications post-traductionnelles des histones

Le premier « reader » des modifications post-traductionnelles des histones a été identifié en 1999. Il s'agissait de la protéine P/CAF (p300/CBP-associated factor). Les auteurs ont montré pour la première fois que le bromodomaine de P/CAF est capable de reconnaître les lysines acétylées. C'est également le premier exemple soutenant l'hypothèse selon laquelle les bromodomains contribuent à l'acétylation des histones en recrutant des histones acétyltransférases (Dhalluin et al., 1999). La liste des « readers » d'histones s'est ensuite très vite agrandie, mais il n'existe pas autant de « readers » que de combinaisons de modifications d'histones. Il existe donc différents mécanismes permettant la lecture et l'analyse des marques épigénétiques (récapitulés figure 2).

L'efficacité d'un « reader » par rapport à une modification donnée peut être influencée par les marques histones adjacentes qui peuvent améliorer ou entraver l'interaction du « reader » avec la marque histone (fig. 2a).

D'autre part, le recrutement d'un « reader » au niveau d'un locus spécifique peut être modulé par l'action combinée de plusieurs « readers » (fig. 2b, c). Ces différents « readers » peuvent reconnaître des marques histones différentes situées sur la même queue histone (fig. 2b). La lecture de la combinaison des marques histones se fait donc en *cis*. Lorsque les « readers » reconnaissent des modifications au sein du même nucléosome, mais sur des queues histones différentes ou au sein de nucléosomes différents, la lecture combinatoire se fait en *trans* (fig. 2c).

Il existe par ailleurs, un niveau bien plus élaboré de lecture combinatoire en *trans*, impliquant des complexes multi-protéiques (fig. 2d). L'interaction entre les différents effecteurs impliqués, qu'ils soient « reader », « eraser » ou « writer », confère un haut degré de spécificité d'action au niveau de la chromatine.

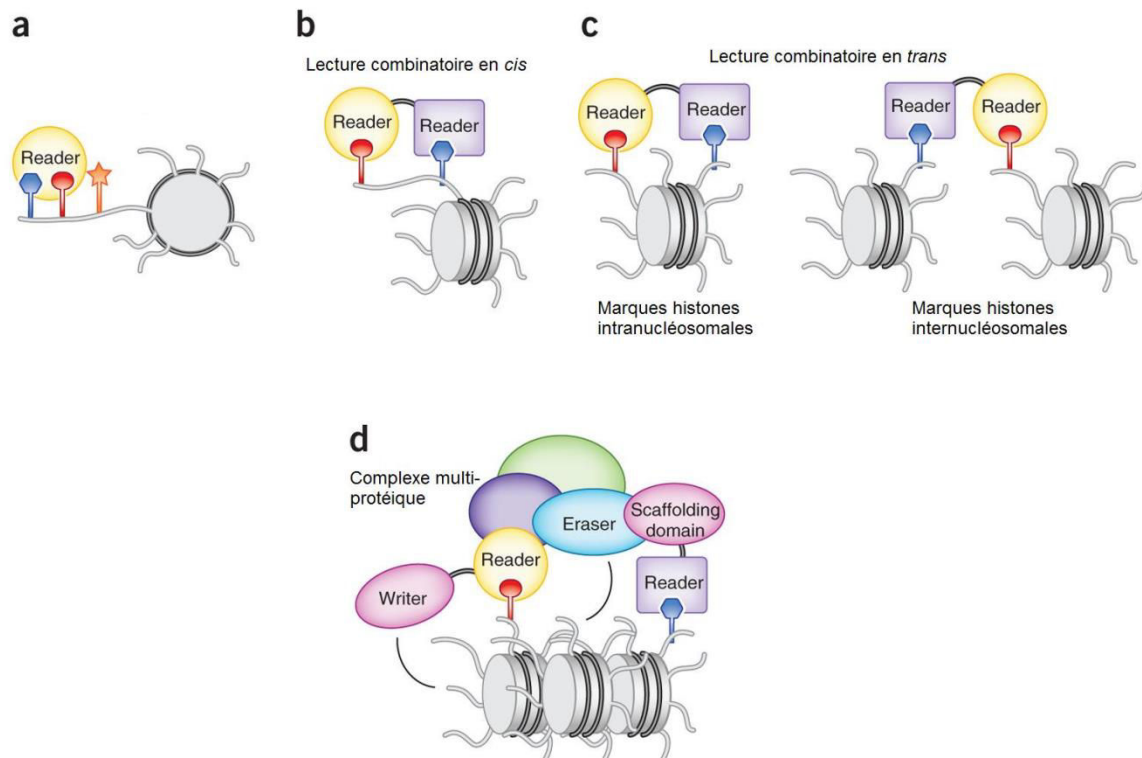


Figure 2. Modes de lecture combinatoire des modifications post-traductionnelles des histones. (a) Reconnaissance des marques histones sur la même queue histone par le même « reader ». (b) Lecture combinée de plusieurs « readers » sur la même queue histone (mécanisme de lecture en *cis*). (c) Lecture combinée de plusieurs « readers » sur des queues histones différentes au sein du même nucléosome ou au sein de nucléosomes différents mais adjacents (mécanisme de lecture en *trans*). (d) Complexe multi-protéique pouvant contenir « writer », « reader » et « eraser ». Certaines protéines sont dotées d'un domaine de repliement servant à ponter les différents effecteurs. Adapté de (Musselman et al., 2012).

1.3.Conclusion

Les différents motifs créés par les modifications post-traductionnelles des histones résultent de l'interaction entre les différents effecteurs que sont les « writers », « readers » et « erasers ». Ces diverses combinaisons établissent le paysage chromatinien et sont associées à de multiples fonctions vitales telles que la régulation de l'expression des gènes, la réplication de l'ADN et sa réparation. Comme tout mécanisme de régulation impliqué dans des fonctions biologiques majeures, la dérégulation de l'équilibre épigénétique conduit à de nombreuses maladies parmi lesquelles dominent les cancers.

Lors de ma thèse, je me suis intéressée tout particulièrement à un groupe de facteurs épigénétiques complexe, tant par leurs fonctions que par leurs mécanismes d'action. Il s'agit des protéines du groupe Polycomb qui illustrent parfaitement l'existence d'effecteurs « writers » et « readers » au sein de complexes multi-protéiques dont le rôle de régulateur transcriptionnel est souvent mis à mal au sein de nombreuses pathologies.

II. Les protéines du groupe Polycomb

2.1 Présentation des protéines du groupe Polycomb

Les protéines du groupe Polycomb (PcG) sont des facteurs épigénétiques initialement découverts chez la *Drosophila* et impliqués dans la répression des gènes *Hox* au cours du développement. Les protéines PcG forment différents complexes protéiques comportant plusieurs sous-unités dotées de fonctions différentes et dont le mode d'action repose sur la modification de la structure de la chromatine. Ces protéines sont capables de générer et de reconnaître de nombreuses modifications post-traductionnelles au niveau des histones. On dénombre à ce jour trois complexes répresseurs comprenant différentes protéines PcG.

2.1.1 PhoRC

Le premier complexe nommé PhoRC (pour **Pho** Repressive Complex) est composé des protéines Pho (Pleiohomeotic) et dSfmbt chez la *Drosophila* (respectivement YY1 et L3MBTL2 chez l'Homme) (fig. 2a). Chez la *Drosophila*, ce complexe recrute les autres protéines PcG par l'intermédiaire de la protéine Pho mais son rôle n'est pas aussi bien établi chez les mammifères puisque la protéine YY1 n'est pas impliquée dans la répression due aux Polycomb (Kahn et al., 2014). On trouve ensuite les complexes répresseurs PRC1 et PRC2 (pour **P**olycomb **R**epressive Complex) décrits plus précisément ci-dessous.

2.1.2 PRC1

On distingue deux principaux complexes PRC1 contenant différentes sous-unités. Le complexe PRC1 canonique comprend les protéines Cbx, Phc, Ring et Pcgf tandis que le complexe non-canonique est formé des protéines Rybp, Ring et Pcgf (fig. 3c et d). Chacune de ces sous-unités pouvant elles-mêmes avoir plusieurs paralogues (cf glossaire) (Turner and Bracken, 2013), il existe plusieurs types de complexes PRC1 canoniques et non-canoniques. Quelle que soit sa nature, l'unité catalytique de PRC1 repose sur les protéines Ring1A ou Ring1B qui catalysent l'ajout d'une ubiquitine sur la lysine 119 de l'histone H2A (H2AK119ub1) (de Napoles et al., 2004).

Selon la composition du complexe PRC1, il existe des différences de fonctions biologiques. Par exemple, il a été montré qu'au sein du complexe PRC1 canonique, la composition en protéines Cbx conditionne le destin des cellules souches embryonnaires (ES) de souris. En effet, la présence de Cbx7 au sein du complexe permet le maintien de la pluripotence des cellules tandis que la présence de Cbx2 et 4 est nécessaire à la différenciation des cellules ES (Morey et al., 2012).

Par ailleurs, en plus des différences que l'on peut observer au sein d'un même type de complexe, on constate également des différences entre les complexes canonique et non-canonique. Ainsi, le groupe de Di Croce a récemment mis en évidence, toujours dans des cellules souches embryonnaires de souris, que les complexes PRC1 canonique et non canoniques coexistent mais ont des fonctions différentes. En effet, le complexe PRC1 canonique contient Cbx7 qui est nécessaire au recrutement de Ring1B à la chromatine. Les gènes ciblés par ce complexe sont fortement réprimés et sont des gènes impliqués dans la différenciation lors du développement précoce. Au contraire, le complexe non-canonique contient la protéine Rybp qui est nécessaire à l'activité enzymatique de PRC1. Ce complexe cible plutôt des gènes impliqués dans la régulation du métabolisme et dans la progression du cycle cellulaire (Morey et al., 2013).

2.1.3 PRC2

Les principales protéines composant le complexe PRC2 sont au nombre de trois : enhancer of zeste (EZH2), embryonic ectoderm development (EED) et suppressor of zeste 12 (SUZ12). L'association de ces trois sous-unités est nécessaire à l'activité catalytique d'EZH2 qui réside en la mono-, di- ou triméthylation de la lysine 27 de l'histone H3 (H3K27me1, 2 ou 3) (Cao and Zhang, 2004; Nekrasov et al., 2005; Pasini et al., 2004). D'autres protéines peuvent se rajouter au complexe de base PRC2 telles que RbAp48/46, PCL1/2/3, AEBP2, Jarid2, C17orf96 et C10orf12 (fig. 3b). Elles augmentent alors l'activité catalytique du complexe (par exemple RbAp46/48 et AEBP2) ou modifient la liaison et l'adressage de PRC2 à la chromatine (JARID2 et PCL) (Vizán et al., 2015).

EZH2 est la seule protéine du complexe PRC2 à avoir un paralogue connu, appelé EZH1. EZH2 et EZH1 sont des protéines exprimées de façon différente en fonction du type cellulaire, et retrouvées dans des complexes PRC2 dont la composition et la fonction sont différentes. EZH2 s'associe généralement à EED et SUZ12 alors qu'EZH1 est retrouvé seul ou avec SUZ12, formant un complexe PRC2 non canonique (Xu et al., 2015).

Bien que les deux protéines aient une activité catalytique et une localisation nucléaire redondantes, EZH2 est plutôt associée à la di- et triméthylation de l'H3K27 au niveau de loci génomiques réprimés, tandis qu'EZH1 est associé à la mono-méthylation de l'H3K27 dans des régions transcriptionnellement actives (Mousavi et al., 2012; Xu et al., 2015).

Lors de la différenciation cellulaire, le ratio entre EZH1 et EZH2 au sein du complexe PRC2 se modifie en faveur d'une augmentation d'EZH1 (Margueron et al., 2008; Mousavi et al., 2012; Xu et al., 2015).

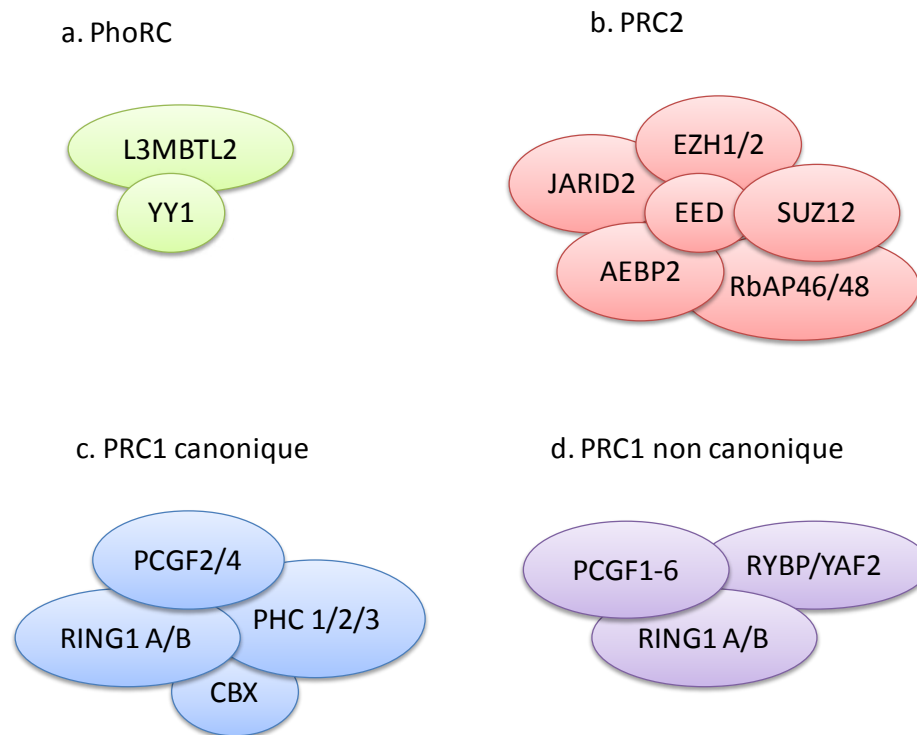


Figure 3. Composition des différents complexes Polycomb chez les Mammifères.

2.2 Mode d'action des protéines PcG

2.2.1 Généralités

Les complexes PRC agissent tous les deux en tant que modificateurs d'histones, c'est-à-dire qu'ils sont capables de modifier de façon post-traductionnelle les histones via la triméthylation de l'H3K27 pour PRC2 et via l'ubiquitination de l'H2AK119 pour PRC1. Ces modifications permettent de réguler la transcription en servant de sites de fixation spécifiques ou en modulant l'affinité des PcG à l'ADN (Musselman et al., 2012).

Les protéines PcG limitent l'accès à l'ADN de la machinerie de transcription en compactant la chromatine (pour revue (Di Croce and Helin, 2013; Schwartz and Pirrotta, 2013)). Par ailleurs, les protéines PcG peuvent induire la répression transcriptionnelle en empêchant la modification épigénétique des histones par des facteurs activateurs tels que les protéines du groupe Trithorax (TrxG). Les protéines TrxG catalysent la triméthylation de l'histone H3 sur les lysines 4 et 36 (H3K4me3 et H3K36me3) au niveau de gènes transcriptionnellement actifs. De nombreuses études ont montré que les protéines PcG avaient un effet antagoniste à celui des protéines TrxG sur la régulation de la transcription (Steffen and Ringrose, 2014).

Les protéines PcG contrecarrent également certaines modifications activatrices au niveau des histones situées dans des régions précises du génome telles que les enhancers. Par exemple, la méthylation de l'H3K27 empêche son acétylation (H3K27me3 et H3K27ac sont des marques antagonistes). Or, l'H3K27ac est une marque que l'on retrouve particulièrement enrichie au niveau de régions enhancer actives (Ferrari et al., 2014).

2.2.2 Modèle de recrutement hiérarchique des protéines PcG

Le modèle hiérarchique de recrutement des protéines PcG a été établi progressivement en fonction de plusieurs observations. Tout d'abord, des études ont montré que la triméthylation de l'H3K27 pouvait servir de plateforme de recrutement à certaines protéines du complexe PRC1 (les protéines PC chez la *Drosophila*, Cbx chez les mammifères) (Cao et al., 2002). Ensuite, il a été observé que lorsque l'activité enzymatique du complexe PRC2 est perdue, l'H3K27 n'est plus triméthylée et la liaison de PRC1 à l'ADN diminue fortement (Boyer et al., 2006; Cao et al., 2002; Wang et al., 2004). Ces résultats ont donné naissance au modèle hiérarchique (ou modèle séquentiel) qui propose que le recrutement de PRC2 à la chromatine induit la triméthylation de l'H3K27. Cette marque est ensuite reconnue par le complexe PRC1 via les protéines Cbx qui ont une affinité avec l'H3K27me3. Le complexe PRC1, via les protéines Ring, catalyse l'ubiquitination de l'histone H2AK119, ce qui a pour effet de compacter la chromatine et de réprimer la transcription (fig. 4a).

Bien que le modèle hiérarchique de recrutement ait dominé le champ d'investigation des protéines du groupe Polycomb pendant plusieurs années, il est toutefois apparu très vite qu'il existait des mécanismes alternatifs de recrutement et donc de fonctionnement de cette famille de protéines. Des études génomiques sur des cellules ES ont tout d'abord révélé que les protéines des complexes PRC1 et PRC2 ne partageaient qu'un nombre limité de sites de liaison à l'ADN (Blackledge et al., 2014; Boyer et al., 2006; Ku et al., 2008). Des expériences de CHIP-on-chip sur des cellules ES de souris indiquent que seulement 25% des sites de démarrage de la transcription (TSS pour Transcription Start Sites) sont co-occupés par PRC1 et PRC2 (protéines PRC1 étudiées : Phc1 et Rnf2 ; protéines PRC2 étudiées : Eed et Suz12) (Boyer et al., 2006).

D'autres analyses de CHIP-sequencing sur les protéines Ring1B et Ezh2 montrent que si 90% des sites de liaison de Ring1B à l'ADN sont également occupés par Ezh2, seulement 50% des sites de liaison d'Ezh2 sont occupés par Ring1B (Ku et al., 2008). Ces résultats montrent que PRC1 et PRC2 ne se lient pas toujours aux mêmes régions, contrairement à ce

qui est prédit par le modèle hiérarchique, ce qui indique que les complexes PRC1 et 2 peuvent avoir des fonctions et des mécanismes de recrutement indépendants.

Dès 2012, des analyses de génomique et de protéomique sur les complexes PRC1 ont permis d'identifier 6 groupes principaux caractérisés par des sous-unités distinctes et qui diffèrent dans leurs sites de liaison à l'ADN, parmi lesquels seulement une petite partie co-localise avec l'H3K23me3 (Gao et al., 2012). De plus, il a été démontré que le recrutement de PRC1 n'est pas uniquement dépendant de l'H3K27me3. En effet, dans des cellules souches embryonnaires de souris déficientes pour les protéines du complexe PRC2, il a été démontré que la marque H2AK119ub pouvait toujours être déposée, entraînant la répression des gènes (Tavares et al., 2012).

Des études plus récentes montrent également que PRC1 peut être impliqué dans le recrutement de PRC2 (Blackledge et al., 2014; Cooper et al., 2014; Kalb et al., 2014) (fig. 4b). En effet, la diminution d'expression (knockdown) de certaines protéines du complexe PRC1, en plus d'induire une diminution drastique de l'H2AK119ub1, entraîne une diminution de liaison de PRC2 à l'ADN. Par ailleurs, l'adressage forcé des protéines du complexe PRC1 à la chromatine entraîne l'apposition de la marque H2AK119ub1 qui recrute le complexe PRC2. De façon intéressante, seuls les complexes non-canoniques peuvent recruter PRC2 (Blackledge et al., 2014; Cooper et al., 2014). Une étude complémentaire a montré par des expériences de pull down sur des extraits nucléaires de *Drosophila* ou de cellules ES de souris, que les composants de PRC2 sont fortement associés avec l'H2AK118ub ou l'H2AK119ub respectivement. Cette étude démontre que l'H2A ubiquitinée sert de site de liaison pour le complexe PRC2 contenant Jarid2-Aebp2 et entraîne ainsi la production d'H3K27me3 (Kalb et al., 2014).

Ces nouvelles données viennent bouleverser le modèle hiérarchique de recrutement des Polycomb et démontrent que les complexes PRC1 et 2 peuvent influencer leur recrutement respectif.

Le modèle classique peut d'ailleurs être discuté dans la mesure où il se focalise sur la triméthylation de l'H3K27 par PRC2 alors que ce complexe est également responsable de la mono- et diméthylation de l'H3K27 (Ferrari et al., 2014). Au contraire de la marque répressive H3K27me3 qui se situe majoritairement au niveau de sites de démarrage de la transcription (TSS), la marque H3K27me1 est principalement trouvée enrichie au niveau du corps de gènes activement transcrits. La marque H3K27me2 se distribue quant à elle sur tout le génome, couvrant environ 70% des histones. Cette marque empêcherait une acétylation aspécifique de l'H3K27 dans le but de prévenir une activation non spécifique de certains

enhancers (Ferrari et al., 2014). Ces résultats assez récents mettent en lumière de nouveaux mécanismes de régulation de l'activité transcriptionnelle des gènes par les protéines du groupe PcG.

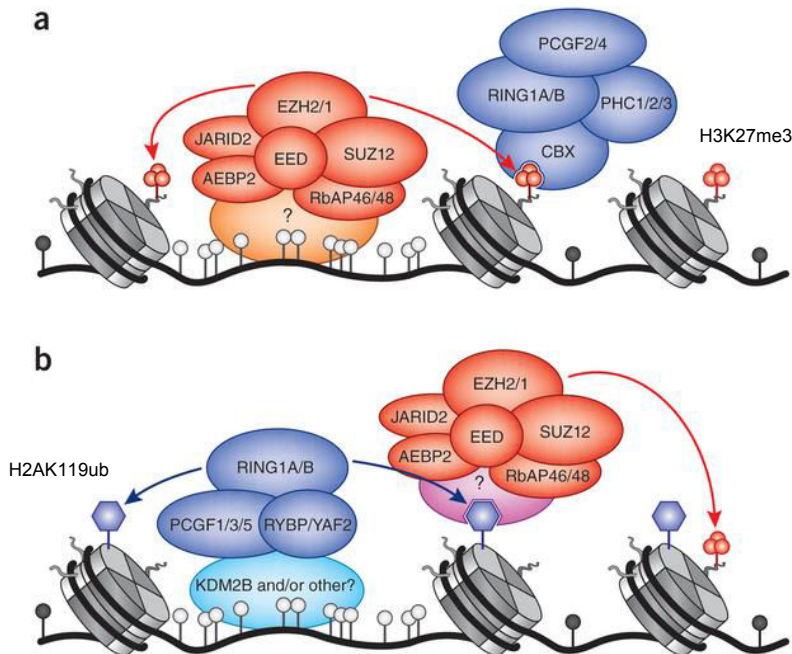


Figure 4 : Différents modèles de recrutement des complexes Polycomb. a) Modèle hiérarchique selon lequel PRC2 est recruté à la chromatine et catalyse, via EZH2 l'apposition de la marque H3K27me3 qui a son tour conduit au recrutement de PRC1 dont l'activité catalytique ubiquitine ligase sur l'H2AK119 est faible. b) Nouveau modèle hiérarchique proposé par Blackledge *et al.* et Cooper *et al.* suggérant que les complexes PRC1 non canoniques sont d'abord recrutés à la chromatine, ce qui induit la monoubiquitination de l'H2AK119 et le recrutement de PRC2.

D'après (Comet and Helin, 2014).

2.3 Recrutement des protéines PcG à la chromatine

Les mécanismes biochimiques par lesquels les protéines PcG inactivent la transcription ont été (et sont encore) largement étudiés. Cependant la façon dont les complexes PRC sont adressés à leurs gènes cibles demeure une question importante.

2.3.1 Les éléments de réponse aux Polycomb chez la *Drosophila*

Des motifs de recrutement spécifiques aux Polycomb ont été mis en évidence pour la première fois chez la *Drosophila* au sein du complexe Bithorax, complexe regroupant trois gènes homéotiques impliqués dans la polarité segmentaire. Ces éléments de réponse aux protéines PcG appelés PREs (pour **P**olycomb **R**esponse **E**lements) sont capables de réprimer un gène rapporteur lorsqu'ils en sont placés à proximité (Simon et al., 1993).

La protéine Pho a été la première protéine identifiée se liant à l'ADN de façon séquence spécifique et capable de conduire au recrutement des protéines PcG au niveau des PREs. En effet, Pho reconnaît une séquence de 17 paires de bases (pb) localisée au sein d'un PRE de 176 pb se trouvant en amont du locus du gène « engrailed », réprimé par les protéines PcG (Brown et al., 1998).

Suite à cette étude, de nombreux PREs contenant des motifs de liaison à divers facteurs de transcription tels que les protéines Gaga (Strutt, 1997), Pho ou Zeste (Hur et al., 2002; Saurin et al., 2001) ont été découverts chez la *Drosophila*.

En se basant sur ces différents motifs, la localisation des PREs au sein du génome chez la *Drosophila* a été prédite (Ringrose et al., 2003) mais d'autres études menées à l'échelle du génome montrent que les complexes PRC1 et 2 ne se lient qu'à quelques-uns de ces PREs (Nègre et al., 2006; Schwartz et al., 2006; Tolhuis et al., 2006).

Cette apparente contradiction trouve plusieurs explications. Tout d'abord, la prédiction de la localisation des PREs se base sur l'hypothèse que tous les PREs identifiés ont la même fonction au même moment sans tenir compte de la complexité du mode d'action des Polycomb dont les cibles varient au cours du développement. Ensuite, la position des PREs par rapport aux gènes régulés est très variable (proche du promoteur ou éloigné de plusieurs kilobases), ce qui pourrait induire un mode de fonctionnement différent de ces PREs.

Enfin, de la même façon que Pho peut recruter PRC1 à la chromatine via la reconnaissance d'un PRE, d'autres facteurs spécifiques de certains loci pourraient influencer l'activité de certains de ces PREs (pour revue (Kassis and Brown, 2013)).

2.3.2 Les éléments de réponse aux Polycomb chez les vertébrés

Il n'existe pas, chez les vertébrés, de PREs à proprement parler mais quelques études ont permis d'identifier des séquences s'en approchant.

Ainsi, un PRE a été identifié chez la souris. La séquence nommée PRE-kr est capable de réguler l'expression du gène murin MafB/Kreisler en recrutant les protéines PcG (Sing et al., 2009).

Un autre PRE d'1.8kb a été identifié chez l'Homme, au niveau des gènes HOXD11 et HOXD12. Les Polycomb sont capables de se lier à cette séquence et l'utilisation d'un système de gène rapporteur montre une répression Polycomb-dépendante (Woo et al., 2010).

Un PRE capable de recruter les Polycomb, notamment la protéine SUZ12, a également été découvert dans des cellules humaines (cellules T) (Cuddapah et al., 2012).

2.3.3 Les îlots CpG

Les CpG sont des di-nucléotides cytosine et guanine reliés par un phosphate. Très peu fréquents dans le génome, les CpG se regroupent en îlots appelés CGI pour CpG Islands.

Dans les cellules ES, la densité des îlots CpG et leur état de méthylation sont de bons prédicteurs du recrutement de PRC2. En effet, sur des cellules ES humaines, il a été montré que 40% des régions auxquelles se lie la protéine SUZ12 sont enrichies en CpG (Lee et al.,

2006). Par ailleurs, dans des cellules souches embryonnaires de souris, l'insertion d'éléments riches en CpG permet de recruter PRC2 (Mendenhall et al., 2010). Une autre étude menée dans des cellules souches embryonnaires murines montre également que la présence d'éléments riches en CpG induit le recrutement de PRC2 (Lynch et al., 2012).

La méthylation des CGI empêche la liaison de PRC2. En effet, des analyses de spectrométrie de masse indiquent que l'incorporation de CGI méthylés au sein de nucléosomes inhibe la liaison des protéines EED et SUZ12, membres de PRC2 (Bartke et al., 2010). Le phénomène d'exclusion mutuelle entre la présence de CGI méthylés et de la marque H3K27me3 a été également démontré dans les génomes de vertébrés. En effet, dans des cellules ES déficientes pour les DNA méthyltransférases Dnmt3a et b, des sites de recrutement des Polycomb sont créés (sites « de novo ») au niveau de séquences riches en CpG qui sont normalement méthylées (Lynch et al., 2012).

De façon générale, chez les vertébrés, le complexe PRC2 est adressé à des régions hypométhylées.

En ce qui concerne PRC1, il a été montré que la protéine de liaison à l'ADN KDM2B reconnaît les CGI non méthylés et forme un complexe avec la protéine Polycomb RING1B. Le recrutement de KDM2B aux promoteurs induit la présence de la marque H2AK119ub, suivie de la liaison de PRC2 à l'ADN et l'inhibition de la transcription des gènes (Farcas et al., 2012).

2.3.4 Recrutement du complexe PRC2 par des facteurs de transcription

Un grand nombre de facteurs de transcription interviennent dans le recrutement des complexes PRC chez les vertébrés. Cela a été montré lors des premières études à l'échelle génomique des protéines Polycomb et de la marque H3K27me3 dans des cellules ES de souris dans lesquelles les auteurs observaient une co-localisation entre la protéine PcG SUZ12 et des facteurs de transcription impliqués dans la pluripotence des cellules tels que Sox2, Oct4 et Nanog (Lee et al., 2006). Plus récemment, une étude menée chez le Xénope a montré que le facteur de transcription Snail2 interagit avec le complexe PRC2 via la protéine Ezh2. Cette interaction a pour effet de moduler l'expression des gènes de la crête neurale en maintenant le niveau d'H3K27me3 (Tien et al., 2015).

Le complexe PRC1 peut également être recruté à la chromatine par des facteurs de transcription. C'est le cas de la protéine Bmi1 qui interagit avec deux facteurs de transcription formant un complexe : Runx1/CBF β (Yu et al., 2012). Dans cette même étude, les auteurs montrent, sur un modèle de cellules mégacaryocytaires murine, que la protéine Ring1b et le

complexe Runx1/CBF β partagent un nombre important de gènes cibles. La déplétion génétique de Runx1 entraîne une diminution concomitante de la liaison de Ring1b à la chromatine.

De façon plus générale, une étude prédictive informatique menée à partir de données publiques a suggéré que la liaison d'un facteur de transcription à l'ADN permet de prédire quelle sera la marque histone associée. Ces résultats renforcent l'hypothèse selon laquelle les facteurs de transcription peuvent servir de plateforme de recrutement aux protéines modificatrices d'histones (Benveniste et al., 2014).

2.3.5 Rôle des modifications d'histones dans le recrutement de PRC2

La structure de la chromatine peut influencer la liaison de PRC2 à l'ADN de deux façons : par le biais de modifications de la densité des nucléosomes, ou via une intercommunication avec les histones modifiées.

a) Densité du nucléosome

Au niveau des sites de fixation des protéines PcG et TrxG, les histones sont remplacées fréquemment par d'autres histones (Mito et al., 2007) et la liaison de PRC2 à l'ADN au niveau des CGI est inversement corrélée à la densité des nucléosomes (Riising et al., 2014).

De façon contradictoire, la liaison et l'activité de PRC2 sont augmentées en présence de di-nucléosomes, comparé aux mono-nucléosomes (Martin et al., 2006) et bien que la densité des nucléosomes au niveau des CGIs diminue avant le recrutement de PRC2, il semblerait que la compaction de ces nucléosomes augmente localement avant la formation d'H3K27me3 (Yuan et al., 2011).

Ces études indiquent donc que les protéines PcG se lient à des régions peu denses en nucléosomes, densité qui est modifiée juste avant la liaison du complexe.

b) Rôle activateur des modifications d'histones

Des modifications d'histones préexistantes telles que H3K27me3, H2AK119ub et H3K9me3 peuvent faciliter le recrutement de PRC2. Ces marques épigénétiques sont partiellement transmises lors de la prolifération cellulaire et sont reconstituées au moyen d'un rétrocontrôle positif. Par exemple, si PRC2 catalyse la marque H3K27me3, le complexe est également capable de s'y lier via la protéine Eed. La liaison d'Eed à l'H3K27me3 induit une activation allostérique du complexe et la propagation de la marque (Margueron et al., 2009; Xu et al., 2010).

L'H3K27me3 peut également servir de plateforme de recrutement au complexe PRC1 via les protéines Cbx (Boyer et al., 2006; Cao et al., 2002; Wang et al., 2004). Par ailleurs,

comme nous l'avons vu plus haut, la marque H2AK119ub1 apposée par le complexe PRC1 peut permettre le recrutement de PRC2 (Blackledge et al., 2014; Cooper et al., 2014; Kalb et al., 2014).

Bien que cela n'influence pas l'activité catalytique d'EZH2, il a été montré que la protéine Eed présente une forte affinité pour l'H3K9me3 (Xu et al., 2010).

Dans les cellules HeLa et dans les cellules ES de souris, PRC2 interagit avec les protéines G9a/GLP qui sont des H3K9 méthyltransférases partageant 25% de ces loci cibles avec PRC2. Au niveau de ces sites de liaison communs, l'H3K27me3 diminue dans des cellules déficientes pour G9a et/ou GLP. La liaison de G9a au niveau d'un site artificiel de recrutement conduit au recrutement d'EZH2 et à la triméthylation de l'H3K27, ce qui n'est pas le cas avec un mutant catalytique de la protéine G9a. De plus, le défaut de recrutement d'Ezh2 dans des cellules ES murines mutantes pour G9a peut être aboli en présence de la protéine sauvage (Mozzetta et al., 2014).

c) Rôle inhibiteur des modifications d'histones

Les modifications des histones associées à l'activation de la transcription telles que l'H1K26me3, l'H3K27ac ou encore l'H3S28ph inhibent le recrutement de PRC2.

i. H1K26me3

Le complexe PRC2 peut être détourné de ses gènes cibles via sa liaison à l'H1K26me3. En effet, l'H1K26me3 entre en compétition avec les marques répressives H3K27me3 et H3K9me3 pour la liaison à la cage aromatique de la protéine Eed. La liaison entre PRC2 et l'H1K26 a pour effet de diminuer l'activité catalytique de ce complexe (Xu et al., 2010).

ii. H3K27ac

L'acétylation et la méthylation de l'histone H3 sur le même résidu, *ie* la lysine 27, sont des modifications exclusives bien qu'elles puissent coexister sur différentes queues histones au sein d'un même nucléosome. Des analyses à l'échelle génomique menées sur des embryons de *Drosophila* et sur des cellules ES murines ont révélé qu'acétylation et méthylation de l'H3K27 sont anti-corrélées (Pasini et al., 2010a; Tie et al., 2009).

Dans les cellules ES de souris, il a été montré que la désacétylation par le complexe NuRD de l'H3K27 entraîne le recrutement d'un complexe PRC2 actif sur le plan catalytique (Reynolds et al., 2012).

iii. H3S28ph

La phosphorylation de la sérine 28 de l'histone H3 facilite l'acétylation de la K27 de cette même histone. Le recrutement de l'H3K28 kinase Msk1 au niveau du promoteur endogène de l' α -globuline dans les cellules HEK293 entraîne l'activation de la transcription de ce gène. Ce

recrutement entraîne une augmentation d'H3S28ph et d'H3K27ac tandis que l'H3K27me3 diminue (Lau and Cheung, 2011).

Par ailleurs, l'effet inhibiteur de l'H3K228ph sur le recrutement de PRC2 a été observé dans les cellules HeLa. Dans ces cellules, en réponse à un stress cellulaire par exemple, les kinases MSK phosphorylent l'H3K28 sur des queues histones déjà triméthylées sur l'H3K27, ce qui a pour résultat de diminuer la liaison des protéines Cbx et Suz12 (Gehani et al., 2010).

Une autre étude basée sur la liaison de PRC2 au promoteur de la *myogénine* au cours de la différenciation des cellules musculaires squelettiques montre que la liaison de Msk1 et l'augmentation de l'H3K28ph lors de l'activation de la transcription déplace EZH2 au niveau du promoteur (Stojic et al., 2011).

d) Interaction de PRC2 avec les ARNs

➤ Les longs ARN non-codants

De nombreuses études ont montré que des ARNs spécifiques pouvaient se lier à PRC2. C'est le cas des longs ARN non codants (ou lncRNA pour **long non coding RNA**) tels que Xist ou HOTAIR.

Lors du phénomène d'inactivation du chromosome X chez la femme, c'est le lncRNA Xist (X-inactive specific transcript) qui recrute le complexe PRC2. Les mécanismes de recrutement de PRC2 par ce lncRNA sont cependant encore débattus (da Rocha et al., 2014; Zhao et al., 2008).

Le lncRNA HOTAIR intervient, lui, dans la répression des gènes hox chez l'humain en interagissant avec PRC2 (Rinn et al., 2007; Tsai et al., 2010). De plus, une étude récente montre que les lncRNAs fonctionnent comme des supports permettant de stabiliser la liaison de certaines protéines à des sous-unités de PRC2 telles qu'Ezh2 et Jarid2 (Kaneko et al., 2014).

➤ Les ARN naissants

En plus de se lier à des ARN de façon séquence spécifique, PRC2 est capable de se lier à des molécules d'ARN de façon moins sélective. L'immunoprécipitation de l'ARN dans des cellules ES murines montre que PRC2 s'associe avec des milliers de molécules d'ARN différentes (Kaneko et al., 2013; Zhao et al., 2010).

Deux études récentes montrent que les ARN liés à PRC2 correspondent parfois à des régions 5' de gènes dont la transcription est active. La première étude montre que PRC2 se lie de façon peu spécifique à l'ARN en cours de transcription. Si le gène cible transcrit est une véritable cible des protéines PcG, le recrutement de PRC2 via l'interaction avec l'ARN

entraîne la reconnaissance d'H3K27me3 précédemment déposée et la restauration de la répression via la stimulation de l'activité histone méthyltransférase de PRC2. A contrario, si PRC2 se lie au transcrit d'un gène actif, en l'absence d'H3K27me3 et en présence d'H3K36me3 et d'H3K4me3, l'activité histone méthyltransférase de PRC2 est inhibée et son affinité pour les nucléosomes réduite. Selon ce modèle, dit du « Junk Mail », la liaison de PRC2 à des ARN permettrait de « scanner » les gènes cibles de PRC2 et de repérer les gènes ayant échappé à la répression transcriptionnelle (Davidovich et al., 2013).

La deuxième étude montre que PRC2 se lie faiblement à une grande partie des promoteurs dans les cellules ES de souris, y compris aux promoteurs actifs et dépourvus d'H3K27me3 pour scanner leur activité (rôle de « sampling »). Cette large activité de surveillance des promoteurs permet à PRC2 d'évaluer l'état d'activation d'un large nombre de gènes. Si le gène est en cours de transcription, PRC2 le détecte grâce à son interaction avec l'extrémité 5' de l'ARN naissant (ie ARN en cours de transcription avant même l'étape d'épissage). Cette interaction provoque une diminution d'H3K27me3 ce qui permet à la chromatine de rester ouverte et au promoteur de rester actif. Lors de la différenciation, des répresseurs transcriptionnels lignée-spécifiques sont surexprimés et inhibent la transcription de leurs gènes cibles. L'absence d'ARN naissants capables de se lier à PRC2 permet au complexe de catalyser la marque répressive H3K27me3 donnant naissance à une hétérochromatine facultative (cf glossaire) (Kaneko et al., 2013).

Une troisième étude vient cependant nuancer ces résultats et montre que la liaison d'EZH2 aux ARN peut être très spécifique (Cifuentes-Rojas et al., 2014).

Ces études suggèrent donc un modèle dans lequel PRC2 se lie à l'ARN pour scanner le génome et évaluer l'activité transcriptionnelle des gènes dans le but de déployer ou de redistribuer le complexe en conséquence.

2.4 Fonctions des protéines et des complexes PcG

2.4.1 Rôle des protéines PcG dans le développement embryonnaire

Comme dit précédemment, les protéines PcG ont été découvertes chez la *Drosophila* comme étant impliquées dans la répression des gènes lors du développement. Les premières études montraient que les protéines PcG contrôlaient la segmentation lors de l'embryogénèse précoce en maintenant la répression spatiale et temporelle des gènes Hox (Duncan, 1982; Lewis, 1978). Des études d'inactivation génique (knockout, KO) de différentes protéines PcG chez la souris ont ensuite montré que, tout comme chez la *Drosophila*, les protéines PcG avaient un rôle important lors du développement via leur implication dans la répression transcriptionnelle

des gènes Hox notamment. En effet ces études de KO ont généré des embryons présentant des défauts lors de la gastrulation. Les embryons KO pour les protéines Eed, Ezh2 et Suz12 meurent lors des phases précoces suivant l'implantation de l'embryon (Faust et al., 1995; O'Carroll et al., 2001; Pasini et al., 2004). La perte de la sous-unité catalytique de PRC1, ie Ring1B, conduit également à la létalité embryonnaire via un arrêt de la gastrulation, tandis que les souris KO pour Ring1A sont viables (de Naples et al., 2004; Voncken et al., 2003). De plus, des souris ayant un allèle hypomorphe pour Ring1B présentent une transformation homéotique postérieure du squelette axial (Suzuki et al., 2002).

Les composants non-canoniques de PRC1 sont également essentiels pour le développement embryonnaire. Les embryons KO pour Rybp meurent après la phase d'implantation de l'embryon (Pirity et al., 2005). L3mbtl2 joue un rôle essentiel dans la gastrulation puisque son KO compromet la prolifération cellulaire et conduit à une différenciation anormale des cellules souches embryonnaires (Qin et al., 2012). Enfin, Kdm2b est nécessaire au bon développement neural embryonnaire via la régulation de la prolifération et de la mort cellulaires chez la souris (Fukuda et al., 2011).

A contrario, la déplétion des protéines Cbx qui sont responsables du recrutement du complexe PRC1 canonique, n'affecte pas le développement embryonnaire mais les souris mutantes pour les protéines Cbx2 et Cbx4 présentent une mortalité post-natale (Core et al., 1997; Liu et al., 2013). De façon intéressante, Cbx4 semble réguler de façon spécifique la prolifération des cellules épithéliales thymiques et le maintien de cet épithélium, mettant en lumière un nouveau rôle des protéines PcG au sein du système immunitaire (Liu et al., 2013).

La délétion des protéines Mel18 et Bmi1 entraîne des défauts de spécification de l'axe antéro-postérieur du squelette (Akasaka et al., 1996; Alkema et al., 1995; Lugt et al., 1994). Le double KO Mel18/Bmi1 induit la mort des souris aux environs de 9.5 jours post-coïtum (Akasaka et al., 2001) et provoque des défauts développementaux plus sévères que chacun des KO séparément, suggérant que Mel18 et Bmi1 ont des fonctions partiellement redondantes.

2.4.2 Rôle des protéines PcG dans la différenciation et l'auto-renouvellement cellulaires

Au-delà de leur rôle majeur lors du développement embryonnaire, les protéines PcG jouent également un rôle dans la différenciation cellulaire.

Au sein des cellules ES, certains facteurs de transcription spécifiques tels que Oct4, Sox2 et Nanog sont exprimés afin de garantir la pluripotence des cellules et de permettre l'auto-renouvellement des cellules (Boyer et al., 2005). Il a été montré que ces facteurs de

transcription sont associés aux protéines PcG afin de réprimer des gènes clés impliqués dans la différenciation cellulaire (Lee et al., 2006).

De façon surprenante, seule la perte de l'activité de PRC1 affecte les capacités d'auto-renouvellement des cellules ES. La perte de l'activité de PRC2 affecte plutôt leurs capacités de différenciation. En effet, lors de l'induction de la différenciation, les cellules déficientes pour PRC2 n'arrivent pas à activer correctement les programmes de transcription lignage-spécifiques (Chamberlain et al., 2008; Endoh et al., 2008; Pasini et al., 2007). Ces résultats ont d'abord été expliqués par l'incapacité des cellules déficientes à maintenir la répression de gènes lignée-spécifiques (Bracken and Helin, 2009). Mais de récentes données ont montré que la perte de l'activité de PRC2 perturbe l'apposition des marques H3K27me1 et H3K27me2 qui se situent respectivement au niveau des régions intra-géniques de gènes hautement transcrits et au niveau d'enhancers qui ne doivent pas être activés (Ferrari et al., 2014).

De plus, la déplétion de certaines sous-unités qui régulent PRC2, tels que Jarid2, a aussi été rapportée comme affectant les programmes de différenciation cellulaire. En effet, les cellules ES déplétées pour Jarid2 ne peuvent plus se différencier, ce qui est cohérent par rapport à sa fonction de recruteur de PRC2 (Landeira et al., 2010; Pasini et al., 2010b).

Les protéines Polycomb ont également un rôle fondamental dans l'homéostasie des cellules souches adultes et des cellules somatiques. Par exemple, dans le système hématopoïétique, le développement des cellules B précoces dépend de l'activité d'Ezh2 qui régule le réarrangement du gène *Igh* (Su et al., 2003). De façon similaire, une étude a montré que des CSH qui n'expriment plus Eed ne sont pas capables de générer des cellules sanguines matures (Xie et al., 2014). Par ailleurs, Bmi1 est requis pour le maintien du stock de CSH via la répression de la transcription du locus *Ink4a/Arf* (Park et al., 2003). Les protéines Cbx quant à elles, régulent étroitement les capacités d'auto-renouvellement et de différenciation des CSH. Cbx7 est présent en grande quantité dans les CSH et sa surexpression peut entraîner des leucémies. Au cours de la différenciation des CSH, la quantité de protéines Cbx7 diminue au profit des protéines Cbx2/4/8, qui, lorsqu'elles sont surexprimées, induisent l'épuisement des cellules souches (Klauke et al., 2013).

Les protéines Polycomb peuvent également intervenir dans d'autres processus cellulaires que l'hématopoïèse. La formation des adipocytes est perturbée en absence d'activité de PRC2 qui ne réprime plus la voie de signalisation Wnt (Wang et al., 2010). PRC2 est également requis lors de la myogenèse au cours de laquelle un important niveau d'expression d'EZH2 dans les précurseurs empêche la transcription prématurée de gènes spécifiques activant la myogenèse (Caretta et al., 2004). La présence de PRC2 est également nécessaire pour inhiber le

programme de différenciation des entérocytes au sein de la muqueuse intestinale (Benoit et al., 2012).

2.5 Les protéines PcG et leur implication dans les cancers

2.5.1 Rôle général des protéines PcG dans les cancers

Les protéines PcG sont dérégulées dans un grand nombre de cancers et de multiples études visent à comprendre leur rôle exact afin d'envisager de nouvelles thérapeutiques. Mais les mécanismes mis en jeu sont multiples et complexes.

a) Implication des protéines du complexe PRC2 dans les cancers

Au sein du complexe PRC2, EZH2 est une protéine fréquemment dérégulée et un paragraphe sera exclusivement consacré à son rôle dans les cancers. En ce qui concerne les autres protéines du complexe, des mutations et délétions au niveau des loci des gènes *EED*, *SUZ12* et *JARID2* sont retrouvées dans les syndromes myélodysplasiques (MDS pour myelodysplastic syndrome) et les leucémies (Ntziachristos et al., 2012; Puda et al., 2012; Score et al., 2012; Ueda et al., 2012; Zhang et al., 2012).

Le rôle de l'activité de PRC2 varie en fonction de la pathologie concernée. En effet, si l'activité de PRC2 est requise lors du développement des leucémies myéloïdes aiguës de type MLL-AF9 (Kim et al., 2013b; Neff et al., 2012; Shi et al., 2013), la perte de son activité contribue au développement des MDS et des leucémies induites par des mutations de la protéine ASXL1 (Abdel-Wahab et al., 2012).

b) Implication des protéines du complexe PRC1 dans les cancers

Le rôle de PRC1 dans les cancers est tout aussi important et complexe que celui de PRC2. Par exemple, la protéine Bmi1 est essentielle dans les leucémies induites par les protéines de fusion AML1-ETO et PLZF-RAR α mais n'intervient pas dans les leucémies MLL-AF9 (Boukarabila et al., 2009; Smith et al., 2011). Cela s'explique par le fait que la protéine de fusion MLL-AF9 active l'expression de *Hoxa7* et *Hoxa9* qui maintiennent la répression du gène suppresseur de tumeur *Ink4a-Arf* en l'absence d'activité de Bmi1 (Smith et al., 2011).

Des études ont ensuite montré que la protéine CBX8 est également impliquée dans le développement des leucémies MLL-AF9 (Tan et al., 2011). Pour autant, il a été montré que la surexpression de Cbx8 dans les CSH et les cellules progénitrices induit l'épuisement du stock de cellules et leur différenciation. A la différence de Cbx8, Cbx7 est préférentiellement exprimé dans les HSC et sa surexpression conduit à une augmentation de l'auto-renouvellement des cellules qui présentent une morphologie de cellules immatures de type blastique pour aboutir enfin au développement de leucémies (Klauke et al., 2013).

Cbx7 est un exemple de protéine PcG pouvant présenter des caractéristiques d'oncogène ou de suppresseur de tumeur en fonction du type de cancer dans lequel elle est impliquée.

Dans les lymphomes et le cancer de la prostate, Cbx7 se conduit comme un oncogène en réprimant le locus *Ink4a/Arf* (Bernard et al., 2005; Scott et al., 2007) tandis que Cbx7 est inactivé dans les glioblastomes (Gargiulo et al., 2013). Une autre étude a également montré le rôle suppresseur de tumeur de Cbx7 puisque le knock-out de la protéine chez la souris adulte induit des adénomes et carcinomes du foie et des poumons (Forzati et al., 2012).

2.6 Cas de la protéine EZH2

EZH2 est une protéine dont le rôle est central au sein du complexe PRC2. En effet, elle est responsable de l'apposition de la marque répressive H3K27me3 et porte seule l'activité catalytique du complexe. EZH2 est donc, comme toute protéine ayant un rôle régulateur majeur, fréquemment trouvée dérégulée dans les cancers.

2.6.1 Rôles d'EZH2 dans les hémopathies et les cancers solides

De nombreuses anomalies contribuent à la dérégulation d'EZH2 dans un large nombre de cancers : mutations, délétions, sur ou sous-expression... de cette protéine en font un suppresseur de tumeur ou un oncogène en fonction du type de maladie considérée.

a) Anomalies d'EZH2 dans les hémopathies : EZH2, oncogène ou suppresseur de tumeur ?

Le rôle oncogénique d'EZH2 a été mis en évidence dans plusieurs études qui rapportent qu'une surexpression d'EZH2 augmente le potentiel prolifératif et l'auto-renouvellement des cellules dans le contexte de maladies lymphoprolifératives T et B (Raaphorst et al., 2000; Sasaki et al., 2011; Visser et al., 2001). De plus, des mutations touchant la tyrosine 641 du domaine catalytique d'EZH2 (domaine SET) ont été décrites dans 7,2% des cas de lymphomes folliculaires et dans 21,7% des cas de lymphomes diffus à grandes cellules B (Morin et al., 2010). A l'origine, ces mutations étaient considérées comme étant des mutations perte de fonction. Des études supplémentaires ont finalement montré que ces mutations étaient des mutations gain de fonction ayant pour effet une accumulation de l'H3K27me3 (Sneeringer et al., 2010).

Par ailleurs, il a été montré que la surexpression d'EZH2 entraîne, chez des patients atteints de syndromes myéloprolifératifs et myélodysplasiques (MPN/MDS), une hyperméthylation de certains gènes suppresseurs de tumeurs. Cette surexpression corrèle avec un mauvais pronostic (Xu et al., 2011).

De façon surprenante, d'autres études suggèrent que la protéine EZH2 peut se comporter en suppresseur de tumeur. En effet, la perte de la marque H3K27me3, qu'elle soit directe via des mutations inactivatrices d'EZH2 ou indirecte via des mutations d'ASXL1 (Abdel-Wahab et al., 2012), peut également contribuer au développement de certaines hémopathies.

Une étude menée chez 614 patients atteints de pathologies myéloïdes a mis en évidence 49 mutations inactivatrices d'EZH2 (Ernst et al., 2010). Les mutations les plus fréquentes se retrouvent chez des individus atteints de MPN/MDS (12%) ou de myélofibrose (13%). Les patients mutés pour EZH2 atteints de MPN/MDS ont une survie globale et une survie sans progression plus faible que les patients non mutés.

Le gène *EZH2* a également été séquencé chez 126 patients atteints de MDS, révélant des mutations inactivatrices somatiques « frameshift », non sens et faux sens tout au long du gène (Nikoloski et al., 2010).

Enfin, il a été montré récemment que la perte d'activité d'EZH2 participe au développement des MDS induites par des mutations de la protéine RUNX bien qu'elle prévienne la transformation du syndrome myélodysplasique en AML (Sashida et al., 2014).

L'ensemble de ces données non exhaustives illustre clairement le rôle dual que peuvent avoir les protéines PcG telle qu'EZH2 qui se comporte comme oncogène ou suppresseur de tumeur en fonction du contexte cellulaire.

b) Anomalies d'EZH2 dans les cancers solides : EZH2, une protéine à double facette
La protéine EZH2 est dérégulée dans de nombreux cancers dits « solides ». Elle est fréquemment surexprimée dans les cancers de la prostate, du sein, des ovaires, du foie, du cerveau, des poumons... (Alford et al., 2012; Bachmann et al., 2006; Behrens et al., 2013; Crea et al., 2010; Kondo et al., 2007; Rao et al., 2010). Dans la plupart des cas, l'expression d'EZH2 est corrélée à une prolifération importante des cellules cancéreuses et donc à un mauvais pronostic. Par exemple, la surexpression *in vivo* d'EZH2 dans les cellules épithéliales mammaires entraîne une hyperplasie de l'épithélium (Li et al., 2009).

De la même façon que dans les hémopathies, EZH2 présente le double rôle d'oncogène et de suppresseur de tumeur dans les cancers solides. Ainsi, la délétion conditionnelle d'*EZH2* chez la souris dans l'épithélium pancréatique entraîne une accélération de l'effet néoplasique de K-Ras (Mallen-St. Clair et al., 2012).

Ces résultats soulignent donc le double rôle d'EZH2 qui fonctionne majoritairement comme un oncogène dans les tumeurs solides mais qui peut exercer également des fonctions de suppresseur de tumeur.

2.6.2 Nouvelles fonctions récemment identifiées d'EZH2

a) EZH2 : un activateur transcriptionnel dans les cancers

Bien que la protéine EZH2 ait initialement été décrite comme ayant une fonction de répresseur transcriptionnel via la triméthylation de l'H3K27, un nombre croissant d'études démontre depuis quelques années qu'EZH2 a des fonctions indépendantes de son activité catalytique sur les histones. Plusieurs études rapportent même un rôle activateur de la transcription, notamment dans les cancers du sein et de la prostate.

Il a été montré qu'EZH2 interagit avec le récepteur aux œstrogènes α et avec la β -caténine. Ce complexe régule l'expression de c-Myc et de la cycline D1 dans les cellules cancéreuses mammaires. De plus, EZH2 agit en tant que plateforme de recrutement pour le complexe Mediator, impliqué dans la régulation de la transcription par la RNA-polymérase II (Shi et al., 2007). L'interaction entre EZH2 et la β -caténine a également été observée dans un modèle de souris transgéniques, dans lesquelles EZH2 a été surexprimée de façon spécifique dans la glande mammaire. Cette interaction mène à l'hyperplasie des cellules épithéliales mammaires (Li et al., 2009). Une autre étude a permis de mettre en évidence la différence de fonctionnement d'EZH2 dans les cellules de cancers du sein de type basal, positives ou non pour le récepteur aux œstrogènes (ER). Dans les cellules ER-négatives, EZH2 interagit avec le complexe RelA/RelB, ce qui a pour conséquence d'activer les gènes cibles du facteur nucléaire NF- κ B qui, à leur tour, activent de façon constitutive ce facteur, créant ainsi une boucle de rétrocontrôle positive. Au contraire, EZH2 interagit avec le récepteur aux œstrogènes, dans les cellules ER-positives, et réprime ainsi l'expression des gènes cibles de NF- κ B en induisant la triméthylation de l'H3K27 au niveau de leur promoteur. Dans ce contexte, c'est de nouveau l'activité méthyltransférase d'EZH2 qui est mise en jeu (Lee et al., 2011).

EZH2 fonctionne également comme un activateur dans les cancers de la prostate résistants à la castration. En effet, EZH2 agit comme un co-activateur de certains facteurs de transcription dont fait partie le récepteur aux androgènes (AR). A noter que c'est la forme phosphorylée d'EZH2 qui est impliquée dans ces fonctions co-activatrices et qui est responsable du changement fonctionnel d'EZH2 (Xu et al., 2012). Cette étude montre également, de façon très intéressante qu'au niveau génomique, certains gènes cibles d'EZH2 sont associés à la marque activatrice H3K4me3 et à la présence de la RNA polymérase II.

b) EZH2 : une enzyme qui ne modifie pas que les histones

EZH2 est une enzyme capable de méthyliser des substrats autres que les histones. En effet, EZH2 peut méthyliser des protéines et réguler ainsi leur activité. Par exemple, EZH2 interagit avec et méthyle STAT3, ce qui augmente la phosphorylation de cette protéine ainsi que son activité dans les glioblastomes multiformes (Kim et al., 2013a). Il a été montré par ailleurs qu'EZH2 mono-méthyle le récepteur nucléaire et suppresseur de tumeur ROR α (retinoic acid-related orphan nuclear receptor α). La forme mono-méthylée de ROR α est reconnue par un complexe contenant une ubiquitine ligase et entraîne sa dégradation (Lee et al., 2012).

Enfin, récemment il a été montré qu'EZH2 a pour substrat Jarid2, qui lui-même régule l'activité catalytique de PRC2 (Sanulli et al., 2015).

2.7 Conclusion

Tous ces éléments placent les protéines Polycomb au centre des recherches actuelles menées en épigénétique. Des efforts sont fait afin d'élucider les mécanismes pouvant expliquer leur(s) action(s) qui sont de plus en plus nombreuses et variées et une partie des défis actuels se focalise sur les tests de nouvelles molécules, le plus souvent inhibitrices, qui bloqueraient les effets oncogènes des Polycomb et notamment de la protéine EZH2.

III. La protéine PLZF : Promyelocytic Leukemia Zinc Finger

3.1 Caractéristiques de PLZF

La protéine PLZF (pour Promyelocytic Leukemia Zinc Finger) est un facteur de transcription décrit comme répresseur, également appelée Zbtb16 (Zinc Finger and BTB domain containing 16) ou Zfp145 (Zinc finger protein 145) ; elle a été identifiée pour la première fois chez un patient atteint de leucémie aiguë promyélocytaire. Chez ce patient, PLZF était fusionné à la protéine RAR α (récepteur de l'acide rétinoïque alpha, **R**etinoic **A**cid **R**eceptor alpha) suite à une translocation chromosomique entre les chromosomes 11 et 17 (t(11;17)(q23;21)) (Chen et al., 1993).

Chez l'Homme, le gène *zfp145* codant pour PLZF se situe sur le chromosome 11q23 au sein d'un groupe de gènes codant tous pour des protéines en doigts de zinc.

Ce gène code chez l'Homme pour une protéine de 673 acides aminés. La partie N-terminale de la protéine contient le domaine BTB/POZ pour **B**ric à **B**rac, **T**ramtrack and **B**road-complex/ **P**ox virus and **Z**inc finger (Zollman et al., 1994) qui est conservé chez les mammifères et joue un rôle important dans la répression transcriptionnelle, la dimérisation de

la protéine et les interactions protéines-protéines (Bardwell and Treisman, 1994; Yoshida et al., 1999).

Ensuite vient le domaine RD2 (**R**epression **D**omain **2**) beaucoup moins bien caractérisé que le domaine BTB/POZ. Des mutations dans ce domaine modifient l'activité transcriptionnelle de PLZF (Kang et al., 2003). Enfin, la partie C-terminale de la protéine contient neuf doigts de zinc de type Krüppel C2H2 et permet la liaison spécifique de PLZF à l'ADN au niveau de ses gènes cibles (Li et al., 1997) (fig. 5). La région promotrice de PLZF commence environ 6kb en amont du codon start chez l'Homme.

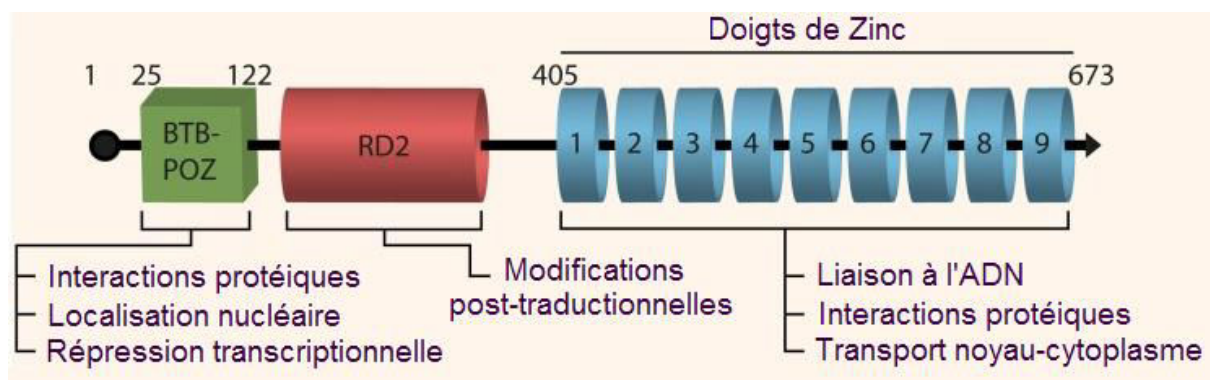


Figure 5. Structure de PLZF. Les différents domaines de PLZF ainsi que leurs principales fonctions sont représentés. Figure adaptée de la revue (Suliman et al., 2012).

3.2 Rôle général et expression de PLZF

Le rôle de PLZF dans le blocage de la différenciation des promyélocytes dans le cas des APL a ouvert la voie à des études visant à déterminer le rôle de cette protéine dans de nombreux processus cellulaires régulant la prolifération et la différenciation cellulaire.

Si les premières études sur PLZF confinaient son expression aux cellules souches et aux cellules progénitrices précoces (Shaknovich et al., 1998), il a ensuite été démontré que PLZF était également exprimé dans le système nerveux central, dans les cellules hématopoïétiques de la moelle osseuse, dans un grand nombre de cellules telles que les cellules glandulaires de la vésicule biliaire, les îlots de Langerhans dans le pancréas, les cellules épithéliales respiratoires, les myocytes cardiaques et du muscle squelettiques, ... (Uhlén et al., 2005).

3.3 Modèles animaux de PLZF

Un modèle de souris dans lequel le gène zfp145 est inactivé a été établi dans les années 2000.

L'exon 2 de *Plzf* a été remplacé par le gène de résistance à la néomycine. Les souris *PLZF*^{-/-} présentent des défauts musculo-squelettiques majeurs au niveau de leurs membres avec une transformation homéotique de structures antérieures en structures postérieures, des déformations du cartilage et du squelette ainsi que des altérations dans la formation des doigts causées par l'inhibition de l'apoptose dans les cellules progénitrices précoces des pattes arrières dans la région de l'autopode, probablement dus à une dérégulation des gènes *Hox* et *Bmp* (Barna et al., 2000).

De plus il a été montré que PLZF réprime l'expression des gènes *HoxD* dans les pattes postérieures des souris via le recrutement d'histones désacétylases (HDACs). Par ailleurs, l'action répressive de PLZF passe par une action sur la chromatine via le recrutement de la protéine PcG Bmi-1 au niveau du locus des gènes *Hox* (Barna et al., 2002).

Une mutation naturelle, nommée « luxoid » (*lu*), est apparue en 1950 chez des souris dont le phénotype était caractérisé par des anomalies des membres et une infertilité des mâles homozygotes. Quelques années plus tard, il a été montré que la protéine PLZF était mutée (mutation non-sens générant une protéine tronquée) dans ce modèle murin, mettant en lumière le rôle de cette protéine dans le maintien de l'état indifférencié des cellules spermatogoniales (Buaas et al., 2004).

Les souris *PLZF*^{-/-} sont également plus sensibles aux infections virales. Cela est dû au fait que dans ces souris, PLZF n'est plus en mesure d'induire l'expression de gènes de réponse à l'interféron. Cette étude met d'ailleurs en lumière la capacité de PLZF à agir comme un activateur de la transcription et non plus comme un répresseur (Xu et al., 2009).

3.4 Fonctions biologiques de PLZF (fig. 6)

3.4.1 Auto-renouvellement versus différenciation

PLZF joue un rôle majeur dans les phénomènes de renouvellement et de maintien des cellules souches et des progéniteurs précoces. L'étude des progéniteurs hématopoïétiques montre que l'expression de PLZF est très importante dans les cellules indifférenciées et diminue au cours de la différenciation (Reid et al., 1995).

De la même façon, dans les modèles murins, l'inactivation de PLZF au niveau des cellules spermatogoniales mène à une sortie de ces cellules de leur état de quiescence. Les cellules se différencient et le pool de cellules indifférenciées n'étant plus maintenu, le tissu testiculaire dégénère (Costoya et al., 2004).

PLZF est également impliqué dans le maintien des cellules souches adultes via la modulation de la protéine Sall4 (Sal-Like protein 4). Sall4 est un facteur de transcription

impliqué dans la pluripotence des cellules souches embryonnaires et dans le développement embryonnaire précoce. Il a été montré que l'interaction de Sall4 avec PLZF régule la balance auto-renouvellement/différenciation. En effet, dans les cellules progénitrices spermatogoniales en cours de différenciation, le niveau de Sall4 augmente, et l'interaction entre Sall4 et Plzf entraîne la séquestration de PLZF dans des zones du noyau dans lesquelles il ne peut plus réprimer la transcription de certains gènes nécessaires à la différenciation cellulaire, tel que Kit. A contrario, lors de l'auto-renouvellement des cellules progénitrices spermatogoniales, c'est Plzf qui séquestre Sall4, l'empêchant de réprimer Sall1, gène largement exprimé dans les cellules progénitrices spermatogoniales ne se différenciant pas (Hobbs et al., 2012).

3.4.2 Cycle cellulaire et apoptose

Un des premiers gènes cibles de PLZF identifié est celui de la cycline A2. Les cyclines sont des protéines qui régulent la fonction de kinases particulières appelées CDKs, pour **Cyclin-Dependant Kinases**. Les CDKs contrôlent la coordination du cycle cellulaire en modulant la fonction de nombreux enzymes et facteurs de transcription requis à chaque étape de la mitose. La Cycline A2 se lie à la kinase CDK2 et l'active, favorisant ainsi les transitions entre les phases G1/S et G2/M du cycle cellulaire. Il a été montré que PLZF est un régulateur négatif du cycle cellulaire. En effet, PLZF se lie au promoteur de la cycline A2 et réprime son expression, bloquant les cellules en phase S du cycle cellulaire (Yeyati et al., 1999).

PLZF joue également un rôle dans la prolifération cellulaire et l'apoptose. En effet, dans les cellules hématopoïétiques Jurkat, il a été montré que PLZF ralentit la prolifération des cellules lorsque la densité cellulaire est peu élevée. De plus, l'induction de l'expression de PLZF dans ces cellules entraîne une diminution d'expression de la Telomerase Reverse Transcriptase (TERT) qui est un gène anti-apoptotique et une augmentation d'expression de gènes pro-apoptotiques tels que TP53INP1, ID2, ID3... (Bernardo et al., 2007).

Une autre étude montre que dans des cellules HeLa, la surexpression de PLZF entraîne un arrêt de la croissance cellulaire dû à des phénomènes d'apoptose plus fréquents. Dans ces cellules, une augmentation de l'activité enzymatique de la caspase 3 est observée, ce qui reflète le clivage de la poly(ADP-ribose) polymérase (PARP), mécanisme impliqué spécifiquement dans l'apoptose (Rho et al., 2007).

3.4.3 Remodelage de la chromatine (via les HDAC)

Les histones désacétylases (HDAC pour Histone Deacetylases) sont des enzymes catalysant le retrait du groupement acétyl sur la queue N-terminale des histones ; groupement dont l'apposition est catalysé par les histones acétyl-transférases (HAT). La désacétylation des histones entraîne la compaction de l'ADN, ce qui a un effet répresseur sur la transcription des gènes.

Il a été montré que la protéine de fusion PLZF-RAR α recrute des co-répresseurs tels que N-Cor (Nuclear co-repressor), l'histone désacétylase 1 (HDAC1) et le complexe répresseur SMRT (Grignani et al., 1998; Hong et al., 1997). Ce qui laisse penser que la répression transcriptionnelle exercée par PLZF passe par l'interaction avec les HDACs qui permettent de contrôler l'accessibilité de PLZF au niveau de ces gènes cibles.

PLZF peut également interagir avec l'HDAC4 dans des cellules normales et leucémiques. Sous l'effet d'une diminution de l'expression ou d'un inhibiteur d'HDAC, la répression des gènes exercée par PLZF est interrompue (Chauchereau et al., 2004). De plus, des expériences de GST pull-down montrent que PLZF peut également interagir directement avec les extrémités N-terminales des HDAC4, 5 et 7 (Lemerrier et al., 2002).

Plus récemment, il a été montré que PLZF pouvait interagir avec l'HDAC3 pour modérer la voie NF- κ B (Sadler et al., 2015).

3.4.4 PLZF et immunité

Les interférons (IFNs) sont des cytokines produites par les cellules après infection ou transformation oncogénique. Ils jouent un rôle crucial aussi bien dans l'immunité innée qu'acquise. L'implication de PLZF dans la réponse immune a été largement étudiée.

De façon générale, PLZF est impliqué dans la réponse à l'IFN qui induit son association à HDAC1 (Xu et al., 2009). Le rôle de PLZF dans la réponse immune, notamment dans les différentes populations de cellules de type NK pour « Natural Killer ».

Les cellules NK sont des cellules de l'immunité innée (« Innate Lymphoid Cells », ILC) et des lymphocytes cytotoxiques ayant un rôle pivot au sein du système immunitaire inné (Walker et al., 2013). En effet, les cellules NK peuvent lyser spontanément des cellules anormales ne présentant pas le complexe majeur d'histocompatibilité de type I (CMH-I).

Les cellules NK issues de souris *PLZF*^{-/-} présentent des fonctions biologiques altérées, comme le montre leur incapacité à lyser les cellules après stimulation (Xu et al., 2009).

PLZF est également un facteur de transcription nécessaire à l'activation et à la fonction des cellules NKT (Natural Killer T) (Savage et al., 2008). Une étude récente montre que PLZF

régule la différenciation terminale des NKT et leur fonction effectrice via son recrutement par le micro-ARN Let-7 dans le thymus (Pobezinsky et al., 2015).

3.4.5 PLZF et hématopoïèse

Après son identification en 1993, PLZF a été décrit comme fortement exprimé dans les cellules progénitrices CD34⁺ issues de moelle osseuse humaine. Le taux d'expression de PLZF diminuant dans ces cellules hématopoïétiques au cours de leur différenciation (Reid et al., 1995), son implication en tant que facteur de transcription contrôlant l'hématopoïèse a très vite été supposée.

Evi-1 est une protéine qui active le promoteur de PLZF dans les étapes les plus précoces du développement hématopoïétique (Takahashi and Licht, 2002).

PLZF assure la régulation de la différenciation des progéniteurs myéloïdes via la modulation de plusieurs facteurs de transcription tels que *GFI-1*, *C/EBPα* et *LEF-1* (Doulatov et al., 2009).

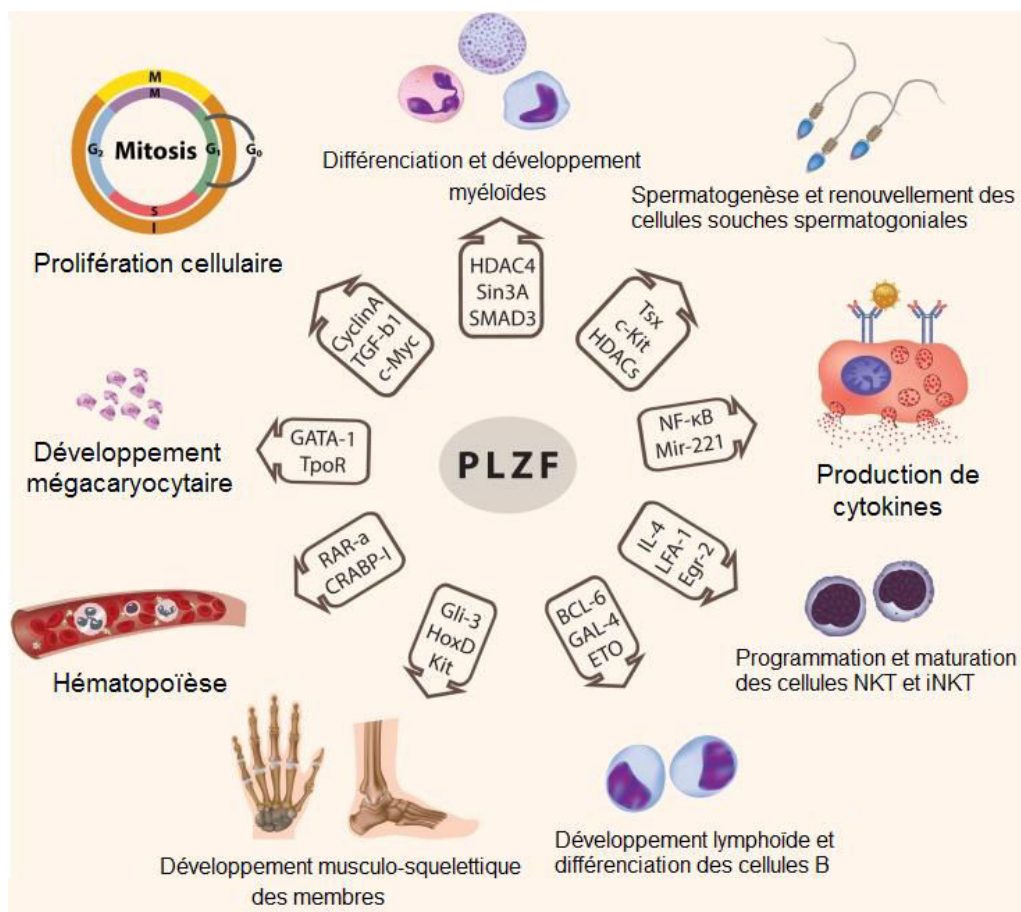


Figure 6. Récapitulatif des principales fonctions biologiques de PLZF et des principaux gènes cibles impliqués. Figure adaptée de la revue (Suliman et al., 2012).

3.5 Rôle activateur de PLZF

Si le rôle largement dominant de PLZF est de réprimer la transcription, des études montrent que PLZF peut parfois activer ses gènes cibles. Ainsi, Melnick *et al.* décrivent une mutation au niveau du domaine BTB/POZ de PLZF qui devient alors activateur (Melnick et al., 2000).

Plus tard, une autre étude montrera que PLZF peut à la fois réprimer des facteurs de transcription impliqués dans la différenciation myéloïde et activer des régulateurs négatifs de cette différenciation au cours de la myélopoïèse (Doulatov et al., 2009).

Enfin, il a été montré que PLZF active un groupe de gènes en réponse à une stimulation à l'Interféron (IFN) (Xu et al., 2009).

3.6 Modifications post-traductionnelles de PLZF

3.6.1 Acétylation

L'acétylation est une modification réversible sur les lysines des protéines dont l'effet porte principalement sur les fonctions et les interactions de la protéine modifiée.

Certaines lysines du doigt de zinc 9 de PLZF sont acétylées par la protéine p300. Cette acétylation augmente la capacité de PLZF à se lier à l'ADN (Guidez et al., 2005).

Plus récemment, une étude a mis en évidence que l'acétylation de PLZF par l'acétyltransférase HAT-1 permet de réguler la réponse immunitaire en favorisant l'association de PLZF à des co-répresseurs modérant la voie NF- κ B (Sadler et al., 2015). Parallèlement, une étude a démontré que la désacétylation de PLZF par les désacétylases HDAC3 et SIRT1 induit une diminution de fixation de PLZF à l'ADN et une dé-répression de ses gènes cibles (McConnell et al., 2015).

3.6.2 Phosphorylation

La phosphorylation consiste en l'ajout d'un groupement phosphate sur les résidus sérine ou thréonine des protéines. PLZF est phosphorylé par la kinase cdc2/Cyclin-B1, ce qui augmente sa capacité de liaison à l'ADN puisqu'un traitement à la phosphatase perturbe l'association de PLZF à l'ADN (Ball et al., 1999). A contrario, la phosphorylation de PLZF par la protéine CDK2 entraîne son ubiquitination et par conséquent sa dégradation, ce réduit donc les effets inhibiteurs de PLZF sur la croissance cellulaire (Costoya et al., 2008).

PLZF peut également être phosphorylé sur la sérine 76 *in vitro* en réponse à une stimulation à l'interféron. L'activité transcriptionnelle de PLZF est alors augmentée, mais dans ce cas précis, PLZF agit comme un activateur de la transcription (Xu et al., 2009).

3.6.3 Sumoylation

Les protéines SUMO (Small Ubiquitin-like **MO**difier) sont des petites protéines qui se lient de façon réversible à leurs protéines cibles afin de réguler leurs fonctions.

Il a été montré *in vivo* que SUMO1 se lie à la lysine 242 de PLZF, au niveau du domaine RD2, entraînant une meilleure liaison de PLZF à l'ADN et une activité transcriptionnelle répressive plus efficace (Kang et al., 2003).

Il a également été montré que la forme sumoylée de PLZF interagit avec la protéine LYRIC/AEG1 qui séquestre ainsi PLZF dans les corps nucléaires et empêche PLZF de se lier à l'ADN levant son effet sur la répression transcriptionnelle (Thirkettle et al., 2009).

3.6.4 Ubiquitination

L'ubiquitine est une petite protéine trouvée uniquement chez les eucaryotes dont la fonction principale est de marquer les protéines à dégrader, les orientant vers le protéasome.

La protéine BTBD6 se lie à PLZF et sert d'adaptateur à la protéine CUL3, une E3 ubiquitine ligase. PLZF est donc envoyée vers le protéasome afin d'y être dégradée, phénomène essentiel pour les progéniteurs qui peuvent ainsi se différencier en neurones (Sobieszczuk et al., 2010).

Il existe une relation antagoniste entre sumoylation et ubiquitination de la lysine 242 du domaine RD2 de PLZF. En fonction du niveau d'espèces réactives de l'oxygène (ROS pour Reactive Oxygen Species) dans les cellules, PLZF sera différemment modifié. En effet au plus sera élevé le niveau de ROS, au moins il y aura de PLZF sumoylée (enzymes de sumoylation inhibées) et au plus il y aura de PLZF ubiquitiné (Kang et al., 2008).

3.7 *PLZF et cancers*

3.7.1 Translocation PLZF-RAR α

Les leucémies aiguës promyélocyaires sont classiquement caractérisées par la mise en évidence de la translocation t(15;17), correspondant à la présence d'un ARN de fusion entre les gènes PML (Pro-Myelocytic Leukemia) et RAR α (récepteur α à l'acide rétinoïque) (Warrell et al., 1993). Un sous-groupe de rares cas de leucémies aiguës promyélocyaires ayant une translocation impliquant également RAR α sur le chromosome 17 mais associé à un autre partenaire (PLZF, sur le chromosome 11) a été identifié (Chen et al., 1993). Chez ces patients, la translocation chromosomique a donné lieu aux protéines de fusion PLZF-RAR α et RAR α -PLZF qui, bien que rares, se sont révélées précieuses comme modèle d'étude, permettant d'identifier PLZF et de comprendre la fonction de la protéine.

Ces protéines de fusion ont causé la progression de la leucémie en maintenant une répression forte sur des gènes impliqués dans la différenciation cellulaire via le recrutement de répresseurs tels que N-CoR, HDAC, mSin3 et SMRT (Chen et al., 1993; Grignani et al., 1998).

Il a également été montré que la protéine PLZF-RAR α interagit avec Bmi1. Cette interaction conduit au recrutement du complexe PRC1 au niveau des éléments de réponse à l'acide rétinoïque (RAREs), indépendamment de la sensibilité à l'ATRA (all-trans retinoic acid). Ce qui a pour résultat de provoquer la transformation des cellules (Boukarabila et al., 2009).

3.7.2 Implication de PLZF dans les cancers (forme non fusionnée)

La protéine PLZF est également impliquée dans de nombreux cancers sans pour autant être impliquée dans une translocation. Par exemple, dans les cancers de la prostate, il a été montré que la lignée DU145 (lignée androgène-indépendante) n'exprime pas PLZF. Expression qui est rétablie consécutivement à l'expression ectopique du récepteur aux androgènes. La surexpression de PLZF dans cette lignée entraîne la diminution d'expression des protéines Pbx1 et HoxC8 et conduit à la réduction du potentiel prolifératif des cellules *in vitro*. Ces observations mettent en évidence le rôle de suppresseur de tumeur de PLZF (Kikugawa et al., 2006). Une deuxième étude plus récente vient confirmer le rôle majeur de PLZF dans les cancers de la prostate. En effet, les auteurs démontrent que la diminution d'expression de PLZF entraîne un phénotype résistant à la castration des cellules cancéreuses. Le rétablissement de l'expression de PLZF est suffisant à inverser ce phénotype (Hsieh et al., 2015). Enfin, une étude a montré que la perte de l'expression de PLZF dans les cancers de la prostate était corrélée à l'agressivité des tumeurs et aux métastases et pouvait donc servir de marqueur afin de personnaliser les traitements (Xiao et al., 2015).

L'implication de PLZF dans les cancers n'est toutefois pas restreinte aux seuls cancers de la prostate. Ainsi, l'expression de PLZF est fréquemment perdue dans les cellules de mésothéliomes malins les plus agressives (délétion chromosomique). La restauration de l'expression de PLZF dans ces cellules diminue leur viabilité en culture, réduit la formation de colonies en expérience de « soft agar » et augmente l'apoptose (Cheung et al., 2010)

De la même façon, l'expression de PLZF est diminuée dans les mélanomes et les cancers du poumon non à petites cellules (Brunner et al., 2008; Felicetti et al., 2008; Wang et al., 2013).

3.8 Conclusion

A l'instar d'autres facteurs de transcription, PLZF peut agir soit comme un activateur soit comme un répresseur transcriptionnel. Si les bases moléculaires de cette bi fonctionnalité ne sont pas encore connues, il apparaît clairement que son activité est liée aux phénomènes épigénétiques. Des mutations ou des modifications post-traductionnelles de PLZF peuvent engendrer des variations de son activité transcriptionnelle via des changements de cofacteurs et dont les répercussions se portent sur le paysage chromatinien.

Contexte et Objectifs de la thèse.

I. Contexte général

L'hématopoïèse est un processus biologique finement contrôlé par un grand nombre de mécanismes dont la dérégulation peut conduire au développement d'hémopathies malignes, telles que les leucémies qui résultent de la transformation des cellules hématopoïétiques. On distingue deux grands groupes de leucémies : les leucémies chroniques, dans lesquelles des cellules différenciées prolifèrent de façon anarchique et les leucémies aiguës où, au contraire, les cellules sont bloquées au stade des progéniteurs immatures. Dans ce dernier cas, les cellules blastiques prolifèrent de façon incontrôlée ce qui provoque leur accumulation notamment au niveau de la moelle osseuse et du sang entraînant ainsi une diminution de l'hématopoïèse normale. Nous nous intéresserons ici plus particulièrement au groupe des leucémies aiguës. En fonction des marqueurs spécifiques de la lignée myéloïde ou lymphoïde (T ou B) exprimés par les cellules, les leucémies sont classées respectivement en leucémies aiguës myéloblastiques (LAM) ou leucémies aiguës lymphoblastiques T (LAL-T) ou B (LAL-B).

Au cours de ces dernières années, de nombreuses perturbations de la régulation épigénétique au sein des cellules hématopoïétiques ont été décrites comme étant impliquées dans le développement des leucémies. Ces dérégulations sont de natures diverses. Bien souvent, il s'agit de mutations pertes de fonction ou activatrices mais dans certains cas, un recrutement aberrant de certains facteurs épigénétiques est observé. En fonction du type d'hémopathies, mutations et recrutement aberrant peuvent concerner la même protéine, comme c'est le cas pour la protéine EZH2.

De nombreuses études se sont intéressées au rôle d'EZH2 dans les cancers et ont pu mettre en évidence un effet oncogénique ou suppresseur de tumeur en fonction du type de cancer étudié. En effet, dans les hémopathies malignes, EZH2 peut être la cible de mutations

gain de fonction ou de mutations perte de fonction. Toutefois, dans le cas des LAM, des mutations affectant EZH2 ou des membres PcG ne sont retrouvées que très rarement. Des études ont montré que dans ce type de leucémies, de nombreux gènes cibles d'EZH2 sont dérégulés bien que son activité répressive soit toujours présente, mettant en évidence d'éventuels défauts de recrutement de cette protéine. Les mécanismes qui initient le recrutement des PRC au niveau de leurs gènes cibles chez les vertébrés ne sont pas bien compris. Une des hypothèses actuelles stipule que des facteurs intermédiaires seraient responsables du recrutement spécifique de PRC2, tels que les longs ARN non-codants (Lnc-RNA) et certains facteurs de transcription séquence-spécifiques. Etudier les mécanismes impliqués dans la régulation des protéines PcG et leur adressage à la chromatine représente donc un enjeu majeur dans la compréhension du développement des leucémies.

C'est dans ce contexte que vient se placer mon travail de thèse qui s'articule autour de 3 projets qui ont eu pour objectif une meilleure compréhension des mécanismes de régulation épigénétique mis en jeu lors de l'hématopoïèse normale et leucémique.

II. Projet principal (article 1)

Cette étude constitue l'axe majeur de mon travail de thèse et a eu pour objectif d'étudier la relation entre le facteur de transcription PLZF et la protéine PcG EZH2. En effet, parmi les facteurs de transcription susceptibles de contrôler l'association des protéines PcG à la chromatine, le facteur de transcription PLZF (Promyelocytic Leukemia Zinc Finger) est un candidat intéressant. Plusieurs arguments sont en faveur d'une interaction fonctionnelle entre PLZF et les PcG. Tout d'abord, le laboratoire a montré une interaction entre BMI-1, membre de PRC1, et PLZF (Boukarabila et al., 2009; Spicuglia et al., 2011). Ensuite, pour mieux comprendre l'action concertée de PLZF et des PcG au cours de la différenciation myéloïde normale et leucémique, le laboratoire a identifié les gènes cibles de PLZF par une approche d'immunoprécipitation de la chromatine suivie d'une hybridation sur puce à promoteurs (ChIP-on-chip). Les résultats obtenus indiquent que la distribution génomique de PLZF concorde avec la distribution génomique de certains composants des PcG (résultats non publiés). J'ai donc souhaité caractériser plus précisément l'interaction entre PLZF et EZH2. Pour cela, j'ai, dans un premier temps, étudié l'interaction fonctionnelle entre PLZF et EZH2 à la chromatine dans un modèle cellulaire de LAM (lignée cellulaire myéloïde KG1). Nous avons donc mené des analyses à l'échelle génomique (expériences d'immunoprécipitation de la chromatine suivie de séquençage haut débit ou ChIP-seq). Ces données nous ont permis de

localiser PLZF et EZH2 au niveau du génome et d'étudier les marques épigénétiques associées à ces protéines.

Dans un deuxième temps, nous avons mis à profit une lignée cellulaire dans laquelle l'expression de PLZF pouvait être diminuée afin d'étudier quels étaient les effets de cette diminution d'expression sur EZH2 et les marques histones.

Enfin, nous avons pu isoler un groupe de gènes : les gènes bivalents, présentant un mode de régulation différent. Les gènes bivalents sont des gènes enrichis à la fois pour la marque activatrice H3K4me3 et pour la marque répressive H3K27me3. Ce sont des gènes réprimés mais dont la transcription est facilement activée en cas de besoin particulier de la cellule (Voigt et al., 2013).

L'objectif de cette partie de mon travail a donc été d'étudier la relation fonctionnelle, à la chromatine, de la protéine PcG EZH2 avec le facteur de transcription PLZF, au sein des cellules hématopoïétiques.

III. Projets collaboratifs (articles 2 et 3)

Au cours de ma thèse, j'ai pu participer à deux projets collaboratifs : l'un mené au sein même du laboratoire et portant sur l'étude du rôle de PLZF dans la régulation de la transcription dans les cellules souches hématopoïétiques (CSH), le deuxième dirigé par une équipe du Centre d'Immunologie de Marseille Luminy dont l'objectif était de mieux comprendre les mécanismes d'activation épigénétiques de l'oncogène TAL1 dans les LAL-T.

a) Etude du rôle de PLZF dans la régulation de la transcription dans les CSH

Comme dit plus haut, les mécanismes épigénétiques jouent un rôle prépondérant dans la régulation de la transcription aussi bien lors du développement d'hémopathies qu'au cours de l'hématopoïèse normale. Il est donc nécessaire de comprendre les mécanismes moléculaires par lesquels les modifications épigénétiques contribuent au contrôle de l'expression des gènes, ainsi qu'à la différenciation et à l'auto-renouvellement des CSH.

Au sein du système hématopoïétique, PLZF joue un rôle important dans le développement de nombreuses cellules hématopoïétiques et certaines données suggèrent que PLZF pourrait être un important régulateur des CSH sans qu'aucun lien direct n'ait encore été démontré. L'objectif de ce travail a été de déterminer, via l'étude de ses gènes cibles, le rôle de PLZF dans la régulation de la différenciation des CSH. Pour cela, un modèle murin portant une mutation naturelle pour PLZF (mutant luxoid) qui engendre une protéine tronquée et non

fonctionnelle a été utilisé afin de comparer l'hématopoïèse normale et l'hématopoïèse en absence de la forme fonctionnelle de PLZF.

Des analyses de micro-array ont été réalisées sur des CSH de la moelle osseuse de souris sauvages ou mutantes afin de mettre en évidence les programmes transcriptionnels régulés par PLZF. Cette analyse montre qu'un certain nombre de gènes voit leur expression modifiée dans les cellules mutées par rapport aux cellules sauvages. De façon intéressante, les gènes surexprimés dans les cellules mutantes sont associés à des gènes impliqués dans le cycle et le métabolisme cellulaires, et ces gènes sont dérégulés pour la marque répressive H3K27me3. J'ai donc mis à profit mes données issues de ChIP-seq afin de montrer que la majorité de ces gènes sont des cibles, de PLZF. Ces données suggèrent que PLZF agit sur le cycle cellulaire en modifiant le paysage épigénétique des gènes régulés.

Cette étude a donc mis en évidence un nouveau rôle régulateur de PLZF au sein du compartiment souche hématopoïétique et montre son implication dans la régulation épigénétique des gènes.

b) Etude de l'activation de l'oncogène TAL1 dans les LAL-T

Parmi les divers oncogènes impliqués dans les LAL-T, TAL1 est l'un des plus fréquemment dérégulés. En effet, une expression ectopique de TAL1 (cas TAL1+) est observée dans plus de 40% des cas de LAL-T. Cependant, dans plus de la moitié des cas TAL1+, aucune altération génomique n'est observée, et les mécanismes mis en jeu sont encore à définir.

Dans un premier temps, les marques épigénétiques associées à la région correspondant au gène *TAL1* ont été étudiées dans des thymocytes murins positifs pour les marqueurs CD4 et CD8 et les résultats montrent qu'une large région comprenant le promoteur et une partie du gène *TAL1* est enrichie pour la marque répressive H3K27me3. Ces observations indiquent donc qu'en conditions normales, le gène *TAL1* n'est pas exprimé. Afin de comprendre comment cette répression épigénétique peut être levée dans des conditions pathologiques, une région d'environ 10 kb a été séquencée chez un patient atteint de LAL-T monoallélique pour TAL1 chez lequel une micro-insertion de 12 pb en amont du promoteur de TAL1 et située en bordure de la marque répressive H3K27me3 a été identifiée. L'étude a été étendue et a révélé que 20% des patients monoalléliques TAL1+ possèdent une insertion de taille variable (patients *TALins*).

De façon intéressante, une lignée cellulaire humaine de LAL-T (les cellules Jurkat), présente justement une insertion monoallélique. Nous avons donc utilisé ces cellules afin de

modéliser les cas de patients *TALins*. J'ai réalisé, à la fois chez les patients et sur les cellules Jurkat, des ChIP-alléliques (ChIP-qPCR pouvant discriminer l'allèle muté de l'allèle sauvage) qui nous ont permis de confirmer que la mutation *TALins* est associée à la diminution de la marque répressive H3K27me3 du locus TAL1 et à l'augmentation de la marque activatrice H3K27ac au même niveau. Cette étude a donc mis en évidence un nouveau mécanisme d'activation de l'oncogène TAL1 qui se traduit par une levée de l'inhibition épigénétique de ce gène suite à une insertion en amont de ce locus.

In fine, j'ai été amenée durant ma thèse à étudier les mécanismes épigénétiques qui régulent le système hématopoïétique ou au contraire qui conduisent à sa dérégulation.

Article de Recherche 1 :

**A non-conventional genomic association
between PLZF and EZH2 revealed
through genome-wide analyses**

A non-conventional genomic association between PLZF and EZH2 revealed through genome-wide analyses

Myriam Koubi, Abdessamad El-Kaoutari, Mathilde Poplineau, Guillaume Tiberi, Andrew Saurin and Estelle Duprez

Les protéines du groupe Polycomb (PcG) sont des facteurs épigénétiques qui jouent un rôle important dans le contrôle de la différenciation et de l'auto-renouvellement cellulaires. Les protéines PcG forment deux complexes majeurs appelés PRC1 et PRC2 (pour **P**olycomb **R**epressive **C**omplex) composés de plusieurs sous-unités. Il existe divers modèles décrivant le fonctionnement des PRC. Le modèle classique, dit hiérarchique, est le mieux décrit. Selon ce modèle, l'histone méthyle transférase EZH2 (**E**nhancer of **Z**este **H**omolog **2**), appartenant au complexe PRC2, amorce la répression transcriptionnelle en triméthylant l'histone H3 sur la lysine 27 (H3K27me3). Cette marque permet alors le recrutement du complexe PRC1 et son maintien au niveau du gène cible qui ne sera donc pas transcrit. Ce modèle est cependant discuté et les mécanismes qui initient le recrutement des PRC au niveau de leurs gènes cibles chez les vertébrés ne sont pas bien compris.

De nombreuses études se sont intéressées au rôle des PcG dans les cancers et ont pu mettre en évidence un effet oncogénique ou suppresseur de tumeur en fonction du type de cancer étudié. Cette dualité est mise en évidence par l'exemple du comportement d'EZH2, notamment dans les hémopathies malignes dans lesquelles EZH2 peut être la cible de mutations gain ou perte de fonction. Dans le cadre de ma thèse, je me suis intéressée au modèle des leucémies aiguës myéloïdes (LAM). Contrairement à d'autres pathologies myéloïdes, des mutations affectant EZH2 ou des membres PcG ne sont retrouvées que très rarement dans les LAM (<1%). Plusieurs études ont montré que dans ce type de leucémies, l'activité répressive d'EZH2 était présente mais déplacée sur des cibles suppressives de tumeurs. Des défauts de recrutement de ces protéines pourraient expliquer les dérégulations observées au sein de ces pathologies. Parmi les facteurs susceptibles de contrôler l'association des PcG à la chromatine, le facteur de transcription PLZF (Promyelocytic Leukemia Zinc Finger) est un candidat intéressant. Plusieurs arguments sont en faveur d'une interaction fonctionnelle entre PLZF et les protéines PcG. Tout d'abord, le laboratoire a montré une interaction entre BMI-1, membre de PRC1, et PLZF. De plus, des analyses de ChIP-on-chip sur des cellules hématopoïétiques ont montré que la distribution génomique de PLZF concorde avec celle de certains composants des protéines PcG (résultats non publiés).

Nous avons donc étudié l'interaction fonctionnelle entre PLZF et EZH2 à la chromatine dans un modèle cellulaire de LAM. Des expériences de co-immunoprécipitation

dans la lignée myéloïde KG1 (cellules CD34 positives) ont montré que PLZF et EZH2 pouvaient interagir *in vivo* via le domaine BTB de PLZF (domaine responsable de l'interaction protéine-protéine).

Afin de mieux appréhender l'interaction de PLZF avec EZH2 et de définir leur action sur la structure chromatinienne, la distribution d'EZH2, de PLZF et des marques histones activatrices et répressives H3K4me3 et H3K27me3 a été étudiée par ChIP-seq. Les résultats obtenus montrent que les sites de fixation de PLZF et EZH2 se retrouvent surreprésentés au niveau des régions promotrices et introniques et que ces deux protéines partagent de nombreux gènes cibles, notamment au niveau des régions promotrices. Ces résultats suggèrent donc une co-localisation génomique de PLZF et EZH2.

Une analyse de type K-mean, qui permet de séparer les gènes en fonction de leurs différents profils d'enrichissement, a mis en évidence deux grands groupes de gènes. Sur le premier groupe de gènes, on retrouve comme attendu EZH2 associé à la marque répressive H3K27me3. Ces gènes ne sont pas enrichis pour PLZF. De façon intéressante, un deuxième groupe de gènes, enrichi pour PLZF et EZH2, est associé à la marque activatrice H3K4me3. De plus, seuls les gènes cibles communs de PLZF et d'EZH2 sont associés à la présence de Polymérase II et d'H3K36me3, caractéristiques d'une activation transcriptionnelle.

Afin d'étudier la relation de ces deux facteurs au niveau de la chromatine, une lignée cellulaire KG1 dans laquelle l'expression de PLZF est modulable a ensuite été générée. Dans cette lignée, l'effet de la diminution d'expression de PLZF sur EZH2 et sur les marques histones a été déterminé en quantifiant l'enrichissement pour EZH2 et les marques histones dans les conditions contrôles et shPLZF. De façon surprenante, la diminution d'expression de PLZF n'a aucun effet sur la liaison d'EZH2 à l'ADN mais diminue de façon significative le niveau d'H3K4me3 sur les gènes les plus enrichis en PLZF. De façon intéressante, la proportion des gènes exprimés augmente en fonction de leur enrichissement en PLZF. Une analyse plus fine des sites de fixation de PLZF en fonction de la région génique montre qu'il y a une plus grande proportion de gènes exprimés lorsque PLZF se trouve au niveau du promoteur que lorsque PLZF se situe au niveau d'une région intragénique (hors promoteur). Pour finir, l'étude des gènes bivalents cibles de PLZF a été menée. Sur cette catégorie particulière de gènes, enrichis à la fois en H3K4me3 et H3K27me3, la diminution d'expression de PLZF induit une diminution d'H3K27me3 tandis que l'H3K4me3 ne varie pas.

L'ensemble de ces résultats met en évidence l'interaction fonctionnelle du facteur de transcription PLZF avec la protéine Polycomb EZH2. De façon globale, PLZF est associé

dans les cellules KG1 à des gènes transcriptionnellement actifs mais une analyse plus fine révèle que la position de PLZF sur le gène (promoteur ou gene body) influence le niveau d'expression des gènes. Nos résultats apportent donc de nouveaux éléments dans la compréhension des mécanismes de la régulation épigénétique impliquant PLZF et EZH2.

Avant-propos concernant l'article 1

Les résultats présentés ci-après le sont sous forme d'un article et il nous a semblé nécessaire d'expliquer ici certains points qui n'ont pu être abordés en détails par la suite et les outils utilisés pour l'analyse. Ce travail a été réalisé en collaboration avec le service bioinformatique du laboratoire (plateforme CIBI).

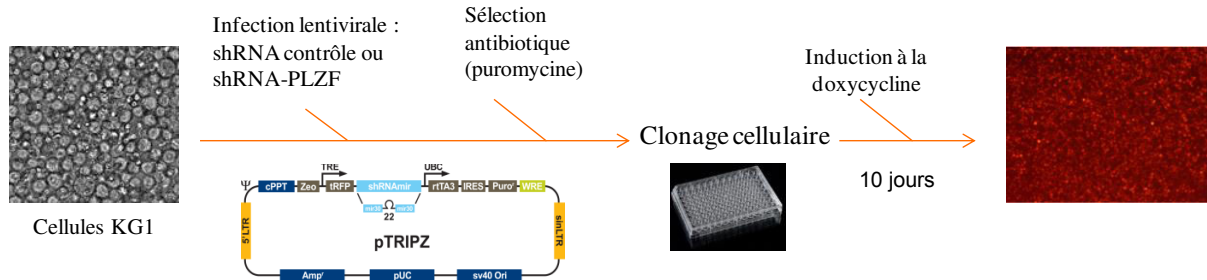
Un des objectifs de mon travail est de mettre en évidence la localisation génomique de PLZF et EZH2. Nous avons pour cela choisi d'utiliser une lignée cellulaire hématopoïétique d'origine myéloïde dans laquelle s'expriment à la fois PLZF et EZH2 : la lignée KG1.

1. Analyse globale de la localisation génomique de PLZF et EZH2 : lignées utilisées

Dans la mesure où nous avons formulé l'hypothèse que PLZF interagit avec EZH2 et afin de mieux caractériser l'éventuelle interaction entre PLZF et EZH2, nous avons voulu travailler avec des cellules ayant une expression modulable de PLZF. Nous avons donc choisi une approche de shRNA capable de diminuer l'expression de PLZF en présence d'un inducteur qui est ici l'analogue de la tétracycline : la doxycycline. Pour cela, j'ai transfecté stablement les cellules afin que la construction contenant le shRNA s'intègre de façon stable dans le génome. Les cellules KG1 étant difficiles à transfecter, les cellules ont été infectées à l'aide de particules lentivirales. Après la phase d'infection, les cellules ayant intégré le plasmide ont été sélectionnées via l'ajout d'un antibiotique, la puromycine (dont la résistance est conférée aux cellules par le plasmide). Ces cellules sélectionnées sont ensuite clonées (clonage cellulaire) afin d'avoir des cellules ayant toutes intégré le plasmide au même endroit dans le génome. J'ai donc dérivé des clones cellulaires, à partir de la lignée KG1, dans lesquels l'expression de PLZF est modulable (shPLZF) ou non (shContrôle). Les clones dans lesquels le shRNA s'exprime après induction, expriment également une protéine fluorescente (fig. 7a). Après avoir établi que l'effet du shRNA était optimal après 7 jours d'induction, les cellules sont récupérées et la diminution d'expression de PLZF est visualisée par western blot (fig. 7b).

Ces lignées ont été exploitées lors de mon travail de thèse lors d'expériences de ChIP-seq dans le but de répondre à la question initiale de la localisation génomique de PLZF et EZH2.

A



B

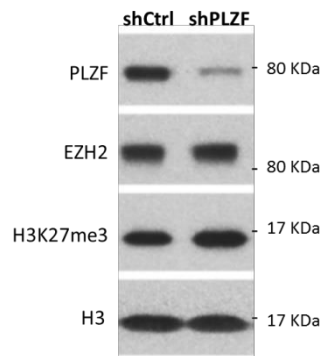


Figure 7. a) Etapes nécessaires à l'obtention des lignées KG1 shContrôle et shPLZF. b) Validation de la diminution d'expression de PLZF dans la lignée KG1-shPLZF.

1. Analyse globale de la localisation génomique de PLZF et EZH2 : expériences d'immunoprécipitation de la chromatine suivie de séquençage haut débit (ChIP-seq)

Afin de localiser précisément sur le génome où se fixent PLZF et EZH2, nous avons réalisé des expériences ChIP-seq. Les ChIPs permettent de déterminer les sites de liaison à l'ADN de protéines spécifiques telles que les facteurs de transcription ou les histones. Brièvement, les cellules sont fixées afin d'établir un pontage entre les protéines et l'ADN. La chromatine est extraite des noyaux et fragmentée (fragments allant de 200 à 500 pb en moyenne). Les complexes protéines/ADN sont ensuite immunoprécipités à l'aide d'un anticorps spécifique dirigé contre PLZF, EZH2 ou dirigé contre une marque histone. L'ADN lié à la protéine d'intérêt est ensuite purifié puis envoyé en séquençage afin d'être identifié (figure 8).

Avant d'envoyer l'ADN à la plateforme de séquençage, j'ai vérifié par ChIP-qPCR la qualité des ChIPs en regardant l'enrichissement d'un ou plusieurs locus contrôles pour la protéine immunoprécipitée.

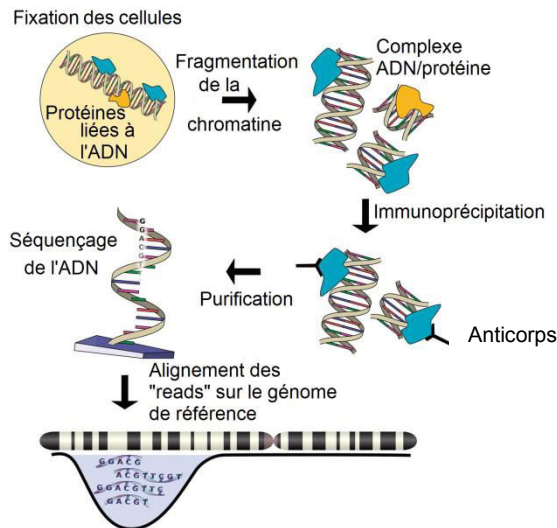


Figure 8. Différentes étapes du ChIP-seq. Les liaisons ADN/protéines sont fixées par le formaldéhyde et après fragmentation de la chromatine, les complexes ADN/protéine d'intérêt sont immunoprécipités à l'aide d'anticorps spécifiques. L'ADN est enfin purifié et envoyé à séquencer pour analyses ultérieures.

D'après (Szalkowski and Schmid, 2011)

La préparation des banques d'ADN et l'étape de séquençage sont réalisées par une plateforme. La préparation des banques d'ADN a pour but d'amplifier puis de purifier l'ADN récupéré lors des ChIPs. Elle se fait à l'aide d'un kit (TruSeq®ChIP Sample Preparation, Illumina) et comprend 5 étapes :

1. Réparation des extrémités des fragments d'ADN
2. Ajout d'une base A à ces extrémités
3. Ligation de deux types d'adaptateurs à ces extrémités
4. Amplification par PCR.
5. Une purification sur gel permettant une sélection de taille (350 pb +/- 50) (figure 9)

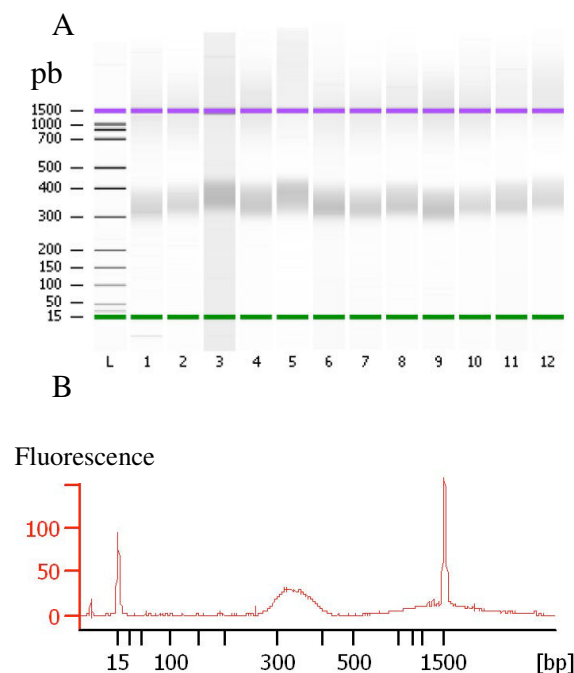


Figure 9. Dosage des banques et contrôles qualités au bioanalyzer. a) Purification des banques après synthèse b) Quantification de l'ADN sur puce Agilent High Sensitivity

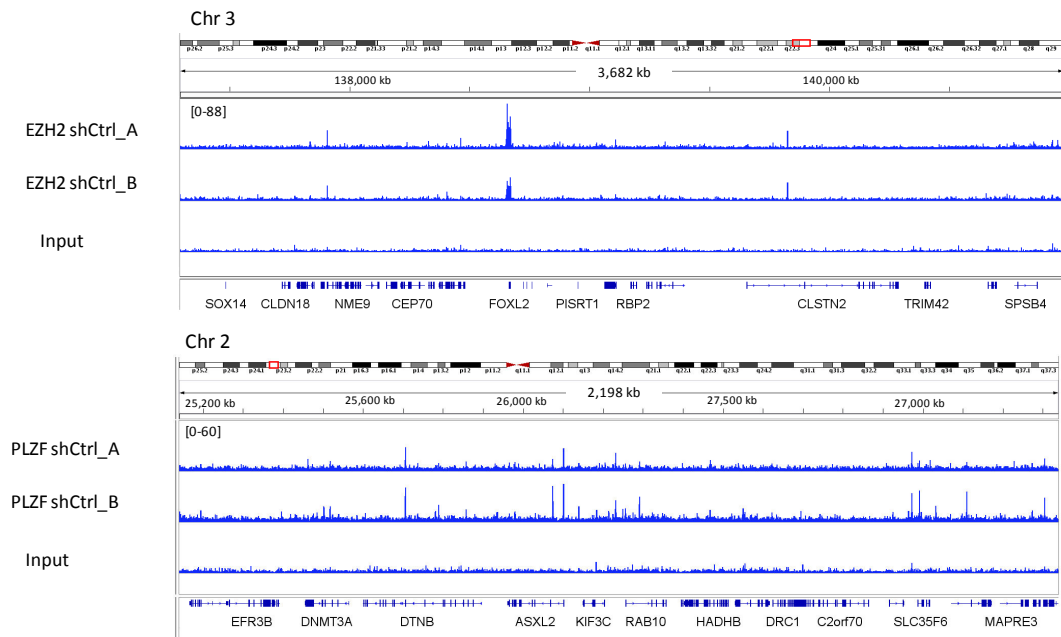
Le séquençage des échantillons a été réalisé en multiplex avec un séquenceur Illumina HiSeq 2000. La qualité de séquençage a été validée après vérification de plusieurs points de contrôle incluant par exemple la recherche de séquences surreprésentées, et le suivi du taux de GC. Nous obtenons entre 20 et 50 millions de « reads » (fragments d'ADN) séquencés par échantillons (cf matériel et méthodes).

Les « reads » obtenus ont ensuite été alignés sur le génome de référence (hg19) à l'aide de l'outil bfast. Le niveau de recouvrement du génome par nos « reads » constitue le signal ou niveau d'enrichissement pour la protéine étudiée. Ainsi, une région génomique possédant un empilement de « reads » traduira un enrichissement de notre cible immunoprécipitée. Plus l'empilement de « reads » est important, plus la région est considérée enrichie pour le facteur de transcription ou la marque histone étudiée.

De façon à rendre la comparaison entre échantillons possible, nous avons procédé à une étape de normalisation du signal d'enrichissement, en ramenant chaque échantillon à un total de 10 millions de « reads ». Après cette normalisation nous avons décidé de regrouper nos réplicats biologiques au vue de l'importante reproductibilité observée (figure 10). Cette méthode nous permet notamment d'augmenter la profondeur de séquençage, et donc la sensibilité de notre expérience.

Il a ensuite fallu identifier les zones de sur-enrichissement de nos protéines sur le génome : c'est l'étape de « peak calling » ou appel de pic. Pour réaliser cette localisation, le signal de l'échantillon issu de l'immunoprécipitation est comparé au signal de l'échantillon input correspondant. L'input correspond à l'ADN total avant immunoprécipitation. Le séquençage de l'input donne donc une indication sur le bruit de fond éventuel. Le signal est considéré comme formant un pic lorsque que son intensité est supérieure à un seuil fixé par rapport au signal adjacent et au signal de l'échantillon input correspondant. La détermination de ce seuil a donc été un paramètre critique. En effet de sa valeur découle le nombre de pics déduits. De plus, il est possible de fixer un taux de pics dits « faux positifs » (FDR pour False Discovery Rate) ainsi qu'une p-value pour les pics identifiés. Plusieurs jeux de seuils, de FDR et de p-values ont été comparés. Leur détermination s'est faite en prenant compte des spécificités techniques du jeu de données et de la cohérence du résultat biologique (vérification expérimentale de la présence/absence de pics au niveau de gènes contrôles). Le peak calling a été réalisé avec la suite d'outils Homer (Hypergeometric Optimization of Motif EnRichment). Une fois l'identification des pics obtenue, il suffit de faire correspondre la localisation du pic avec la position des gènes dans le génome pour en déduire une liste de gènes cibles. Les listes de gènes cibles obtenues seront fournies à part dans un fichier excel.

A



B

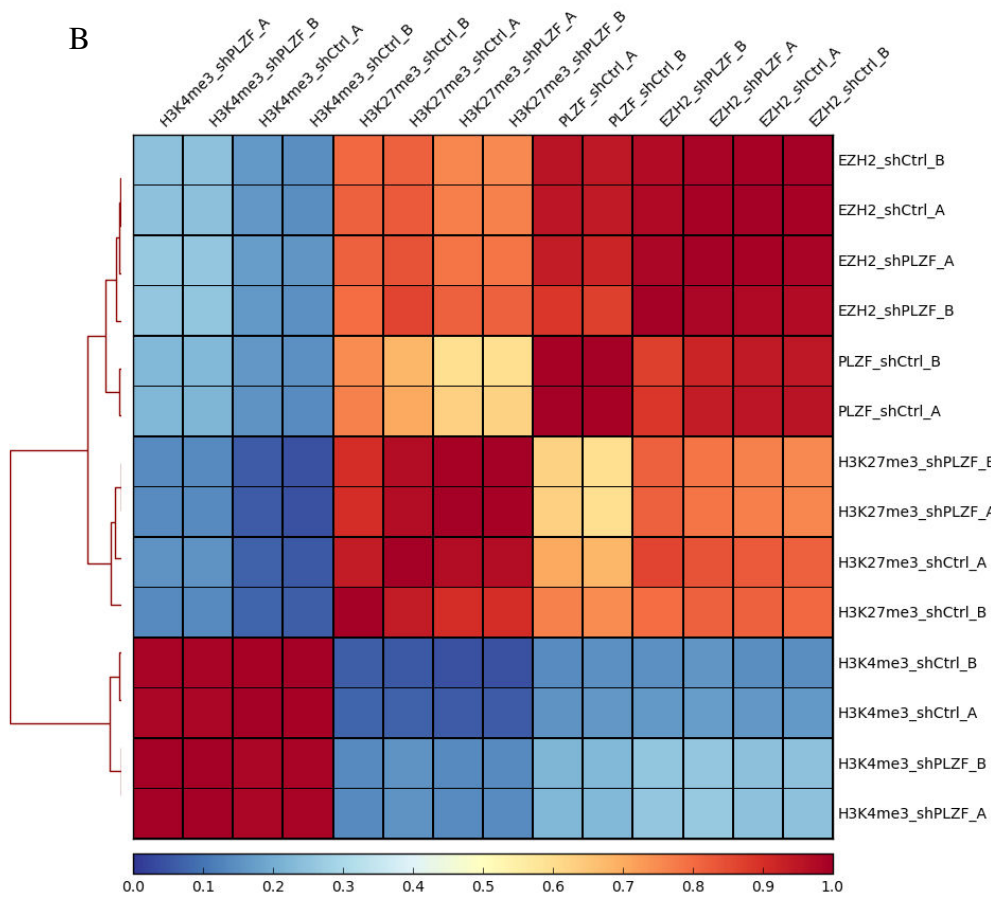


Figure 10. **a)** Visualisation sur le « genome browser » IGV des alignements effectués après séquençage et illustrant la cohérence des réplicats. **b)** Calcul du coefficient de corrélation de Pearson entre les différents réplicats biologiques.

Koubi et al., *in preparation*

**A non-conventional genomic association between PLZF and EZH2 revealed
through genome-wide analyses**

Myriam Koubi¹, Abdessamad El-Kaoutari¹, Mathilde Poplineau¹, Guillaume Tiberi¹, Andrew Saurin² and Estelle Duprez¹

1 CRCM (Epigenetic Factors in Normal and Malignant Hematopoiesis), Inserm, U1068;
Institut Paoli-Calmettes; Aix-Marseille Université, UM105; CNRS, UMR7258, Marseille,
France

2 Aix Marseille Université, CNRS, IBDM, UMR 7288, Marseille 13288, Cedex 09, France

- Correspondance : estelle.duprez@inserm.fr

Introduction

Polycomb group proteins (PcG) have been associated with stable and heritable repression of the Hox gene loci, through either covalently modifying or locally remodelling nucleosome structure. More recently, genome-wide approaches have highlighted the role of the PcG proteins in a more global regulation of the genome and demonstrated the dynamics and flexibility of PcG recruitment during development and differentiation (Boyer et al., 2006) (Bracken et al., 2006) (Lee et al., 2006). Linked to this activity of controlling cell identity, many independent studies have established a major role for PcG proteins in cancer (Sauvageau and Sauvageau, 2010); (Laugesen and Helin, 2014). PcG proteins function as multi-subunit complexes through the concerted participation of at least two major complexes (PRC1/PRC2) allowing the initiation and maintenance of the H3K27me3 repressive epigenetic mark. Importantly, the PRC2 complex bears the histone methyl transferase EZH2 that initiates the repression by trimethylation of lysine 27 of the histone H3 (H3K27me3) (Margueron and Reinberg, 2011).

Depending on the cellular context of the tumour subtype, PcG proteins may serve as oncogenes or tumour suppressors, suggesting an unrevealed tissue-specificity of action (Koppens and van Lohuizen, 2015). This has been highlighted by the dual role of EZH2 in hematopoietic malignancy. The initial discovery of activating mutations in the catalytic EZH2 set domain associated with two types of B-cell lymphomas was consistent with a straightforward role of EZH2 as an oncogene (Morin et al., 2011) (McCabe et al., 2012). This oncogenic activity was modelled in mice where loss of *Ezh2* perturbs the progression of AML by promoting the differentiation of leukemic cells in mice transplanted with MLL-AF9 transformed granulo-monocytic progenitors (Tanaka et al., 2012). Conversely, EZH2 loss-of-function mutations were recently described in myeloproliferative neoplasms and T-lymphoid leukemias (Hock, 2012) (Nikoloski et al., 2010) (Ntziachristos et al., 2012) suggesting that EZH2 acts differently in various disease entities due to specific loci targeting and/or specific oncogenic networks. Can we find a molecular mechanism to explain this duality? In the context of PRC2, EZH2 typically deposits two or three methyl groups on histone H3K27 (H3K27me2/3), and induces a repressed chromatin state (Cao et al., 2002). Many investigations have been orientated towards the characterisation of factors that could regulate EZH2 activity or recruitment. Indeed, Transcription factors or TFs (such as YY1) or LncRNA (such as HOTAIR) have been reported to recruit EZH2 to chromatin (Morey and Helin, 2010)

(Simon and Kingston, 2013). However the nature and the effect of these interactions are not fully understood. Partial explanation of PRC2 specificity may reside in variation of PRC2 composition and post-translational modifications that could influence EZH2 function. For example, it has been reported that PRC2 methylation of Jarid2 controls PRC2 catalytic activity (Sanulli et al., 2015). More conceptually challenging, increasing evidence suggests that EZH2 when not associated with canonical PRC2 component, occupies active genes independently of its H3K27me3 activity (Lee et al., 2011), (Mousavi et al., 2012), (Xu et al., 2012). Indeed, recent studies show that in hormone-refractory breast cancer and castration-resistant prostate cancer (CRPC), EZH2 can switch to a transcriptional activator and acts independently of the PRC2 complex (Lee et al., 2011) (Shi et al., 2007). In addition, in CRPC, EZH2 acts as a co-activator via the methylation of androgen receptor (AR) or AR-associated proteins. Interestingly in this study, at the genome wide level, EZH2 “solo peaks” (peaks without the repressive mark H3K27me3) were enriched for the active histone marks H3K4me2 and H3K4me3 and RNA polymerase II (Pol II) (Xu et al., 2012).

Previous work from the lab pointed out PLZF as a potential regulator of EZH2 functions (unpublished data). PLZF is a master regulator with effects on growth, self-renewal and differentiation with a well-recognised activity on regulation of haematopoietic cell commitment (Dick and Doulatov, 2009); (Suliman et al., 2012). Due to its essential role in Hox gene regulation, limb development (Barna, 2002) and stem cell maintenance (Kotaja and Sassone-Corsi, 2004) (Vincent-Fabert et al., *in revision*), PLZF and Polycomb Group (PcG) functions were suggested to synergise. Our previous work identified a molecular mechanism underlying this synergy, showing the recruitment of the PcG proteins by the interaction of the BTB/POZ domain of PLZF. This interaction impacts on the activity of the PLZF/RARA fusion by modifying its oncogenic property (Boukarabila, 2009).

In this work we studied whether PLZF and EZH2 interact and the effect they have on chromatin regulation. We show that PLZF interacts with EZH2 *in vivo* in myeloid cells and that together, they regulate a group of common genes at the chromatin level. Interestingly, EZH2 and PLZF common target genes are mostly associated with the active histone mark H3K4me3 and active transcription. Modulating PLZF expression, while not changing EZH2 chromatin binding does modify the chromatin landscape particularly at bivalent genes for which PLZF has a repressive effect. These results bring new elements for the understanding of epigenetic regulation mechanisms involving PLZF and EZH2.

Results

PLZF interacts with EZH2 through its BTB-POZ domain

The potential interplay between PLZF and EZH2 was investigated by co-immunoprecipitations (coIPs) *in vivo* in the myeloid cell line KG1 in which PLZF and EZH2 are endogenously expressed. Western blot analysis on proteins co-precipitating with EZH2 demonstrate that PLZF associates with EZH2 and SUZ12, a known partner of EZH2 in the PRC2 complex (**Fig. 1A**). Specificity of the association was shown by the absence of the histone demethylase LSD1 (a CoREST protein) in EZH2 precipitate (**Fig. 1A**). Reverse IP with an anti-PLZF antibody confirmed the association of PLZF and the PRC2 components EZH2 and SUZ12. The BTB-POZ domain of PLZF is known to be involved in macromolecular protein interaction (Suliman et al.). Involvement of this domain in PLZF-EZH2 association was investigated through co-IPs in HEK293T cells over expressing a MYC-tagged EZH2 with either WT or truncated PLZF. We found that PLZF WT and PLZF Δ Zn3-9, a mutant removing the C-terminal seven zinc finger domains, co-precipitated EZH2, whereas PLZF Δ BTB that lacks the BTB/POZ domain does not (**Fig. 1B**), demonstrating that PLZF molecularly interacts with EZH2 through its BTB-POZ domain. To assess the importance of the BTB domain for PLZF and EZH2 interplay on the transcriptional output we used the 293T-Rex cell system that contains the luciferase reporter gene controlled by a 5x Gal4 UAS sequence. We stably transfected the 293T-Rex cells with inducible PLZF and PLZF mutants in which the PLZF DNA binding domain (zinc fingers 3-9 that are not used for EZH2 interaction (Fig. 1B)) has been swapped with the Gal4 DNA binding domain (**Fig. S1A**). Upon induction, all PLZF-Gal4 fusions induced transcriptional repression (**Fig. S1B**). However, monitoring EZH2 activity by measuring H3K27me3 level, we found that only the PLZF form that has kept the BTB domain led to an increase of H3K27me3 upon induction (**Fig. S1C**). This result highlights the importance of the PLZF BTB domain for EZH2 recruitment and shows that recruitment of PLZF leads to transcriptional repression and an increase of H3K27me3 through co-recruitment of EZH2 via the PLZF BTB domain.

PLZF shares some genomic targets with EZH2

To ascertain whether the PLZF associates with EZH2 on chromatin *in vivo*, we profiled PLZF and EZH2 chromatin occupancy by chromatin immunoprecipitation (ChIP) coupled to high-throughput DNA sequencing (ChIP-seq) in the KG1 cell line. Sequence reads from PLZF and

EZH2 ChIPs and input controls were mapped to the hg19 human genome assembly and statistically significant enrichments of PLZF and EZH2 were identified by HOMER. PLZF and EZH2 peaks were analysed for location with respect to gene features using RefSeq gene annotations (see Supplemental Information), which showed that PLZF and EZH2 binding was over-represented at gene promoter and intronic regions (**Fig. 2A; Table S1**). PLZF and EZH2 co-occupy the transcription start site (TSS) at numerous promoters (**Fig. 2B**). Percentage of PLZF and EZH2 co-occupancy was even greater in the gene body with 35% of the PLZF target was also bound by EZH2 (**Fig. 2B**). Gene ontology analysis using the Homer gene ontology tools revealed that the genes bound by PLZF were enriched for genes associated with cell signalling and immune system process whereas gene bound by EZH2 were enriched for genes associated with developmental process and cell mobility (**Table S2**). Among these genes, already described PLZF target genes (McConnell et al., 2003), (Doulatov et al., 2009) were found and three examples of PLZF and EZH2 ChIP-seq coverage are shown in **Figure 2C**. We next validated some of PLZF-unique, PLZF/EZH2-common or EZH2-unique bound loci by performing PLZF and EZH2 ChIP followed by quantitative PCR (ChIP-qPCR). These experiments revealed that PLZF antibody immunoprecipitated PLZF unique regions with the same efficiency than PLZF/EZH2 common regions (**Fig. 2D**, left panel). In contrast, although detectable by ChIP-qPCR, EZH2 binding was lower in the case of PLZF/EZH2 common genes (**Fig. 2D**, right panel) in comparison to EZH2-unique. This indicates that EZH2 unique and EZH2 when co-localising with PLZF sites are quantifiably different. Motif search using HOMER at highest-ranked (by *p*-value) 500 PLZF or EZH2 ChIP-seq peaks identified enrichment of SPI-1 and NFY binding motifs respectively. When motif search was done with the PLZF/EZH2 common peaks we found that ETS1 (SPI-1) consensus motif was markedly overrepresented suggesting that PLZF/EZH2 peaks were influenced by the PLZF part of the complex.

PLZF and PLZF/EZH2 are bound to active genes

To further explore the association of PLZF and EZH2 on chromatin, we performed enrichment analyses followed by *k*-mean clustering of EZH2, PLZF, H3K27me3, H3K4me3, RNA PolII and H3K36me3 on refSeq genes (Fig 3A). This clustering reveal that PLZF enriched genes were mostly associated with the active mark H3K4me3 and the elongation mark H3K36me3 a chromatin profile associated with gene activation. Addition, this clustering separated EZH2-targeted genes into two distinct categories, based on the either the presence or absence of PLZF. The cluster of EZH2-targets lacking PLZF-binding was

enriched for H3K27me3 and lacks PolII, H3K4me3 and H3K36me3 enrichment (**Fig. 3A**; cluster highlighted in red). The genes co-targeted by EZH2 and PLZF were highly enriched in PolII, H3K4me3 and H3K36me3 but lacking H3K27me3 (**Fig. 3A**; highlighted in green). A third set of genes contained no enrichment for any of the factors tested (**Fig. 3A**; highlighted in blue).

We next assessed the expression status of these three groups of genes using publically available expression profiles on KG1 cells (GSM1316691). This analysis showed that the genes occupied by EZH2 and H3K27me3-enriched (red cluster) were poorly expressed and similar to genes containing no tested factors (blue cluster), whereas genes co-targeted by EZH2 and PLZF (green cluster) were highly expressed (**Fig. 3B**). Thus, PLZF or PLZF/EZH2 are mostly associated with gene activation in KG1 cells. Considering that PLZF has been described as a transcriptional repressor, based on reporter-gene assays (Kelly and Daniel, 2006) (see Fig. S1), these results were somewhat surprising. To rule out the possibility of a result biased by the nature of our cell line (KG1 cell line is a leukemic-derived cell line), we compared histone mark profiles and expression level of our defined PLZF-target genes in leukemic KG1 cells with ChIP-seq data available on normal CD34 positive hematopoietic stem cells (CD34⁺HSC) (GSM308470). H3K27me3, H3K4me3 and H3K27Ac CD34⁺HSC ChIP-seq data were clustered according to EZH2 and PLZF enrichment in KG1 cells, revealing highly comparable enrichment profiles for the histone marks in both KG1 and CD34⁺HSC cells (Fig. S2), demonstrating that PLZF target genes in KG1 cells are associated with a chromatin profile that is permissive to transcription in CD34⁺HSC. We next analysed gene expression data of both KG1 and CD34⁺HSC cells according to the PLZF enrichment determined in KG1 cells, through grouping PLZF-enriched genes (green cluster) into 20 sets of 200 genes containing decreasing amounts of PLZF enriched at the transcription start site (TSS) (**Fig. 3C**). Strikingly, the results showed that independently of the nature of the cells tested, genes highly enriched for PLZF in KG1 cells were mostly expressed in both cell types (**Fig. 3C**). To further explore the impact of EZH2 association or gene localisation on PLZF transcriptional outputs, we integrated the genomic localisation of PLZF (TSS or gene body) and the presence or absence of EZH2 at the TSS or gene body with the gene expression data. EZH2 association did not change PLZF transcriptional output in either KG1 or CD34⁺HSC (**Fig. 3D**). In contrast, localisation of PLZF along the gene had an impact on gene expression witnessed by genes that are bound by PLZF at gene body are more frequently repressed (**Fig. 3D**).

Together these results show that fixation of PLZF at promoters is mainly associated to active transcription and that binding of EZH2 at the same promoter does not modify PLZF transcription activity. Additionally, these data highlight that EZH2 at the promoter is frequently associated with transcriptionally active histone marks.

PLZF/EZH2 targets are not enriched for EZH1

It has recently been reported that EZH1 and SUZ12 form a non-canonical PRC2 complex that is associated to transcriptional activation and H3K4me3 histone marks during hematopoietic differentiation (Xu et al., 2015). In view of this study, we wanted to ascertain whether EZH1 could be associated to PLZF/EZH2 active genes. We first measured EZH1 binding at PLZF/EZH2 target genes by ChIP-qPCR. A panel of EZH2-unique, PLZF/EZH2-common and PLZF-unique genes was tested. We could not draw any correlation between the presence of either PLZF or EZH2, or both PLZF/EZH2 with a high EZH1 binding (**Fig. 4A**), suggesting that EZH1 binding was independent of PLZF or EZH2. H3K4me3 and H3K27me3 enrichment was tested at these genes (**Fig. 4B**). High H3K4me3 level was observed non-specifically at either EZH1 or EZH2 target genes (**Fig. 4B**), making difficult to assign H3K4me3 to either one or the other EZH protein. Additionally, coIP experiments showed that while PLZF was not able to interact with EZH1, and EZH2-EZH1 interaction was observed (**Fig. 4C**). These data shows that EZH2 and EZH1 activities are difficult to distinguish in this system, although they could be found in the same protein complex, which concurs with a previous study showing that EZH1 recruits EZH2 to chromatin (Son et al., 2013). Nonetheless, contrary to EZH2, we found no evidence of EZH1 being a PLZF partner in gene activation and conclude that PLZF active transcriptional activity is not due to EZH1 association.

Knockdown of PLZF results in decrease of H3K4me3 linked to decreased expression

To further understand the role of PLZF and EZH2 at the chromatin level, we performed shRNA-directed knockdown (KD) experiments. KD of PLZF was obtained by infecting KG1 cells with lentivirus containing shRNA directed against PLZF. KG1 clones efficiently down regulating PLZF were selected (**Fig. 5A**). Down regulation of PLZF did not modulate EZH2 expression nor globally modify H3K27me3 level (**Fig. 5A**). However, PLZF KD diminished PLZF binding at target genes (**Fig. 5B**). We next assessed whether PLZF KD modified the epigenetic landscape at categorised genes. PLZF-bound target genes were ranked according to PLZF enrichment at the TSS and EZH2, H3K4me3 and H3K27me3 ChIP signal were

compared, showing that genes strongly bound by PLZF at the TSS are highly marked by H3K4me3 and EZH2 and lacking H3K27me3 (**Fig. S3**). To assess the potential role of PLZF on the chromatin landscape we determine the genomic profiles of EZH2 together with the H3K4me3 histone mark upon PLZF KD (**Fig. 5C**). Quantification of ChIP-seq signal demonstrated that down-regulation of PLZF had no effect on EZH2 binding on high or low PLZF targets (**Fig. 5C**). However, PLZF KD diminished H3K4me3 level with a more pronounced effect on high PLZF targets (**Fig. 5C**). These analyses show that PLZF KD reduces H3K4me3 at PLZF target genes.

PLZF targets bivalent genes differently

Bivalent domains that combine the activating H3K4me3 mark and the repressive H3K27me3 mark are considered to poise expression of developmental genes (Bernstein et al., 2006). To study the involvement of PLZF in the regulation of these development genes, we analyzed bivalent promoters for PLZF occupancy. We found that 28% of bivalent promoters were bound by PLZF (**Fig. 6A**). Interestingly, expression analysis of this subset of bivalent genes in comparison to all PLZF target promoter genes showed an over representation of repressed genes (**Fig. 6B**). In addition, analysis of the H3K4me3/H3K27me3 signature under PLZF KD showed a marked decrease in H3K27me3 levels of bivalent genes (**Fig. 6C**; $p=0.0002$). This effect was restricted to the H3K27me3 mark as, contrary to non-bivalent genes (**Fig. 5C**), no significant modification of H3K4me3, EZH2 or Pol II level was observed in the bivalent subset of genes (**Fig. 6C** and **Fig. S4**). We next investigated the effect of PLZF KD on PLZF target gene expression. In line with the modification of H3K4me3 level and bivalent marks, the majority of PLZF target genes were down regulated upon PLZF KD, by contrast bivalent genes were up regulated upon PLZF KD (**Fig. 6D and E**).

Collectively, whereas PLZF is largely associated to H3K4me3 patterns and positively influences gene expression, at bivalent target genes PLZF functions as a transcriptional repressor, shown by PLZF KD resulting in a decrease of H3K27me3 and an increase in gene expression.

Discussion

Gene regulation is a central cellular process governing cell fate in development and differentiation. It relies to a large extent on transcriptional regulation, mediated by TFs and chromatin regulators. However, interplay between transcription factors and epigenetic modifications is still poorly understood.

Like many transcription factors, PLZF acts as either a repressor or an activator of transcription. However, it is the PLZF repressive activity that has been more recognized and extensively documented. Indeed many structural studies have shown that PLZF can repress transcription through sequence specific recruitment of co-repressors such as N-CoR/HDAC or PcG complexes (Lin et al., 1998), (Barna et al., 2002), (Boukarabila et al., 2009) implicating the BTB/POZ domain of the proteins as the main repressive activity domain (Melnick et al., 2000). In our study, we also confirm that PLZF is by “default” a transcriptional repressor when assayed using a Gal4/UAS reporter system. However, in the myeloid cellular context, we revealed that PLZF is mostly associated with active transcription. Indeed, in myeloid development, PLZF has been shown to activate the transcription of some of its bound promoters (Doulatov et al., 2009). While this transcriptional activation property of PLZF may be specific to the hematopoietic tissue, it will be important to study PLZF activity globally in other tissues, such as spermatogenesis, where it plays an important role during differentiation.

The dual effect of PLZF on transcription could be due to many factors influencing its partner binding or its specific DNA motif recognitions. We ruled out the possibility that in our system PLZF is recognising “non canonical” DNA sequence as consensus genomic motif as the 500 most bound PLZF promoter in our system is very similar to the consensus -A-T/G-G/C-T-A/C-A/C-A-G-G-T- found previously by *in vitro* analysis (Li et al., 1997). Others factors that could modify PLZF activity are its post-translational modifications. Previous studies have demonstrated that the transcriptional repression activity of PLZF is specifically dependent on the HAT activity of p300 or HAT1, which acetylates PLZF (Guidez et al., 2005) (Sadler et al., 2015) and that SUMOylation in K242, K387 and K396, extensively modulates PLZF-mediated transcriptional repression (Chao et al., 2007). We do not know if these two pathways are defective in KG1 cells and could afford an explanation for the lack of repressive activity of PLZF.

Here, we showed that PLZF immunoprecipitates EZH2 and SUZ12, two key components of the PRC2 components, but not EZH1 or LSD1, suggesting that PLZF is mainly associated with EZH2-containing PRC2. This was informative in the light of recent

literature that revealed a non-canonical function of PRC2-EZH1 complex that positively regulates gene expression (Xu et al., 2015). However, the presence or absence of EZH2 had no influence on the transcriptional output of PLZF target genes, suggesting that PLZF transcriptional activity was not modulated by EZH2. More surprisingly, we revealed that EZH2 associates with both active and repressive chromatin. EZH2-bound promoters, when co-occupied by PLZF are mainly associated with the active H3K4me3 histone mark and related coding regions. These genes are also enriched for the H3K36me3, a histone mark associated with elongating Pol II, and are expressed. Conversely, EZH2 promoters lacking PLZF are associated with H3K27me3 and poor levels of expression. The observation of EZH2 being associated with active chromatin is not without precedence. Previous reports described EZH2 as a transcriptional co-activator in the context of prostate cancer (Lee, 2011); (Xu et al., 2012), and linked PRC2 to transcriptionally active genes (Mousavi et al., 2012) as well as association with Pol II (Brookes et al., 2012) and nascent RNA in ES cell (Kaneko et al., 2013). It seems evident that EZH2 enzymatic activity that leads to H3K27me3 deposition can be modulated, although the molecular bases of this modulation are not fully understood. A recent structural study underlined the EZH2 PRC2-mediated H3K27 trimethylation activity, but also revealed the complex regulation of PRC2 activity by a plethora of protein factors and noncoding RNAs in cells (Jiao and Liu, 2015).

Could PLZF be an influencing factor of EZH2 activity? Our KD experience showed that down-regulating PLZF had little influence on EZH2 binding, suggesting that PLZF is not involved in EZH2 recruitment *per se*. However, our results showed that PLZF KD globally decreases H3K4me3 levels at PLZF-bound promoters as well as decreasing H3K27me3 levels specifically at bivalent PLZF-bound promoters, suggesting that loss of PLZF decreases EZH2 activity only at bivalent PLZF bound promoters. By analogy with what has been proposed for RBB4 (Schmitges et al., 2011) or lncRNA (Davidovich et al., 2013), where both factors have been proposed to be sensors of EZH2 activity, PLZF could be also a modulator of EZH2 activity. PLZF could recognise active marks and depending if the active mark is associated with repressive mark (bivalent genes), PLZF could play a role in modulating PRC2 catalytic activity.

Acknowledgments

Authors thanks Dr Vincent Geli and Dr Ari Melnick for discussions. This work was supported by the Association pour la recherche sur le cancer (ARC), by the Association Laurette Fugain (ALF), by ITMO-cancer and INCA (PLBio). MK was supported by the Ministère de l'Enseignement Supérieur and la Ligue Contre le Cancer (LCC). SE was supported by the cancéropôle PACA. MP was supported by ITMO-cancer. We gratefully acknowledge the Montpellier MGX and MIT Biomicro Center Sequencing facility for the ChIP-seq sequencing.

Material and methods

Cell culture and Knockdown

Human myeloblastic cell line KG1 and 293T cells were maintained at exponential growth in RPMI or DMEM respectively supplemented with 10% foetal calf serum and 1% sodium pyruvate. EZH2 and PLZF knockdown were achieved using a doxycycline-induced short hairpin RNA (sh-RNA)-targeting EZH2 or PLZF (pTRIPZ-EZH2, Openbiosystem # V2THS_63066, pTRIPZ-PLZF, Openbiosystem # V2LHS_72096). A non-silencing sh-RNA (pTRIPZ-NS, openbiosystem # RHS4743) was used as control. Cells containing the pTRIPZ were selected on puromycine (1,75 μ g/mL). Knockdown of PLZF and EZH2 was obtained by the addition of doxycycline (2 μ g/ml) to the cells during 7 to 10 days.

Nuclear extracts, coIPs and western blotting.

Nuclear proteins extraction of KG1 and 293T cells was done using the dounce homogenizer with high salt concentration. For Co-IP in KG1 cells, nuclear extracts were diluted in hypotonic buffer and incubated 3 hours with 2 μ g of anti-PLZF (Santa Cruz Biotechnology, H300) or anti-EZH2 (Active Motif, pAb, cat 39901 or AC22 antibody) antibody in the presence of Protein G beads using (Dynabeads, Life Technologies) and resulting complexes were washed, denatured and eluted according to the manufacturer's instructions.

Nuclear extracts of 293T cells were IP with the ProFound c-Myc tag IP/coIP kit (No 23 622) following the manufacturer's instructions. WBs were performed as previously described (Boukarabila et al., 2009). Briefly, nuclear extracts were separated on polyacrylamide gels, blotted on nitrocellulose membrane and incubated with different primary antibodies, listed below. Anti-EZH2 BD Biosciences # 612666 (1/1000), anti-PLZF H300 (Santa Cruz

Biotechnology, 1/2000), anti-LSD1 (Active Motif, cat 39186, 1/5000), anti-BMI1 (Abcam, 1.T.21, 1/500), EZH1 (Abcam, ab13665, 1/500), anti-MLL1 (Millipore, N4.4, 1/400), anti-WDR5 (Active Motif, cat 61485, 1/1000), anti-H3K27me3 (Millipore, cat 07-449, 1/500), anti-H3K4me3 (Diagenode, pAb-003-050, 1/1000), anti-Tubulin (Sigma Aldrich, DM4A, 1/5000), anti-H3 (Active Motif, 39163, 1/10000).

Chromatin Immunoprecipitation

ChIP was performed as described before in Boukarabila et al. (2009) with some modifications. After chromatin extraction, DNA was sonicated to obtain 100-500 base pair fragments (ultrasonic processor, Fischer Scientific, ref 75041, 40% amplitude, 30 s on and 30 s off for 20 minutes). For ChIP using PLZF and EZH2 antibodies chromatin extracted from 2x10⁷ cells per condition was used. For ChIP using H3K4me3 or H3K27me3 antibodies chromatin extracted from 5x10⁶ cells per condition was used. After immunoprecipitation, DNA was purified with the I-Pure kit (Diagenode) for sequencing analyses or with chelex-100 (Bio-Rad Laboratories) of q-PCR analyses.

Quantitative (q)-RT-PCR analysis

Quantification of ChIPed DNA or cDNA was performed by quantitative real-time PCR. DNA target sequences were selected based on our ChIP-Seq analyses in KG1 cells and primers were designed for amplifying the DNA location of interest with Primer-BLAST. Real-time PCR was performed using Power SYBR Green PCR Master Mix (Applied Biosystems) according to the manufacturer's instructions with the 7500 Fast Real Time PCR system (Life Technologies). For ChIPed DNA, IgG control "cycle over the threshold" Ct values were subtracted to Input or IP Ct values and converted into bound value by $2^{-(IP\ Ct\ or\ input\ Ct - IgG\ IP\ Ct)}$. Data are expressed as % of bound /input. For cDNA relative expression levels were determined by the $2^{\Delta\Delta CT}$ method using HPRT or PBG-D as control genes. Experiments were performed with two individual clones treated or not with the doxycycline. Technical triplicates are presented as mean values $\pm$ SD. Primers are given under request.

Sequencing and data processing

Libraries and sequencing were performed with a *HiSeq* 2000 next-generation sequencing platform (Illumina) using a 40 bp single-end protocol, either at the Montpellier GenomiX (MGX) or at Harvard sequencing MIT BIO. Sequence data quality in each sample was assessed using FastQC (a quality control tool for high throughput sequence data,

<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads with Phred quality score less than 30 were filtered out and sample replicates were merged into a single fastq file. Sequences were mapped to the human genome (hg19) using BFAST alignment tool (ref: PMID 19907642) with default parameters.

Mapped sequencing data (BAM files) are in the process to be deposited in the NCBI Gene Expression Omnibus (GEO; [http:// www.ncbi.nih.gov/geo/](http://www.ncbi.nih.gov/geo/)).

High confidence binding sites were determined through peak calling using HOMER tool (Heinz et al., 2010) with the “factor setting” for EZH2, PLZF and PolII and the “histone setting” for histone marks (H3K27me3 and H3K4me3). False-discovery-rate (FDR) cutoff was set at <0.05 and both overall and local p-value cutoff at 0.01. To increase the accuracy of the identified ChIP enrichment sites, only peaks with ≥ 3 -fold tag count enrichment compared to both input control and local tag density (on region of 10kb around putative peak) were considered for downstream analyses. For all ChIP-seq datasets, tag counts were normalized by the total number of mapped tags such that each directory contains 10 million reads. For peak annotation, we used HOMER annotate Peaks tool on BED files containing the retained peaks of each sample. We identified therefore all target genes and the distance of every peak to the nearest Transcription Start Site. Binding sites were at considered the promoters if they were found at +/- 1kb of the TSS. Binding sites were considered at the gene body if they could be detected between 1-kb downstream of the TSS and the Transcription End Site (TES).

Heatmap visualisation of ChIP enrichment at EZH2 and PLZF-enriched non-redundant transcripts was performed using ngs.plot (version 2.47; ref: <http://www.biomedcentral.com/1471-2164/15/284>).

Plotting of the enrichment profiles were done by ranking the genes using K-means clustering method (here set up at 6 clusters). The average profiles were calculated using the mean read coverage among the gene bodies in each resulted cluster. Quantification of ChIP enrichment for profile plots was performed using the mean (factor binding) or median (histone marks) enrichment signal for each gene. The regions used to calculate enrichment signals were different for EZH2 and for H3K4me3 heatmap figure and correspond to “promoter -proximal regions” (defined as -2000 bp to the TSS) for EZH2 and an H3K4me3 region (first 33% of the gene body) for H3K4me3. The calculated enrichment signal for each gene was then mean smoothed using a 40-gene sliding window across all genes and average profile plots generated in R.

Expression data

To explore gene expression in both KG1 and HSC CD34⁺ cells, we retrieved publicly available expression data. To determine if a gene was expressed or not (present or absent), we used the “ABS_CALL” values determined by the Affymetrix GeneChip with “Detection Call” algorithm described by (Liu et al., 2002). The corresponding probe set is present (flagged “P”, $p\text{-value} \leq 0,05$), absent (flagged “A”, $p\text{-value} > 0,066$) or marginal (flag “M”, $0,05 \leq p\text{-value} \leq 0.065$).

Data access

In addition to our own ChIP-seq data, previously published data sets were retrieved from the Gene Expression Omnibus (GEO) database for re-analysis and comparisons. Sequencing datasets were thus retrieved for: human hematopoietic cells enriched within the immature cell surface marker CD34⁺ (HSC-CD34⁺) H3K27me3 (GSM772951), HSC-CD34⁺ H3K4me3 (GSM773041), (HSC-CD34⁺) H3K27ac. In addition, expression data from the HSC CD34⁺ cells of healthy donor (GSM308470) and expression data from human untreated KG1 cells (GSM1316691) were extracted.

References

- Barna, M., Merghoub, T., Costoya, J.A., Ruggero, D., Branford, M., Bergia, A., Samori, B., and Pandolfi, P.P. (2002). Plzf mediates transcriptional repression of HoxD gene expression through chromatin remodeling. *Dev Cell* 3, 499-510.
- Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., *et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125, 315-326.
- Boukarabila, H., Saurin, A.J., Batsche, E., Mossadegh, N., van Lohuizen, M., Otte, A.P., Pradel, J., Muchardt, C., Sieweke, M., and Duprez, E. (2009). The PRC1 Polycomb group complex interacts with PLZF/RARA to mediate leukemic transformation. *Genes Dev* 23, 1195-1206.
- Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., *et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441, 349-353.
- Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H., and Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev* 20, 1123-1136.
- Brookes E, *et al.* (2012) Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. *Cell Stem Cell*; 10:157–170.
- Chao, T.T., Chang, C.C., and Shih, H.M. (2007). SUMO modification modulates the transrepression activity of PLZF. *Biochem Biophys Res Commun* 358, 475-482.
- Davidovich, C., Zheng, L., Goodrich, K.J., and Cech, T.R. (2013). Promiscuous RNA binding by Polycomb repressive complex 2. *Nat Struct Mol Biol* 20, 1250-1257.
- Deneberg, S., Guardiola, P., Lennartsson, A., Qu, Y., Gaidzik, V., Blanchet, O., Karimi, M., Bengtzen, S., Nahi, H., Uggla, B., *et al.* (2011). Prognostic DNA methylation patterns in cytogenetically normal acute myeloid leukemia are predefined by stem cell chromatin marks. *Blood* 118, 5573-5582.
- Doulatov, S., Notta, F., Rice, K.L., Howell, L., Zelent, A., Licht, J.D., and Dick, J.E. (2009). PLZF is a regulator of homeostatic and cytokine-induced myeloid development. *Genes Dev* 23, 2076-2087.

- Guidez, F., Howell, L., Isalan, M., Cebrat, M., Alani, R.M., Ivins, S., Hormaeche, I., McConnell, M.J., Pierce, S., Cole, P.A., *et al.* (2005). Histone acetyltransferase activity of p300 is required for transcriptional repression by the promyelocytic leukemia zinc finger protein. *Mol Cell Biol* 25, 5552-5566.
- 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.
- Hock, H. (2012). A complex Polycomb issue: the two faces of EZH2 in cancer. *Genes Dev* 26, 751-755.
- Jiao, L., and Liu, X. (2015). Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. *Science* 350, aac4383.
- Kaneko, S., Son, J., Shen, S.S., Reinberg, D., and Bonasio, R. (2013). PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nat Struct Mol Biol* 20, 1258-1264.
- Kelly, K.F., and Daniel, J.M. (2006). POZ for effect--POZ-ZF transcription factors in cancer and development. *Trends Cell Biol* 16, 578-587.
- Koppens, M., and van Lohuizen, M. (2015). Context-dependent actions of Polycomb repressors in cancer. *Oncogene*.
- Laugesen, A., and Helin, K. (2014). Chromatin repressive complexes in stem cells, development, and cancer. *Cell Stem Cell* 14, 735-751.
- Lee, S.T., Li, Z., Wu, Z., Aau, M., Guan, P., Karuturi, R.K.M., Liou, Y.C., and Yu, Q. (2011). Context-specific regulation of NF- κ B target gene expression by EZH2 in breast cancers. *Mol Cell* 43, 798-810.
- Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., *et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 125, 301-313.
- Li, J.Y., English, M.A., Ball, H.J., Yeyati, P.L., Waxman, S., and Licht, J.D. (1997). Sequence-specific DNA binding and transcriptional regulation by the promyelocytic leukemia zinc finger protein. *J Biol Chem* 272, 22447-22455.
- Lin, R.J., Nagy, L., Inoue, S., Shao, W., Miller, W.H., Jr., and Evans, R.M. (1998). Role of the histone deacetylase complex in acute promyelocytic leukaemia. *Nature* 391, 811-814.
- Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature* 469, 343-349.
- McCabe, M.T., Ott, H.M., Ganji, G., Korenchuk, S., Thompson, C., Van Aller, G.S., Liu, Y., Graves, A.P., Della Pietra, A., 3rd, Diaz, E., *et al.* (2012). EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* 492, 108-112.
- McConnell, M.J., Chevallier, N., Berkofsky-Fessler, W., Giltane, J.M., Malani, R.B., Staudt, L.M., and Licht, J.D. (2003). Growth suppression by acute promyelocytic leukemia-associated protein PLZF is mediated by repression of c-myc expression. *Mol Cell Biol* 23, 9375-9388.
- Melnick, A., Ahmad, K.F., Arai, S., Polinger, A., Ball, H., Borden, K.L., Carlile, G.W., Prive, G.G., and Licht, J.D. (2000). In-depth mutational analysis of the promyelocytic leukemia zinc finger BTB/POZ domain reveals motifs and residues required for biological and transcriptional functions. *Mol Cell Biol* 20, 6550-6567.
- Morey, L., and Helin, K. (2010). Polycomb group protein-mediated repression of transcription. *Trends Biochem Sci* 35, 323-332.
- Morin, R.D., Mendez-Lago, M., Mungall, A.J., Goya, R., Mungall, K.L., Corbett, R.D., Johnson, N.A., Severson, T.M., Chiu, R., Field, M., *et al.* (2011). Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* 476, 298-303.
- Mousavi, K., Zare, H., Wang, A.H., and Sartorelli, V. (2012). Polycomb protein Ezh1 promotes RNA polymerase II elongation. *Mol Cell* 45, 255-262.
- Nikoloski, G., Langemeijer, S.M., Kuiper, R.P., Knops, R., Massop, M., Tonissen, E.R., van der Heijden, A., Scheele, T.N., Vandenberghe, P., de Witte, T., *et al.* (2010). Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. *Nat Genet* 42, 665-667.
- Ntziachristos, P., Tsiganos, A., Vlierberghe, P.V., Nedjic, J., Trimarchi, T., Flaherty, M.S., Ferres-Marco, D., da Ros, V., Tang, Z., Siegle, J., *et al.* (2012). Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. *Nat Med* 18, 298-303.

- Sadler, A.J., Suliman, B.A., Yu, L., Yuan, X., Wang, D., Irving, A.T., Sarvestani, S.T., Banerjee, A., Mansell, A.S., Liu, J.P., *et al.* (2015). The acetyltransferase HAT1 moderates the NF-kappaB response by regulating the transcription factor PLZF. *Nat Commun* 6, 6795.
- Schmitges, F.W., Prusty, A.B., Faty, M., Stutzer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., *et al.* (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol Cell* 42, 330-341.
- Simon, J.A., and Kingston, R.E. (2013). Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol Cell* 49, 808-824.
- Son, J., Shen, S.S., Margueron, R., and Reinberg, D. (2013). Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. *Genes Dev* 27, 2663-2677.
- Suliman, B.A., Xu, D., and Williams, B.R. (2012). The promyelocytic leukemia zinc finger protein: two decades of molecular oncology. *Front Oncol* 2, 74.
- Tanaka, S., Miyagi, S., Sashida, G., Chiba, T., Yuan, J., Mochizuki-Kashio, M., Suzuki, Y., Sugano, S., Nakaseko, C., Yokote, K., *et al.* (2012). Ezh2 augments leukemogenicity by reinforcing differentiation blockage in acute myeloid leukemia. *Blood* 120, 1107-1117.
- Xu, J., Shao, Z., Li, D., Xie, H., Kim, W., Huang, J., Taylor, J.E., Pinello, L., Glass, K., Jaffe, J.D., *et al.* (2015). Developmental control of polycomb subunit composition by GATA factors mediates a switch to non-canonical functions. *Mol Cell* 57, 304-316.
- Xu, K., Wu, Z.J., Groner, A.C., He, H.H., Cai, C., Lis, R.T., Wu, X., Stack, E.C., Loda, M., Liu, T., *et al.* (2012). EZH2 oncogenic activity in castration-resistant prostate cancer cells is Polycomb-independent. *Science* 338, 1465-1469.

Figure legends

Figure 1: PLZF coimmunoprecipitates with EZH2 through its BTB-POZ domain. **A** Nuclear extracts of KG1 cells were immunoprecipitated either with an anti-EZH2 (left panel) or anti-PLZF (right panel) antibody. Immunoprecipitates (IPs) were immunoblotted with an anti- EZH2, an anti-PLZF, an anti-SUZ12 and an anti LSD1. Inputs (In) represent 5% of total input. **B** Nuclei of 293T cells co-transfected with PLZF constructs (WT or mutants) and Myc-EZH2 were immunoprecipitated with an anti-Myc coupled resin. In represent 2% of total input.

Figure 2: PLZF and EZH2 colocolize at the genome. **A** Genomic distribution of PLZF and EZH2 peaks obtained from PLZF and EZH2 ChIP-seq in KG1 cells. **B** Venn diagram showing the overlap between genes enriched for PLZF and EZH2 at the TSS ($\pm$ 1kb) or in the Genebody (>1 kb from TSS). **C** UCSC browser views of EZH2 and PLZF ChIP-seq data. Three loci are shown. The y-axis corresponds to ChIP-seq signal intensity. **D** ChIP-qPCR analysis of PLZF, EZH2 positive genes are underlined and indicated as EZH2⁺ (left panel) and EZH2 PLZF positive genes are underlined and indicated as PLZF⁺ (right panel) of selected target genes. Percentages of input DNA are shown as a mean $\pm$ SD of two independent experiments (n=3). **E** Result of de novo motif search at PLZF, EZH2 or PLZF-EZH2 bound regions.

Figure 3: PLZF bound regions are associated with H3K4me3 histone mark and are expressed in KG1 cells. **A** K-mean clustering (n=6) of refSeq transcripts according to PLZF, EZH2, H3K4me3, H3K27me3, PolII and H3K36me3 enrichment in KG1 cells at 25% of the gene length before the TSS to 25% after the TTS. All genes were scaled to have the same length. **B** mRNA expression values are shown for the 3 clusters defined in a: red for the EZH2/ H3K27me3 positive genes; green for the PLZF/EZH2 positive genes and blue for Null genes. **C-D** Percentage of present and absent genes extracted from transcriptomic data of KG1 cells (GSM1316691) and CD34+HSC (GSM308470). Data was analysed by absolute call that indicates if a transcript was present (P), absent (A), marginal (M) for defined groups. In **C** Refseq genes were sorted and grouped (20 groups; number of genes=200 per group) according to their downward PLZF genomic level in KG1 cells at $\pm$ 500 kb from TSS. In **D** PLZF bound genes were classified in five groups: (1) PLZF only TSS, (2) common PLZF and EZH2 at TSS, (3) common PLZF at TSS and EZH2 at gene body, (4) PLZF only at gene body, (5) common PLZF and EZH2 at gene body.

Figure 4: EZH2/PLZF targets are not enriched for EZH1. **A** ChIP-qPCR analysis of EZH1 binding on a panel of EZH2/PLZF target genes. **B** ChIP-qPCR analysis of H3K4me3 and H3K27me3 levels of EZH2/PLZF target genes. Percentages of input DNA are shown as a mean of at least three technical replicates $\pm$ SD. Representative results from at least two independent experiments. **C** EZH1 does not immunoprecipitate with PLZF but with EZH2 in KG1 cells.

Figure 5: Knockdown of PLZF results in decrease of global H3K4me3. **A** Western blot of PLZF, EZH2 protein content and H3K27me3 upon shMock or shPLZF conditions. **B** ChIP showing a decrease of PLZF binding on PLZF target genes in the KD condition. **C** quantification of EZH2 and H3K4me3 level after PLZF KD. PLZF targets were grouped in quartile Heat map and average profile plots of mean log2 ChIP enrichment at $\pm$ 2kb of the TSS are shown for ChIPs performed on KG1 shMock cells (-; green) and KG1 shPLZF cells (+; red).

Figure 6: PLZF targets bivalent genes differently. **A** Overlap of PLZF bound regions with bivalent marked genes in KG1 cells. **B** Comparison of expression level of all PLZF bound regions (1) with bivalent only PLZF targeted genes (2). **C** H3K4me3, H3K27me3, EZH2 and PolII levels around the TSS of all or bivalent marked PLZF target genes in shMock or shPLZF KG1 cells. The regions used to calculate enrichment signals correspond to +/- 5% of the gene length around the TSS for H3K4me3, EZH2 and PolII and -10% of gene length upstream the TSS for H3K27me3. **D-E** Gene expression analysis on PLZF target genes after PLZF KD on two independent clones. Relative level of expression was evaluated using the $2^{\Delta\Delta CT}$ method with PBG-D (D) or HPRT (E) as the reference gene. Data are presented as a mean $\pm$ SD of three technical replicates.

Figure 1

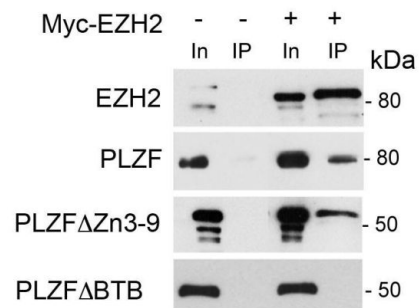
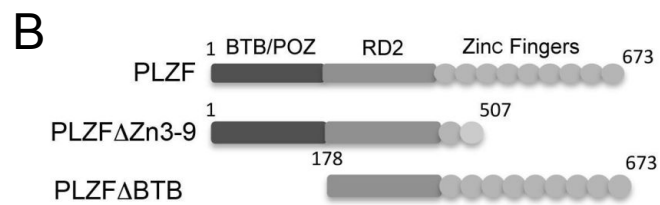
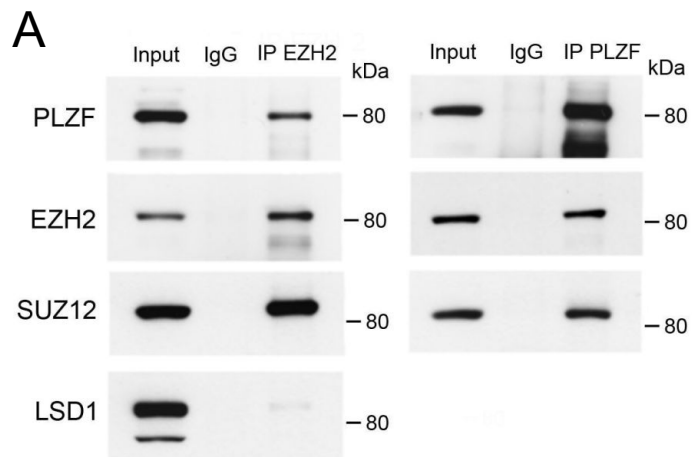


Figure 2

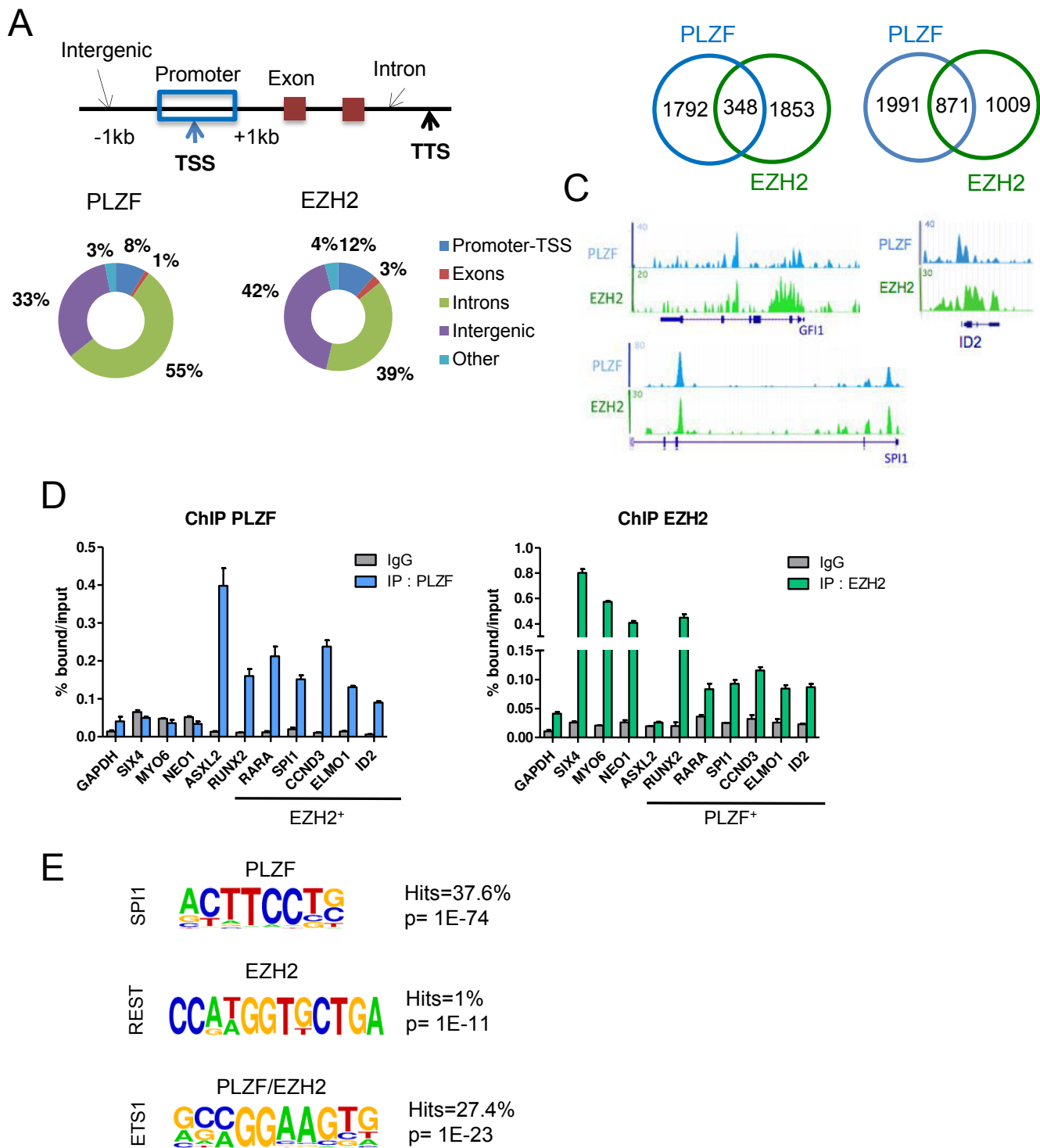


Figure 3

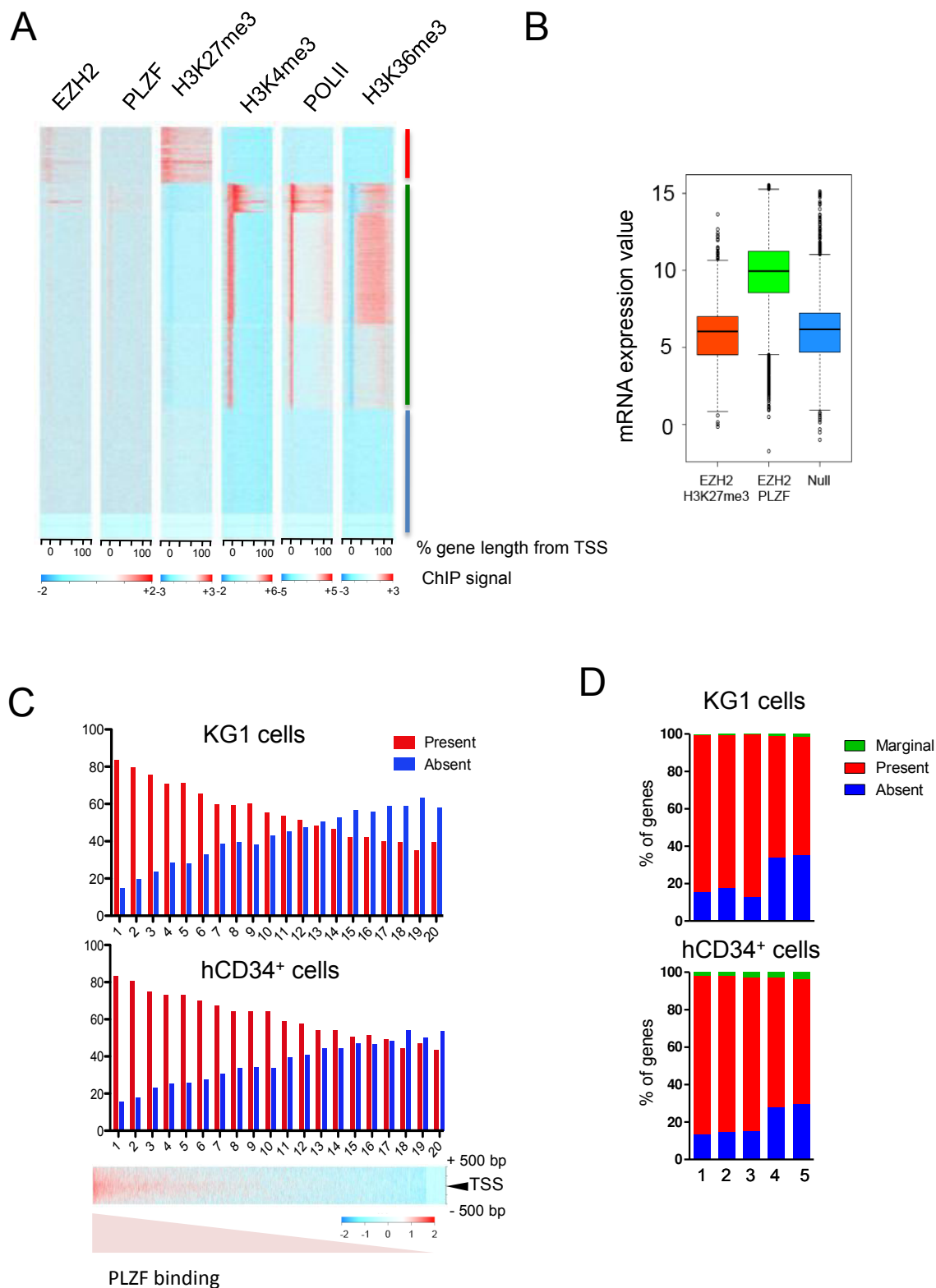
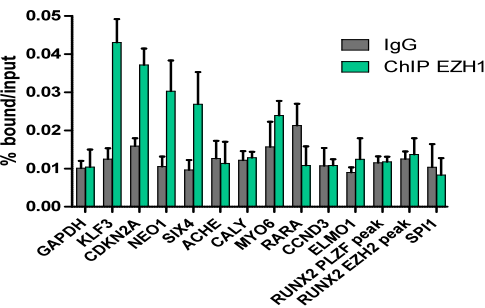
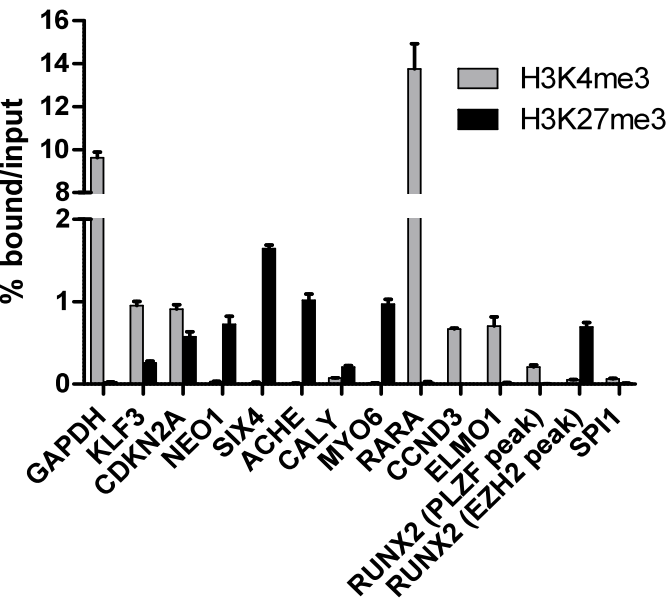


Figure 4

A



B



C

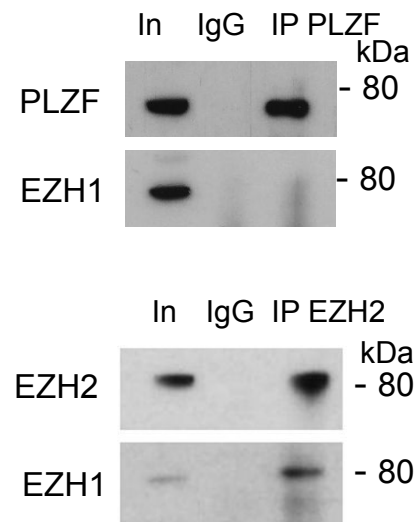


Figure 5

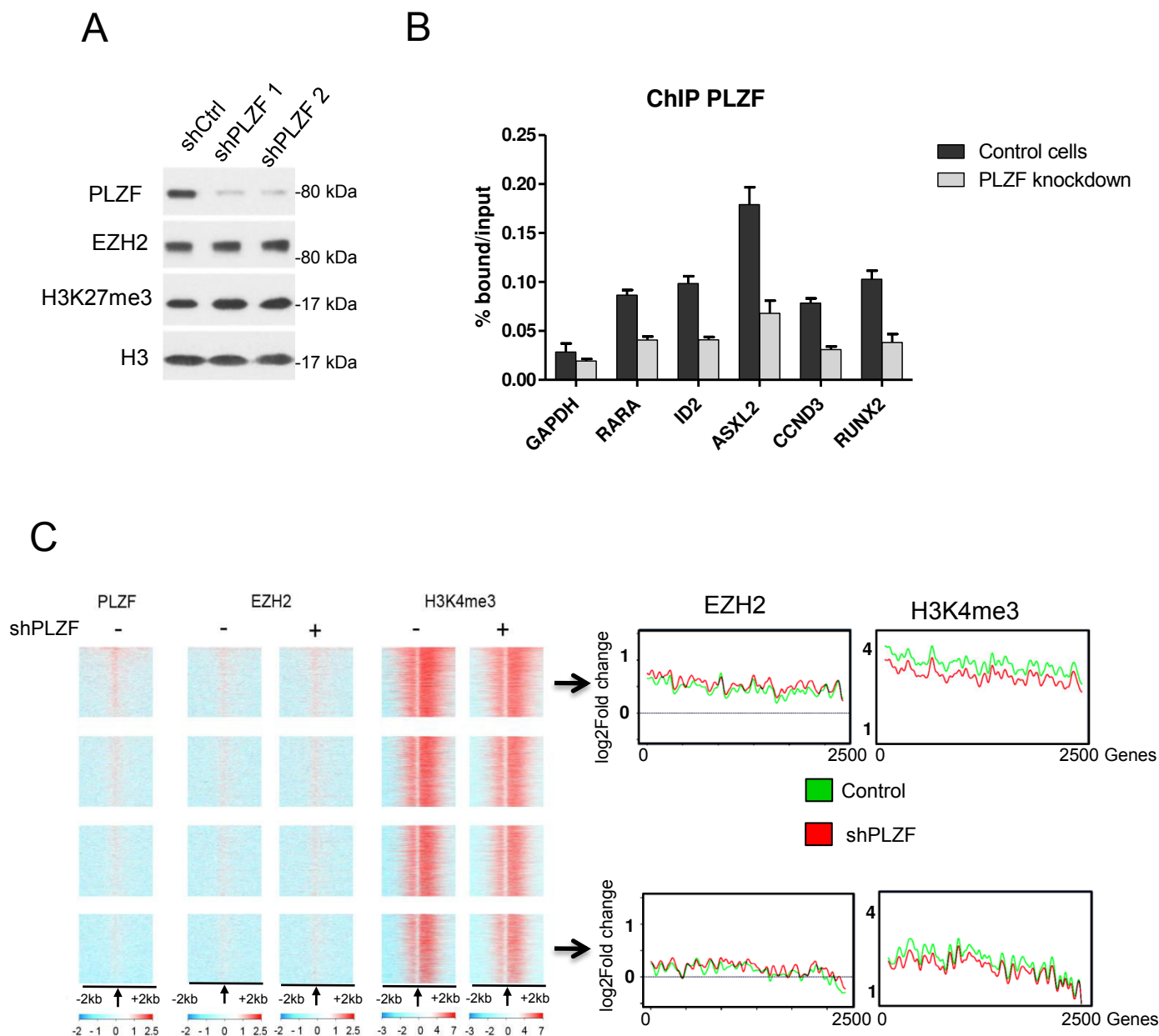


Figure 6

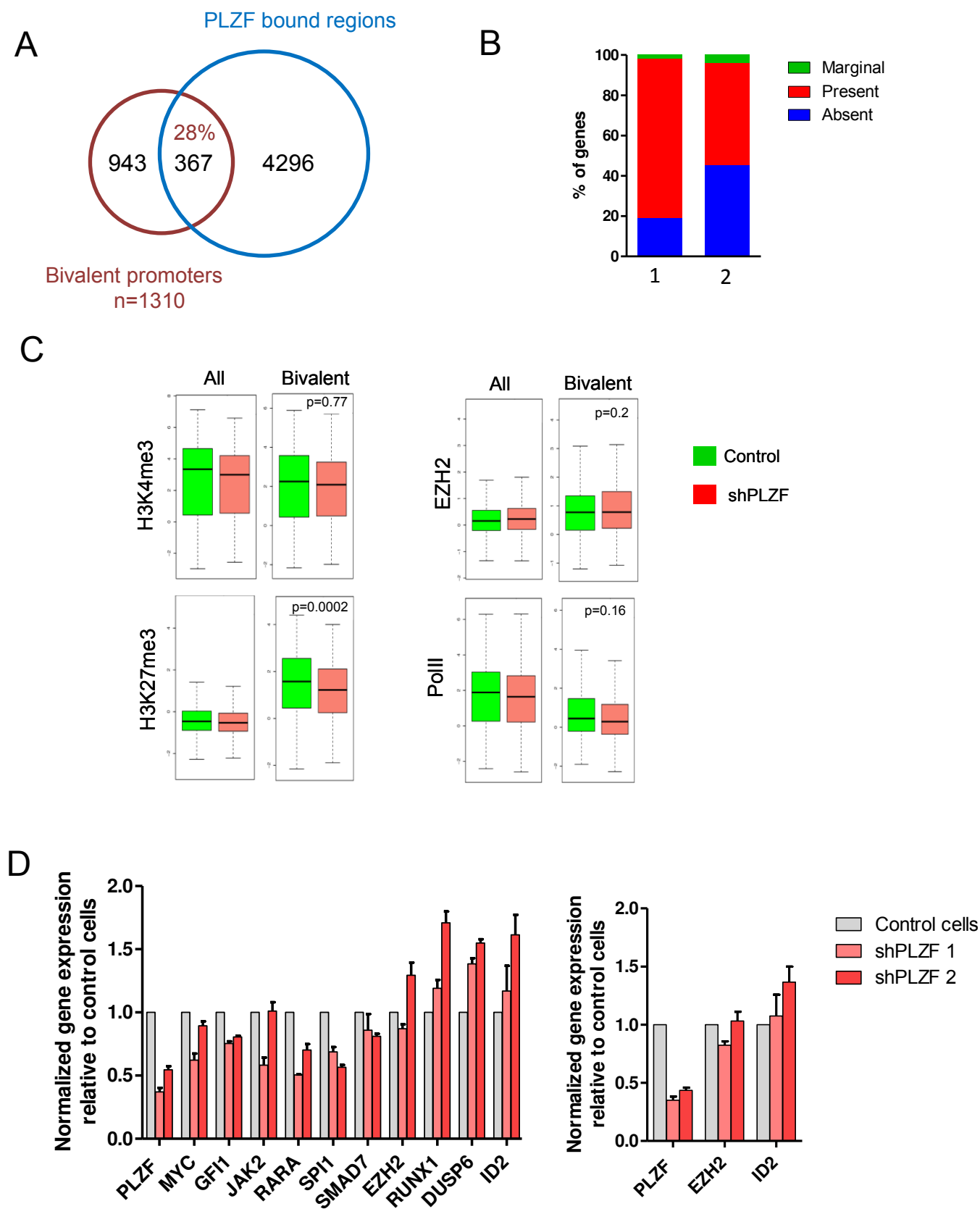


Table S1: Distribution patterns of PLZF and EZH2 peaks across 5 categories of genomic locations

	Number of peaks				
	Promoter TSS	Exon	Intron	Intergenic	Other
PLZF	1645	242	10570	6311	596
EZH2	1776	404	6107	6560	621

Table S2: Enriched GO terms in biological process for PLZF (A) and EZH2 (B) regulated genes.

A)

Term_ID	Description	log10 pvalue	Representative
GO:0010605	negative regulation of macromolecule metabolic process	-5.2921	negative regulation of macromolecule metabolism
GO:0010604	positive regulation of macromolecule metabolic process	-2.7717	negative regulation of macromolecule metabolism
GO:0048522	positive regulation of cellular process	-3.3149	negative regulation of macromolecule metabolism
GO:0048519	negative regulation of biological process	-6.8564	negative regulation of macromolecule metabolism
GO:0048518	positive regulation of biological process	-4.1327	negative regulation of macromolecule metabolism
GO:0009892	negative regulation of metabolic process	-4.6129	negative regulation of macromolecule metabolism
GO:0051239	regulation of multicellular organismal process	-3.0137	negative regulation of macromolecule metabolism
GO:0006334	nucleosome assembly	-2.5466	negative regulation of macromolecule metabolism
GO:0031399	regulation of protein modification process	-2.3162	negative regulation of macromolecule metabolism
GO:0060255	regulation of macromolecule metabolic process	-4.1648	negative regulation of macromolecule metabolism
GO:0019222	regulation of metabolic process	-2.6234	negative regulation of macromolecule metabolism
GO:0032501	multicellular organismal process	-6.1185	multicellular organismal process
GO:0032502	developmental process	-4.7234	developmental process
GO:0040011	locomotion	-4.5456	locomotion
GO:0048598	embryonic morphogenesis	-5.7903	embryonic morphogenesis
GO:0050793	regulation of developmental process	-2.3885	embryonic morphogenesis
GO:0007275	multicellular organismal development	-7.9433	embryonic morphogenesis
GO:0009888	tissue development	-3.1405	embryonic morphogenesis
GO:0009887	organ morphogenesis	-4.4561	embryonic morphogenesis
GO:0051240	positive regulation of multicellular organismal process	-2.4893	embryonic morphogenesis
GO:0000904	cell morphogenesis involved in differentiation	-2.3901	embryonic morphogenesis
GO:0061458	reproductive system development	-3.9657	embryonic morphogenesis
GO:0009790	embryo development	-4.6321	embryonic morphogenesis
GO:0048608	reproductive structure development	-4.0083	embryonic morphogenesis
GO:0006928	cellular component movement	-3.9768	cellular component movement
GO:0010595	positive regulation of endothelial cell migration	-2.3951	cellular component movement
GO:0051674	localization of cell	-2.7017	cellular component movement
GO:0016477	cell migration	-3.1972	cellular component movement

B)

Term_ID	Description	log10 pvalue	Representative
GO:0010605	negative regulation of macromolecule metabolic process	-5.2921	negative regulation of macromolecule metabolism
GO:0010604	positive regulation of macromolecule metabolic process	-2.7717	negative regulation of macromolecule metabolism
GO:0048522	positive regulation of cellular process	-3.3149	negative regulation of macromolecule metabolism
GO:0048519	negative regulation of biological process	-6.8564	negative regulation of macromolecule metabolism
GO:0048518	positive regulation of biological process	-4.1327	negative regulation of macromolecule metabolism
GO:0009892	negative regulation of metabolic process	-4.6129	negative regulation of macromolecule metabolism
GO:0051239	regulation of multicellular organismal process	-3.0137	negative regulation of macromolecule metabolism
GO:0006334	nucleosome assembly	-2.5466	negative regulation of macromolecule metabolism
GO:0031399	regulation of protein modification process	-2.3162	negative regulation of macromolecule metabolism
GO:0060255	regulation of macromolecule metabolic process	-4.1648	negative regulation of macromolecule metabolism
GO:0019222	regulation of metabolic process	-2.6234	negative regulation of macromolecule metabolism
GO:0032501	multicellular organismal process	-6.1185	multicellular organismal process
GO:0032502	developmental process	-4.7234	developmental process
GO:0040011	locomotion	-4.5456	locomotion
GO:0048598	embryonic morphogenesis	-5.7903	embryonic morphogenesis
GO:0050793	regulation of developmental process	-2.3885	embryonic morphogenesis
GO:0007275	multicellular organismal development	-7.9433	embryonic morphogenesis
GO:0009888	tissue development	-3.1405	embryonic morphogenesis
GO:0009887	organ morphogenesis	-4.4561	embryonic morphogenesis
GO:0051240	positive regulation of multicellular organismal process	-2.4893	embryonic morphogenesis
GO:0000904	cell morphogenesis involved in differentiation	-2.3901	embryonic morphogenesis
GO:0061458	reproductive system development	-3.9657	embryonic morphogenesis
GO:0009790	embryo development	-4.6321	embryonic morphogenesis
GO:0048608	reproductive structure development	-4.0083	embryonic morphogenesis
GO:0006928	cellular component movement	-3.9768	cellular component movement
GO:0010595	positive regulation of endothelial cell migration	-2.3951	cellular component movement
GO:0051674	localization of cell	-2.7017	cellular component movement
GO:0016477	cell migration	-3.1972	cellular component movement

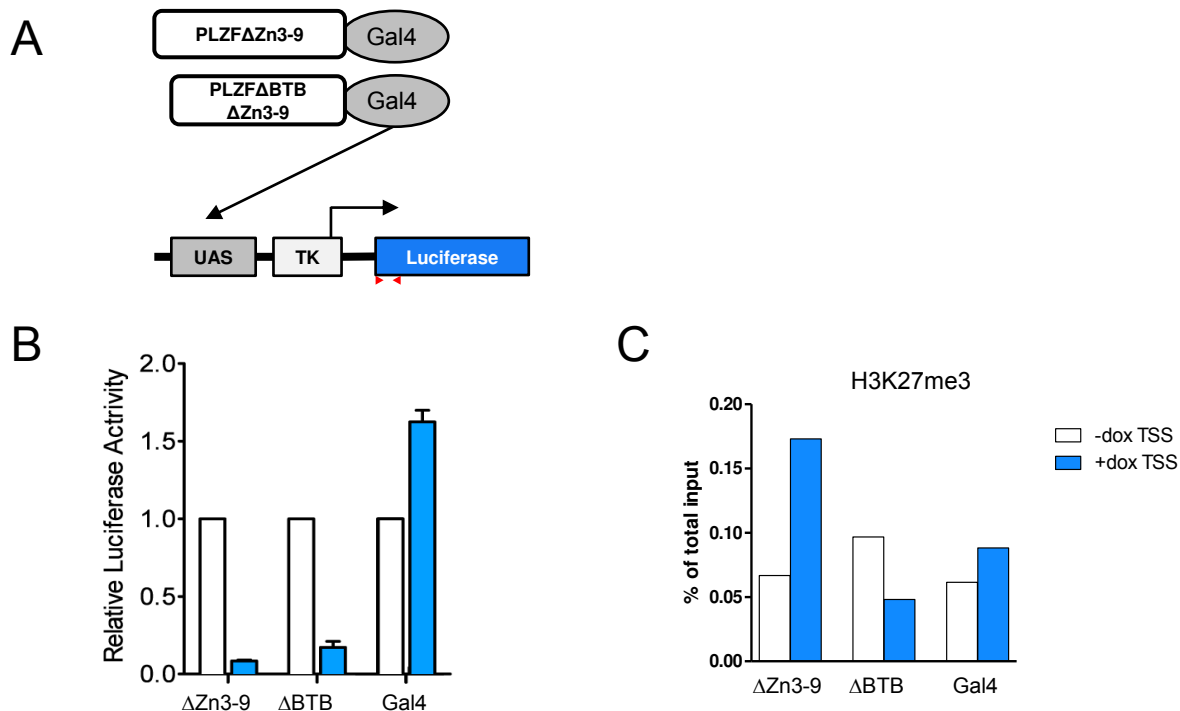


Figure S1. (A) Schematic of the luciferase reporter under the control of TK promoter, with Gal4 DNA-binding domain fused to PLZF domain. (B) Luciferase was measured 24h after dox induction on cellular clones expressing PLZFΔZn3-9Gal4 and PLZFΔBTBΔZn3-9Gal4. (C) H3K27me3 was immunoprecipitated and DNA enrichment was measured at the luciferase transgene. Amplification localisation is indicated by red arrow in A.

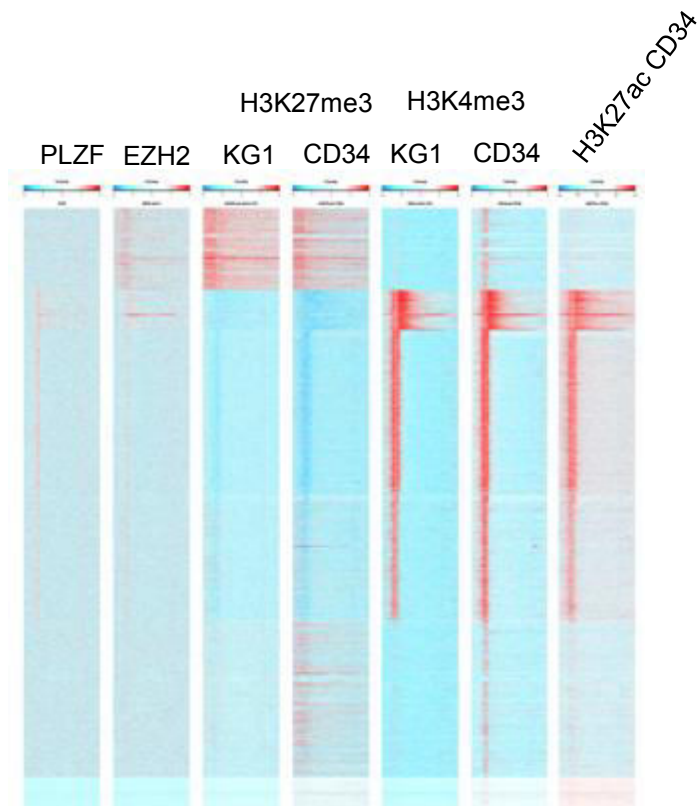


Figure S2: Comparison of epigenetic profiles of KG1 cells and CD34+ HSC on genes ranked by the *k*-mean clustering of figure 3A. Quantification is done from -2 kb of the TSS to +2 kb of the TTS. All genes were scaled to have the same length.

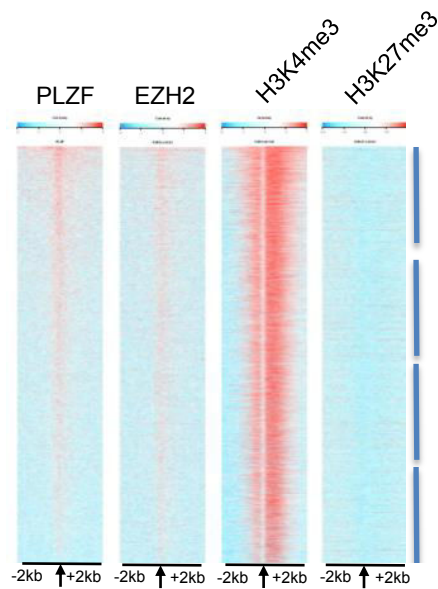


Figure S3: Supplemental to figure 5. PLZF bound regions were clustered according to total promoter PLZF enrichment levels (+/- 2kb) and heatmaps of ChIP at promoter of EZH2, H3K4me3 and H3K27me3 in KG1 cells are shown. The arrow positions the TSS. The blue lines indicate sub-groups that have been analysed in figure 5.

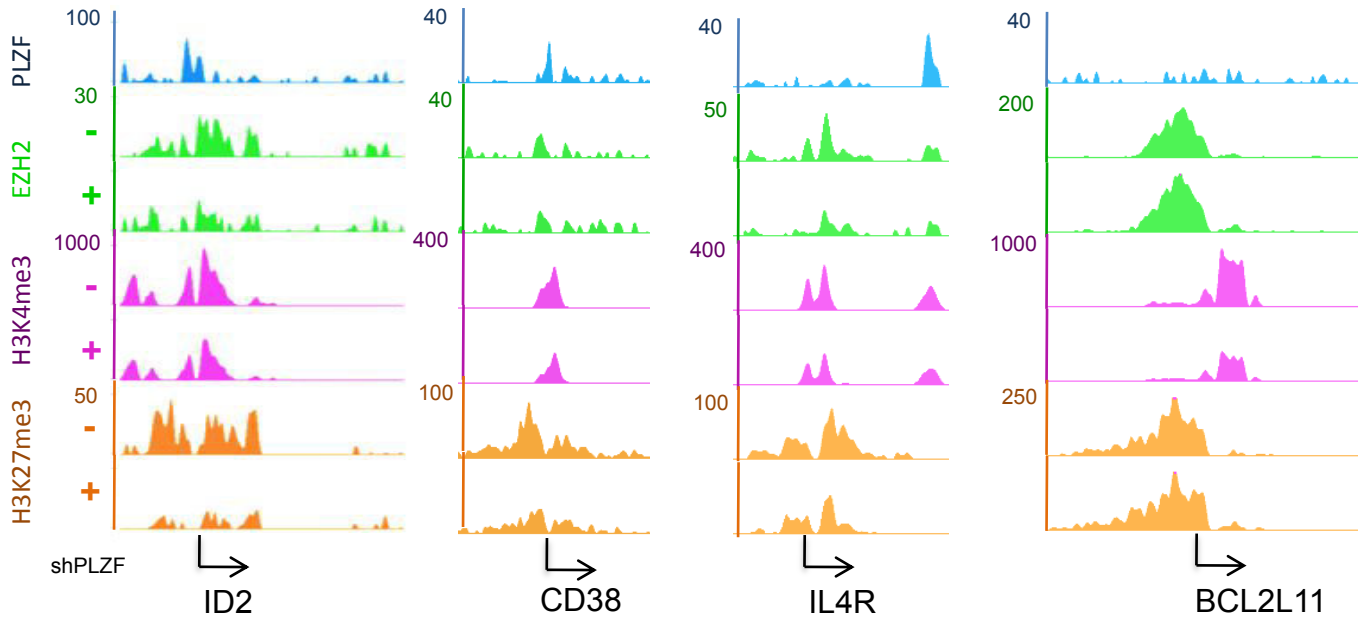


Figure S4: Supplemental to figure 6. ChIP-seq analysis of PLZF, EZH2, H3K4me3 and H3K27me3 in shMok (-) and shPLZF (+) KG1 cells. UCSC browser views of PLZF, EZH2 H3K4me3 and H3K27me3 ChIP-seq data of selected bivalent genes.

Article de Recherche 2 :

**PLZF mutation results in cell cycle
alteration and accelerated mouse
hematopoietic stem cell aging**

PLZF mutation results in cell cycle alteration and accelerated mouse hematopoietic stem cell aging

Christelle Vincent-Fabert, Nadine Platet*, Amelle Vandeveld*, Mathilde Poplineau, Myriam Koubi, Pascal Finetti, Guillaume Tiberi, Anne-Marie Imbert, François Bertucci and Estelle Duprez

* : equal contribution

Au-delà de son implication dans les leucémies aiguës promyélocyaires, PLZF joue un rôle majeur dans de nombreux processus cellulaires allant de la spermatogénèse à la formation des membres inférieurs chez la souris.

Au sein du système hématopoïétique, PLZF a d'abord été considéré comme un facteur de transcription myéloïde capable de réguler la production de granulocytes. Des études complémentaires ont ensuite démontré un rôle plus général de cette protéine dans le développement de nombreuses autres cellules hématopoïétiques.

Certaines données suggèrent que PLZF pourrait être un important régulateur des cellules souches hématopoïétiques mais aucun lien direct n'a encore été démontré.

PLZF fonctionne comme un régulateur transcriptionnel capable de reconnaître des motifs de liaison à l'ADN spécifiques et de recruter des co-régulateurs au niveau des promoteurs de ses gènes cibles. L'objectif de ce travail a été de déterminer, via l'étude de ses gènes cibles, le rôle de PLZF dans la régulation de la différenciation des cellules souches hématopoïétiques (CSH). Pour cela, un modèle murin portant une mutation naturelle pour PLZF (mutant luxoid) qui engendre une protéine tronquée et non fonctionnelle a été utilisé.

Afin de mettre en évidence les programmes transcriptionnels régulés par PLZF, des analyses de micro-array ont été réalisées sur des cellules souches sauvages ou mutantes de la moelle osseuse (cellules LSK, positives pour les marqueurs cKit et Sca et négatives pour le marqueur Lin). L'analyse supervisée montre qu'un certain nombre de gènes voit leur expression modifiée dans les cellules mutées par rapport aux cellules sauvages, ce qui souligne le rôle de PLZF dans la régulation de la transcription au sein des cellules souches hématopoïétiques. De façon intéressante, les gènes surexprimés dans les cellules mutantes sont associés à des gènes impliqués dans le cycle cellulaire et le métabolisme tandis que les gènes sous-exprimés sont impliqués dans les fonctions lymphoïdes.

Par ailleurs, la mutation de PLZF entraîne une diminution de clonogénicité et modifie l'activité des cellules souches hématopoïétiques du compartiment « Long term » (LT-HSC pour Long Term Hematopoietic Stem Cells).

Cette modification est révélée par des expériences de transplantation en séries qui montrent un épuisement des CSH PLZF mutantes et une perte de leur potentiel de différenciation B en faveur de la différenciation myéloïde. Ces résultats suggèrent donc que la mutation de PLZF

altère le phénotype des CSH en favorisant la différenciation myéloïde. La régulation du stock de CSH passe par la régulation du cycle cellulaire qui s'altère lors du vieillissement et des expériences visant à étudier les effets de la mutation de PLZF dans les CSH au niveau du cycle cellulaire ont montré que les cellules mutantes s'accumulent en phase S, ce qui souligne l'implication de PLZF dans la régulation du cycle cellulaire.

Ainsi, sur le plan phénotypique comme sur le plan moléculaire, il semble que les souris *Zbtb16^{lu/lu}* présentent les caractéristiques d'un vieillissement accéléré, suggérant ainsi l'implication de PLZF dans des mécanismes de contrôle du vieillissement cellulaire.

Afin de comprendre par quel mécanisme PLZF régule le cycle cellulaire au sein des CSH, l'expression de gènes régulés par PLZF a été étudiée à partir des données d'expression. Dans les cellules mutantes, une augmentation d'expression des gènes *Klf2* et des cyclines *A2*, *E2* et *D3* est observée en parallèle d'une diminution de l'inhibiteur du cycle cellulaire *p21*. La majorité de ces gènes, sont des cibles, au niveau génomique, de PLZF et sont enrichis en *H3K27me3*, une marque épigénétique répressive qui varie entre les cellules *Zbtb16^{lu/lu}* et sauvages. Ces données suggèrent donc que PLZF agit sur le cycle cellulaire en modifiant le paysage épigénétique des gènes régulés.

Cette étude a donc mis en évidence un nouveau rôle régulateur de PLZF au sein du compartiment souche hématopoïétique. En effet, PLZF régule le cycle cellulaire et permet le maintien du stock de LT-HSC via son rôle de facteur de transcription et son implication dans la régulation épigénétique des gènes.

PLZF mutation results in cell cycle alteration and accelerated mouse hematopoietic stem cell aging

Christelle Vincent-Fabert^{1,2,3,4}, Nadine Platet^{1,2,3,4*}, Amelle Vandeveld^{1,2,3,4*}, Mathilde
Poplineau^{1,2,3,4}, Myriam Koubi^{1,2,3,4}, Pascal Finetti^{1,2,6}, Guillaume Tiberi^{1,2,3,4}, Anne-Marie
Imbert^{1,2,5}, François Bertucci^{1,2,6} and Estelle Duprez^{1,2,3,4*}

1 Inserm, U1068, CRCM, Marseille, F-13009, France

2 Institut Paoli-Calmettes, Marseille, F-13009, France

3 Aix-Marseille Univ, UM 105, F-13284, Marseille, France

4 CNRS, UMR7258, CRCM, Marseille, F-13009, France

5 Inserm CBT-1409, CIC Biothérapies, Marseille, F-13009, France

6 Department of Medical Oncology, Institut Paoli-Calmettes, Marseille, France

* These authors contributed equally

- Correspondance : estelle.duprez@inserm.fr

Text : 4119 words

Abstract : 180 words

6 figures

37 references

KEYPOINTS

- PLZF controls HSC cell cycle
- Inactivation of PLZF accelerates HSC aging

ABSTRACT

Hematopoietic stem cells (HSCs) give rise to all blood populations due to their long-term self-renewal and multipotent differentiation capacities. Because they have to persist throughout an organism's life span, HSCs tightly regulate balance between proliferation and quiescence. Here, we investigated the role of the transcription factor, PLZF in HSC fate using the *Zbtb16*^{lu/lu} mouse model, harboring a natural spontaneous mutation that inactivates PLZF. We found that *Zbtb16*^{lu/lu} HSCs harbor a transcriptional signature associated with a loss of stemness and cell-cycle deregulation. This signature was not associated with HSC deficiency at young age. However, regenerative stress revealed that *Zbtb16*^{lu/lu} HSCs had a lineage skewing potential from lymphopoiesis towards myelopoiesis, an increase in the long-term-HSC pool and a decreased repopulation potential. Furthermore, old PLZF-mutant HSCs present an amplified aging phenotype suggesting a regulatory role for PLZF in preventing HSC aging. Finally, cell-cycle analysis revealed an important role of PLZF in G1-S transition regulation of HSCs. Our study reveals a new role for PLZF in regulating HSC aging that is linked to cell-cycle regulation and positions PLZF as a key player in controlling HSC homeostasis.

Introduction

Characteristics of stem cells are their capacity for self-renewal and multipotent differentiation, which allow them to maintain homeostasis throughout an organism's life. Stem cells are maintained in a body for a long period, where they have developed specific programs against a variety of physiological stresses in order to maintain their lifelong integrity. Hematopoietic stem cells (HSCs) are classified as being in either quiescent or cycling states, with quiescence being likely one of the mechanisms developed by the HSCs in defence against physiological stress. Under homeostatic conditions, adult HSCs are largely quiescent, in a state of dormancy that is believed to be critical for their biological role ¹. For physiological blood replenishment or to respond to hematopoietic stress, HSCs enter into the cell cycle, either committing to differentiation or re-entering the G0/G1 phases following self-renewal. The balance between HSC quiescence and proliferation has been extensively studied and has revealed to be tightly, but flexibly, regulated by a complex network of intrinsic and extrinsic mechanisms ². However, despite the identification of a growing number of molecules that regulate HSC cell cycle, the link between HSC proliferation and self-renewal is poorly understood.

As with most tissues that have a high cellular turnover, hematopoietic tissue is sensitive to aging, which results in multiple functional defects that accumulate in HSCs from older individuals ³. Aging HSCs have a reduction in their regenerative capacity and support decreased lymphoid lineage output together with increased myeloid output ^{4,5}. This age-related myeloid-biased differentiation results in a reduction of the adaptive immune system that promotes immunosenescence, and in an increase in the occurrence of myeloid malignancies ⁶. Surprisingly, the number of phenotypically-defined HSCs increases with age; while it has long been known that aging has a major impact on the frequency and cell-cycle kinetics of hematopoietic cell compartments ⁷, mechanisms underlying this phenomenon remain poorly understood.

PLZF, also known as Zbtb16, was first described as a transcriptional repressor rearranged in cases of acute promyelocytic leukemias associated with the t(11;17) translocation ⁸. Since then, PLZF has been shown to be involved in major developmental and biological processes such as spermatogenesis, hind limb formation and hematopoiesis regulation ⁹. Within the hematopoietic system, PLZF was first considered to be a myeloid transcription factor able to regulate the production of granulocytes ¹⁰, with subsequent studies highlighting its essential role in the development of multiple hematopoietic cells ¹¹⁻¹³. In addition, PLZF is emerging as

an important regulator of innate immune cell function by restraining inflammatory signalling programs^{14,15}. Together, these studies place PLZF at the centre of immune cell commitment and differentiation. PLZF functions as a transcriptional regulator that recognizes specific DNA-binding motifs and recruits co-regulator complexes to the promoter of its target genes⁹. The identification of some of its target genes has allowed a better understanding of PLZF function. For example, PLZF is directly involved in tumour suppression by its transcriptional repressive effect on the c-myc oncogene^{16,17} and this repression induces an inhibition of cellular growth in connection with quiescence, linking PLZF function to cell cycle regulation. Published data indirectly suggest that PLZF could be an important regulator of the HSC, although a direct demonstration is still lacking. To this end, we studied the role of the transcription factor PLZF in the regulation of HSC fate.

Experimental procedures

Mice

C57BL/6-CD45.1 mice were purchased from Charles River Laboratories. B6-CD45.1/CD45.2 mice were bred and maintained in the CRCM mouse facility. *Zbtb16*^{lu/lu} C57BL/6-CD45.2 mice were derived from frozen *Zbtb16*^{wt/lu} embryos purchased from Jackson Laboratories. Homozygous *Zbtb16*^{lu/lu} mutant mice were obtained by crossing *Zbtb16*^{wt/lu} heterozygous mice. The mice were genotyped using a custom TaqMan® Sample-to-SNP™ Kit and GTXpress™ Master Mix (Applied Biosystem).

Flow cytometry and cell sorting

Femurs and tibias from mice were collected and bone marrow (BM) cells were flushed in PBS 3% FCS. Spleen and thymus were also extracted and crushed gently to dissociate the spleen and thymus fractions. The lineage-positive cells were stained using the Lineage Cell Detection Kit (Miltenyi Biotec). The strategy and antibody combination are presented respectively in Figure S1B and Table S3. Flow cytometry analyses were performed using BD-LSRII SORP cytometer (BD Biosciences). Results were analysed using BD-DIVA Version 6.1.2 software (BD Bioscience). For cell sorting, lineage-negative (Lin⁻) cells were enriched using Lineage Cell Depletion Kit (Miltenyi Biotec) according to the manufacturer's instructions. Cell sorting was performed on a 6-laser, 18-detector FACS AriaII SORP (BD Biosciences).

***In vitro* cell culture**

For LT-CIC assay, Lin⁻ BM cells were isolated and seeded in 24-well plates pre-coated with 0,2% gelatin and MS5 layer cells up to 5 weeks in myelocult medium supplemented with hydrocortisone (10⁻⁴M) (Stemcell Technologies). After 3 or 5 weeks of culture on stroma cells, 9,000 hematopoietic cells were plated on 35 mm dishes in methylcellulose medium M3434 (StemCell Technologies) and colonies were scored after a week.

For single-cell CAFC assay, individual wells of flat-bottom 96-well plates, pre-coated with gelatin and MS5 cells, were seeded with single sorted HSC (KSL, Flt3⁻, CD150⁺, CD48⁻) using the single cell deposition of the AriaII SORP cell sorter. Cells were grown in myelocult medium supplemented with hydrocortisone (Stemcell Technologies) up to 5 weeks, with changing the medium every week. Cobblestone area positive wells were scored every week.

Transplantation experiments

For competitive transplantation assays, 1x10⁶ total BM cells of Zbtb16^{lu/lu} mice (C57BL/6-CD45.2) were mixed with 1x10⁶ total BM cells of WT C57BL/6 mice (CD45.1/2) and injected intravenously into 6- to 8-week-old recipient mice (C57BL/6-CD45.1), lethally irradiated at a dose of 8 Gy using the Rad source RS 2000 X-Ray irradiator.

For serial transplantation assays, 2x10⁶ total BM cells were isolated from Zbtb16^{lu/lu} or WT littermates (C57BL/6-CD45.2) and transplanted intravenously three times as previously described.

Cell cycle analysis and apoptosis

Zbtb16^{lu/lu} and WT Lin⁻ fractions were stained for HSCs (KSL, Flt3⁻, CD150⁺) and subsequently incubated with Ki67 and Dapi. For BrdU analysis, mice were injected intraperitoneally with 3 mg of BrdU, 18 hours before isolating BM cells. BM Lin⁻ cells were stained for HSCs (KSL, Flt3⁻, CD150⁺). Phases of cell cycle were analyzed by measuring BrdU- and 7-AAD-incorporation, using the FITC BrdU Flow Kit (BD Pharmingen). For myelosuppression induction, Zbtb16^{lu/lu} and WT mice were injected intraperitoneally with a single dose of 5-Fluorouracil (5-Fu) at a concentration of 150 mg/kg, 5 days before BrdU injection. For apoptosis studies, BM Lin⁻ cells were stained for HSCs (KSL, CD150⁺, CD48⁻) and resuspended in Annexin V Binding Buffer and stained with anti-Annexin V (BD Pharmingen) and with 1 µg/ml Dapi (Invitrogen) for 15 min at 4°C.

Chromatin Immunoprecipitation assay

Chromatin Immunoprecipitation (ChIP) was essentially performed as described before¹⁸. HSPC or Lin⁻ cells were treated with 1% formaldehyde (RT, 8 min), and chromatin was extracted then sonicated for 25 min (full power, 30s on 30s off) using the bath sonication system Bioruptor® (Diagenode). For ChIP using PLZF antibody, (H300; Santa Cruz), chromatin extracted from 1x10⁷ KG1 cells was used. For ChIP using H3K27me3 (Active Motif 39155) or H3K4me3 (Diagenode; C15410003), chromatin from 100 000 cells per condition was used. Quantification of the ChIPed DNA was done by quantitative PCR. Primer used are listed in Supplementary Information

Gene expression profiling

DNA-microarrays were used to define and compare the transcriptional profiles of KSL cells isolated from Zbtb16^{lu/lu} mice and WT mice (n=6). For details, see also Supplemental Experimental Procedures. Raw transcriptomic data were deposited in ArrayExpress (<http://www.ebi.ac.uk/arrayexpress/>, accession number: E-MTAB-3682).

Statistics

Statistical analyses were performed with Mann-Whitney non-parametric tests.

Results

PLZF mutation results in deregulation of transcriptional programs associated with stem cell function and cell cycle

To uncover transcriptional programs that are regulated by PLZF in the early hematopoietic compartment, we performed micro array-based gene expression analyses of wild type or PLZF-mutant (Zbtb16^{lu/lu}; Figure S1A and ref [19]) on cKit, Sca positive and Lin⁻ fraction of the bone marrow (BM) (KSLs). Supervised analysis identified 1428 genes differentially expressed including 735 and 692 genes significantly downregulated or upregulated respectively in Zbtb16^{lu/lu} KSLs (Table S1), highlighting a transcriptional function for PLZF in the stem hematopoietic compartment. Gene ontology (GO) enriched terms associated with the genes upregulated in the Zbtb16^{lu/lu} mutant group included metabolic and cell cycle related processes (Table S2); within these genes we highlighted several chromatin-associated regulators (Figure 1A). Interestingly, the top ten GO biological processes associated with

down-regulated genes were mainly involved in lymphoid functions (Table S2). These genes also encompassed already described PLZF targets such as *Id2* and *hoxA10* that are involved in stem cell maintenance (Figure 1B). Consistent with this analysis, gene set enrichment analysis (GSEA) demonstrated a negative correlation between stemness and lymphoid signatures and the *Zbtb16*^{lu/lu} genotype (Figure 1C). We also found a positive correlation between cell cycle signatures (in particular genes involved in G1-S transition) and the *Zbtb16*^{lu/lu} genotype. We then compared the expression of PLZF using qRT-PCR in the different subpopulations of mouse BM cells. *PLZF* mRNA levels were higher in the immature cells (Lin⁻ subset) than the differentiated cells (Lin⁺ subset). Among immature cells, we confirmed the expression of PLZF in phenotypically defined GMP and found significant expression of PLZF in highly LT-HSC-enriched CD150⁺ CD48⁻ KSL, suggesting a role for PLZF at early stages of murine hematopoiesis (Figure 1D). Together, these findings strongly suggest that the transcriptional activity of PLZF is required for HSC function.

The *Zbtb16*^{lu/lu} mutation reduces the frequency of clonogenic cells in the HSC compartment

Since transcriptomic analyses pointed to a possible loss of *Zbtb16*^{lu/lu} LT-HSC function, we next assessed the differentiation capacity of the *Zbtb16*^{lu/lu} immature cells in methylcellulose supplemented with cytokines. Colony assays showed a reduction of clonogenic capacity in the Lin⁻ *Zbtb16*^{lu/lu} fraction in comparison to wild type (Figure 2A, left panel). Morphology showed an increase of mature cells upon the M-CSF cytokine treatment (Figure 2A, right panel). HSC function was tested with the *in vitro* cobblestone-area forming cell (CAFC) assay onto pre-established stromal feeder cells, an assay that identifies LT-HSC activity in close to physiological conditions³. *Zbtb16*^{lu/lu} Lin⁻ BM cells showed a decrease in colony forming units after 3 or 5 weeks of co-culture compared to WT cells (Figure 2B), suggesting a LT-HSC defect in the Lin⁻ *Zbtb16*^{lu/lu} pool. To better compare LT-HSC activity between the two genotypes, we single-cell purified *Zbtb16*^{lu/lu} or WT CD48⁻ CD150⁺ LT-HSC onto stromal feeder cells and scored the cells for their ability to generate cobblestone colonies for up to 5 weeks. *Zbtb16*^{lu/lu} CD48⁻ CD150⁺ LT-HSC cells showed a reduced percentage of cells with CAFC activity that persists from 2 to 5 weeks compared to WT cells (Figure 2C) suggesting a loss of stem cell properties upon PLZF mutation.

The *Zbtb16*^{lu/lu} mutation results in aging-like phenotypes in young HSC

To assess the functional role of PLZF in HSCs, we compared hematopoiesis in WT and Zbtb16^{lu/lu} mice (Figure S1B). BM cellularity was two-fold lower in mutant bones due to their reduced size¹⁹). We observed no difference in the relative numbers of phenotypically defined progenitors or HSCs in WT and Zbtb16^{lu/lu} BM (Figure S2B-D), with the exception of the GMP population that was slightly increased in Zbtb16^{lu/lu} mice. This did not affect the frequencies of mature cells in the different hematopoietic tissues (Figure S2E-H).

To better characterize Zbtb16^{lu/lu} HSCs, we performed serial transplantation assays (Figure 3A). Donor cell analysis following serial transplantations revealed a reduction of Zbtb16^{lu/lu}-derived cells in comparison to WT-derived cells in tertiary recipients (Figure 3B). This reduction in Zbtb16^{lu/lu}-derived cells was more pronounced in the blood than in the BM, where only a 10% reduction was observed. This suggests that Zbtb16^{lu/lu} HSCs could sustain serial transplantations only with a reduced repopulation ability of mature blood cells. In order to understand if this functional defect was linked to a quantitative default in Zbtb16^{lu/lu} HSC pool, we analyzed by FACS the HSC populations after reconstitution. This showed that Zbtb16^{lu/lu} transplanted mice did not harbour a reduction but rather an expansion of the HSC pool after two transplantations (Figure 3C-D, *left panel*). In addition and in comparison with WT HSCs, Zbtb16^{lu/lu} HSCs were proportionally enriched in immature LT-HSC at the expense of the MPP1 fraction (Figure 3D, *right panel* and Figure S3A). Finally, reconstitution of common lymphoid progenitor (CLP) was compromised, while GMPs were increased in tertiary recipients of Zbtb16^{lu/lu} BM (Figure 3E). This deregulation of the progenitor pool resulted in a loss of B-cells in favour of myeloid cells in peripheral blood, BM and spleen of the Zbtb16^{lu/lu} tertiary reconstituted mice (Figure 3F and Figure S3B-C). Together, these findings suggest that mutation in PLZF alters the phenotype of HSCs through increasing the LT-HSC pool and by biasing the HSC differentiation potential towards the myeloid compartment. These observations are consistent with the GSEA analysis that revealed a loss in B-cell potential in mutant HSCs (Figure 1C). Thus, taken together, phenotypic and molecular analyses of Zbtb16^{lu/lu} HSCs suggest that Zbtb16^{lu/lu} HSCs display biological characteristics associated with accelerated aging and suggest a role for PLZF in the control of age-related mechanisms.

PLZF prevents HSC aging.

To better characterize the role of PLZF in HSC aging, we compared the HSC compartments in one-year-old Zbtb16^{lu/lu} and WT mice. First, we analysed PLZF expression in LT-HSCs.

We found that PLZF mRNA levels were decreased in LT-HSCs of old compared with young WT mice (Figure 4A). The size of the HSC compartment remained unchanged in old *Zbtb16^{lu/lu}* mice compared to their WT littermates (data not shown). However, among the HSC population, LT-HSCs were significantly expanded in aged *Zbtb16^{lu/lu}* mice compared to WT (Figure 4B). These data indicate that the decrease in PLZF activity is associated with aging. To confirm that the aging of HSC observed in the *Zbtb16^{lu/lu}* mice is due to HSC-intrinsic effects of PLZF, and not to the role of PLZF in regulating the HSC niche, we analyzed the BM of WT recipient mice transplanted with donor *Zbtb16^{lu/lu}* cells, one year after the day of transplantation (Figure 4C-D). Primary and secondary transplanted *Zbtb16^{lu/lu}* HSCs, that have aged in a WT environment, were still more enriched with the LT-HSC pool than the transplanted WT HSCs, suggesting that PLZF-mutation intrinsically affects HSC aging in an environment-independent manner. In addition, a marked loss of *Zbtb16^{lu/lu}* MPP1 was observed one year after the first transplantation (Figure 4C), and expansion of the mature myeloid over the lymphoid compartment in BM and peripheral blood PB was observed one year after the secondary transplantation (Figure 4D). Thus, these two aging characteristics were observed one transplant earlier when mice were analyzed at 52 rather than at 14 weeks after transplantation (comparison of panels 4C-D to panels 3D-F). Altogether, these recurrent and consistent observations strongly suggest that *Zbtb16^{lu/lu}* promotes premature-aging-like syndrome in mice after regenerative stress.

To further confirm age related changes in the phenotype of *Zbtb16^{lu/lu}* BM cells, we compared the competitive repopulation ability of young *Zbtb16^{lu/lu}* BM cells with old WT BM cells. Donor BMs, either two-month-old CD45.2 *Zbtb16^{lu/lu}* BM or one-year-old CD45.2 WT BM were transplanted with an equal number of CD45.1/2 competitor WT BM into lethally irradiated CD45.1 recipients (Figure 4E). The contribution of donor BM cells to hematopoiesis was monitored by FACS analysis of peripheral blood PB over 21 weeks (Figure 4F-G). We could monitor an advantage of the donor *Zbtb16^{lu/lu}* over WT competitor cells in PB at 21 weeks post-transplantation (Figure 4F). This competitive advantage was due to an elevated contribution of the *Zbtb16^{lu/lu}* to the myeloid compartment (Figure 4F) This competitive advantage of the myeloid cells detected in the blood was also observed when one-year-old WT BM competitive transplantation was analyzed (Figure 4G), confirming that young *Zbtb16^{lu/lu}* BM resembles aged BM. In addition, when BM HSC composition was analysed at 21 weeks, we observed that young *Zbtb16^{lu/lu}* cells, similarly to aged cells, contributed more to the LT-HSC compartment relatively to ST-HSC and MPP1 fractions (Figure 4H). Together, these results suggest that HSCs isolated from 2 month-old PLZF-

mutant HSCs shared biological features with HSCs isolated from one-year-old WT counterparts.

PLZF controls the S-phase of HSCs

Given that regulating cell cycle entry is a way to preserve HSC integrity and that PLZF is known to control the cell cycle ²⁰, we analyzed the effect of the *Zbtb16*^{lu/lu} mutation in HSC cell cycle regulation. The KSL fraction was stained with a combination of Ki67 and DAPI to distinguish immature cells in G0, G1 and S-G2/M cell cycle phases. We observed an increased proportion of the *Zbtb16*^{lu/lu} HSPC CD150⁺ and MPP2 in S-G2/M phases that could reflect a higher proliferative state of *Zbtb16*^{lu/lu} HSCs (Figure 5A). To better characterise the accumulation in S-G2/M phase of *Zbtb16*^{lu/lu} HSCs, we performed BrdU injection and measured BrdU incorporation (Figure S4A). Analysis of the mice 18-hours after BrdU injection revealed a significant increase of BrdU⁺ cells in the HSPC and HSPC CD150⁺ fractions of the *Zbtb16*^{lu/lu}, relative to WT BM (Figure 5B), reflecting a higher sensitivity and BrdU incorporation in young *Zbtb16*^{lu/lu} HSCs. BrdU incorporation assays, coupled with 7-AAD staining, demonstrated an accumulation of the *Zbtb16*^{lu/lu} HSPC CD150⁺ cells in S-phase (Figure 5C; Figure S4A), this was also the case of the *Zbtb16*^{lu/lu} MPP2 (Figure S4B-C). Old HSCs isolated from WT animals are characterized by an overall decline in cell cycle activity, with old HSCs undergoing fewer cell divisions ²¹. We analysed the cell cycle of one-year-old *Zbtb16*^{lu/lu} HSCs. The old WT or *Zbtb16*^{lu/lu} HSCs were found more in the G0/G1 phase than their young counterparts (Figure 5C-D), confirming the decrease of dividing HSCs with age in response to BrdU injection. Old *Zbtb16*^{lu/lu} HSCs were significantly more frequently in the G0/G1 phases than old WT HSCs, witnessing the pronounced aged-like phenotype in *Zbtb16*^{lu/lu} HSCs (Figure 5D). This also suggests that regulation of cell-cycle checkpoint mechanisms is different in old versus young HSCs. To test whether PLZF mutation affects cell cycle of HSCs during hematopoietic recovery, we induced myelosuppression by 5-FU injection and analyzed HSC cell cycle status. Results showed that after myelosuppression, *Zbtb16*^{lu/lu} HSCs respond by cycling more than their WT counterparts (Figure 5E).

As we did not detect an increase of the HSC compartment at steady state in *Zbtb16*^{lu/lu} mice, we measured the level of apoptosis in *Zbtb16*^{lu/lu} HSPC CD150⁺ cells, where we detected a noticeable increase of apoptosis in *Zbtb16*^{lu/lu} HSPC CD150⁺ cells (Figure 5F) that could account for the limited accumulation of *Zbtb16*^{lu/lu} HSCs. Taken together, these data indicate that PLZF controls HSC cell cycle activity and is involved in HSC survival of young mice.

PLZF modifies chromatin landscape at its target genes.

PLZF is involved in epigenetic regulation at its target genes ²². To characterise the effect of PLZF in epigenetic regulation of the HSC, we first validated the effect of Zbtb16^{lu/lu} mutation on the expression of selected target genes in the LT-HSC populations. Quantitative PCR analysis confirmed that Zbtb16^{lu/lu} LT-HSC presented an increase in expression level of Klf2 and mitotic cyclin A2, E2 and D3 and a decrease in expression of the cell-cycle inhibitor p21, as compared with WT (Figure 6A). Among the genes that were deregulated by PLZF mutation, we determined genes that were bound by PLZF at the genomic level by chromatin immunoprecipitation. Klf2, Cyclin D3 and Cyclin E2 were bound by PLZF (Figure 6B). As these genes are upregulated upon PLZF mutation, these genes are likely repressed by PLZF in WT. We also identified Id2 down-regulated upon Zbtb16^{lu/lu}, as a direct target of PLZF. We showed that PLZF target genes were characterised mainly by a variation level of H3K27me3 upon PLZF mutation suggesting that PLZF influences the H3K27me3 mark at its target gene promoters. These data suggest that PLZF specifically restrains LT-HSC cell cycle by modifying chromatin landscape at cell cycle regulator genes.

Discussion

The study presented here reveals an important and unsuspected role for the transcription factor PLZF in HSC aging, and provides important insight into the link between cell cycle regulation and hematopoiesis changes over the lifespan of an individual. We showed that the luxoid mutation that leads to a truncated form of PLZF results in deregulation of HSC cell cycle status in young mice, characterized by an increase of the S phase at the expense of the G0/G1 phase at steady state or after myelosuppression. This cell cycle deregulation was associated with a deregulation of c-Myc, Cdkn1a, and cyclins A2, E2 and D3; all cell-cycle regulators that are important for HSC integrity ². This effect on proliferation and on expression of mitotic genes is consistent with previous studies on cell lines, showing that overexpression of PLZF controlled G₁/S and G₂/M checkpoints of the cell cycle ^{23, 24}. In the previous studies, PLZF control of the cell cycle was ascribed to its direct binding of the cMyc or cyclin A promoters ²⁴ and to its interaction with Retinoblastoma protein, inducing the inhibition of E2F pathway ²⁵. Here, we identify a critical role of PLZF in controlling HSC cell cycle by directly regulating the promoter of cyclins involved in S phase progression.

Notably, this increasing effect on S-phase had no effect on the number or function of young HSCs, as young PLZF-mutant hematopoiesis was normal. One explanation could be that HSC homeostasis was maintained due to space constraints that promote the concomitant induction of apoptosis observed in PLZF-mutant HSC. This increase in apoptotic PLZF-mutant HSCs, was previously observed in the germ line ^{26,19}, where deletion of PLZF increased germ cell proliferative activity in early postnatal testes, increased cell death and tubule degeneration at later stages, suggesting an important role for PLZF in pluripotency. Although genetic and differentiation programs are different within the hematopoietic system, similarities exist and can lead to speculate that PLZF prevents hematopoietic stem cell exhaustion.

Since quiescence is believed to be a key mechanism by which HSCs preserve their self-renewal properties, one can assume that mutants that promote HSC proliferation would lead to a loss of HSC functions. For example, mutant mouse models that result in sustained hyperproliferation of HSCs display a severe hematopoietic defect, (for review ²⁷). However, in the case of mutant PLZF that results in increased HSC proliferation, a clear hematopoietic defect is not as straightforward. Our *in vitro* studies show a loss of clonogenic activity in the Zbtb16^{lu/lu} HSC compartment, suggesting that Zbtb16^{lu/lu} HSCs are functionally compromised. However, transplantation experiments suggest that young PLZF-mutant HSCs, despite their activation, maintain their self-renewal potential as assessed by transplantation experiments. This paradoxical phenotype is not without precedence and was reported for Myb- and Egr1-deficient HSCs that displayed an increase in cell cycle activity but maintained their self-renewal potential ^{28,29}. Interestingly, PLZF shares some targets with Erg1, such as Cdkn1a or Gfi1 ²⁹. The transcription factor PLZF could contribute to balance self-renewal and differentiation through restricting proliferation, providing new insights into how PLZF regulates HSC expansion.

Activated PLZF-mutant HSCs eventually harbour HSC deficiency linked to aging, where we showed that after serial transplantations, Zbtb16^{lu/lu} HSCs lose their self-renewal potential, have a skewed advantage towards myeloid differentiation at the expense of the lymphoid lineage and accumulate in the bone marrow. All these characteristics are those observed in HSCs isolated from aging animals ^{6, 30}, which implies an intrinsic role for PLZF in protecting HSC over time from phenotypic and functional changes associated with aging. Additionally, we demonstrated that one-year-old Zbtb16^{lu/lu} HSCs have more pronounced age-like phenotypes than WT HSCs, suggesting that alteration of PLZF regulatory pathway promotes HSC decline, a phenotype linked to HSC aging. As PLZF is a key factor for HSC cell cycle regulation, it is likely that control of cell cycle progression is a mechanism used by PLZF to

prevent HSC aging. Additionally, transcriptomic analysis revealed that mutation in PLZF deregulates biological pathways that are linked to promoting aging. Among them, deregulation of nucleic acid and protein catabolic processes by PLZF in HSC is interesting as elevation in cyclin D3 level was linked to aging by promoting age specific translational dysfunction in old liver³¹.

Although PLZF is described as a tumour suppressor³², PLZF mutant mice when older display no increased myeloproliferative disorders associated with aging of the HSC compartment, when compared to wild type, which may be due to the requirement of PLZF in the development of myeloid malignancy³³. We show that PLZF can be associated with active or repressed genes, a dual transcriptional activity that can explain the versatile functions of PLZF³⁴. How PLZF achieves this dual transcriptional activity is at present unclear, although the differential recruitment of epigenetic regulators by PLZF, like the Polycomb Group, likely plays a major determining role. Post-translational modification of PLZF or its rearrangement with RARA change its transcriptional activity through differential epigenetic co-factor recruitment^{35, 18}. As epigenetic processes have been shown to regulate age-dependent functional changes^{36, 37}, it will be important to identify the epigenetic cofactors of PLZF that have the ability to regulate gene expression during aging.

In conclusion, we describe a novel function for the transcription factor PLZF in HSCs through regulating the cell cycle and preventing their aging.

Acknowledgments

This work was supported by the Fondation de France 201200029166, by the Fondation pour la Recherche Médicale (FRM) grant INE201000518534, and by the Fondation ARC pour la recherche sur le cancer. CVF was supported by the “Fondation ARC pour la recherche sur le cancer” and the Cancéropole PACA. AV was supported by the Ligue Nationale Contre le Cancer. MP was supported by ITMO-cancer. GT was supported by the FRM. We are grateful to ML Thibult and F Mallet for assistance with the use of the cytometry and cell sorting facility; F Bardin, R Castellano and L Pouyet for help with irradiated mice injections; JC Orsoni and P Gibier for assistance in animal facility; A Saurin & C Chabannon for critically reading the manuscript.

Authorship contributions

All authors read and approved the final manuscript. CVF designed the study, performed

experiments, analysed the data. NP, AV, MP, MK and AMI performed experiments and analysed the data; FP, GT and FB performed transcriptome analysis. ED initiated, directed and funded the project; ED and CVF actively wrote the paper.

Conflict-of-interest disclosure:

The authors declare no competing financial interests

References

Figure legends

Figure 1. Gene expression profiling of $Zbtb16^{lu/lu}$ HSCs shows loss of LT-HSC potential and cell cycle deregulation. KSLs from $Zbtb16^{lu/lu}$ and WT mice were subjected to gene expression analysis (A) Selection of genes upregulated in $Zbtb16^{lu/lu}$ KSLs (B) Selection of genes down regulated in $Zbtb16^{lu/lu}$ KSLs (C) Gene set enrichment analysis (GSEA) of $Zbtb16^{lu/lu}$ KSLs compared to WT KSLs (D) Expression of PLZF mRNA was measured by qRT-PCR in sorted bone marrow (BM) subpopulations from 8-week-old WT mice. Lineage positive cells (Lin⁺); Lineage negative cells (Lin⁻); Lin⁻, cKit⁺, Sca1⁻ (LK); Lin⁻, Sca1⁺, cKit⁺ (KSL); LK CD16/32⁺ CD34⁺ (GMP) and KSL CD150⁺ Flt3⁻ CD48⁻ (LT-HSC).

Figure 2. $Zbtb16^{lu/lu}$ HSCs harbour defects in colony formation. (A) Colony assay for Lin⁻ BM cells from WT and $Zbtb16^{lu/lu}$ mice. Left; number of colonies obtained from $Zbtb16^{lu/lu}$ or WT Lin⁻ BM cells after plating on methyl cellulose supplemented with indicated cytokines. Results represents the mean of three independent platings *, P<0.05; ***, P<0.001 Right; Representative May Grunwald Giemsa staining of a colony 7 days after plating with MCSF. (B) Number of colonies obtained from $Zbtb16^{lu/lu}$ or WT Lin⁻ BM cells after plating on methyl cellulose following growth on MS5 layer cells for 3 and 5 weeks. Shown is the mean of 3 independents culture on MS5 layer. **, P<0.01 (C) $Zbtb16^{lu/lu}$ or WT LT-HSC CD48⁻ CD150⁺ cells were isolated and seeded individually onto MS5 cells and scored weekly for the presence of cobblestone areas. Data represents a minimum of 400 wells per genotype from a total of 5 mice/ genotype.

Figure 3. PLZF mutation promotes aging-like phenotypes. (A) Serial transplantation assay. Donor BM CD45.2 $Zbtb16^{lu/lu}$ or WT cells were injected into lethally irradiated CD45.1 recipient mice and were serially transplanted. (B) Analyses of CD45.2⁺ donor cells in BM, Spleen, Thymus and Blood following each transplantation. (C) Representative FACS analyses and (D) quantitative and statistical analyses of long-term reconstitution in KSL and stem cell compartments post secondary transplant. (E) Quantitative analysis of long-term reconstitution in progenitor cell compartment post tertiary transplant. (F) Quantitative analysis of mature compartments in BM and PB 14 weeks post tertiary transplant. Bar graphs represent mean +/- SEM. Data is based on two (primary, second and third transplants) experimental repeats with 6 recipient mice per group (n=12). *, P<0.05; **, P< 0.01 ***, P<0.001.

Figure 4. PLZF prevents HSC aging.

(A) Expression of *Plzf* mRNA was measured by qRT-PCR in young (2 months) and old (1 year) WT LT-HSC. Results are presented as relative to *Rpl6* expression level. (B) Quantitative analysis of the HSC compartment in old *Zbtb16^{lu/lu}* and WT mice (n=5). (C-D) Analysis of hematopoietic compartments one-year post primary and secondary transplants (n=4-5). (C) Percentage of LT-HSC, ST-HSC and MPP1 among CD45.2⁺ HSC one-year post primary transplant. (D) Distribution of CD45.2⁺ Myeloid and CD45.2⁺ B-cell in BM and PB one year post secondary transplant (n=4). Data are represented as a mean +SEM. *, P<0.05; **, P< 0.01. (E-H) Competitive transplantation assays. (E) Donor CD45.2 young *Zbtb16^{lu/lu}* or WT aged BM cells were transplanted with an equal number of CD45.1/2 competitor WT BM cells into lethally irradiated CD45.1 recipients. (F-G) Monitoring of total donor-derived cells in peripheral blood of recipients and contribution of *Zbtb16^{lu/lu}* (F) or old WT (G) BM to myeloid (CD11b) and B (B220) lineages (n=6). Data are represented as a mean +/-SEM. *, P<0.05; **, P< 0.01. (H) quantitative and statistical analyses of long-term reconstitution in stem cell compartments 21 weeks post transplantation. Bar graphs represent mean +/- SEM. (n=6).

Figure 5. PLZF controls the S-phase of HSC. (A) Distribution of the HSPC CD150⁺ (left) and MPP2 cell fractions in G0, G1 and S-G2/M phases. Bar graphs represent average +/- SEM. Data is collected over three independent experiments (n=7). (B) Percentage of BrdU⁺ cells 16 hrs after injection in young *Zbtb16^{lu/lu}* and WT mice. Analyses of HSPC, HSPC CD150⁺, MPP2 and LK compartments are presented. Each symbol represents a mouse (C-D) Distribution of *Zbtb16^{lu/lu}* and WT HSPC CD150⁺, from young mice (C) and old mice (D), in G1, S and G2/M phases after BrdU injection. Bar graphs represent average +/- SEM. Data is collected over two independent experiments (n=7). (E) Cell cycle analysis of *Zbtb16^{lu/lu}* and WT HSPC CD150⁺ 6 days after myelosuppression obtained by a single dose of 5-FU (150 mg/kg) (n=5) Bar graphs represent average +/- SEM. Data is collected over two independent experiments (n=5).*, P<0.05; **, P< 0.01. (F) Analysis of apoptosis in *Zbtb16^{lu/lu}* and WT HSPC CD150⁺ from 8-10 week old mice. Each symbol represents a mouse and horizontal bars indicate the mean. ***, P< 0.01

Figure 6. PLZF directly regulates cell cycle genes. (A) q-PCR analysis of cell-cycle-related gene expression in *Zbtb16^{lu/lu}* and WT LT-HSCs sorted from young mice. Results are presented as relative to the lowest expression level. (B) PLZF ChIP analysis of PLZF target genes in a myeloid cell line (C) H3K27me3 and H3K4me3 ChIP analysis of *Id2*, *Klf2*, *Ccne2*, *Ccnd3* and *Cdkn1a* genes on chromatin fragments prepared from WT and *Zbtb16^{lu/lu}* HSPCs. Data are expressed as percentage of enrichment relative to *Gapdh*.

References

1. Wilson A, Laurenti E, Oser G, et al. Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell*. 2008;135:1118-1129.
2. Pietras EM, Warr MR, Passegue E. Cell cycle regulation in hematopoietic stem cells. *The Journal of cell biology*. 2011;195:709-720.
3. Dykstra B, Olthof S, Schreuder J, Ritsema M, de Haan G. Clonal analysis reveals multiple functional defects of aged murine hematopoietic stem cells. *J Exp Med*. 2011;208:2691-2703.
4. Sudo K, Ema H, Morita Y, Nakauchi H. Age-associated characteristics of murine

- hematopoietic stem cells. *The Journal of experimental medicine*. 2000;192:1273-1280.
5. Rossi DJ, Bryder D, Zahn JM, et al. Cell intrinsic alterations underlie hematopoietic stem cell aging. *Proc Natl Acad Sci U S A*. 2005;102:9194-9199.
6. Geiger H, de Haan G, Florian MC. The ageing haematopoietic stem cell compartment. *Nature reviews Immunology*. 2013;13:376-389.
7. de Haan G, Nijhof W, Van Zant G. Mouse strain-dependent changes in frequency and proliferation of hematopoietic stem cells during aging: correlation between lifespan and cycling activity. *Blood*. 1997;89:1543-1550.
8. Chen Z, Brand NJ, Chen A, et al. Fusion between a novel Kruppel-like zinc finger gene and the retinoic acid receptor-alpha locus due to a variant t(11;17) translocation associated with acute promyelocytic leukaemia. *Embo J*. 1993;12:1161-1167.
9. Suliman BA, Xu D, Williams BR. The promyelocytic leukemia zinc finger protein: two decades of molecular oncology. *Front Oncol*. 2012;2:74.
10. Melnick A, Licht JD. Deconstructing a disease: RARalpha, its fusion partners, and their roles in the pathogenesis of acute promyelocytic leukemia. *Blood*. 1999;93:3167-3215.
11. Savage AK, Constantinides MG, Han J, et al. The transcription factor PLZF directs the effector program of the NKT cell lineage. *Immunity*. 2008;29:391-403.
12. Labbaye C, Spinello I, Quaranta MT, et al. A three-step pathway comprising PLZF/miR-146a/CXCR4 controls megakaryopoiesis. *Nat Cell Biol*. 2008;10:788-801.
13. Ozato K. PLZF outreach: a finger in interferon's pie. *Immunity*. 2009;30:757-758.
14. Sadler AJ, Rossello FJ, Yu L, et al. BTB-ZF transcriptional regulator PLZF modifies chromatin to restrain inflammatory signaling programs. *Proc Natl Acad Sci U S A*. 2015;112:1535-1540.
15. Sadler AJ, Suliman BA, Yu L, et al. The acetyltransferase HAT1 moderates the NF-kappaB response by regulating the transcription factor PLZF. *Nat Commun*. 2015;6:6795.
16. McConnell MJ, Chevallier N, Berkofsky-Fessler W, et al. Growth suppression by acute promyelocytic leukemia-associated protein PLZF is mediated by repression of c-myc expression. *Mol Cell Biol*. 2003;23:9375-9388.
17. Rice KL, Hormaeche I, Doulatov S, et al. Comprehensive genomic screens identify a role for PLZF-RARalpha as a positive regulator of cell proliferation via direct regulation of c-MYC. *Blood*. 2009;114:5499-5511.
18. Boukarabila H, Saurin AJ, Batsche E, et al. The PRC1 Polycomb group complex interacts with PLZF/RARA to mediate leukemic transformation. *Genes Dev*. 2009;23:1195-1206.
19. Buaas FW, Kirsh AL, Sharma M, et al. Plzf is required in adult male germ cells for stem cell self-renewal. *Nat Genet*. 2004;36:647-652.
20. Costoya JA, Pandolfi PP. The role of promyelocytic leukemia zinc finger and promyelocytic leukemia in leukemogenesis and development. *Curr Opin Hematol*. 2001;8:212-217.
21. Janzen V, Forkert R, Fleming HE, et al. Stem-cell ageing modified by the cyclin-dependent kinase inhibitor p16INK4a. *Nature*. 2006;443:421-426.
22. Puszyk W, Down T, Grimwade D, et al. The epigenetic regulator PLZF represses L1 retrotransposition in germ and progenitor cells. *EMBO J*. 2013;32:1941-1952.
23. Shakhovich R, Yeyati PL, Ivins S, et al. The promyelocytic leukemia zinc finger protein affects myeloid cell growth, differentiation, and apoptosis. *Mol Cell Biol*. 1998;18:5533-5545.
24. Yeyati PL, Shakhovich R, Boterashvili S, et al. Leukemia translocation protein PLZF inhibits cell growth and expression of cyclin A. *Oncogene*. 1999;18:925-934.

25. Petrie K, Guidez F, Zhu J, et al. Retinoblastoma protein and the leukemia-associated PLZF transcription factor interact to repress target gene promoters. *Oncogene*. 2008;27:5260-5266.
26. Costoya JA, Hobbs RM, Barna M, et al. Essential role of Plzf in maintenance of spermatogonial stem cells. *Nat Genet*. 2004;36:653-659.
27. Rossi L, Lin KK, Boles NC, et al. Less is more: unveiling the functional core of hematopoietic stem cells through knockout mice. *Cell Stem Cell*. 2012;11:302-317.
28. Sandberg ML, Sutton SE, Pletcher MT, et al. c-Myb and p300 regulate hematopoietic stem cell proliferation and differentiation. *Developmental cell*. 2005;8:153-166.
29. Min IM, Pietramaggiori G, Kim FS, Passegue E, Stevenson KE, Wagers AJ. The transcription factor EGR1 controls both the proliferation and localization of hematopoietic stem cells. *Cell Stem Cell*. 2008;2:380-391.
30. Beerman I, Bhattacharya D, Zandi S, et al. Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. *Proceedings of the National Academy of Sciences of the United States of America*. 2010;107:5465-5470.
31. Jin J, Wang GL, Salisbury E, Timchenko L, Timchenko NA. GSK3beta-cyclin D3-CUGBP1-eIF2 pathway in aging and in myotonic dystrophy. *Cell Cycle*. 2009;8:2356-2359.
32. Hobbs RM, Pandolfi PP. Shape-shifting and tumor suppression by PLZF. *Oncotarget*. 2010;1:3-5.
33. Ono R, Masuya M, Nakajima H, et al. Plzf drives MLL-fusion-mediated leukemogenesis specifically in long-term hematopoietic stem cells. *Blood*. 2013;122:1271-1283.
34. Kolesnichenko M, Vogt PK. Understanding PLZF: two transcriptional targets, REDD1 and smooth muscle alpha-actin, define new questions in growth control, senescence, self-renewal and tumor suppression. *Cell Cycle*. 2011;10:771-775.
35. Guidez F, Howell L, Isalan M, et al. Histone acetyltransferase activity of p300 is required for transcriptional repression by the promyelocytic leukemia zinc finger protein. *Mol Cell Biol*. 2005;25:5552-5566.
36. Florian MC, Dorr K, Niebel A, et al. Cdc42 activity regulates hematopoietic stem cell aging and rejuvenation. *Cell stem cell*. 2012;10:520-530.
37. Sun D, Luo M, Jeong M, et al. Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal. *Cell Stem Cell*. 2014;14:673-688.

Figure 1

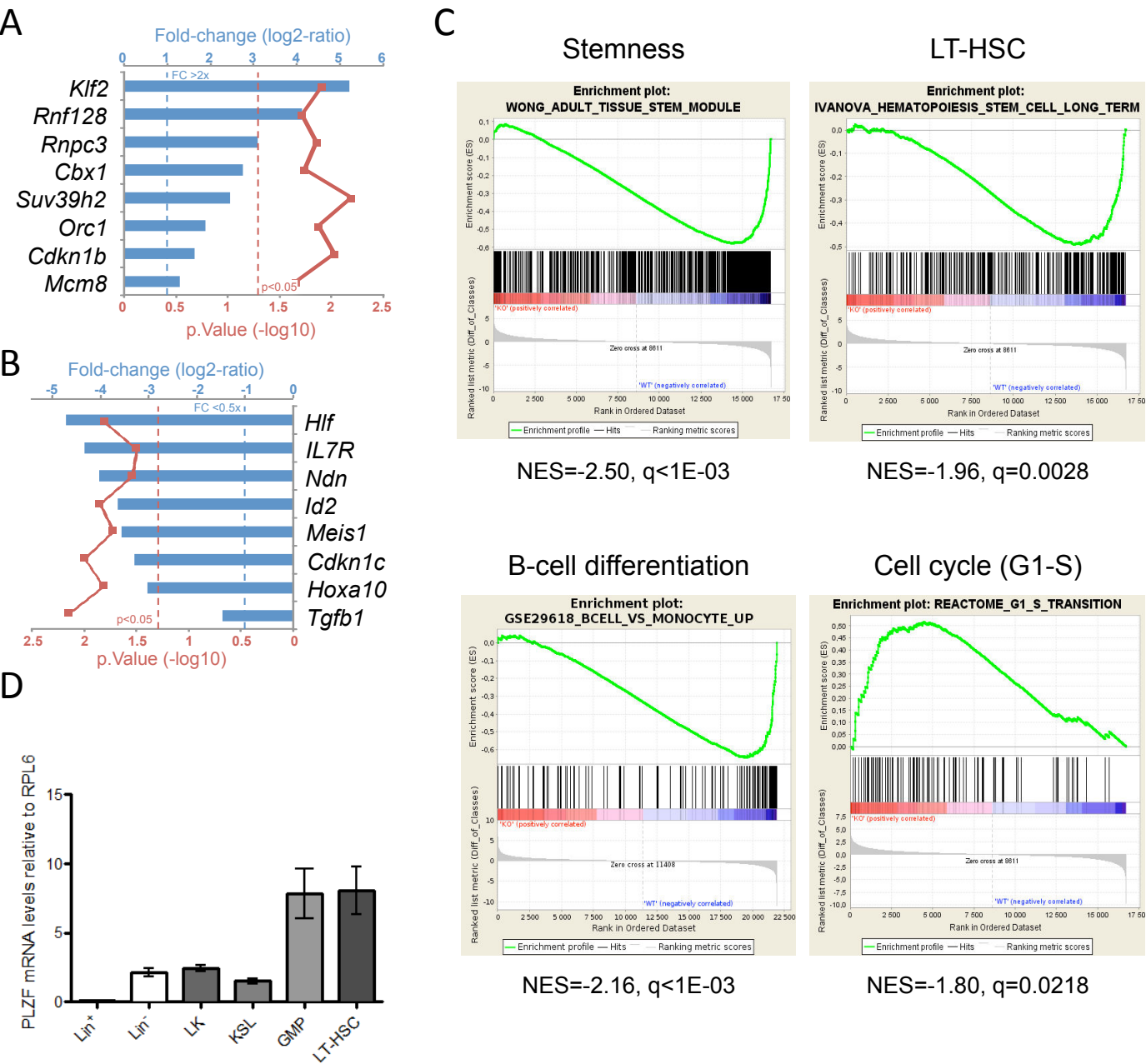
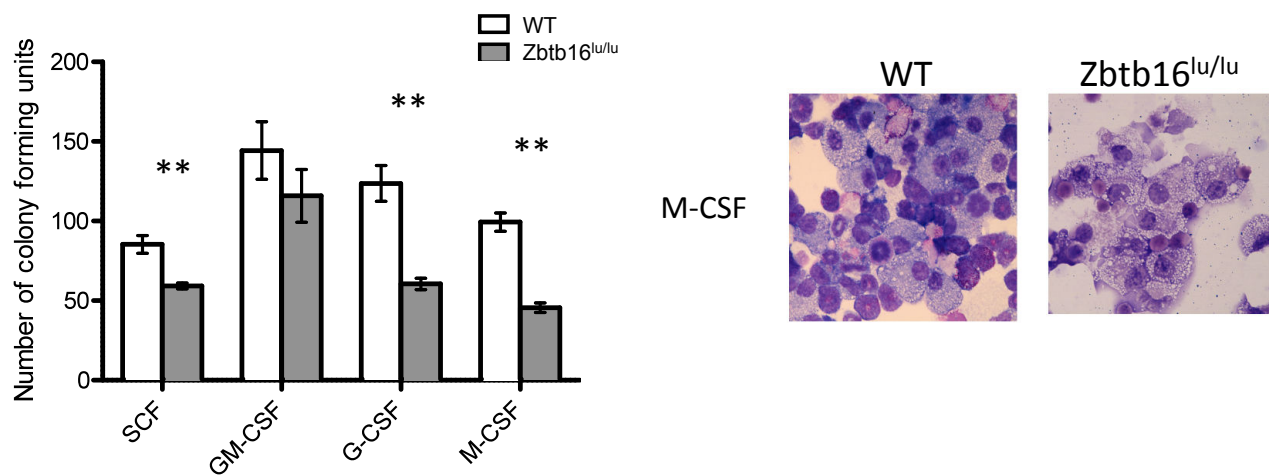
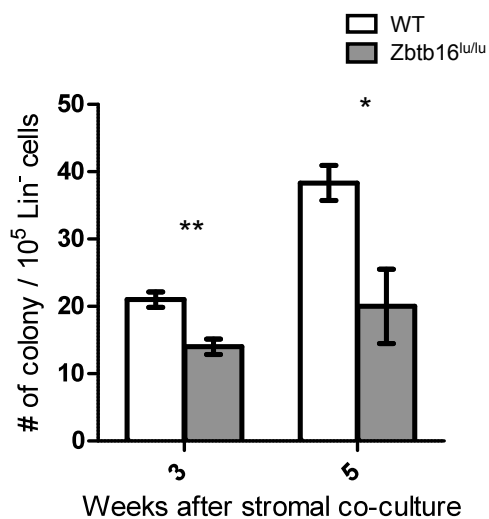


Figure 2

A



B



C

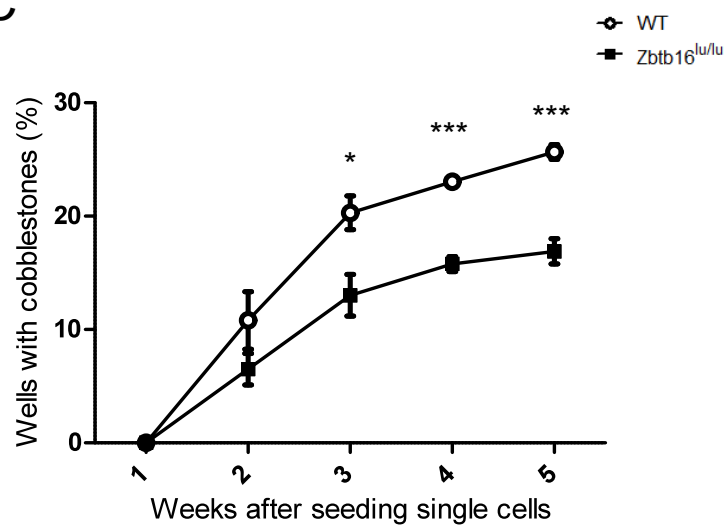


Figure 3

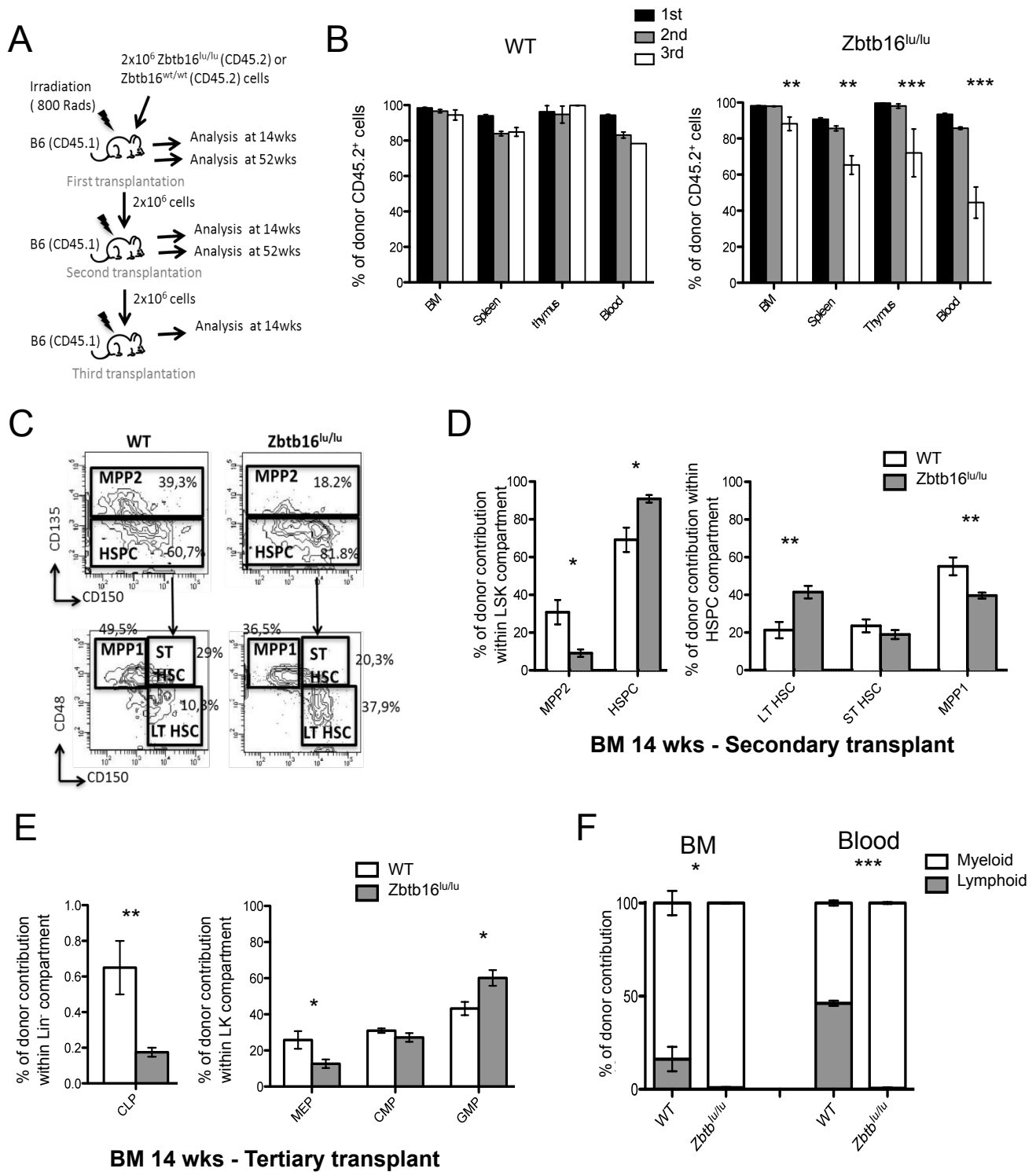


Figure 4

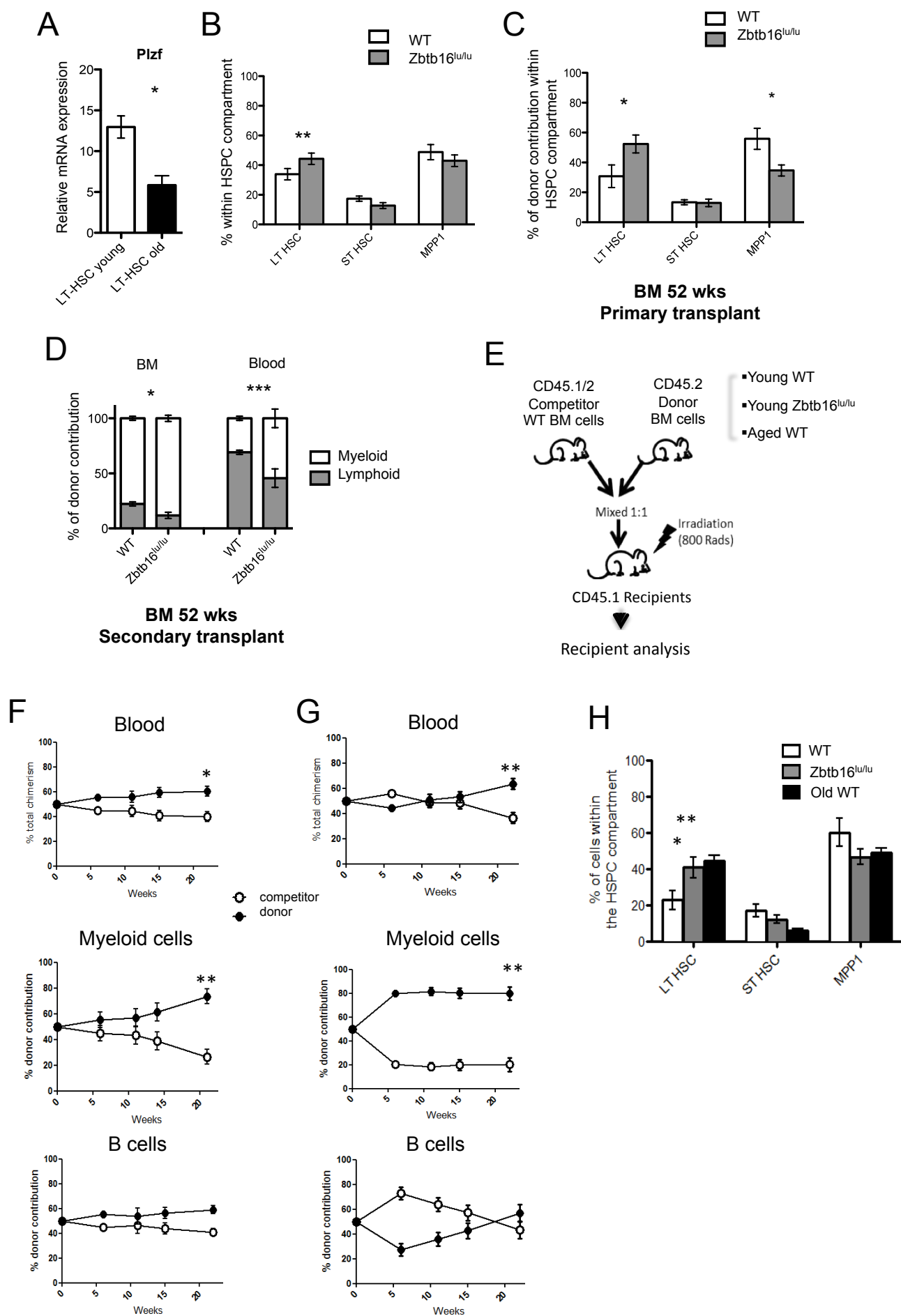


Figure 5

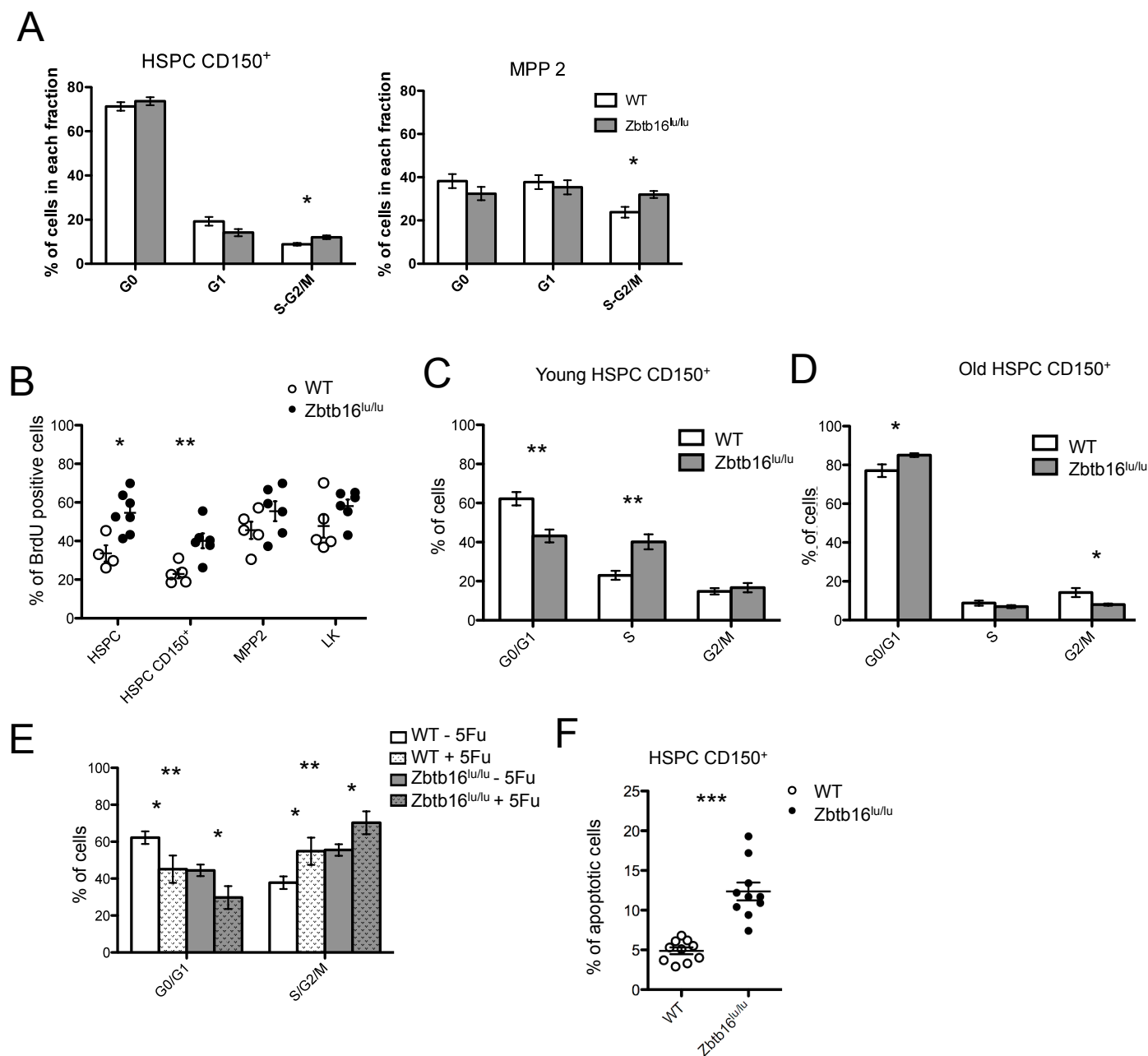
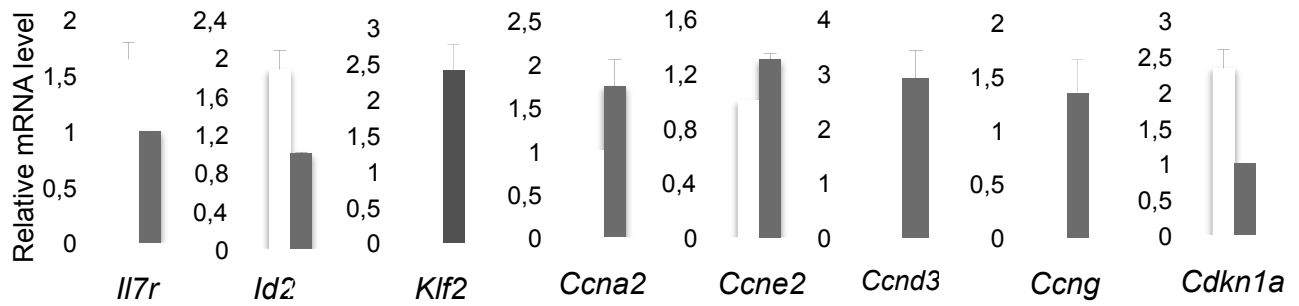


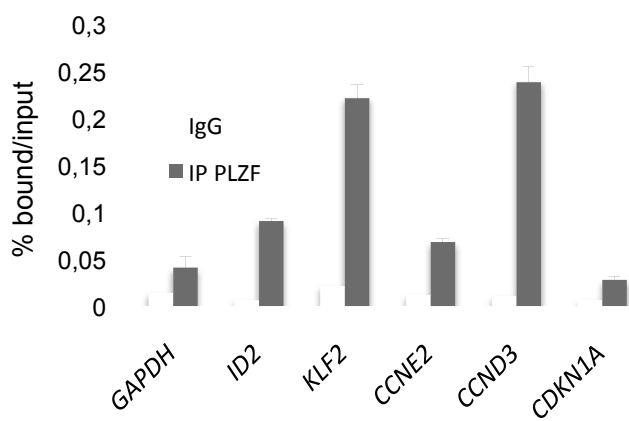
Figure 6

WT
■ Zbtb16^{lu/lu}

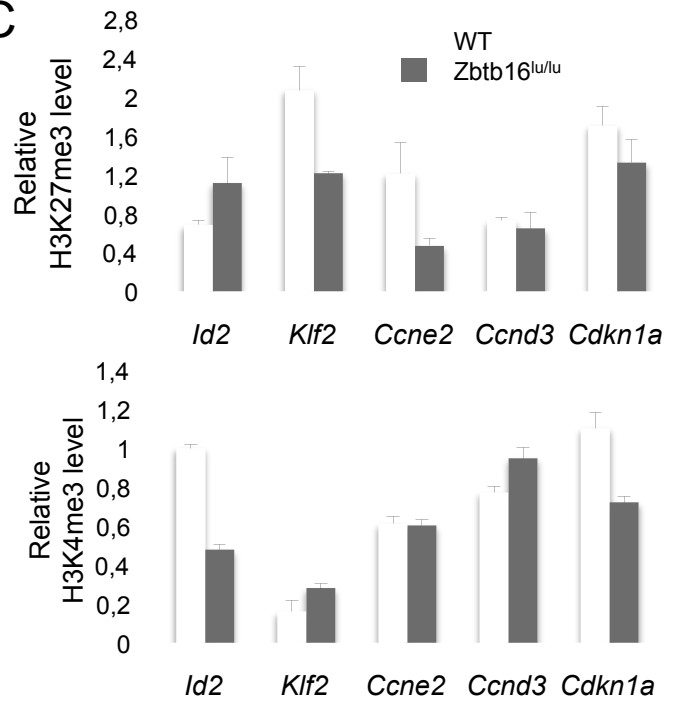
A



B



C



Supplemental Experimental Procedures

Gene expression profiling

DNA-microarrays were used to define and compare the transcriptional profiles of KSL cells isolated from *Zbtb16*^{lu/lu} mice (n=6) and WT mice (n=6). Total RNA was extracted from sorted KSL cells using the RNeasy Micro Kit (Qiagen) and quality was tested using an Agilent Bioanalyzer. RNA from 3 mice of the same genotype was pooled, leading to duplicate RNAs (2 pools) for each phenotype. Experiments were done as recommended by the manufacturer (Affymetrix, Santa Clara) from 100 ng RNA for each sample using the Affymetrix GeneChip Mouse Genome 430-2.0 arrays. Expression data were analysed by the GC-RMA (Robust Multichip Average) method in R using Bioconductor and associated packages. Filtered (14,573 genes) and log2-transformed data were submitted to supervised analysis comparing expression profiles from 2 groups: the 2 pooled KSL cell samples from *Zbtb16*^{lu/lu} mice *versus* the 2 pooled KSL cell samples from WT mice. Supervised analysis retained the genes differentially expressed between the two groups with a significance threshold of $p < 0.05$ (Student's t test) with False Discovery Rate (FDR < 25%) correction, and a fold change (FC) of mean expression levels superior or equal to $|2|$. To help in the interpretation, the lists of differential genes were interrogated using the Database for Annotation, Visualization and Integrated Discovery (DAVID: <http://david.abcc.ncifcrf.gov>). We also determined if gene signatures from the C2 and C7 gene sets of the MSigDB database: <http://www.broadinstitute.org/gsea/msigdb/index.jsp>) were overrepresented in one group using the gene set enrichment analysis (GSEA) algorithm (Subramanian et al., 2005) with the following parameters: 1000 gene sets permutations, log2 ratio as metrics for ranking genes, significance threshold $p < 0.05$, and FDR < 25%. Raw transcriptomic data were deposited in ArrayExpress (<http://www.ebi.ac.uk/arrayexpress/>, accession number: E-MTAB-3682).

Quantitative (q)-RT-PCR analysis

Total RNA was isolated from BM-sorted KSL, LK, GMP and LT-HSC population, harvested from 3 to 5 mice, using the RNeasy Micro Kit (Qiagen). Total RNA isolated from Lin⁺ and Lin⁻ fractions was isolated using RNeasy Mini Kit (Qiagen). cDNA was generated using the Transcriptor High Fidelity cDNA Synthesis Kit (Roche Applied Science).

Quantitative (q)RT-PCR was performed using the Power SYBR® Green PCR Master Mix (Applied Biosystem) and detected with a 7500 Fast Real-Time PCR System (Life

Technologies). Relative expression levels were determined by the delta Ct method. Expression levels of *beta-2-Microglobulin* and *Rpl6* housekeeping genes were used for normalization.

Primers used for cDNA q-PCR analysis.

Primers	Sequences
Beta-2-microglobulin	Fw: 5'-TTTCCCGTTCTTCAGCATTT Rev : 5'-TGGTGCTTGTCTCACTGACC
Rpl6	Fw : 5'-TGAAGCTTCTGAAAATGCCTA Rev : 5'-CTGGAGCGTAGCCTTCTAC
Klf2	Fw : 5'-CTGCGTACACACACAGGTGAG Rev : 5'-GCACAAGTGGCACTGAAAGGG
Id2	Fw : 5'-GGACTCGCATCCCACTATCG Rev : 5'-TCAGATGCCTGCAAGGACAG
Cdkn1a	Fw : 5'-GTACTTCCTCTGCCCTCTG Rev: 5'-TCTGCGCTTGGAGTATAGA
Ccne2	Fw : 5'-GGAGGAATCAGCCCTTGGCAT Rev : 5'-CCTGTGAACATGCCCAGGTTA
Ccnd3	Fw : 5'-CATGGCAGTTGCGGGAGT Rev : 5'-TTTTGACCAAAGCCTGCCG
Il7r	Fw : 5'-CTTCTGGAGGCAGTGGAGTG Rev : 5'-TGCTTCTTTGCGATAAACGA
Ccna2	Fw : 5'-GAGAATGTCAACCCCGAAAA Rev : 5'-GCAGTGACATGCTATCGTT
Plzf	Fw : 5'-AGAGGGAGCTGTTTCAGCAAG Rev : 5'-TCATCCCACTGTGCAGTTTC

Primers used for ChIP-qPCR in KG1 cells.

Primers	Sequences
Gapdh	Fw: 5'-ACTCACCTGCCCTCAATATC Rev : 5'-AGACAGTGTGCCTTTCATTCCAT
Klf2	Fw : 5'-CCCTGGAAACAGCTGAGAGGGT Rev : 5'-TCATCAGCAGCGGAAGTGAG
Id2	Fw : 5'-CCCACACTAAGCCTGTCGTC Rev : 5'-AGACGAACTCACTCTACAAACGT
Cdkn1a	Fw : 5'-TCACAGGTGTTTCTGCGGC Rev: 5'-CCGCGCACTTAGAGACACC
Ccne2	Fw : 5'-TTTGCTTCCTCCCACTGCAA Rev : 5'-CACTTCCCAGCCACCCTATG
Ccnd3	Fw : 5'-GACACCAAGCCCGACAGTAA Rev : 5'-TTCCTGTCAGTTAACCCGGC

Primers used for ChIP qPCR in HSPCs

Primers	Sequences
Gapdh	Fw: 5'-CTTCGGGCCACGCTAATCTCATTT Rev : 5'-AACTCACCCGTTACACCGACCTT
Klf2	Fw : 5'-CTGTCAATCACAGGCGAGGT Rev : 5'-CTGATTCAAGGAAAGCGGGC
Id2	Fw : 5'-CAGCGCACACTGTACTCAA Rev : 5'-GCTCCCGTCCATCATACCAG
Cdkn1a	Fw : 5'-GCGAGGAGGTGACTCATTGT Rev: 5'-CCCGCCAAATAGGTCACTGT
Ccne2	Fw : 5'-GCTAAGGGAACACGTCTGG Rev : 5'-AACCTCGGGAACCCCTGAAT
Ccnd3	Fw : 5'-GGACACTCGCTTTGTTTGGG Rev : 5'-TGGCAGATGAACTACCGAG

Supplemental References

Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545-15550.

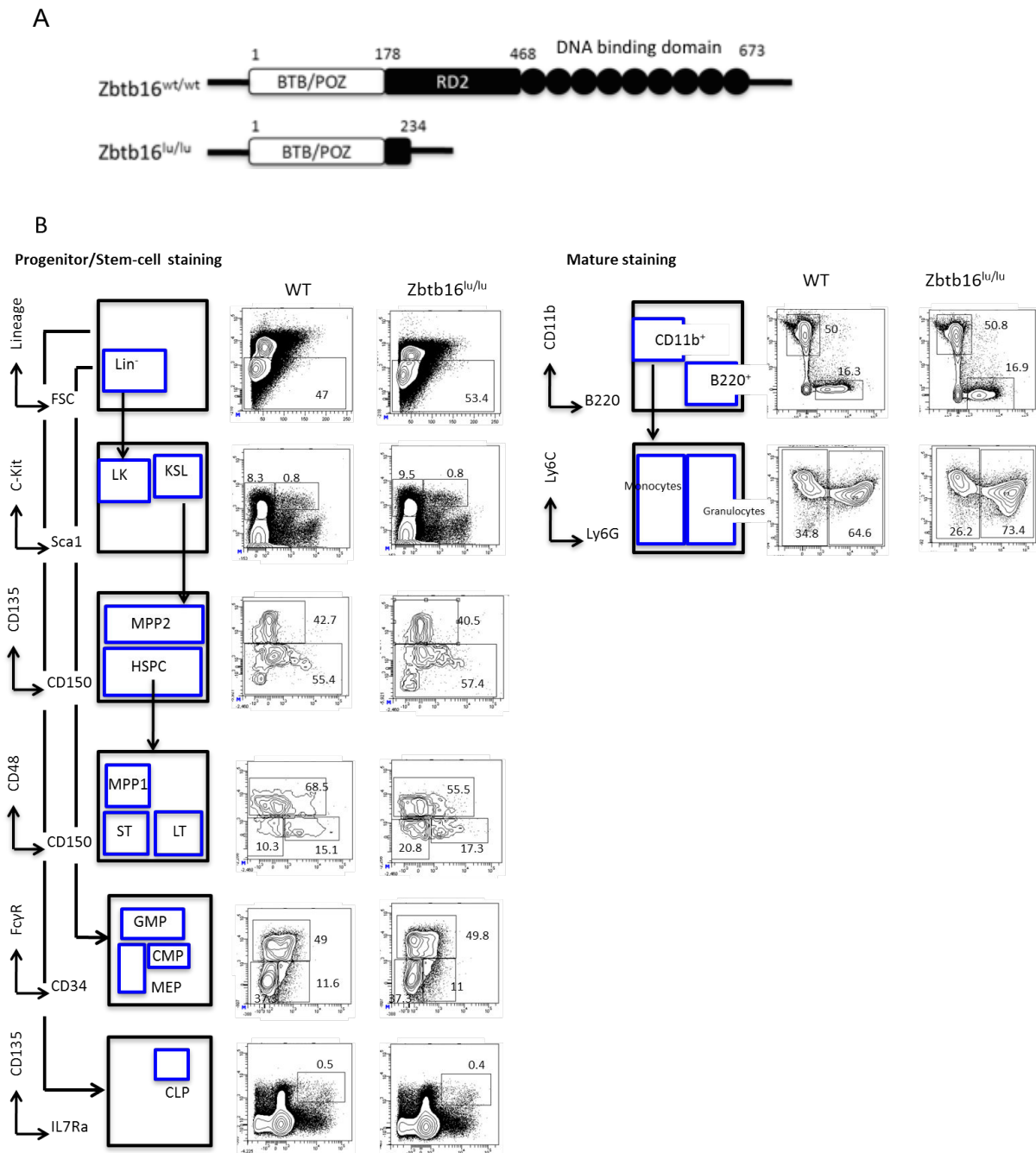


Figure S1:

(A) Representation of the PLZF WT and mutant form. The mutation leads to a truncated protein without the DNA binding domain. (B) Surface markers used for FACS to identify hematopoietic populations studied. Phenotypic populations were defined as KSL (Lin⁻/cKit⁺/Sca1⁺), multipotent progenitors MPP2 (Lin⁻/cKit⁺/Sca1⁺/Flt3⁺/CD150⁻), hematopoietic stem progenitor cells HSPCs (Lin⁻/cKit⁺/Sca1⁺/Flt3⁺/CD150⁺), long-term HSCs LT-HSCs (Lin⁻/cKit⁺/Sca1⁺/Flt3⁺/CD150⁺/CD48⁻), short-term HSCs ST-HSCs (Lin⁻/cKit⁺/Sca1⁺/Flt3⁺/CD150⁺/CD48⁺), MPP1 cells (Lin⁻/cKit^{hi}/Sca1⁺/CD150⁻/CD48⁺), common lymphoid progenitors CLPs (Lin⁻/IL7R⁺/Flt3⁺), myeloid progenitors LK (Lin⁻/cKit⁺/Sca1⁻), common myeloid progenitors CMPs (Lin⁻/cKit⁺/Sca1⁻/CD34⁺/FcγR^{lo}), granulocyte-macrophage progenitors GMPs (Lin⁻/cKit⁺/Sca1⁻/CD34⁺/FcγR^{hi}), and megakaryocyte-erythroid progenitors MEPs (Lin⁻/cKit⁺/Sca1⁻/CD34⁺/FcγR^{lo}). In the mature compartment, myeloid cells are defined as CD11b⁺ cells, granulocytes are CD11b⁺/Gr-1⁺, B cells are B220⁺.

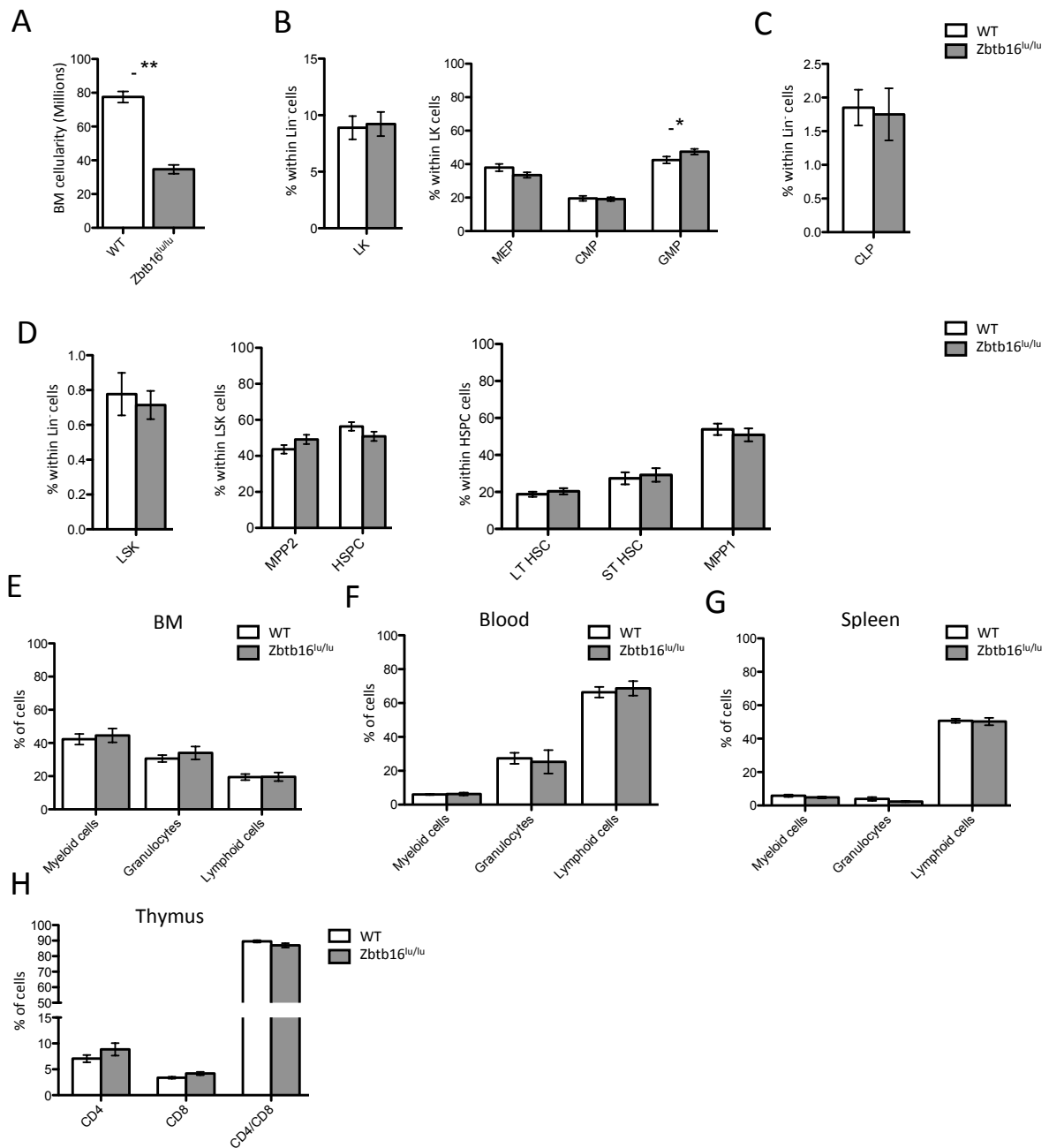


Figure S2: Hematopoiesis analysis in WT and Zbtb16^{lu/lu} mice at steady state.

(A) Total BM mononuclear cell number isolated from 2 tibias and 2 femurs of WT and Zbtb16^{lu/lu} mice. (B) Frequencies of myeloid progenitors (LKs) in Lin⁻ BM and distribution of the myeloid progenitor subsets among LKs: megakaryocytic-erythroid progenitors (MEPs), common myeloid progenitors (CMPs), granulocytic macrophage progenitors (GMPs). (C) Frequencies of the common lymphoid progenitors (CLP). (D) Frequencies of the LSK, MPP2 and HSC population and distribution of LT-HSCs, ST-HSCs and MPP1 among HSCs. (E-G) Frequencies of myeloid cells (CD11b⁺), granulocytes (CD11b⁺/Gr1⁺) and lymphoid cells (B220⁺) in young (2-month-old) adult bone marrow (E), Blood (F) and Spleen (G). (H) Frequencies of T cells (CD4⁺, CD8⁺, CD4⁺/CD8⁺) in young adult thymus. Bar graphs represent averages $\pm$ SEM (n=15). * p<0,05 ; ** p<0,01.

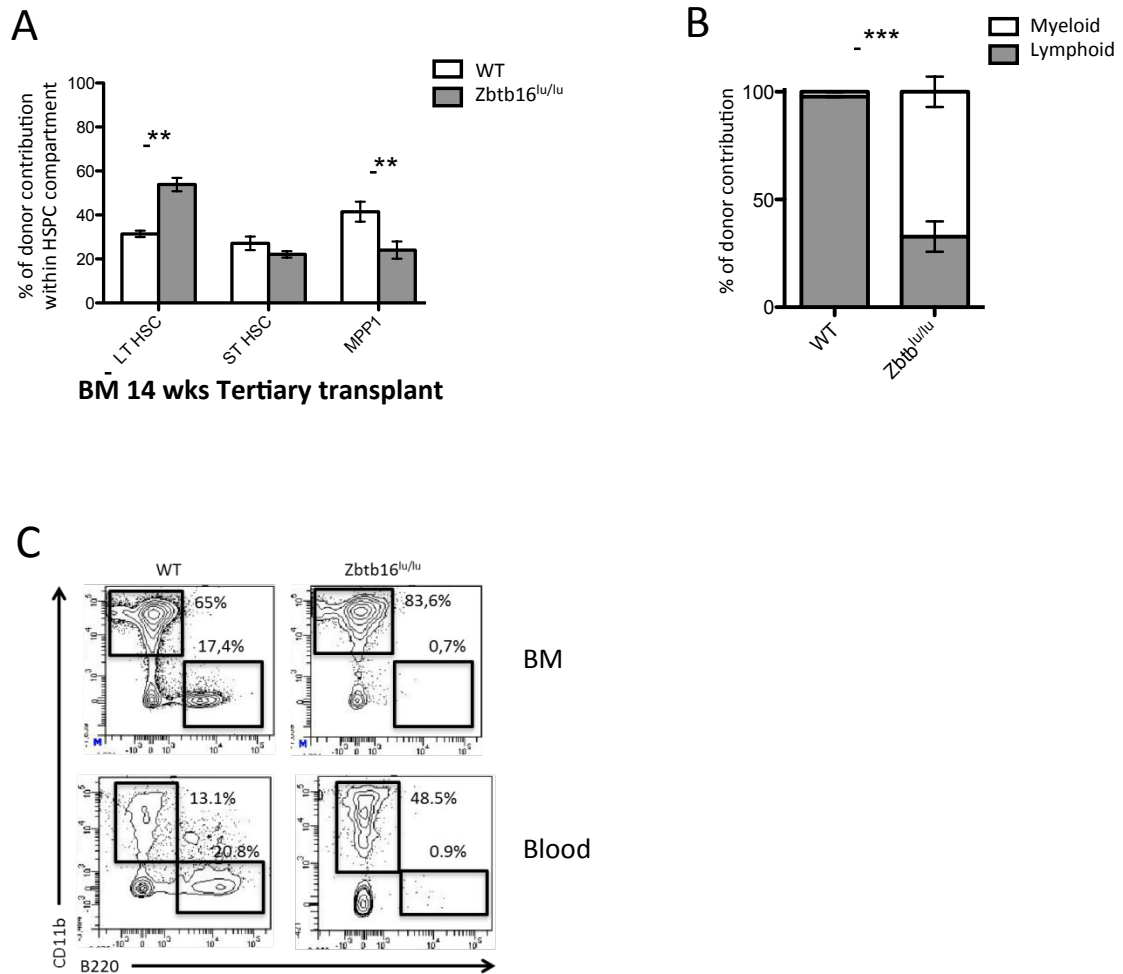


Figure S3: Long-term reconstitution analysis in tertiary transplanted mice.

(A) Long-term reconstitution analysis of stem cells in BM after tertiary transplantation. (B) Long-term reconstitution analysis of mature cells in spleen after tertiary transplantation. (C) FACS analyses of long-term reconstitution of mature compartment in Bone Marrow and Blood after three transplantations. * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$.

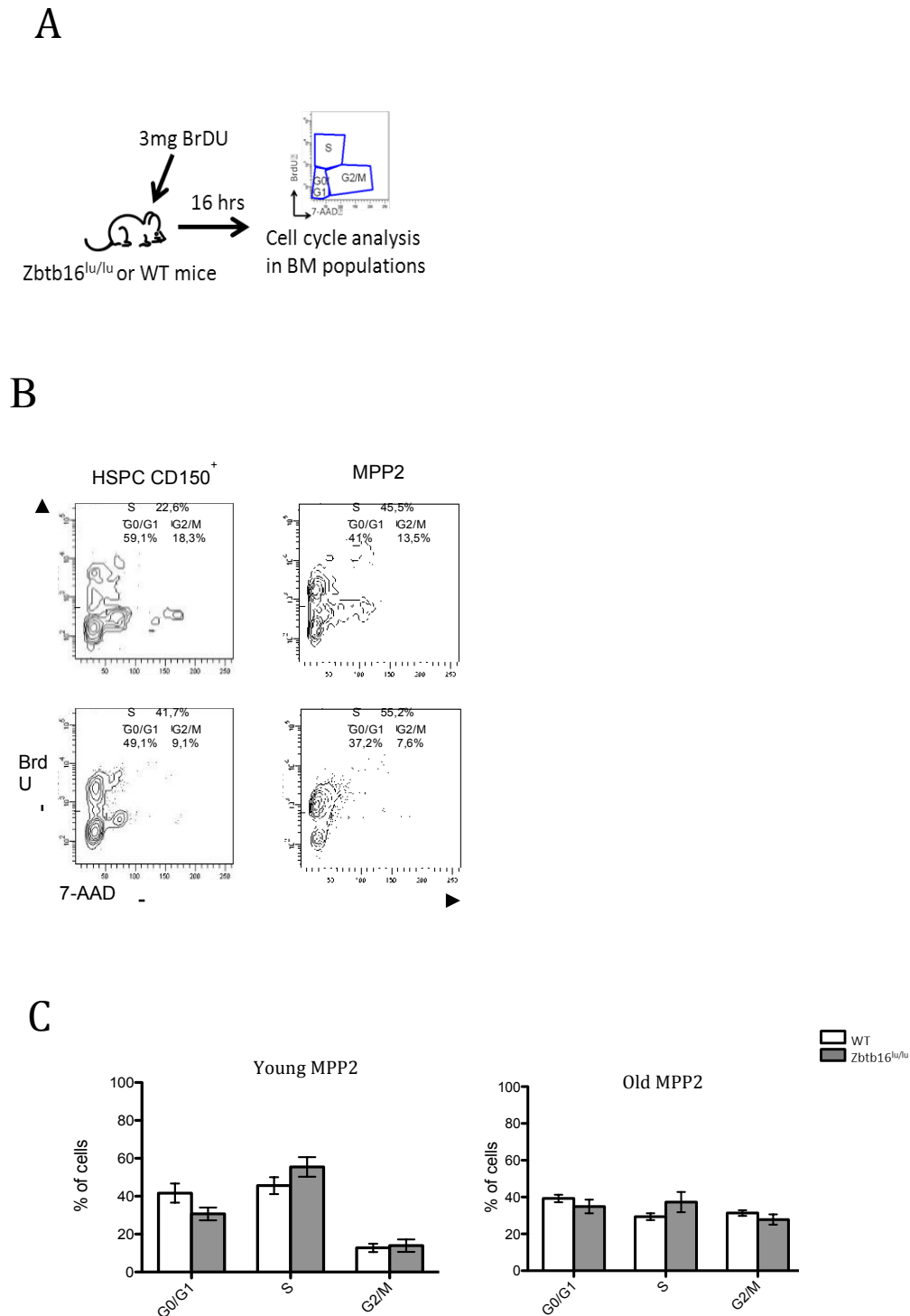


Figure S4 supplemental to Figure 5: Cell cycle analysis in WT and Zbtb16^{lu/lu} mice after BrdU injection.

(A) 3 mg of BrdU was injected by i.p. 18 hours before sacrifice and cell cycle analysis was done by looking at BrdU and 7-AAD labeling on Lin⁻ sorted cells. (B) Representative FACS analysis of the distribution of LT-HSC CD150⁺ and MPP2 in G1, S and G2/M phases in young Zbtb16^{lu/lu} mice, in comparison to WT counterparts. (C) Quantitative analysis of cell cycle in MPP2 from young (n=6-7) and old (n=5) WT or Zbtb16^{lu/lu} mice.

Article de Recherche 3 :

**Site- and allele-specific polycomb
dysregulation in T-cell leukaemia**

Site- and allele-specific polycomb dysregulation in T-cell leukaemia

Jean-Marc Navarro*, Aurore Touzart*, Lydie C. Pradel*, Marie Loosveld*, Myriam Koubi, Romain Fenouil, Sandrine Le Noir, Muhammad Ahmad Maqbool, Ester Morgado, Claude Gregoire, Sebastien Jaeger, Emilie Mamessier, Charles Pignon, Salima Hacein-Bey-Abina, Bernard Malissen, Marta Gut, Ivo G. Gut, Herve Dombret, Elizabeth A. Macintyre, Steven J. Howe, H. Bobby Gaspar, Adrian J. Thrasher, Norbert Ifrah, Dominique Payet-Bornet**, Estelle Duprez**, Jean-Christophe Andrau**, Vahid Asnafi** and Bertrand Nadel**

* and ** : equal contribution

Les Leucémies Aiguës Lymphoblastiques T (LAL-T) résultent de la prolifération maligne de précurseurs lymphoïdes T bloqués à divers stades de maturation. Les LAL-T constituent un groupe de leucémies très hétérogène, dans lequel différents oncogènes sont activés dans des combinaisons diverses. TAL1 (T-cell Acute Lymphocytic Leukemia 1) est un oncogène dont l'expression ectopique due à des translocations ou des délétions est observée dans 30 à 40% des cas de LAL-T. Cependant, dans plus de la moitié de ces cas TAL1+, aucune anomalie génomique de TAL1 n'est observée (cas dits « non résolus »). L'expression ectopique de TAL1 est alors soit monoallélique, correspondant à une altération en *cis* du gène, soit bi-allélique, reflétant une dérégulation indirecte en *trans* (Ferrando, 2004). Une des hypothèses du projet de l'équipe du Dr Nadel repose sur l'idée que des dérégulations épigénétiques interviennent au niveau du locus TAL1.

Dans un premier temps, les marques épigénétiques associées à la région correspondant au gène *TAL1* ont été étudiées dans des thymocytes murins doubles positifs (ie positifs pour les marqueurs CD4 et CD8). Pour cela, des données publiques générées par des expériences d'immunoprécipitation de la chromatine suivie de séquençage haut débit (ChIP-seq) ont été exploitées. Ces données montrent qu'une large région comprenant le promoteur et une partie du gène *TAL1* est enrichie pour la marque répressive H3K27me3 catalysée par la protéine Polycomb EZH2. Ces observations indiquent donc qu'en conditions normales, le gène *TAL1* n'est pas exprimé. Afin de comprendre comment cette répression épigénétique peut être levée dans des conditions pathologiques, une région d'environ 10 kb a été séquencée chez un patient atteint de LAL-T monoallélique pour TAL1. Une micro-insertion de 12 pb en amont du promoteur de TAL1 et située en bordure de la marque répressive H3K27me3 qui recouvre une partie du gène *TAL1* a été identifiée. Afin de mieux identifier le rôle de cette micro-insertion (TALins) dans l'inhibition de la répression épigénétique de TAL1, l'étude a été étendue et une région de 700pb autour du site d'insertion a été séquencée dans des cas de patients TAL1+ monoallélique non résolus. Ce séquençage a révélé que 20% de ces cas possèdent une insertion de taille variable. Par ChIP-allélique (ChIP-qPCR pouvant discriminer l'allèle muté de l'allèle sauvage) nous avons confirmé que la mutation TALins est

associée à la diminution de la marque répressive H3K27me3 du locus TAL1 et à l'augmentation de la marque activatrice H3K27ac au même niveau. Pour finir, les pronostics cliniques de patients présentant une altération biallélique ont été comparés à ceux des patients présentant une altération monoallélique. Il en ressort que le mode d'activation de TAL1 (mono ou biallélique) présente un impact sur la survie des patients puisque les patients bialléliques ont un meilleur pronostic clinique.

Cette étude a donc mis en évidence un nouveau mécanisme d'activation de l'oncogène TAL1 qui se traduit par une levée de l'inhibition épigénétique de ce gène suite à une insertion en amont de ce locus.

ARTICLE

Received 25 Nov 2014 | Accepted 11 Dec 2014 | Published 23 Jan 2015

DOI: 10.1038/ncomms7094

OPEN

Site- and allele-specific polycomb dysregulation in T-cell leukaemia

Jean-Marc Navarro^{1,2,3,*}, Aurore Touzart^{4,*}, Lydie C. Pradel^{1,2,3,*†}, Marie Loosveld^{1,2,3,5,*}, Myriam Koubi⁶, Romain Fenouil^{1,2,3}, Sandrine Le Noir⁴, Muhammad Ahmad Maqbool^{1,2,3,†}, Ester Morgado^{1,2,3}, Claude Gregoire^{1,2,3}, Sebastien Jaeger^{1,2,3}, Emilie Mamessier^{1,2,3}, Charles Pignon^{1,2,3}, Salima Hacein-Bey-Abina⁷, Bernard Malissen^{1,2,3}, Marta Gut⁸, Ivo G. Gut⁸, Hervé Dombret⁹, Elizabeth A. Macintyre⁴, Steven J. Howe¹⁰, H. Bobby Gaspar¹⁰, Adrian J. Thrasher¹⁰, Norbert Ifrah¹¹, Dominique Payet-Bornet^{1,2,3,§}, Estelle Duprez^{5,§}, Jean-Christophe Andrau^{1,2,3,§,†}, Vahid Asnafi^{4,§} & Bertrand Nadel^{1,2,3,§}

T-cell acute lymphoblastic leukaemias (T-ALL) are aggressive malignant proliferations characterized by high relapse rates and great genetic heterogeneity. *TAL1* is amongst the most frequently deregulated oncogenes. Yet, over half of the *TAL1*⁺ cases lack *TAL1* lesions, suggesting unrecognized (epi)genetic deregulation mechanisms. Here we show that *TAL1* is normally silenced in the T-cell lineage, and that the polycomb H3K27me3-repressive mark is focally diminished in *TAL1*⁺ T-ALLs. Sequencing reveals that >20% of monoallelic *TAL1*⁺ patients without previously known alterations display microinsertions or RAG1/2-mediated episomal reintegration in a single site 5' to *TAL1*. Using 'allelic-ChIP' and CrispR assays, we demonstrate that such insertions induce a selective switch from H3K27me3 to H3K27ac at the inserted but not the germline allele. We also show that, despite a considerable mechanistic diversity, the mode of oncogenic *TAL1* activation, rather than expression levels, impact on clinical outcome. Altogether, these studies establish site-specific epigenetic desilencing as a mechanism of oncogenic activation.

¹Center of Immunology of Marseille Luminy, Aix-Marseille University, Parc Scientifique de Luminy case 906, 13288 Marseille, France. ²INSERM U1104, 13288 Marseille, France. ³CNRS UMR7280, 13288 Marseille, France. ⁴Université Paris Descartes Sorbonne Cité, Institut Necker-Enfants Malades (INEM), Institut national de recherche médicale (INSERM) U1151, and Laboratory of Onco-Hematology, Assistance Publique-Hôpitaux de Paris (AP-HP), Hôpital Necker Enfants-Malades, 75015 Paris, France. ⁵Laboratoire Hématologie, APHM, 13385 Marseille, France. ⁶CRCM Inserm U1068, Institut Paoli Calmettes; Aix-Marseille Université, UM 105; CNRS UMR7258, 13009 Marseille, France. ⁷Biotherapy Department, INSERM U429, Hôpital Necker-Enfants-Malades, 149 rue de Sèvres, 75015 Paris, France. ⁸Centro Nacional de Análisis Genómico, 08028 Barcelona, Spain. ⁹Department of Hematology, AP-HP Hôpital St-Louis, 75010 Paris, France. ¹⁰Centre for Immunodeficiency, Molecular and Cellular Immunology, Institute of Child Health, University College London, London WC1N 1EH, UK. ¹¹Services des Maladies du sang CHU and UMR Inserm U 892/CNRS 6299, 49933 Angers, France. * These authors contributed equally to this work. § Joint supervisors. † Present addresses: TAGC/INSERM U 1090, 13288 Marseille, France (L.C.P.); Institut de Génétique Moléculaire de Montpellier (IGMM), CNRS UMR5535, 1919 Route de Mende, 34293 Montpellier, France (M.A.M. and J.-C.A.). Correspondence and requests for materials should be addressed to V.A. (email: vahid.asnafi@nck.aphp.fr) or to B.N. (email: nadel@ciml.univ-mrs.fr).

T-cell acute lymphoblastic leukaemia's (T-ALL) are malignant proliferations of immature T-cell progenitors. Although the outcome of T-ALLs has greatly improved in the last 10 years, ~30% of cases relapse within the first 2 years following diagnosis; moreover, acute- and long-term toxicities remain important issues for long-term survivors, underlining the critical need of better risk stratification of T-ALL patients, and the implementation of more adapted and/or targeted therapies. A major obstacle in the molecular dissection of these processes is that T-ALLs constitute a particularly heterogeneous group of disease, characterized by complex combinations of multigenic aberrations and oncogenic cooperation. The deregulation of over 40 distinct oncogenes and tumour suppressors has been reported, occurring through a large diversity of genomic aberrations and epigenetic mechanisms^{1,2}. Chromosomal alterations mostly consist of translocations, inversions and microdeletions occurring at the vicinity of proto-oncogenes and leading to their deregulated expression.

TAL1 is one of the most frequently deregulated T-ALL oncogenes³. In physiological conditions, *TAL1* is a regulatory gene that promotes access to alternative fates in haematopoiesis. Expressed in early haematopoiesis, its expression is maintained in the erythroid lineage, but normally irreversibly epigenetically silenced in the T-cell lineage^{4–6} (Supplementary Fig. 1). Events leading to the illegitimate maintenance or re-expression of *TAL1* in the T-cell lineage are thought to constitute strong drivers of T-ALL leukemogenesis. Known *TAL1* dysregulation mechanisms consist of t(1;14)(p34;q11) translocations (1–2%) and SIL-TAL deletions (15–20%). Yet, over half of the *TAL1*⁺ cases lack *TAL1* lesions, suggesting unrecognized (epi)genetic deregulation pathways⁷. In such 'unresolved cases', *TAL1* expression can be monoallelic, compatible with a direct alteration in *cis* within or around the *TAL1* gene, or biallelic, likely reflecting indirect deregulation in *trans*⁷.

Here we sought to gain insights into the (epi)genetic deregulation pathways of *TAL1* 'unresolved' cases. Using ChIP-seq and an 'allelic-ChIP' assay allowing to discriminate histone marks separately on each allele, we describe a new process of oncogene activation, whereby the targeted eviction of polycomb H3K27me3 marks and concurrent recruitment of H3K27ac marks by micro- and macroinsertional events, trigger the monoallelic desilencing of *TAL1*. Incidentally, we also report the first example of oncogenic activation by recombination-activating gene (RAG)-mediated episomal reinsertion, a very elusive event predicted *in vitro* as a source of oncogenic activation over 10 years ago^{8,9}, but never identified before in human cancer. We further show that this new epigenetic desilencing process is a recurrent event in T-ALL, accounting for >20% of unexplained cases of monoallelic *TAL1* oncogene activation. Finally, we demonstrate that the mode of activation (monoallelic in *cis* versus biallelic in *trans*) rather than the level of *TAL1* expression impacts on prognosis, with *cis*-mediated alterations significantly associated with adverse clinical outcome.

Results

Microinsertions induce epigenetic switch and *TAL1* expression.

To investigate the (epi)genetic deregulation pathways of *TAL1* 'unresolved' cases, we first took advantage of chromatin immunoprecipitation coupled to a high-throughput sequencing (ChIP-seq) data set describing active/inactive epigenetic marks in mouse developing thymocytes^{6,10}. A large region starting upstream of the *TAL1* promoter and encompassing part of the gene body was enriched in H3K27me3, suggesting the involvement of polycomb complex (PcG) repressive activity in *TAL1* silencing (Fig. 1a). This profile was also observed in human

TAL1[−] peripheral CD4⁺ T cells¹¹, but absent in *TAL1*-expressing cells¹² (CD34⁺ haematopoietic stem cells, HSC; erythroblasts; Fig. 1b), in line with the gradual deposition of H3K27me3-repressive marks during T-cell lineage specification⁶ (Supplementary Fig. 1). To determine if deregulation of this silencing occurs in *TAL1*⁺ T-ALLs, ChIP-seq was also performed in the Jurkat (*TAL1*⁺) T-ALL cell-line and a control *TAL1*[−] sample (DA). Sample DA displayed the expected H3K27me3 deposition, coherent with *TAL1* silencing in T cells (Fig. 1b). Surprisingly, however, low levels of H3K27me3 marks were also observed in Jurkat. Since, in contrast to normal *TAL1*-expressing cells, Jurkat displays monoallelic expression of *TAL1* (Supplementary Fig. 2), we reasoned that a monoallelic chromosomal alteration might have prevented H3K27me3 deposition on one of the two alleles, resulting in an averaged, intermediate ChIP-seq signal. To test this possibility, we sequenced ~10 kb upstream of *TAL1* (Supplementary Fig. 3), covering the regulatory region enriched in H3K27me3 marks. Amid single-nucleotide polymorphisms (SNPs), we found a monoallelic alteration consisting of a 12-bp microinsertion ~7 kb upstream of the *TAL1* p1a promoter, in a region prone to DNA looping with *TAL1* promoters¹³. Interestingly, this insertion mapped at the border of the repressive H3K27me3 pattern (Fig. 1a,b), in line with the possibility that it disrupted normal epigenetic silencing of the *TAL1* locus. To further investigate the association of the insertion with potential allelic distortion of the repressive H3K27me3 mark, we performed an 'allelic-ChIP' that discriminates H3K27me3 marks at the inserted versus germline allele (Fig. 2a). A significant decrease of H3K27me3 signal was observed at the inserted allele compared with the germline allele ($P < 0.005$). Accordingly, knockdown (by shEZH2) or inhibition (using the EZH2 inhibitor GSK126) (ref. 14) of the Polycomb H3K27me3 methyl transferase EZH2, allowed partial reversion of H3K27me3 deposition at the germline (non-inserted) allele. Considering the cooperative role of histone modifications in expression regulation, we further analysed the relevance of acetylation in presence or absence of histone deacetylase inhibitor (sodium butyrate)¹⁵. ChIP-seq showed a slight enrichment of H3K27ac with significant overall increase on sodium butyrate treatment (Fig. 2b). Similar to H3K27me3, H3K27ac marks extended to the insertion site, suggesting dual epigenetic regulation and differential allelic recruitment. Indeed, allelic quantification through tag retrieval and allelic-ChIP revealed a difference of H3K27ac between inserted and GL alleles in a pattern symmetrical to and functionally coherent with H3K27me3. Acetylation levels were histone deacetylase inhibitor dependent. Overall, this suggested that Jurkat's microinsertion contributed to site- and allele-specific switch from H3K27me3 to H3K27ac deposition, leading to the maintenance of *TAL1* expression through T-cell differentiation.

A recurrent epigenetic mechanism of *TAL1* deregulation. To determine if similar structural abnormalities occurred recurrently in *TAL1* patients, 134 primary T-ALL samples were analysed by high-density Affymetrix SNP array-6 analysis; an ~700-bp region surrounding the Jurkat insertion site was also sequenced in a subset of 93 samples and six cell lines; in parallel, the literature was reviewed for cases with unexplained *TAL1* activation. While no macromolecular *TAL1* alteration was identified by SNP array (0/134), sequencing revealed seven new cases of similar micro-insertions (1–9 bp), all precisely located at the Jurkat insertion breakpoint (Fig. 3a). Such insertions were not present in the germline from 2/2 available patients tested (Supplementary Fig. 4). No additional mutation/indel could be found in the

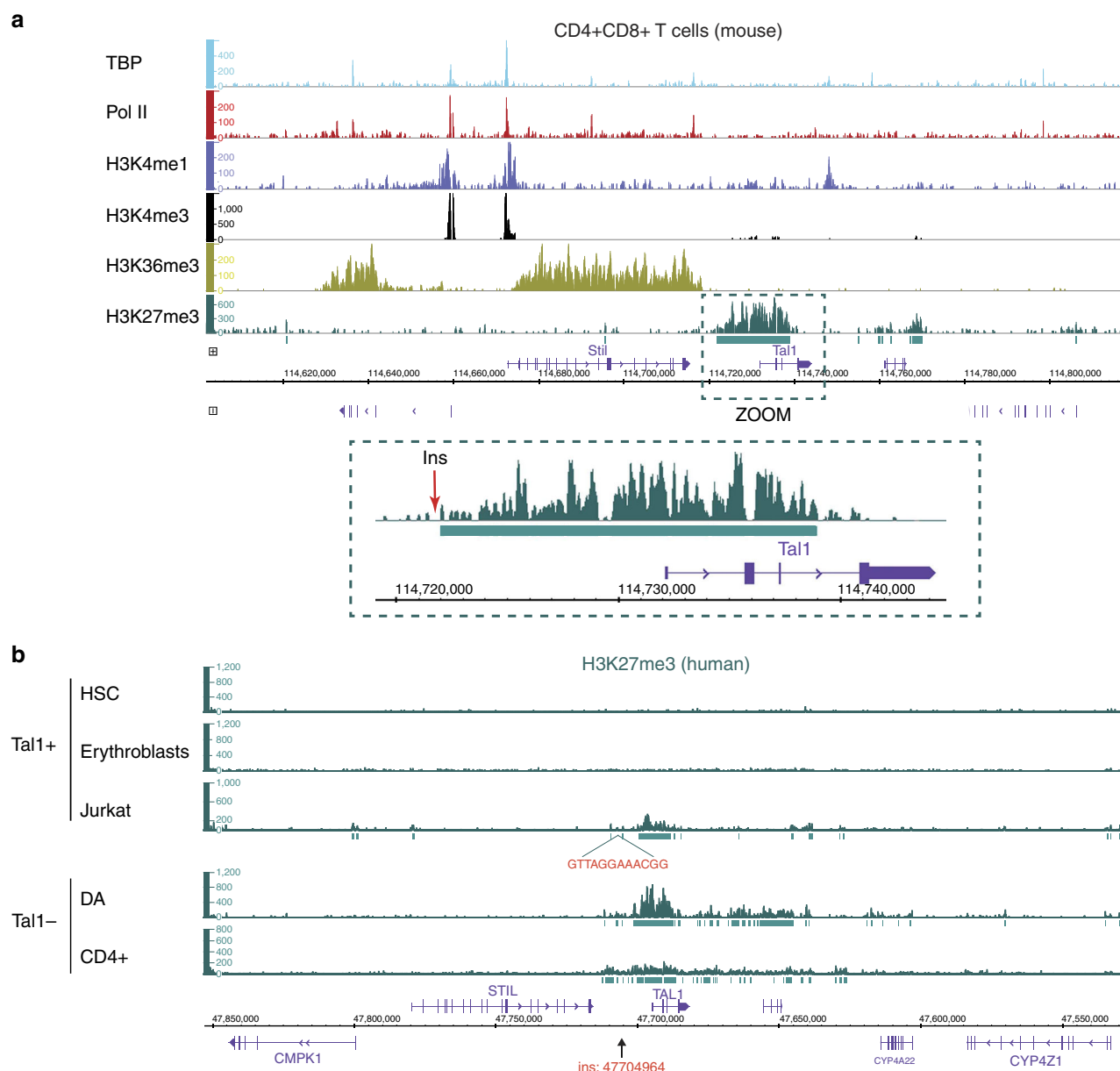


Figure 1 | ChIP-seq analysis of the *TAL1* locus in *TAL1*-expressing and *TAL1*-repressed cells. (a) ChIP-seq profiles of chromatin marks in normal mouse CD4⁺CD8⁺ thymocytes. The *TAL1* genomic area on chromosome 4 (including the adjacent *STIL* gene, which is physiologically expressed at that developmental stage) is shown for TATA-binding protein-GTF recruited to promoters (TBP), Polymerase II (Pol II), H3K4me1 (active chromatin mark for enhancer/regulatory regions), H3K4me3 (active chromatin marks for promoters), H3K36me3 (active chromatin marks for gene bodies) and H3K27me3 (repressive chromatin mark dependent on the PcG). The *TAL1* region is zoomed, and the insertion breakpoint localization to the human orthologue region (114,722,607) is indicated by a red arrow. **(b)** ChIP-seq profiles of polycomb repressive chromatin mark H3K27me3 in normal human TAL1⁺ (HSC; erythroblasts) and TAL1⁻ (peripheral CD4⁺ T cells) lineages, compared with TAL1⁺ (Jurkat) and TAL1⁻ (DA) T-ALL cells. The *TAL1* gene and surrounding area on chromosome 1 (including the *STIL* gene) is shown. The insertion breakpoint localization (47704964, HG19 coordinates) is indicated by an arrow. All profiles were input subtracted, except for HSC and erythroblasts for which input data were not available¹². Significantly enriched areas are represented as green rectangles under the lanes (MACS2 peaks).

surrounding ~700 bp region in tumour samples. Significantly, insertions were exclusively found in TAL1⁺ patients; moreover, among patients with informative SNPs in the *TAL1* 3' UTR allowing distinction of mono- from biallelic expression⁷ ($n = 60$), insertions were exclusively found in monoallelic cases (4/19, >20%), in agreement with a *cis*-mode of *TAL1* activation (Fig. 3b).

Oncogenic RAG1/2-mediated episomal reintegration. One additional candidate was recovered by data mining¹⁶. In this case

(patient #OC), TAL1^{high} activation concurred with the insertion of a large piece of chromosome 7 disrupting the *TAL1* locus. Strikingly, breakpoint mapping by ligation-mediated PCR revealed that the insertion occurred at the very same insertion site, although this time with few nucleotide deletions on each side of the breakpoint (Fig. 3a). Detailed analysis of the junctions revealed the occurrence of RAG1/2-mediated reinsertion of an ~370-kb TCR β episomal circle (TREC β , excised during normal V(D)J recombination, Fig. 4 and Supplementary Fig. 5). This establishes the first example of oncogenic RAG1/2-mediated

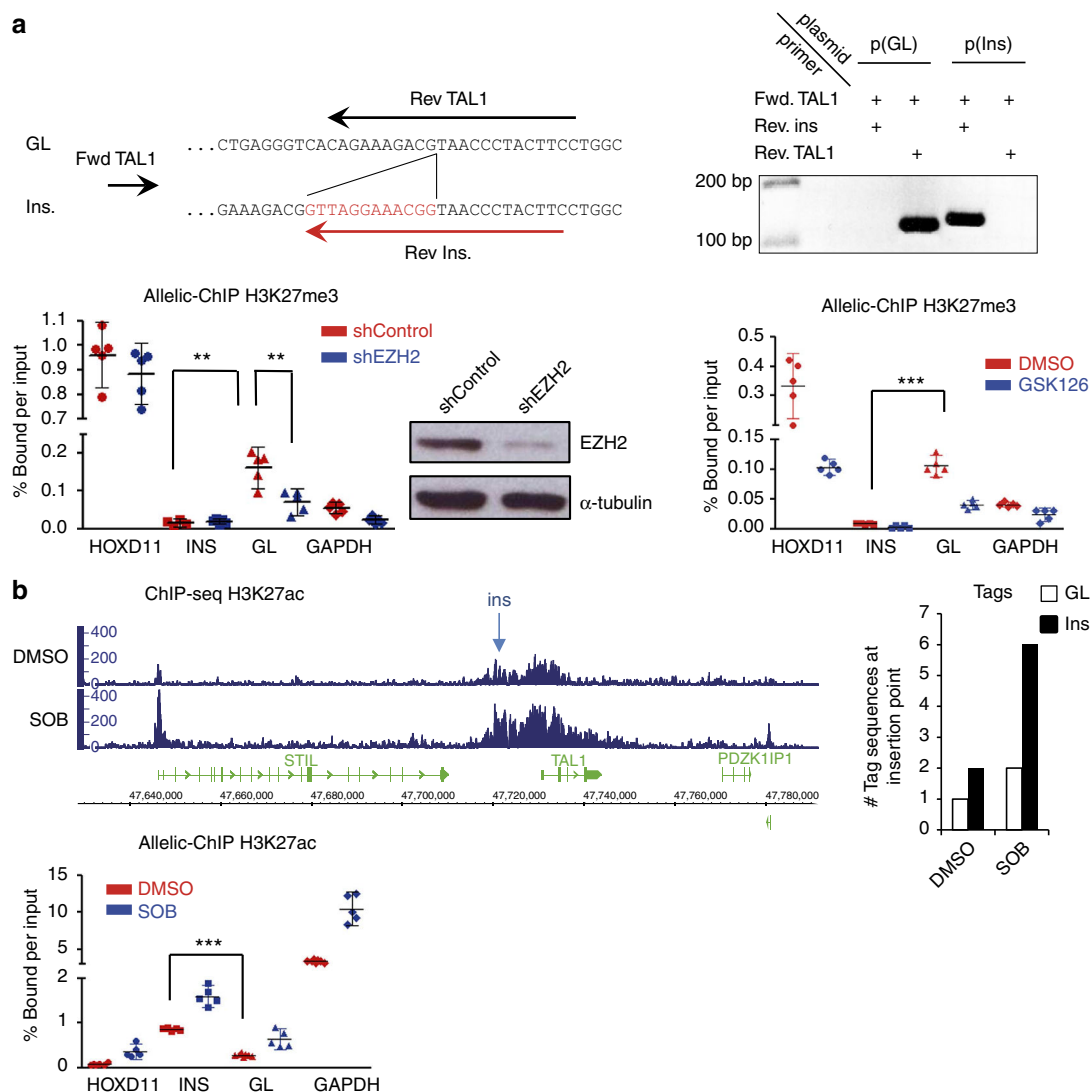


Figure 2 | Site- and allele-specific analysis of histone methylation/acetylation marks at the insertion breakpoint in Jurkat. (a) Allelic-ChIP assay of H3K27me3 marks. Top panel: the assay to discriminate the germline (GL) from the inserted allele (Ins.) by substituting one of the GL primers allowing detection of the GL configuration at the insertion site (Rev. TAL1), with an insertion-specific primer allowing detection of the inserted configuration (Rev. Ins, overlapping the 12-bp insertion). Primer pairs were tested on GL p(GL) or inserted p(Ins) cloned fragments (and on cell lines containing (Jurkat) or not (DND) the insertion) to exclusively amplify each configuration, and do not crossreact. Bottom middle panel: western blot of EZH2 protein content on shMock or EZH2 knockdown conditions. Allelic-ChIP assays were performed in presence of a non-silencing sh-RNA (shControl) or a sh-RNA-targeting EZH2 (shEZH2) (left panel) or after the incubation of Jurkat cells with GSK126 (0.5 μ M, 72h) or vehicle (dimethyl sulphoxide, DMSO; right panel). GAPDH and HoxD11 were used as controls for activated/repressed genes, modulated according to the polycomb-dependent H3K27me3 marks. Note that EZH2 knockdown/inhibition triggered only partial decrease of H3K27me3 marks at the PcG-repressed HoxD11 control gene, possibly due to incomplete knockdown/inhibition and/or redundancy of polycomb components in the adult lymphoid lineage⁴². (b) Enrichment of acetylation marks at the *TAL1* locus. H3K27Ac ChIP was performed with Jurkat cells incubated with vehicle (DMSO) or the histone deacetylase inhibitor sodium butyrate (SOB) (5 mM, 4h); DNA was then analysed by ChIP-seq (left panel) or by allelic-ChIP (bottom panel). For the ChIP-Seq, quantification of the number of tag sequences at the insertion point is shown (right panel). *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$, unpaired *t*-test. Errors bars represent 95% confidence interval.

reintegration, demonstrating that TRECs may indeed contribute to oncogenesis¹⁷. Owing to the large TREC β size, we considered the possibility that, unlike other microinsertions, a promoter located in the episome could have initiated a >7-kb-long fusion transcript encompassing *TAL1* (Supplementary Fig. 6). However, reverse transcription-PCR (RT-PCR) exon walking and 5' RACE assays indicated that transcripts initiated from the *TAL1* p4 promoter, excluding this possibility.

Epigenetic modulation and *TAL1* gene expression. Allelic-ChIP was then performed on patients #TAMFA and #OC. Similar to

Jurkat, significant enrichment of the repressive H3K27me3 mark was consistently observed in germline compared with inserted alleles (Fig. 5a,b). The amplitude of the allelic distortion appeared higher in OC than in Jurkat and TAMFA, possibly due to the large difference in the insertion size. Of note, *TAL1* transcription levels were also higher in OC (Fig. 3b). Using CrispR DNA editing, we next mimicked site-specific insertion and disruption of the region 7 kb 5' of *TAL1* in the TAL-negative PEER cell line (Fig. 5b). In clone #2.4 recapitulating the 12-bp Jurkat insertion at its 3' end, an approximately fivefold increase of *TAL1* could be observed. While we cannot formally exclude the possibility that the selection cassette contributed to the fivefold

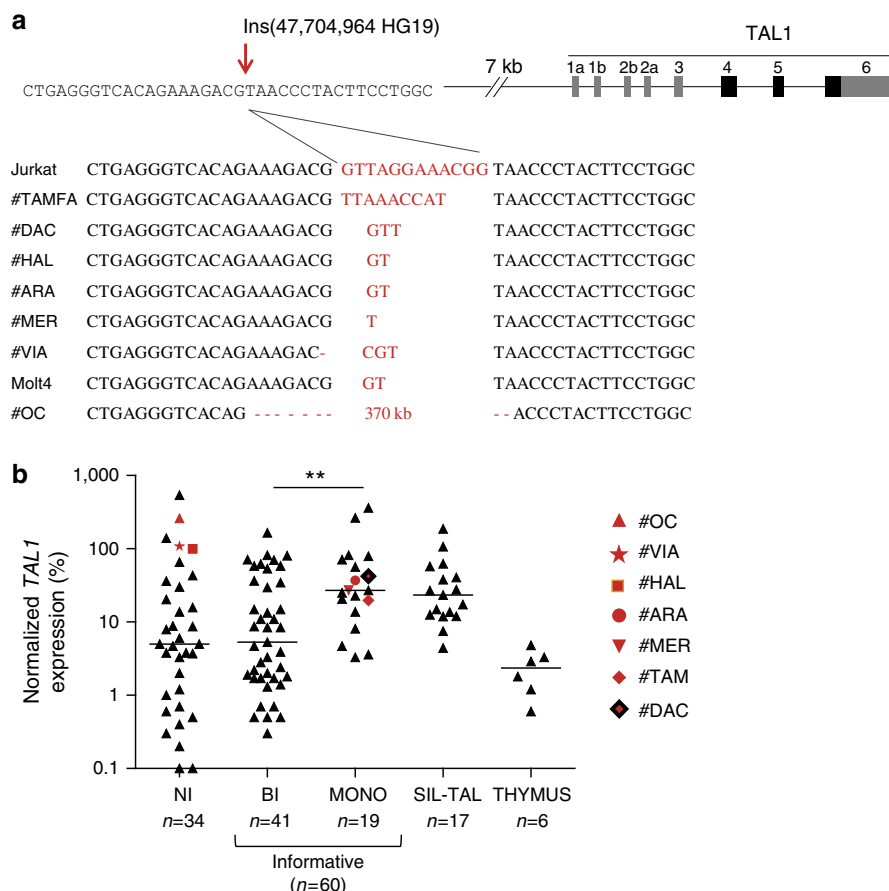


Figure 3 | Micro- and episomal insertions are recurrently found in monoallelic $TAL1^+$ ‘unresolved cases’. (a) Nucleotide sequences of episomal/microinsertions. All insertions were specifically and exclusively located at the indicated genomic position, and are pictured in red. Nucleotide deletions are indicated by a red dash. No SNPs are referenced at this position (Supplementary Fig. 8); (b) Relative $TAL1$ expression in T-ALL patients ($n=111$) according to biallelic (BI), or monoallelic (MONO) expression; patients with micro/episomal insertions are indicated in red; SIL-TAL cases are shown separately; informative cases: the presence of SNPs in $TAL1$ 3’UTR allows to determine if the expression is mono- or biallelic. The average physiological $TAL1$ levels in thymus is shown as reference (Thymus); Horizontal bars indicate median expression levels; **indicates significant difference between BI and MONO expression (Mann-Whitney U-test, $P<0.01$); note that a number of biallelic patients are reaching/below physiological thymus levels, and might result from the presence of residual $TAL1$ -expressing erythroblasts among tumoral cells⁴³; $TAL1$ expression was analyzed by Taqman assay and is normalized to ABL (see Methods).

change in clone #2.4, a 55-fold increase was observed in clone #5.10, in which an ~ 1.3 kb deletion 5’ of the insertion site mimicked locus disruption in patient #OC; furthermore, this was accompanied by an allelic switch from H3K27 methylation to acetylation. This provides direct evidence for a causal relationship between site mutagenesis, epigenetic modulation and $TAL1$ gene expression.

The mode of $TAL1$ activation impacts on clinical outcome.

Patients with identified insertions were globally of adverse prognosis. We sought to determine if clinical outcome correlated with quantitative or qualitative aspects of $TAL1$ deregulation. A cohort of 165 adult T-ALL treated prospectively in the GRAALL (Group for Research in Adult Acute Lymphoblastic Leukaemia) trial was split into $TAL1$ expression quartiles, and compared for disease-free (DFS) and overall survival (OS). The seven patients with identified insertions (three of whom were GRAALL treated), all belonged to the high-expression quartiles (Q3–4). However, no significant difference in survival was observed between the quartiles (Fig. 6a,b), suggesting that quantitative $TAL1$ expression does not correlate with clinical outcome. We next tested whether *cis*-mediated $TAL1$ alterations leading to monoallelic expression (including or not the SIL- $TAL1^+$ cases) affected the clinical

outcome compared with *trans*-mediated events, associated with biallelic $TAL1$ expression. Clinical outcome was indeed found to be significantly improved in the biallelic group (DFS, $P=0.04$; OS, $P=0.03$; Fig. 6c–f). Although numbers are low, monoallelic cases retained an inferior OS trend in multivariate analysis (including age, leukocytosis; $P=0.07$, Cox analysis). Despite genetic heterogeneity, monoallelic cases also displayed higher blasts counts at diagnosis than biallelic cases and a significantly lower frequency of deregulation of recurrent oncogenes such as *TLX1*, *CALM-AF10* and *TLX3* (Supplementary Table 1). This supports the emerging notion that the mode of alteration may shape the oncogenic landscape in a more profound manner than (and potentially override the effect of) transcriptional levels, and that this may eventually impact on the tumour’s clinical behaviour¹⁸. Deciphering the mechanisms underlying the (epi)genetic deregulation of biallelic $TAL1^+$ cases will be instrumental to resolving this issue^{19–22} (Supplementary Fig. 7 and Supplementary Table 2).

Discussion

Establishing the detailed maps of the complex oncogenic networks involved in T-ALL has contributed to major genetic discoveries, and has been of prime importance for further

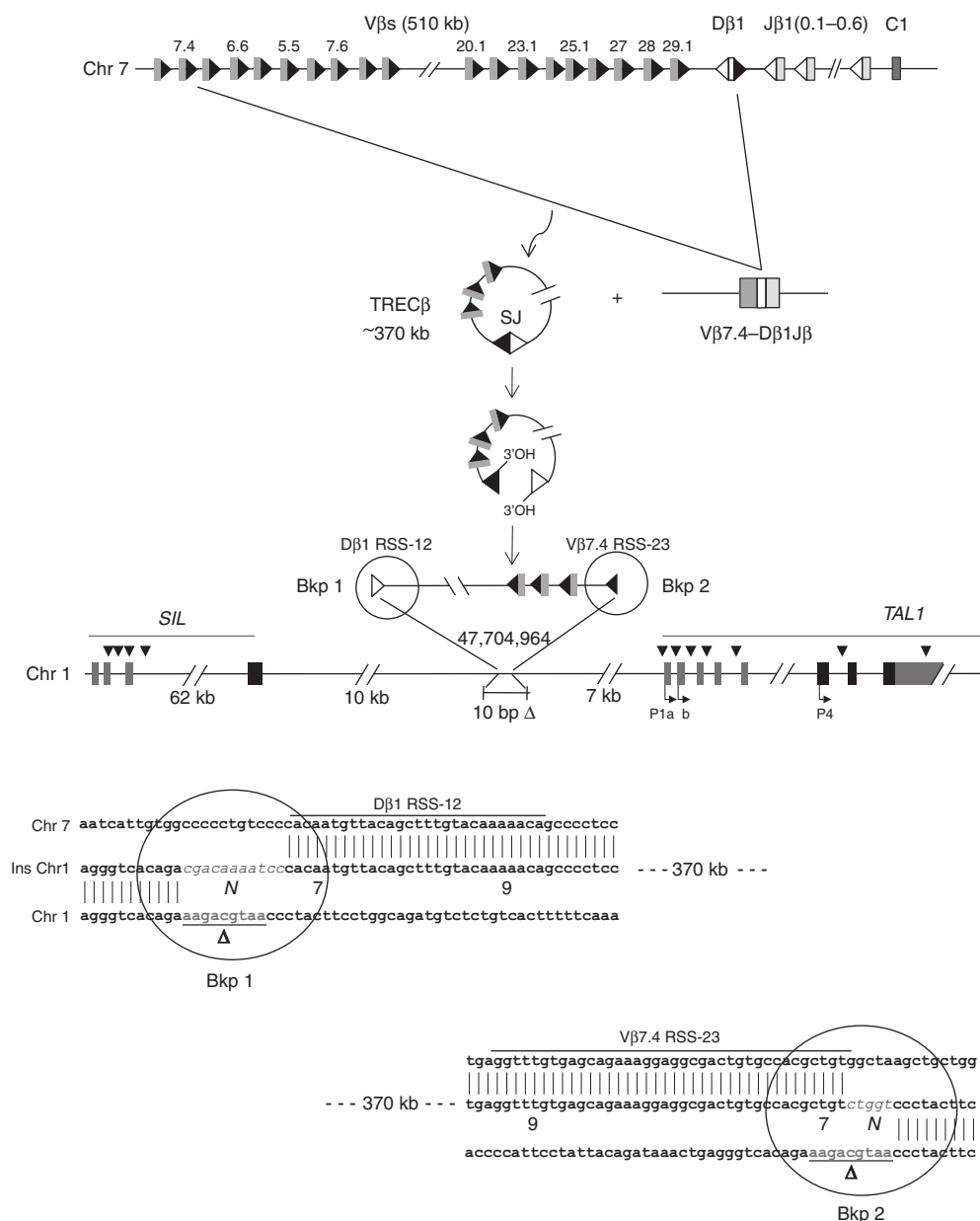


Figure 4 | Schematic representation of the episomal reintegration in Patient OC. The TCRβ locus is displayed (top lane, not to scale). A functional Vβ7.4-to-Dβ1 rearrangement generating an excised TRECβ, and containing a (Vβ7.4/Dβ1) signal joint (SJ) is represented. The episome might have been open at the SJ by a nick-nick process⁴⁴ generating 3' hydroxyl ends before integration in chromosome 1. The episome is integrated in reverse orientation 10 kb downstream of the *STIL* gene, and 7 kb upstream of the *TAL1* gene (middle lane). A 10-bp deletion (Δ, underlined) occurred at the insertion site. Localization of cryptic RSSs used by illegitimate V(D)J-mediated *SIL*-*TAL1* deletion, and by t(1;14) TCRδ/*TAL1* translocations are indicated by black arrow heads. *TAL1* promoters (P1a, P1b, P4) are indicated. The breakpoints sequences (Bkp1/2) are shown (bottom lane). N, N regions; Vβ7.4 and Dβ1 RSSs are indicated, with heptamers (7) and nonamers (9) depicted.

therapeutic improvement. Over three decades of intense efforts in genomic research have allowed unravelling the extraordinary diversity of the mechanisms by which oncogenes are deregulated in this disease¹. Yet, a large number of major oncogene deregulations still remain unexplained to date. Among the diversity of mechanisms involved, V(D)J recombination-mediated alterations (translocations, microdeletions) constitute the hallmarks of T-ALLs^{23–26}. Interestingly, despite arrays of biochemical and functional evidence that the reintegration of excised episomal circles (TRECs) by the V(D)J recombinase (RAG1/2) might constitute a potent source of genomic instability, such events remained so far unreported in human cancer patients^{8,9,27–30}. Here we report the first case of such an

oncogenic RAG1/2-mediated episomal reintegration, demonstrating that TRECs can indeed contribute to human oncogenesis (Supplementary Fig. 1). Most intriguingly, this insertion occurred in a pediatric SCID-X1 patient who developed a leukaemia secondary to retroviral reinsertion (in front of LMO2, a known *TAL1*-cooperating oncogene) following gene therapy¹⁶. The screening of two other SCID-X1 patients' leukaemic samples^{31,32}, and of a large collection of T-ALLs did not reveal additional episomal insertions, and the reason for the extraordinary coincidence of two rare oncogenic integration events in this patient remains unanswered.

This and the other insertional mutagenesis T-ALL cases described here also revealed a novel oncogenic activation

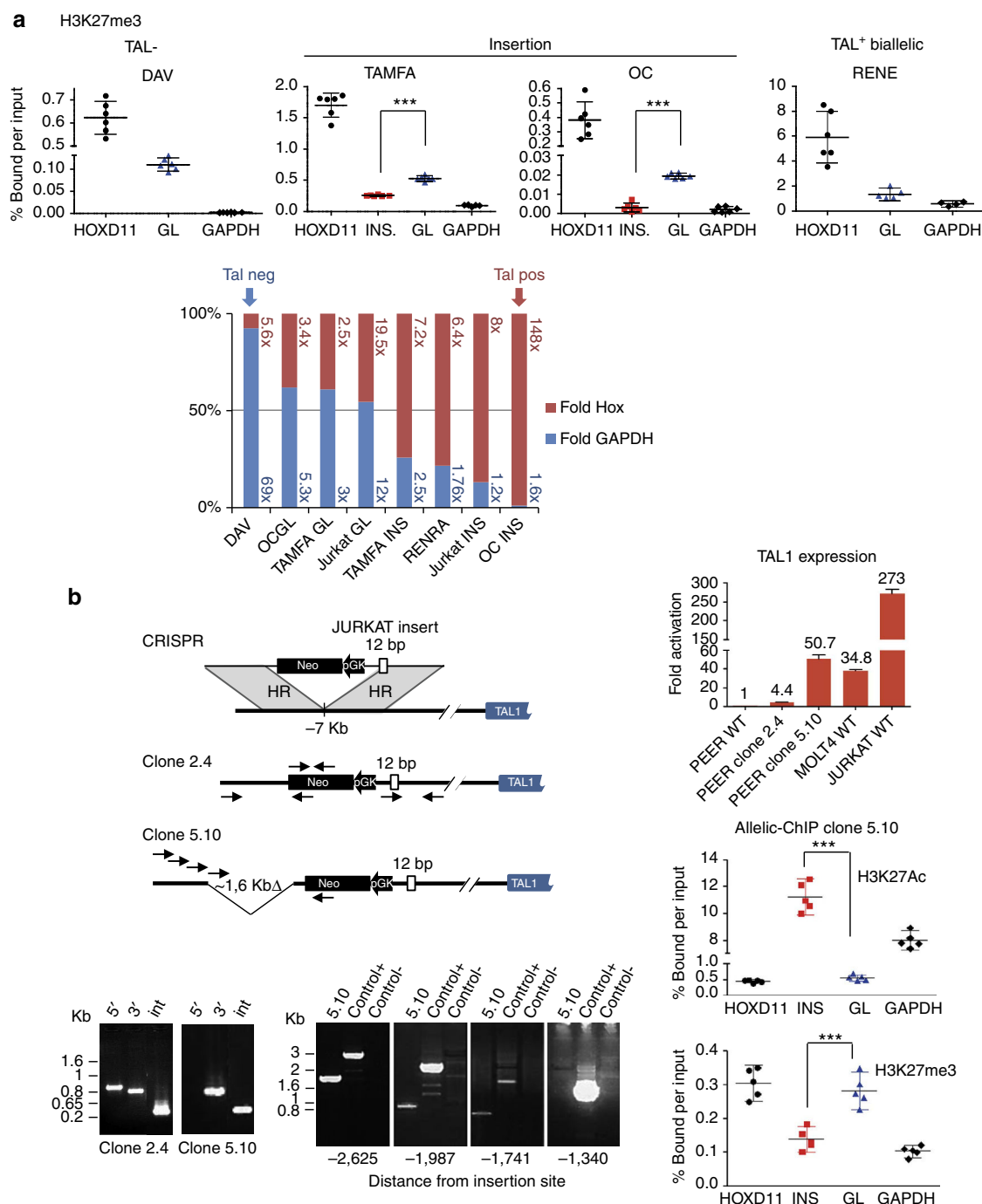


Figure 5 | Insertional mutagenesis is associated with epigenetic modulation and *TAL1* gene expression. (a) Allelic-ChIP analysis of H3K27me3 marks at the insertion breakpoint in primary patients. See legend to Fig. 2a. Marks at the GL alleles in one TAL1⁺ biallelic (RENE) and in one TAL1⁻ (DAV) patients were performed as controls. ****P* < 0.001, unpaired *t*-test; Relative-fold plots: Inserted and/or GL allele ChIP values were calculated as fold increases relative to GAPDH (numbers in blue) or HoxD11 (numbers in red), and folds plotted as relative percentages (GAPDH relative folds: blue histograms; HOXD11 relative folds: red histograms). Blue histograms over 50% indicate higher differences with the expressed than the repressed control genes and correspond to (partially) repressed *TAL1* expression; conversely, red histograms over 50% indicate higher differences with the repressed than the expressed control genes and correspond to (partially) derepressed *TAL1* expression; histograms are ordered according to decreasing *TAL1* repression. (b) Epigenetic modulation and *TAL1* gene expression by DNA editing mimicking insertional mutagenesis. Left panel: schematic representation of the CRISPR design for homologous recombination at the *TAL1* locus, and configuration of two edited clones in the PEER cell line. The locations of PCR primers (plain arrows) for detecting successful targeted events and for genome walking are indicated. Bottom left panel: successful homologous recombination was confirmed by PCR of the expected genome-donor and donor-insert boundaries. Top right panel: RQ-PCR analysis of *TAL1* expression after editing. Transcripts were normalized to ABL and reported as relative values to non-edited PEER cells. Four PCR replicates were performed on 1 (clone 2.4, due to impaired growth) or 2 (clone 5.10) independent RNA extractions. Bottom right panel: allelic-ChIP assays of H3K27me3 and H3K27ac marks in edited clone 5.10. See legend to Fig. 2a. ****P* < 0.001, unpaired *t*-test.

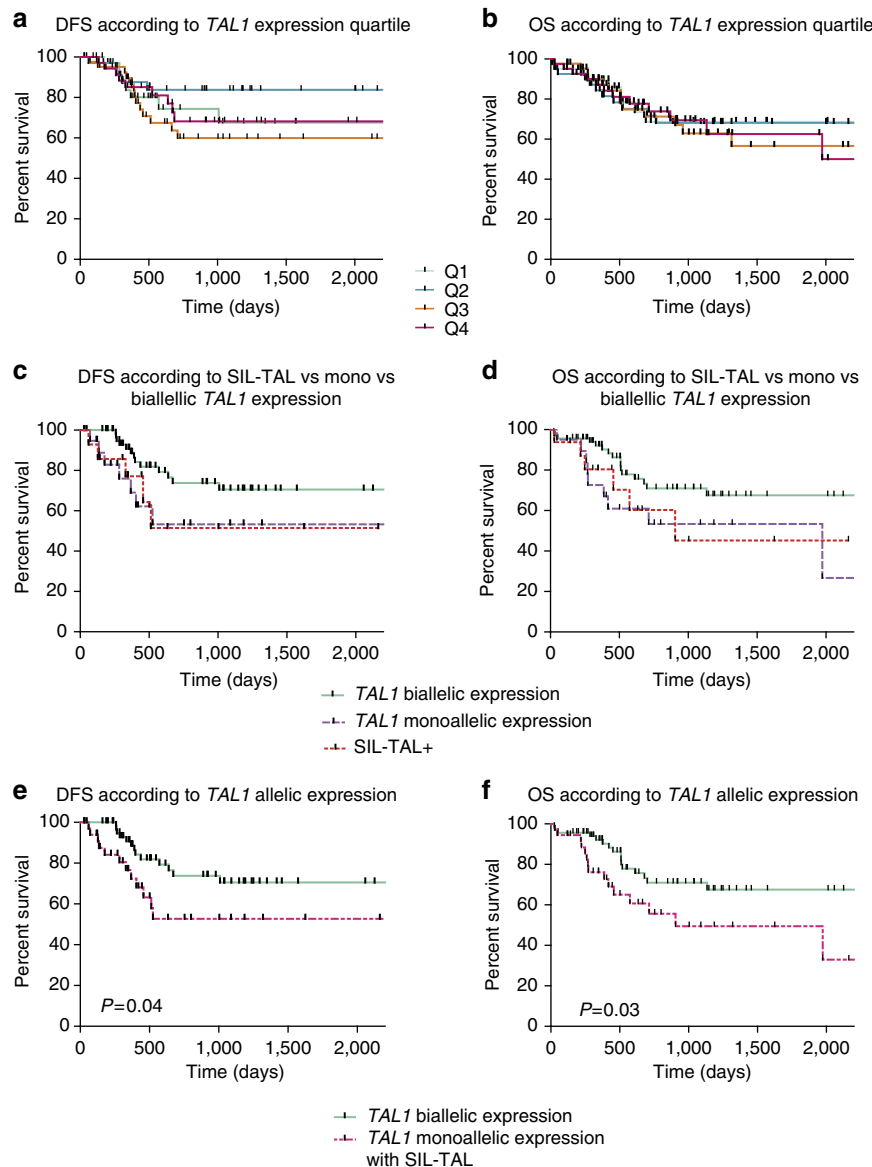


Figure 6 | Survival analysis. Kaplan-Meier analysis showing DFS and OS of 165 protocolar patients treated in the GRAALL trial according to: (a,b) *TAL1* expression quartiles; (c-f) the mode of *TAL1* expression. *P* values are indicated, log-rank (Mantle-Cox) test.

pathway, whereby a genetic alteration drives a site-specific and monoallelic epigenetic deregulation. We demonstrate that such insertions drove a switch from H3K27me3 to H3K27ac deposition, leading to the maintenance and/or re-expression of *TAL1* expression through T-cell differentiation. Interestingly, the difference in *TAL1* expression levels observed in mutants from the gene editing assay (Fig. 5) suggest that while small insertions might be sufficient to prevent the deposition of PcG repressive marks during T-cell lineage specification (thus permitting H3K27ac switch and maintenance of *TAL1* expression), further disruption of the region 5' of the insertion site (by deletion or insertional uncoupling) might be necessary to impose desilencing once *TAL1* extinction is established in the T-cell lineage⁶ (Supplementary Fig. 1). These kinetics are coherent with thymocyte ontogeny in patient #OC, in which TRECβ rearrangement/reintegration (DN2-3) likely occurred after *TAL1* silencing (DN1-2). Altogether, our data are in line with current models of permanent gene extinction of transcription factors during T-lineage commitment⁶ (Supplementary Fig. 1) and further identify locus control regions involved in deposition

and/or maintenance of *TAL1* silencing. Their genetic disruption constitute a recurrent epigenetic mechanism of *TAL1* deregulation in T-ALLs, contributing to a substantial fraction (>20%) of the *TAL1*⁺ monoallelic 'unresolved cases', and associated with adverse prognosis. That a *cis*-deregulation regrouping as diverse mechanisms as SIL-TAL deletions, translocations or insertional desilencing impact more on prognosis than *TAL1* expression levels underlines the fundamental oncogenic difference between a deregulation targeting a single locus, and the wider effect of *trans*-acting factors. Transcription factors indeed often bind to a large number of target genes (hundreds to thousands) and their deregulation (whether gain or loss) will likely affect a complex set of cellular functions, some of which might antagonize tumour progression, or resistance to treatment. Recently, reports identifying loss-of-function mutations in polycomb-related components^{19–21,33} have provided the framework by which global epigenetic modification might trigger the indirect (and biallelic) activation of numerous target genes, likely including a complex and conflicting set of oncogenes and tumour suppressors. In humans, PcG are recruited

to and repress specific regions in the genome through as yet undefined set(s) of DNA-binding transcription factors and long non-coding RNAs³⁴. The insertional mutagenesis described here identifies a site- and allele-specific switch from H3K27me3 recruitment/maintenance to H3K27ac, providing new avenues to decipher the mechanisms and DNA-binding intermediates involved³. A complex interplay between transcription factors and PcG would be in line with recent findings that Notch1 activation antagonizes PRC2 silencing of Notch1 target genes in T-ALL oncogenesis¹⁹. In intricate T-ALL networks where both *NOTCH* and *TAL1* deregulation can coexist, we find that while *NOTCH* patients are associated with a favourable prognosis, monoallelic, but not biallelic *TAL1*, expression tends to convert the clinical outcome towards more adverse prognosis in the Notch subgroup, and to further aggravate the bad prognosis of Notch1^{WT} patients. This suggests that distinct modes of deregulation of the same epigenetic complex might coexist in a tumour cell, leading to complex and potentially conflicting clinical outcomes which ought to be clarified when considering epigenetic inhibitors for new lines of treatment^{15,33,35}.

Methods

ChIP. ChIP was performed as previously described¹⁰. In brief, the cells were chemically crosslinked by the addition of one-tenth volume of fresh 11% formaldehyde solution for 10 min at room temperature. Following the quenching of the reaction with glycine (250 mM final concentration, 5 min, room temperature), cells were rinsed twice with $1 \times$ PBS and flash frozen in liquid nitrogen and stored at -80°C before use. Cells were resuspended, lysed and sonicated to solubilize and shear crosslinked DNA. Sonication was conducted using a Bioruptor (Next Gen, Diagenode) for 15 min (30 s on, 30 s off), resulting in sheared DNA between 100 and 400 bp with the bulk at ~ 250 bp. The resulting whole-cell extract was incubated overnight at 4°C with 100 μl of Dynal Protein G magnetic beads that had been preincubated with the appropriate antibody. The anti-H3K27me3 antibodies used were: # 07-449, Millipore (2 μg), and # ab6002, Abcam (1 μg); the anti-H3K27ac antibodies used were: #ab4729, Abcam (2 μg) and # 39133, active motif (5 μg). Beads were washed eight times with RIPA buffer and one time with TE containing 50 mM NaCl. Bound complexes were eluted from the beads by heating at 65°C with occasional vortexing, and crosslinking was reversed by overnight incubation at 65°C . Whole-cell extract DNA (reserved from the sonication step) was also treated for crosslink reversal. Immunoprecipitated DNA and whole-cell extract DNA were then purified by treatment with RNaseA, proteinase K and multiple phenol:chloroform: isoamyl alcohol extractions.

ChIP-seq. Before sequencing, ChIP DNA was quantified using the picogreen method (Invitrogen, USA) and quality controlled on a 2100 Bioanalyzer (Agilent). At least 1 ng of ChIP or input DNA was used for library preparation according to the Illumina ChIP-seq protocol. After end repair and adapter ligation, fragments were size selected on a gel before preamplification and clustering. The resulting fragments were again verified on a 2100 Bioanalyzer before clustering and 25 or 36 cycle sequencing on a Genome Analyzer II (Illumina, USA) according to the manufacturer's instructions. Raw data were bowtie aligned and the tags elongated and further processed to wiggle files as described³⁶. ChIP-seq experiments were performed on two independent biological replicates (samples for primary cells or culture for cell lines). Data presented in Fig. 1 are available under accession number GSE29362 for TBP, Pol II, H3K4me1, H3K4me3 and H3K36me3 or GSE38577 for H3K27me3(1A), and for H3K27me3 in human cells (1B) GSE12646 for HSC and erythroblast, GSE12889 for CD4+ and GSE59257 for the data newly generated (Jurkat and DA) in this article. Technical replicates were merged before alignment with Bowtie and the resulting BAM files were used as treatment in MACS2. Input data sets were used as control when available (Jurkat, DA). Enriched regions for genomic tracks of Fig. 1a,b were extracted as BED files using MACS2 peak detection algorithm with the following parameters: genome sizes = 2.70×10^9 for human and 1.87×10^9 for mouse, bandwidth = 300, model fold = (5, 50), q value cutoff = 5.00×10^{-2} , larger scaled towards smaller, λ range = 1,000–10,000 bps, Broad on.

Assessment of wt versus inserted ChIP-seq signal. To identify a difference in H3K27ac enrichment in ChIP-seq between alleles with and without insertion (12 bp) in Fig. 2b, we aligned all 25 bp reads of each experiment to this specific region of insertion using the R package 'Biostrings'. This alignment was performed against a region of 52 bp around the insertion site allowing for up to two mismatches, and the number of hits on forward and reverse strands was cumulated. For the mutant allele, the Jurkat insert sequence (5'-CCGTTTCCTAAG-3') was inserted in the reference sequence before alignment, extending the initial alignment region to a size of 64 bp.

Allelic-ChIP. Input and IP genomic DNAs were analysed by RT-PCR using power SybrGreen on a 7,500 Fast Real-time PCR system (Applied Biosystems). IgG control 'cycle over the threshold' Ct values were subtracted to Input or IP Ct values and converted into bound value by $2^{-(\text{IP Ct or input Ct} - \text{IgG IP Ct})}$. Allelic-ChIP was carried out using allele-specific primers (by substituting one of the germline primers with a primer located in or overlapping the insertion). EZH2 knockdown was achieved using a doxycycline-induced short hairpin RNA (sh-RNA)-targeting EZH2 (pTRIPZ-EZH2, openbiosystem # V21HS-63066). A non-silencing sh-RNA (pTRIPZ-NS) was used as control. Jurkat cells were electroporated and cells containing the pTRIPZ were selected on puromycin. Knockdown of EZH2 was obtained by the addition of doxycycline ($2 \mu\text{g ml}^{-1}$) to the cells during 10 days. Western blot was performed using the anti-EZH2 BD Biosciences # 612666.

Sequencing and SNP array. For Jurkat mapping, a region of 10 kb 5' of *TAL1* exon 1 was mapped on both alleles by LRPCR/cloning and standard Sanger sequencing as previously described³⁷ (see Supplementary Fig. 2 for details on mapping strategy). Identification of allelic variants (SNPs versus somatic indels/mutations) was performed with vector NTI using alignment against reference *TAL1* sequence and variants (<http://www.ncbi.nlm.nih.gov/SNP>; <http://projects.tcag.ca/variation>, Supplementary Fig. 8). For the sequencing screen on T-ALL patients and cell lines, a region of ~ 700 bp surrounding the Jurkat insertion site was directly PCR/sequenced on both strands in a subset of 93 samples and six cell lines. Heterogeneous sequences (ambiguous reading due to allelic differences) were systematically cloned, sequenced and analysed as above. For SNP array, hybridization on Affymetrix Cytogenetics SNP Array-6 was performed according to the manufacturer's recommendations. Data analysis was performed with Chromosome Analysis Suite software using the following settings: the CGH log2 copy number ratio for heterozygous deletion was defined as 0.5 to 1.5, whereas log2 copy number ratios < 1.5 were defined as homozygous deletions. Gene copy number (GCN) aberrations were compared with the Database of Genomic variants (<http://projects.tcag.ca/variation>) to study only non-variant GCV aberrations.

Patients. Diagnostic samples from a consecutive series T-ALLs from 165 adults (older than 16 years) included in GRAALL-03/05 trial (registration #NCT00327678 and #NCT00222027) were analysed for *TAL1* expression. Sample collection and analyses were approved by the local ethical committee. Informed consent was obtained from the patients or relatives in accordance with the Declaration of Helsinki, with the institutional review board approval of all involved hospitals. Diagnosis of T-ALL was based on the World Health Organization 2008 criteria, defined by expression of cytoplasmic and/or surface CD3, and negativity of CD19 and MPO, as reported³⁸. To be included, samples had to contain at least 80% of lymphoblasts. Immuno-geno/phenotyping and oncogene quantification were performed as previously described^{38,39}.

Cell lines. Cell lines used in this study were purchased from the ATCC collection and were mycoplasma free.

RQ-PCR. RNA was reverse transcribed using MMLV (Invitrogen). We used a TaqMan assay to quantify *TAL1* transcript with the following primers: *TAL1* F: 5'-ACA-ATC-GAG-TGA-AGA-GGA-GAC-CTT-C-3', *TAL1* Probe: fam-5'-CTA-TGA-GAT-GGA-GAT-GGA-GAT-TAC-TGA-TG-3'-tamra, *TAL1* R: 5'-ACG-CCG-CAC-AAC-TTT-GGT-G-3', 40 cycles were run on ABI 7500HT (Applied Biosystem) as described⁴⁰. *TAL1* transcript quantification was performed after normalization with the housekeeping gene *ABL* using the ΔCt method and results calculated according to the following formula $2^{\Delta(\text{CtABL}-\text{CtTAL1})}$.

***TAL1* allelic expression analysis.** Allelic expression was performed as previously described⁷. In brief, polymorphic markers in the 3' UTR of the *TAL1* gene were identified by PCR amplification and direct sequencing of 100 ng of genomic DNA. Allelic expression analysis was performed by PCR amplification and by direct sequencing of RT-PCR products from heterozygous patient samples. Three different PCRs were made to cover nine most frequent SNP among the 11 SNP previously described⁷.

Statistical analysis. *ChIP.* The power of t -test was estimated a priori using pwrR-package, and the expected variations between conditions evaluated. Differences in ChIP data between inserted and GL *TAL1* alleles were analysed by unpaired t -test. Samples collections constituted of five or six technical replicates were first checked for normal distribution using Kolmogorov-Smirnov test and the equality of variances was tested using F test. Results of t -test are shown as two-tailed P values. The statistical power of executed t -tests was at least 80%. Errors bars on histograms represent s.e.m.

OS/DFS. Patients' characteristics were compared using the Fisher's exact test. Median comparisons were performed using the Mann-Whitney U -test. OS and DFS were calculated from the date of prephase initiation. Events accounting for DFS were induction failure and first relapse from any cause in first CR. OS and DFS were estimated by the Kaplan-Meier method and then compared by

the log-rank test. All calculations were performed using the SPSS software, version 15.0 and the GraphPad Prism, La Jolla, CA, USA.

Ligation-mediated PCR. Genomic DNA was extracted using QIAamp deoxyribonucleic acid Blood mini kit according to the manufacturer's instructions (Qiagen). DNA (500 ng) was used for LM-PCR; DNA was digested by *DraI*, *EcoRV*, *PvuII*, *SmaI*, *SspI* or *StuI* restriction endonucleases. Purified digested DNAs were ligated with an adapter (composed of two complementary primers GWA⁺ and GWAB⁻). A primary PCR amplification was performed using an adapter-specific primer (AP1) and primer specific for the different TCR β gene segments (ext). A secondary PCR was performed using nested AP2 and TCR β primers (int), and analysed on 1% agarose gel. Non-germline PCR products were purified and sequenced. The functional and non-functional V(D)J rearrangements from patient OC were obtained using β 2.7-extB/intB and AP1/2 primers. The breakpoints corresponding to the episomal insertion were then obtained using V β 7.4-extB/intB, D β 1-extA/intA and AP1/2 primers, and validated by direct PCR using TAL1.OC.1B and V β 7.4-intB primers.

RT-PCR exon walking and 5' RACE. Total RNA was extracted using a column-based system RNeasy mini kit (Qiagen) according to the manufacturer's instructions. Reverse transcription was performed with SuperScript III Reverse transcriptase (Invitrogen) and random primers (Applied Biosystem). cDNAs were analysed by real-time quantitative PCR (RT-PCR) using power SybrGreen on an ABI PRISM 7500 (Applied Biosystems). All PCRs were performed in duplicate. 5'-RACE was performed using 2 μ g of total RNA and the 5'/3' RACE kit, 2nd generation (Roche). Modifications from the instruction manufacturer were the generation a poly(G) tailing of first strand cDNA and the use of an oligo d(C) anchor primer. PCR was performed using the Pfu Ultra II fusion HS DNA polymerase (Agilent technologies).

Genome editing in T-ALL cell line by type II CRISPR system. PEER T-ALL cells line were cultured in RPMI medium (Life Technologies) containing 20% fetal calf serum, 1% L-glutamine, 1% sodium pyruvate and 100 U ml⁻¹ penicillin/streptomycin (Life Technologies) at 37 °C in the presence of 5% CO₂. The day of transfection, 1 million cells were nucleofected according to the manufacturer's instruction (Lonza), with 500 ng DNA donor sequence containing Neomycin-resistant gene and 2 μ g of the Cas9/gRNA expression vector (Addgene #42230). The chimeric guide RNA targeted *TAL1* insertion site, and was cloned according to Cong *et al.*⁴¹. One day after nucleofection, cells were plated in 96-wells plate at 10⁴ cells per well and incubated in presence of 1,200 μ g ml⁻¹ geneticin G418 (Life Technologies) for 2 weeks. After selection and growing, a PCR was conducted to amplify the targeted region with genomic DNA derived from the surviving clones, and amplicons were separated on a 1% agarose gel then extracted with GEL/PCR clean up wizard (Promega) and sequenced (MWG-Biotech). CRISPR guide RNA: 5'-GAAAGACGTAACCTACTTCC-3'.

Primers list is available on request.

References

- Teitell, M. A. & Pandolfi, P. P. Molecular genetics of acute lymphoblastic leukemia. *Annu. Rev. Pathol.* **4**, 175–198 (2009).
- Van Vlierberghe, P., Pieters, R., Beverloo, H. B. & Meijerink, J. P. Molecular-genetic insights in paediatric T-cell acute lymphoblastic leukaemia. *Br. J. Haematol.* **143**, 153–168 (2008).
- Sanda, T. *et al.* Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. *Cancer Cell* **22**, 209–221 (2012).
- Mouthon, M. A. *et al.* Expression of tal-1 and GATA-binding proteins during human hematopoiesis. *Blood* **81**, 647–655 (1993).
- Herblot, S., Steff, A. M., Hugo, P., Aplan, P. D. & Hoang, T. SCL and LMO1 alter thymocyte differentiation: inhibition of E2A-HEB function and pre-T alpha chain expression. *Nat. Immunol.* **1**, 138–144 (2000).
- Zhang, J. A., Mortazavi, A., Williams, B. A., Wold, B. J. & Rothenberg, E. V. Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. *Cell* **149**, 467–482 (2012).
- Ferrando, A. A. *et al.* Biallelic transcriptional activation of oncogenic transcription factors in T-cell acute lymphoblastic leukemia. *Blood* **103**, 1909–1911 (2004).
- Hiom, K., Melek, M. & Gellert, M. DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell* **94**, 463–470 (1998).
- Agrawal, A., Eastman, Q. M. & Schatz, D. G. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* **394**, 744–751 (1998).
- Koch, F. *et al.* Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat. Struct. Mol. Biol.* **18**, 956–963 (2011).
- Cuddapah, S. *et al.* Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* **19**, 24–32 (2009).
- Cui, K. *et al.* Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell Stem Cell* **4**, 80–93 (2009).
- Zhou, Y. *et al.* Chromatin looping defines expression of TAL1, its flanking genes and regulation in T-ALL. *Blood* **122**, 4199–4209 (2013).
- McCabe, M. T. *et al.* EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* **492**, 108–112 (2012).
- Cardoso, B. A. *et al.* TAL1/SCL is downregulated upon histone deacetylase inhibition in T-cell acute lymphoblastic leukemia cells. *Leukaemia* **25**, 1578–1586 (2011).
- Howe, S. J. *et al.* Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J. Clin. Invest.* **118**, 3143–3150 (2008).
- Brandt, V. L. & Roth, D. B. V(D)J recombination: how to tame a transposase. *Immunol. Rev.* **200**, 249–260 (2004).
- Clappier, E. *et al.* The C-MYB locus is involved in chromosomal translocation and genomic duplications in human T-cell acute leukemia (T-ALL), the translocation defining a new T-ALL subtype in very young children. *Blood* **110**, 1251–1261 (2007).
- Ntziachristos, P. *et al.* Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. *Nat. Med.* **18**, 298–301 (2012).
- Simon, C. *et al.* A key role for EZH2 and associated genes in mouse and human adult T-cell acute leukemia. *Genes Dev.* (2012).
- Zhang, J. *et al.* The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* **481**, 157–163 (2012).
- Grossmann, V. *et al.* EZH2 mutations and their association with PICALM-MLLT10 positive acute leukaemia. *Br. J. Haematol.* **157**, 387–390 (2012).
- Van Vlierberghe, P. & Ferrando, A. The molecular basis of T cell acute lymphoblastic leukemia. *J. Clin. Invest.* **122**, 3398–3406 (2012).
- Marculescu, R., Le, T., Simon, P., Jaeger, U. & Nadel, B. V(D)J-mediated translocations in lymphoid neoplasms: a functional assessment of genomic instability by cryptic sites. *J. Exp. Med.* **195**, 85–98 (2002).
- Mendes, R. D. *et al.* PTEN microdeletions in T-cell acute lymphoblastic leukemia are caused by illegitimate RAG-mediated recombination events. *Blood* **124**, 567–578 (2014).
- Marculescu, R. *et al.* Distinct t(7;9)(q34;q32) breakpoints in healthy individuals and individuals with T-ALL. *Nat. Genet.* **33**, 342–344 (2003).
- Vanura, K. *et al.* In vivo reinsertion of excised episomes by the V(D)J recombinase: a potential threat to genomic stability. *PLoS Biol.* **5**, e43 (2007).
- Reddy, Y. V. R., Perkins, E. J. & Ramsden, D. A. Genomic instability due to V(D)J recombination-associated transposition. *Genes Dev.* **20**, 1575–1582 (2006).
- Curry, J. D. *et al.* Chromosomal reinsertion of broken RSS ends during T cell development. *J. Exp. Med.* **204**, 2293–2303 (2007).
- Messier, T. L., O'Neill, J. P., Hou, S. M., Nicklas, J. A. & Finette, B. A. In vivo transposition mediated by V(D)J recombinase in human T lymphocytes. *EMBO J.* **22**, 1381–1388 (2003).
- Hacein-Bey-Abina, S. *et al.* Sustained correction of X-linked severe combined immunodeficiency by *ex vivo* gene therapy. *New Engl. J. Med.* **346**, 1185–1193 (2002).
- Hacein-Bey-Abina, S. *et al.* LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302**, 415–419 (2003).
- Ntziachristos, P. *et al.* Contrasting roles of histone 3 lysine 27 demethylases in acute lymphoblastic leukaemia. *Nature* **514**, 513–517 (2014).
- Bracken, A. P. & Helin, K. Polycomb group proteins: navigators of lineage pathways led astray in cancer. *Nat. Rev. Cancer* **9**, 773–784 (2009).
- Van der Meulen, J., Van Roy, N., Van Vlierberghe, P. & Speleman, F. The epigenetic landscape of T-cell acute lymphoblastic leukemia. *Int. J. Biochem. Cell Biol.* (2014).
- Fenouil, R. *et al.* CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Res.* **22**, 2399–2408 (2012).
- Roulland, S. *et al.* Follicular lymphoma-like B cells in healthy individuals: a novel intermediate step in early lymphomagenesis. *J. Exp. Med.* **203**, 2425–2431 (2006).
- Asnafi, V. *et al.* Analysis of TCR, pT alpha, and RAG-1 in T-acute lymphoblastic leukemias improves understanding of early human T-lymphoid lineage commitment. *Blood* **101**, 2693–2703 (2003).
- Asnafi, V. *et al.* Age-related phenotypic and oncogenic differences in T-acute lymphoblastic leukemias may reflect thymic atrophy. *Blood* (2004).
- Bergeron, J. *et al.* Prognostic and oncogenic relevance of TLX1/HOX11 expression level in T-ALLs. *Blood* **110**, 2324–2330 (2007).
- Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
- Mochizuki-Kashio, M. *et al.* Dependency on the polycomb gene *Ezh2* distinguishes fetal from adult hematopoietic stem cells. *Blood* **118**, 6553–6561 (2011).

43. Delabesse, E. *et al.* TAL1 expression does not occur in the majority of T-ALL blasts. *Br. J. Haematol.* **102**, 449–457 (1998).
44. Neiditch, M. B., Lee, G. S., Huye, L. E., Brandt, V. L. & Roth, D. B. The V(D)J recombinase efficiently cleaves and transposes signal joints. *Mol. Cell* **9**, 871–878 (2002).

Acknowledgements

This work was supported by grants to BN laboratory from la Fondation pour la Recherche Médicale (#INE2003114116), Le Conseil Général des Bouches du Rhône, le Canceropôle PACA, CALYM consortium, l'Institut National du Cancer (INCa PLBIO09), la Fondation de France (#2008001490) and institutional grants from INSERM, CNRS and AMU. M.L. is a recipient of a fellowship from INCa (#ASC12035ASA). B.N. is the recipient of a CHRT INSERM/Assistance Publique-Hôpitaux de Marseille (HP-HM). Work in the Necker laboratory was supported by grants from the 'Association Laurette Fugain', the Institut National du Cancer and the 'Comité départemental de la Ligue Contre le Cancer'. S.L.N. was supported by the 'Institut National du Cancer (INCa-DHOS) and a grant from the 'Association pour la Recherche sur le Cancer (ARC)'. R.F. is supported by a Centre National de la Recherche Scientifique grant. M.A.M. was supported by an ANR chromaTin and by the Fondation pour la Recherche Médicale (FRM AJE20130728183). The sequencing costs for ChIP-seq experiments were partially granted by an ESGI consortium grant from the Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 262055. E.D. gratefully acknowledges support from the Association pour la Recherche sur le Cancer and the Fondation pour la Recherche Médicale. We thank participants in the GRAALL-2003 and -2005 study groups for collecting and providing biological samples and data sets. The GRAALL was supported by grants P0200701 and P030425/AOM03081 from the Programme Hospitalier de Recherche Clinique, Ministère de l'Emploi et de la Solidarité, France and the Swiss Federal Government in Switzerland. Samples were collected and processed by the AP-HP 'Direction de Recherche Clinique' Tumour Bank at Necker-Enfants Malades. We are grateful to V. Lheritier for her help in collecting clinical samples, L. Spinelli for statistical advice, T. Guignard for Macs2

analyses and C Pignon for technical help. We are grateful to Dr T Cathomen for expert advices in DNA editing.

Author contributions

ChIP-seq experiments were performed and analysed by L.C.P., R.F., M.A.M., M.G., I.G.G. and J.-C.A.; allelic-ChIP experiments were performed and analysed by J.-M.N., L.C.P., M.L., M.K., C.P., D.P.-B. and E.D.; T-ALL sample characterizations were performed and analysed by A.T., J.-M.N., L.C.P., S.L.N., S.J., L.S., E.A.M. and V.A.; TALEN/CRISPR assays were performed and analysed by J.-M.N., M.L. and D.P.-B. with the advice of C.G. and B.M.; episomal reintegration characterization was performed by J.-M.N., E.A.M., S.H.-B.-A., S.J.H., H.B.G. and A.J.T.; survival analysis was carried out by A.T., H.D., E.A.M., N.I. and V.A.; V.A. and B.N. conceived and directed the project, and B.N. wrote the paper.

Additional information

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

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

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

How to cite this article: Navarro, J.-M. *et al.* Site- and allele-specific polycomb dysregulation in T-cell leukaemia. *Nat. Commun.* 6:6094 doi: 10.1038/ncomms7094 (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/>

Discussion & Perspectives

I. Interaction entre PLZF et EZH2

Les protéines du groupe Polycomb (PcG) sont des facteurs épigénétiques dont le rôle est de maintenir la répression de leurs gènes cibles au niveau de la chromatine via la modification des protéines histones. EZH2 est une protéine clé dans les mécanismes de régulation puisqu'elle catalyse la mise en place de la marque répressive H3K27me3. Dans le cadre de ma thèse, je me suis intéressée, entre autre, au modèle des leucémies aiguës myéloïdes (LAM) dans lesquelles, contrairement à d'autres pathologies myéloïdes, des mutations affectant EZH2 ou des membres Polycomb ne sont retrouvées que très rarement (Greenblatt and Nimer, 2014). Des études ont montré que dans ce type de leucémies, de nombreux gènes cibles d'EZH2 sont dérégulés bien que son activité répressive soit toujours présente, mettant en évidence d'éventuels défauts de recrutement de cette protéine. Parmi les facteurs de transcription susceptibles de réguler l'association des protéines PcG à la chromatine, le facteur de transcription et suppresseur de tumeur PLZF représentait un candidat intéressant. En effet, des études ont montré une interaction entre BMI-1, membre de PRC1, et PLZF (Barna et al., 2002; Boukarabila et al., 2009; Spicuglia et al., 2011).

Nous avons donc étudié l'interaction fonctionnelle entre PLZF et EZH2 à la chromatine dans un modèle cellulaire de LAM. Des expériences de co-immunoprécipitation dans la lignée myéloïde KG1 (cellules CD34 positives) complétées par des expériences de « domain mapping » ont montré que PLZF et EZH2 pouvaient interagir *in vivo* via le domaine BTB de PLZF. Le domaine BTB est un domaine qui est responsable dans un grand nombre de cas des interactions entre PLZF et ses partenaires (Suliman et al., 2012). Cependant dans le cas de notre étude, nous ne connaissons pas la nature de l'interaction (directe ou non) de PLZF et EZH2. Des expériences de double hybride ont été menées afin de répondre à cette question. A ce jour, les résultats ne sont pas assez satisfaisants sur le plan qualité pour tirer de conclusion définitive, mais il semblerait que l'interaction entre les deux protéines ne soit pas directe. Des expériences supplémentaires sont en cours afin d'apporter une réponse définitive sur ce point et permettraient d'éclaircir la relation entre PLZF et EZH2.

Par ailleurs, lors de la réalisation de mes travaux, une étude de protéomique faite sur des cellules NKT murines a montré une interaction indirecte entre PLZF et EZH2 via l'ubiquitine ligase Cul3 (Mathew et al., 2012). Ces éléments ont renforcé l'intérêt d'étudier l'interaction de PLZF et EZH2 au niveau de la chromatine.

II. Expériences de ChIP-seq : pertinence des données et artéfacts éventuels liés à l'analyse

Afin de caractériser la localisation génomique de PLZF et EZH2, nous avons effectué des expériences de ChIP suivies de séquençage à haut débit. Lorsque j'ai commencé le projet, ce type d'étude n'avait jamais été réalisé au sein du laboratoire (CRCM). Si la partie séquençage a été confiée à une plateforme, j'ai dû mettre au point une démarche expérimentale concernant la préparation des échantillons puis l'analyse des données avec l'aide de la plateforme bioinformatique (voir page 42). Une des étapes critiques dans l'analyse des données de ChIP-seq est de définir une liste de gènes cibles. Il existe de nombreux algorithmes permettant de récupérer ces listes à partir des données de séquençage faisant appel à la notion de « peak calling » (ou appel de pic). En effet, un grand nombre de logiciels a été développé, proposant chacun des paramètres propres qui permettent de transformer une intensité de signal en un pic qui sera considéré comme une région liée par un facteur donné. Nous avons utilisé diverses méthodes de « peak calling » et avons retenu le peak caller du logiciel Homer (Heinz et al., 2010). Cet outil présente l'avantage de pouvoir utiliser le même peak caller pour les facteurs de transcription et pour les marques histones. En fonction du logiciel utilisé et même en fonction des paramètres et seuils que nous fixons, les résultats peuvent varier considérablement. Ainsi, par exemple, le gène *KLF2* n'est pas présent dans le peak calling que nous avons retenu mais c'est cependant un gène que j'ai réussi à valider expérimentalement et qui apparaît comme étant cible de PLZF lorsqu'on modifie les paramètres du peak calling. Ces méthodes génèrent donc logiquement des faux positifs et des faux négatifs. Afin de choisir à partir de quelle liste de gènes nous allions mener la suite de l'étude, nous nous sommes basés sur les données de la littérature (ie les paramètres classiquement utilisés pour ce type de données), sur notre expérience « visuelle » (les pics appelés apparaissaient-ils sur le genome browser), et sur les données publiées. Ainsi, pour EZH2, plusieurs gènes cibles que nous avons identifiés ont déjà été décrits. Il s'agit par exemple de certains gènes *HOX*, du gène suppresseur de tumeur *CDKN2A* ou encore du gène *KLF2* qui se trouvent justement dérégulés dans les cancers (Ezhkova et al., 2009; Taniguchi et al., 2012). Les gènes cibles de PLZF sont un peu moins décrits mais nous avons retrouvé dans notre peak calling certains gènes tels que *DUSP6*, *GFI1* ou encore *ID2* qui ont été décrits comme étant la cible de PLZF dans les KG1a (Doulatov et al., 2009). Ces données nous ont donc rassurés sur la pertinence de notre analyse et sur la qualité de nos expériences.

III. EZH2 et activation des gènes

Au cours de notre étude, nous avons mis en évidence deux types de gènes cibles de la protéine EZH2 : les gènes cibles d'EZH2 associés à la marque répressive H3K27me3 et les gènes cibles d'EZH2 associés à PLZF et à la marque activatrice H3K4me3. Si la marque H3K4me3 est synonyme d'activation des gènes, elle ne signifie pas forcément que les gènes sont transcriptionnellement actifs. C'est pourquoi nous avons complété notre analyse en étudiant l'ARN polymérase II (responsable de la transcription des gènes) et sa marque associée, l'H3K36me3. Ces données nous ont permis de constater que les gènes EZH2/PLZF/H3K4me3 sont effectivement des gènes exprimés.

Si la grande majorité des études portant sur EZH2 associe cette protéine à la marque H3K27me3 et à la répression des gènes, quelques études dans lesquelles EZH2 joue un rôle d'activateur transcriptionnel se démarquent. La plupart de ces études rapportent un rôle activateur d'EZH2 dans des contextes pathologiques, notamment dans les cancers du sein et de la prostate (voir la section de l'introduction intitulée « EZH2 : un activateur transcriptionnel dans les cancers »). Toutefois, nos résultats, obtenus dans des cellules leucémiques, montrent que les gènes régulés par PLZF et EZH2 sont également exprimés dans les cellules CD34 positives non leucémiques ; indiquant qu'EZH2 peut également être associée à une transcription active dans un contexte non pathologique. En effet, dès 1996, une équipe se penche sur la possibilité que la protéine E(z) (EZH2 chez la Drosophile), puisse jouer le rôle d'une protéine du groupe trithorax. En effet, cette étude montre qu'E(z) peut agir comme un membre trithorax ou polycomb en fonction de signaux spatiaux ou temporels ou en fonction du locus ciblé (LaJeunesse and Shearn, 1996). D'autres études plus récentes montrent une association des complexes PRC1 et PRC2 à l'ARN polymérase II ou à des promoteurs actifs dans les cellules ES (Brookes et al., 2012; Kaneko et al., 2013; Mousavi et al., 2012). Il est donc évident que l'activité enzymatique d'EZH2 peut être modulée par différents mécanismes moléculaires dont les bases ne sont encore pas complètement connues.

De façon intéressante, nous avons pu observer par ChIP-qPCR et sur les profils de ChIP-seq que le niveau d'H3K4me3 est beaucoup plus faible sur les cibles communes PLZF/EZH2 que sur les cibles PLZF seules. Par ailleurs, les gènes ciblés uniquement par EZH2 sont enrichis en H3K27me3. Il semblerait cependant qu'il n'y ait pas de différence d'activité transcriptionnelle lorsque les gènes sont associés à EZH2/PLZF ou à PLZF seul, comme le montre notre analyse des gènes présents et absents. Le plus faible niveau d'H3K4me3 observé au niveau des gènes PLZF/EZH2 pourrait s'expliquer par une activité

H3K4 méthyltransférase moins performante en présence d'EZH2. Pour étayer cette hypothèse nous avons étudié des interactions potentielles entre PLZF, EZH2 et MLL1, une H3K4 méthyltransférase impliquée dans de nombreux cancers (Li and Ernst, 2014). Nous avons constaté, via des expériences préliminaires de co-immunoprécipitation, que MLL1 interagit avec EZH2 mais n'interagit pas avec PLZF. L'association entre une protéine Polycomb et une protéine Trithorax qui ont des fonctions antagonistes (Steffen and Ringrose, 2014), pourrait s'expliquer par un effet régulateur négatif d'une des deux protéines sur l'autre. Une des possibilités serait qu'EZH2 empêche l'accès de MLL1 à PLZF ce qui expliquerait le niveau plus faible d'H3K4me3. Il faudrait faire des expériences de compétition afin de voir quelle protéine est la plus affine pour PLZF. Par ailleurs, la protéine MLL1 n'est pas la seule à catalyser la marque H3K4me3, il faudrait donc étudier les autres protéines (Ernst and Vakoc, 2012).

Enfin, plusieurs études ont montré que PRC2 pouvait être recruté au niveau de gènes activement transcrits via les protéines PCL (PolyComb like) qui peuvent se lier à l'H3K36me3 (Ballaré et al., 2012; Cai et al., 2013). Par exemple, il a été montré qu'une protéine PCL, la protéine PHF19, interagit à la fois avec PRC2 et avec l'H3K36me3 déméthylase NO66. Il est donc possible que les protéines PCL recrutent PRC2 afin d'initier la répression de gènes actifs (Brien et al., 2012). Ce mécanisme pourrait être appliqué dans notre système dans lequel EZH2 serait recruté à la chromatine par un facteur non encore déterminé pour contrer les effets activateurs de PLZF. Ce qui expliquerait par ricocher la diminution d'H3K4me3 sur les cibles communes PLZF/EZH2 par rapport aux cibles PLZF seul. Cette hypothèse est soutenue par le fait que nous montrons ici qu'une deuxième protéine du complexe PRC2, la protéine SUZ12, interagit avec PLZF, suggérant que c'est le complexe répresseur PRC2 qui est recruté et pas la protéine EZH2 seule.

A contrario, une étude récente montre que selon sa localisation génomique, EZH2 catalyse la mono, di ou triméthylation de l'H3K27, ce qui implique des fonctions différentes au niveau transcriptionnel (Ferrari et al., 2014). En effet, d'après cette étude, EZH2 catalyse la marque H3K27me2, qui est une marque globalement diffuse mais dont le rôle est d'empêcher l'activation inappropriée de certains enhancers. Dans ce cas, la marque H3K27me2 a un rôle répresseur. Mais EZH2 intervient également dans la catalyse de la marque H3K27me1 qui se situe dans les régions intragéniques de gènes actifs au niveau transcriptionnel. La marque H3K27me1 étant favorisée par celle de l'H3K36me3 qui inhiberait la marque répressive H3K27me2. Ce modèle a été décrit chez la souris et il aurait été intéressant de voir, si dans notre système, les gènes communs PLZF et EZH2 sont associés à la marque H3K27me1. Il

serait envisageable, dans un tel cas, qu'EZH2 soit à l'origine de la réactivation transcriptionnelle des gènes.

IV. Quel est le rôle de la protéine EZH1 ?

Notre étude à l'échelle génomique de l'interaction entre PLZF et EZH2 révèle un certain nombre de gènes associés à la marque H3K36me3, synonyme de la présence de la RNA-PolII, responsable de la transcription.

Un article a mis évidence la co-occupation au niveau génomique d'EZH1 avec les marques H3K4me3, H3K36me3 et l'ARN PolII (Mousavi et al., 2012). Bien qu'ayant utilisé un anticorps spécifique d'EZH2 dans notre étude, nous avons voulu écarter l'hypothèse selon laquelle les gènes associés à PLZF étaient également des cibles d'EZH1 et non EZH2. Mais les approches de ChIP-qPCR et les co-IP que nous avons utilisées n'ont pas montré de corrélation entre PLZF et EZH1. Cependant, dans notre système, on note qu'il existe bien une interaction entre EZH1 et EZH2 qui donc ne forment pas de complexes PRC2 exclusifs l'un par rapport à l'autre. Il a cependant été montré qu'EZH1 et EZH2 régulent la répression transcriptionnelle via des mécanismes différents (Margueron et al., 2008), il aurait donc été intéressant de faire une étude comparative de la localisation génomique et des gènes cibles d'EZH1 et d'EZH2.

V. Double rôle de PLZF sur la transcription de ses gènes cibles

Comme beaucoup de facteurs de transcription, PLZF peut agir en répresseur ou en activateur de la transcription. Historiquement PLZF a été décrit comme étant un facteur recrutant des co-répresseurs tels que les HDACs ou certaines protéines Polycomb (Barna et al., 2002; Boukarabila et al., 2009; Lin et al., 1998). Toutefois, nous montrons ici que dans le contexte myéloïde, la protéine PLZF est associée à une transcription active.

Le rôle activateur de PLZF a été décrit et attribué aussi bien à des mutations de PLZF (Melnick et al., 2000) qu'à son rôle biologique (Doulatov et al., 2009; Xu et al., 2009). Nous avons, au vu de nos résultats, fait séquencer l'exon 2 de PLZF et n'avons trouvé aucune mutation. Nous nous sommes donc questionné sur les mécanismes biologiques pouvant expliquer ce double rôle.

Dans le développement myéloïde et au sein du système immunitaire, le rôle activateur de PLZF a déjà été décrit (Doulatov et al., 2009; Xu et al., 2009). Nous pouvons donc penser que l'effet activateur ou répresseur de PLZF est tissu-spécifique. Quelques éléments viennent

cependant démentir cette hypothèse dans la mesure où, dans le système immunitaire, PLZF peut à la fois induire l'expression de certains gènes de réponse à l'interféron (Xu et al., 2009) et réprimer les gènes de la réponse inflammatoire (Sadler et al., 2015). Par ailleurs, une étude de ChIP-seq menée sur des cellules spermatogoniales immatures de souris montre que PLZF peut activer ou réprimer certains gènes, démontrant le double rôle de PLZF au sein d'une même population (résumé de thèse du Dr Song Yuntao : <http://search.proquest.com/docview/1541549288>). De plus, l'étude du rôle de PLZF dans l'hématopoïèse normale (Vincent-Fabert *et al.*, en révision), suggère que PLZF peut activer un set de gènes, bien que majoritairement les gènes soient réprimés et associés à la marque répressive H3K27me3. Ces données suggèrent donc que le double rôle de PLZF n'est pas tissu ou cellule-spécifique.

Le double effet transcriptionnel de PLZF pourrait cependant s'expliquer par une régulation très fine de ses modifications post-traductionnelles. En effet, comme décrit dans l'introduction, PLZF peut être l'objet de différentes modifications post-traductionnelles influençant son activité. Ainsi, l'acétylation de PLZF augmente ses capacités répressives (Guidez et al., 2005; Sadler et al., 2015), tandis sa phosphorylation sur la sérine 76 lui confère un effet activateur (Xu et al., 2009). Si nous reprenons l'exemple de la régulation des gènes par PLZF en réponse à l'interféron, il est possible de supposer que la stimulation à l'interféron induise la désacétylation de PLZF en provoquant l'interaction de PLZF avec l'HDAC1. Ce qui a pour effet de rendre prioritaire la forme phosphorylée de PLZF dans les cellules et d'induire son effet activateur (Ozato, 2009).

Dans notre système cellulaire, nous savons déjà que PLZF est sujet à des modifications post-traductionnelles (observations sur western blot de bandes spécifiques de PLZF à une taille supérieure à celle de la bande majoritaire). Il serait donc très intéressant d'étudier dans notre système cellulaire de quelle(s) modification(s) PLZF est l'objet.

Par ailleurs, notre travail montre que dans les cellules myéloïdes, PLZF a un effet répresseur sur la transcription d'un petit groupe de gènes : les gènes bivalents. Ces gènes sont ciblés par EZH2 et caractérisés par un enrichissement concomitant des marques H3K4me3 et H3K27me3 (Voigt et al., 2013). De façon intéressante, sur ces gènes, la diminution d'expression de PLZF entraîne une diminution d'H3K27me3. Il a été montré que les complexes PRC1 et PRC2 sont recrutés à la chromatine au niveau des gènes bivalents. Dans ce contexte précis, et sachant que la diminution de l'expression de PLZF induit la diminution de l'H3K27me3 pour ces gènes, nous pouvons penser que, dans ce cas, PLZF joue le rôle de recruteur d'EZH2.

Enfin, notre étude révèle que PLZF et EZH2 co-localisent non seulement au niveau des promoteurs et des régions intragéniques mais aussi au niveau des régions intergéniques. Ces observations que je n'ai pu exploiter pendant mon travail de thèse seront la base d'études futures. En effet, il sera intéressant d'étudier la présence des marques histones et de PLZF et EZH2 au niveau d'autres régions génomiques, telles que, par exemple, les régions enhancers. Ces régions se sont avérées être la clé de l'identité cellulaire, puisque leur activité permet l'expression tissulaire spécifique de gènes (Whyte et al., 2013). De plus, ces régions se trouvent fréquemment dérégulées dans les cancers (Shi et al., 2013), et voir paragraphe suivant.

VI. Exemple d'une dérégulation d'enhancer par insertion : cas du gène TAL1 (article 3)

La protéine TAL1 est un oncogène qui joue un rôle majeur dans les LAL-T puisque 40% des patients la surexpriment (cas LAL-T TAL1+). L'expression ectopique de TAL1 est alors soit mono-allélique, correspondant à une altération en *cis* du gène, soit bi-allélique, reflétant une dérégulation indirecte en *trans* (Ferrando, 2004). Toutefois, dans environ la moitié des cas de LAL-T TAL1+, le mode d'activation de TAL1 est inexpliqué.

1. Activation mono-allélique de TAL1

L'étude menée en collaboration avec l'équipe du Dr B. Nadel nous a permis de mettre en évidence un nouveau mécanisme d'activation de TAL1 dans les LAL-T, consistant en une dérépression épigénétique du locus *TAL1*. En effet, il existe une microinsertion en amont du gène *TAL1* (TALins) qui entraîne la perte de la marque répressive H3K27me3 au niveau de ce locus. De façon intéressante, plusieurs études ont rapporté l'activation transcriptionnelle de TAL1 suite à son interaction avec des complexes permettant le remodelage chromatinien tels que mSin3a et HP1 (Huang and Brandt, 2000; Wen et al., 2005).

2. Création d'un super enhancer en amont du gène TAL1

La séquence TALins est impliquée dans la dérégulation de la répression du locus *TAL1* via une perte de l'H3K27me3. Ce locus représente donc a priori un très bon modèle pour étudier le mécanisme de ciblage par PRC2. Nous avons donc recherché par la technique de CHIP la présence des protéines EZH2 ou SUZ12 au niveau des allèles sauvages ou mutés.

Aucune des deux protéines n'a été retrouvée au niveau du locus TAL1 (données non publiées), ce qui pourrait s'expliquer par le fait que l'apposition de la marque H3K27me3 sur ce locus est générée lors de la maturation des cellules T. Il est donc possible qu'une fois les cellules différenciées, le complexe PRC2 ne soit plus présent au niveau du locus. Mais il est également possible que le complexe PRC2 ne se situe pas exactement au niveau du lieu de l'insertion mais plus loin sur le locus. On pourrait donc penser que l'insertion, au lieu d'empêcher le recrutement du complexe PRC2, permettrait plutôt le recrutement d'activateurs, de déméthylases et d'acétyltransférases.

Une étude concomitante à la nôtre a alors montré que ce site d'insertion créé en amont du locus TAL1 était un site de fixation pour le facteur de transcription MYB (Mansour et al., 2014). L'insertion de ce site de fixation est à la base de la formation *de novo* d'un super-enhancer. Un super enhancer est une structure caractérisée par un fort enrichissement des marques H3K27ac et H3K4me1 sur une région étendue. La fixation de MYB au site d'insertion entraîne le recrutement du co-activateur transcriptionnel CREB binding protein (CBP), qui facilite l'H3K27ac (Tie et al., 2014), et d'un complexe constitué de GATA3, RUNX1 et de la protéine TAL1 elle-même (Mansour et al., 2014; Sanda et al., 2012).

Enfin, dans cette étude, nous montrons que la mutation TALins explique environ 20% des cas de LAL-T TAL1+ mono-allélique non résolu. Pour les cas restants, le séquençage complet du locus pourrait révéler l'existence d'autres mutations.

Conclusion

L'hématopoïèse est un processus biologique finement contrôlé par un grand nombre de mécanismes dont la dérégulation peut conduire au développement d'hémopathies malignes, telles que les leucémies qui résultent de la transformation des cellules hématopoïétiques.

Au cours de ces dernières années, de nombreuses perturbations de la régulation épigénétique au sein des cellules hématopoïétiques ont été décrites comme étant impliquées dans le développement des leucémies. C'est pourquoi il est nécessaire d'étudier les mécanismes engagés dans ces phénomènes. Mon travail de thèse met en évidence toute la complexité de la régulation épigénétique au sein de l'hématopoïèse. Il existe une balance permanente entre répression et activation transcriptionnelle comme l'illustre l'exemple de l'interaction au niveau génomique du facteur de transcription PLZF avec le facteur épigénétique EZH2.

Toutefois, chaque système est différent et la spécificité tissulaire est un phénomène à prendre en compte spécialement pour le développement de thérapie. En effet, si nous avons pointé du doigt l'effet majoritairement activateur de PLZF et EZH2 dans notre lignée cellulaire KG1, les résultats obtenus sur des CSH murines sont différents. Nous avons ainsi pu mettre en évidence le rôle répresseur de PLZF sur les gènes du cycle cellulaire. Au cours de cette étude nous avons également montré que le rôle régulateur de PLZF passait par la modification de l'environnement épigénétique des gènes régulés.

Enfin, la transformation leucémique peut passer par le détournement d'enhancer et souligne l'importance des interactions longues distances dans les mécanismes de régulation des gènes.

Bibliographie

- Abdel-Wahab, O., Adli, M., LaFave, L.M., Gao, J., Hricik, T., Shih, A.H., Pandey, S., Patel, J.P., Chung, Y.R., Koche, R., et al. (2012). ASXL1 Mutations Promote Myeloid Transformation through Loss of PRC2-Mediated Gene Repression. *Cancer Cell* 22, 180–193.
- Akasaka, T., Kanno, M., Balling, R., Mieza, M.A., Taniguchi, M., and Koseki, H. (1996). A role for mel-18, a Polycomb group-related vertebrate gene, during theanteroposterior specification of the axial skeleton. *Development* 122, 1513–1522.
- Akasaka, T., Lohuizen, M. van, Lugt, N. van der, Mizutani-Koseki, Y., Kanno, M., Taniguchi, M., Vidal, M., Alkema, M., Berns, A., and Koseki, H. (2001). Mice doubly deficient for the Polycomb Group genes Mel18 and Bmi1 reveal synergy and requirement for maintenance but not initiation of Hox gene expression. *Development* 128, 1587–1597.
- Alford, S.H., Toy, K., Merajver, S.D., and Kleer, C.G. (2012). Increased risk for distant metastasis in patients with familial early-stage breast cancer and high EZH2 expression. *Breast Cancer Res. Treat.* 132, 429–437.
- Alkema, M.J., van der Lugt, N.M.T., Bobeldijk, R.C., Berns, A., and van Lohuizen, M. (1995). Transformation of axial skeleton due to overexpression of bmi-1 in transgenic mice. *Nature* 374, 724–727.
- Bachmann, I.M., Halvorsen, O.J., Collett, K., Stefansson, I.M., Straume, O., Haukaas, S.A., Salvesen, H.B., Otte, A.P., and Akslen, L.A. (2006). EZH2 expression is associated with high proliferation rate and aggressive tumor subgroups in cutaneous melanoma and cancers of the endometrium, prostate, and breast. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* 24, 268–273.
- Ball, H.J., Melnick, A., Shaknovich, R., Kohanski, R.A., and Licht, J.D. (1999). The promyelocytic leukemia zinc finger (PLZF) protein binds DNA in a high molecular weight complex associated with cdc2 kinase. *Nucleic Acids Res.* 27, 4106–4113.
- Ballaré, C., Lange, M., Lapinaite, A., Martin, G.M., Morey, L., Pascual, G., Liefke, R., Simon, B., Shi, Y., Gozani, O., et al. (2012). Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nat. Struct. Mol. Biol.* 19, 1257–1265.
- Bardwell, V.J., and Treisman, R. (1994). The POZ domain: a conserved protein-protein interaction motif. *Genes Dev.* 8, 1664–1677.
- Barna, M., Hawe, N., Niswander, L., and Pandolfi, P.P. (2000). Plzf regulates limb and axial skeletal patterning. *Nat. Genet.* 25, 166–172.
- Barna, M., Merghoub, T., Costoya, J.A., Ruggero, D., Branford, M., Bergia, A., Samori, B., and Pandolfi, P.P. (2002). Plzf mediates transcriptional repression of HoxD gene expression through chromatin remodeling. *Dev. Cell* 3, 499–510.
- Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S.C., Mann, M., and Kouzarides, T. (2010). Nucleosome-Interacting Proteins Regulated by DNA and Histone Methylation. *Cell* 143, 470–484.
- Behrens, C., Solis, L.M., Lin, H., Yuan, P., Tang, X., Kadara, H., Riquelme, E., Galindo, H., Moran, C.A., Kalhor, N., et al. (2013). EZH2 protein expression associates with the early pathogenesis, tumor progression, and prognosis of non-small cell lung carcinoma. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 19, 6556–6565.

- Benoit, Y.D., Lepage, M.B., Khalfaoui, T., Tremblay, E., Basora, N., Carrier, J.C., Gudas, L.J., and Beaulieu, J.-F. (2012). Polycomb repressive complex 2 impedes intestinal cell terminal differentiation. *J. Cell Sci.* *125*, 3454–3463.
- Benveniste, D., Sonntag, H.-J., Sanguinetti, G., and Sproul, D. (2014). Transcription factor binding predicts histone modifications in human cell lines. *Proc. Natl. Acad. Sci.* *111*, 13367–13372.
- Berger, S.L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A. (2009). An operational definition of epigenetics. *Genes Dev.* *23*, 781–783.
- Bernard, D., Martinez-Leal, J.F., Rizzo, S., Martinez, D., Hudson, D., Visakorpi, T., Peters, G., Carnero, A., Beach, D., and Gil, J. (2005). CBX7 controls the growth of normal and tumor-derived prostate cells by repressing the Ink4a/Arf locus. *Oncogene* *24*, 5543–5551.
- Bernardo, M.V., Yelo, E., Gimeno, L., Campillo, J.A., and Parrado, A. (2007). Identification of apoptosis-related PLZF target genes. *Biochem. Biophys. Res. Commun.* *359*, 317–322.
- Blackledge, N.P., Farcas, A.M., Kondo, T., King, H.W., McGouran, J.F., Hanssen, L.L.P., Ito, S., Cooper, S., Kondo, K., Koseki, Y., et al. (2014). Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. *Cell* *157*, 1445–1459.
- Boukarabila, H., Saurin, A.J., Batsché, E., Mossadegh, N., Lohuizen, M. van, Otte, A.P., Pradel, J., Muchardt, C., Sieweke, M., and Duprez, E. (2009). The PRC1 Polycomb group complex interacts with PLZF/RARA to mediate leukemic transformation. *Genes Dev.* *23*, 1195–1206.
- Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. *Cell* *122*, 947–956.
- Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* *441*, 349–353.
- Bracken, A.P., and Helin, K. (2009). Polycomb group proteins: navigators of lineage pathways led astray in cancer. *Nat. Rev. Cancer* *9*, 773–784.
- Brien, G.L., Gambero, G., O’Connell, D.J., Jerman, E., Turner, S.A., Egan, C.M., Dunne, E.J., Jurgens, M.C., Wynne, K., Piao, L., et al. (2012). Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. *Nat. Struct. Mol. Biol.* *19*, 1273–1281.
- Brookes, E., de Santiago, I., Hebenstreit, D., Morris, K.J., Carroll, T., Xie, S.Q., Stock, J.K., Heidemann, M., Eick, D., Nozaki, N., et al. (2012). Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. *Cell Stem Cell* *10*, 157–170.
- Brown, J.L., Mucci, D., Whiteley, M., Dirksen, M.-L., and Kassis, J.A. (1998). The *Drosophila* Polycomb Group Gene pleiohomeotic Encodes a DNA Binding Protein with Homology to the Transcription Factor YY1. *Mol. Cell* *1*, 1057–1064.

- Brunner, G., Reitz, M., Schwipper, V., Tilkorn, H., Lippold, A., Biess, B., Suter, L., and Atzpodien, J. (2008). Increased expression of the tumor suppressor PLZF is a continuous predictor of long-term survival in malignant melanoma patients. *Cancer Biother. Radiopharm.* 23, 451–459.
- Buaas, F.W., Kirsh, A.L., Sharma, M., McLean, D.J., Morris, J.L., Griswold, M.D., de Rooij, D.G., and Braun, R.E. (2004). Plzf is required in adult male germ cells for stem cell self-renewal. *Nat. Genet.* 36, 647–652.
- Cai, L., Rothbart, S.B., Lu, R., Xu, B., Chen, W.-Y., Tripathy, A., Rockowitz, S., Zheng, D., Patel, D.J., Allis, C.D., et al. (2013). An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. *Mol. Cell* 49, 571–582.
- Cao, R., and Zhang, Y. (2004). SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Mol. Cell* 15, 57–67.
- Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* 298, 1039–1043.
- Caretti, G., Di Padova, M., Micales, B., Lyons, G.E., and Sartorelli, V. (2004). The Polycomb Ezh2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation. *Genes Dev.* 18, 2627–2638.
- Chamberlain, S.J., Yee, D., and Magnuson, T. (2008). Polycomb Repressive Complex 2 Is Dispensable for Maintenance of Embryonic Stem Cell Pluripotency. *STEM CELLS* 26, 1496–1505.
- Chauchereau, A., Mathieu, M., de Saintignon, J., Ferreira, R., Pritchard, L.L., Mishal, Z., Dejean, A., and Harel-Bellan, A. (2004). HDAC4 mediates transcriptional repression by the acute promyelocytic leukaemia-associated protein PLZF. *Oncogene* 23, 8777–8784.
- Chen, Z., Brand, N.J., Chen, A., Chen, S.J., Tong, J.H., Wang, Z.Y., Waxman, S., and Zelent, A. (1993). Fusion between a novel Krüppel-like zinc finger gene and the retinoic acid receptor-alpha locus due to a variant t(11;17) translocation associated with acute promyelocytic leukaemia. *EMBO J.* 12, 1161–1167.
- Cheung, M., Pei, J., Pei, Y., Jhanwar, S.C., Pass, H.I., and Testa, J.R. (2010). The promyelocytic leukemia zinc finger gene, PLZF, is frequently downregulated in malignant mesothelioma cells and contributes to cell survival. *Oncogene* 29, 1633–1640.
- Cifuentes-Rojas, C., Hernandez, A.J., Sarma, K., and Lee, J.T. (2014). Regulatory Interactions between RNA and Polycomb Repressive Complex 2. *Mol. Cell* 55, 171–185.
- Comet, I., and Helin, K. (2014). Revolution in the Polycomb hierarchy. *Nat. Struct. Mol. Biol.* 21, 573–575.
- Cooper, S., Dienstbier, M., Hassan, R., Schermelleh, L., Sharif, J., Blackledge, N.P., De Marco, V., Elderkin, S., Koseki, H., Klose, R., et al. (2014). Targeting Polycomb to Pericentric Heterochromatin in Embryonic Stem Cells Reveals a Role for H2AK119u1 in PRC2 Recruitment. *Cell Rep.* 7, 1456–1470.

- Core, N., Bel, S., Gaunt, S.J., Aurrand-Lions, M., Pearce, J., Fisher, A., and Djabali, M. (1997). Altered cellular proliferation and mesoderm patterning in Polycomb-M33-deficient mice. *Development* 124, 721–729.
- Costoya, J.A., Hobbs, R.M., Barna, M., Cattoretti, G., Manova, K., Sukhwani, M., Orwig, K.E., Wolgemuth, D.J., and Pandolfi, P.P. (2004). Essential role of Plzf in maintenance of spermatogonial stem cells. *Nat. Genet.* 36, 653–659.
- Costoya, J.A., Hobbs, R.M., and Pandolfi, P.P. (2008). Cyclin-dependent kinase antagonizes promyelocytic leukemia zinc-finger through phosphorylation. *Oncogene* 27, 3789–3796.
- Crea, F., Hurt, E.M., and Farrar, W.L. (2010). Clinical significance of Polycomb gene expression in brain tumors. *Mol. Cancer* 9, 265.
- Cuddapah, S., Roh, T.-Y., Cui, K., Jose, C.C., Fuller, M.T., Zhao, K., and Chen, X. (2012). A novel human polycomb binding site acts as a functional polycomb response element in *Drosophila*. *PloS One* 7, e36365.
- Davidovich, C., Zheng, L., Goodrich, K.J., and Cech, T.R. (2013). Promiscuous RNA binding by Polycomb repressive complex 2. *Nat. Struct. Mol. Biol.* 20, 1250–1257.
- Dhalluin, C., Carlson, J.E., Zeng, L., He, C., Aggarwal, A.K., and Zhou, M.M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. *Nature* 399, 491–496.
- Di Croce, L., and Helin, K. (2013). Transcriptional regulation by Polycomb group proteins. *Nat. Struct. Mol. Biol.* 20, 1147–1155.
- Doulatov, S., Notta, F., Rice, K.L., Howell, L., Zelent, A., Licht, J.D., and Dick, J.E. (2009a). PLZF is a regulator of homeostatic and cytokine-induced myeloid development. *Genes Dev.* 23, 2076–2087.
- Doulatov, S., Notta, F., Rice, K.L., Howell, L., Zelent, A., Licht, J.D., and Dick, J.E. (2009b). PLZF is a regulator of homeostatic and cytokine-induced myeloid development. *Genes Dev.* 23, 2076–2087.
- Duncan, I.M. (1982). Polycomblike: A Gene That Appears to Be Required for the Normal Expression of the Bithorax and Antennapedia Gene Complexes of *DROSOPHILA MELANOGASTER*. *Genetics* 102, 49–70.
- Endoh, M., Endo, T.A., Endoh, T., Fujimura, Y., Ohara, O., Toyoda, T., Otte, A.P., Okano, M., Brockdorff, N., Vidal, M., et al. (2008). Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Dev. Camb. Engl.* 135, 1513–1524.
- Ernst, P., and Vakoc, C.R. (2012). WRAD: enabler of the SET1-family of H3K4 methyltransferases. *Brief. Funct. Genomics* 11, 217–226.
- Ernst, T., Chase, A.J., Score, J., Hidalgo-Curtis, C.E., Bryant, C., Jones, A.V., Waghorn, K., Zoi, K., Ross, F.M., Reiter, A., et al. (2010). Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat. Genet.* 42, 722–726.

- Ezhkova, E., Pasolli, H.A., Parker, J.S., Stokes, N., Su, I. -hsi., Hannon, G., Tarakhovsky, A., and Fuchs, E. (2009). Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. *Cell* 136, 1122–1135.
- Farcas, A.M., Blackledge, N.P., Sudbery, I., Long, H.K., McGouran, J.F., Rose, N.R., Lee, S., Sims, D., Cerase, A., Sheahan, T.W., et al. (2012). KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *eLife* 1, e00205.
- Faust, C., Schumacher, A., Holdener, B., and Magnuson, T. (1995). The *eed* mutation disrupts anterior mesoderm production in mice. *Development* 121, 273–285.
- Felicetti, F., Errico, M.C., Bottero, L., Segnalini, P., Stoppacciaro, A., Biffoni, M., Felli, N., Mattia, G., Petrini, M., Colombo, M.P., et al. (2008). The promyelocytic leukemia zinc finger-microRNA-221/-222 pathway controls melanoma progression through multiple oncogenic mechanisms. *Cancer Res.* 68, 2745–2754.
- Ferrando, A.A. (2004). Biallelic transcriptional activation of oncogenic transcription factors in T-cell acute lymphoblastic leukemia. *Blood* 103, 1909–1911.
- Ferrari, K.J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stützer, A., Fischle, W., Bonaldi, T., and Pasini, D. (2014). Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. *Mol. Cell* 53, 49–62.
- Forzati, F., Federico, A., Pallante, P., Abbate, A., Esposito, F., Malapelle, U., Sepe, R., Palma, G., Troncone, G., Scarfò, M., et al. (2012). CBX7 is a tumor suppressor in mice and humans. *J. Clin. Invest.* 122, 612–623.
- Fukuda, T., Tokunaga, A., Sakamoto, R., and Yoshida, N. (2011). Fbxl10/Kdm2b deficiency accelerates neural progenitor cell death and leads to exencephaly. *Mol. Cell. Neurosci.* 46, 614–624.
- Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. *Mol. Cell* 45, 344–356.
- Gargiulo, G., Cesaroni, M., Serresi, M., de Vries, N., Hulsman, D., Bruggeman, S.W., Lancini, C., and van Lohuizen, M. (2013). In vivo RNAi screen for BMI1 targets identifies TGF- β /BMP-ER stress pathways as key regulators of neural- and malignant glioma-stem cell homeostasis. *Cancer Cell* 23, 660–676.
- Gehani, S.S., Agrawal-Singh, S., Dietrich, N., Christophersen, N.S., Helin, K., and Hansen, K. (2010). Polycomb Group Protein Displacement and Gene Activation through MSK-Dependent H3K27me3S28 Phosphorylation. *Mol. Cell* 39, 886–900.
- Greenblatt, S.M., and Nimer, S.D. (2014). Chromatin modifiers and the promise of epigenetic therapy in acute leukemia. *Leukemia* 28, 1396–1406.
- Grignani, F., De Matteis, S., Nervi, C., Tomassoni, L., Gelmetti, V., Cioce, M., Fanelli, M., Ruthardt, M., Ferrara, F.F., Zamir, I., et al. (1998). Fusion proteins of the retinoic acid receptor- α recruit histone deacetylase in promyelocytic leukaemia. *Nature* 391, 815–818.

Guidez, F., Howell, L., Isalan, M., Cebrat, M., Alani, R.M., Ivins, S., Hormaeche, I., McConnell, M.J., Pierce, S., Cole, P.A., et al. (2005). Histone Acetyltransferase Activity of p300 Is Required for Transcriptional Repression by the Promyelocytic Leukemia Zinc Finger Protein. *Mol. Cell. Biol.* 25, 5552–5566.

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.

Hobbs, R.M., Fagoonee, S., Papa, A., Webster, K., Altruda, F., Nishinakamura, R., Chai, L., and Pandolfi, P.P. (2012). Functional antagonism between Sall4 and Plzf define germline progenitors. *Cell Stem Cell* 10, 284–298.

Hong, S.-H., David, G., Wong, C.-W., Dejean, A., and Privalsky, M.L. (1997). SMRT corepressor interacts with PLZF and with the PML-retinoic acid receptor α (RAR α) and PLZF-RAR α oncoproteins associated with acute promyelocytic leukemia. *Proc. Natl. Acad. Sci. U. S. A.* 94, 9028–9033.

Hsieh, C.-L., Botta, G., Gao, S., Li, T., Van Allen, E.M., Treacy, D.J., Cai, C., He, H.H., Sweeney, C.J., Brown, M., et al. (2015). PLZF, a tumor suppressor genetically lost in metastatic castration-resistant prostate cancer, is a mediator of resistance to androgen deprivation therapy. *Cancer Res.* 75, 1944–1948.

Huang, S., and Brandt, S.J. (2000). mSin3A regulates murine erythroleukemia cell differentiation through association with the TAL1 (or SCL) transcription factor. *Mol. Cell. Biol.* 20, 2248–2259.

Hur, M.-W., Laney, J.D., Jeon, S.-H., Ali, J., and Biggin, M.D. (2002). Zeste maintains repression of Ubx transgenes: support for a new model of Polycomb repression. *Dev. Camb. Engl.* 129, 1339–1343.

Kahn, T.G., Stenberg, P., Pirrotta, V., and Schwartz, Y.B. (2014). Combinatorial Interactions Are Required for the Efficient Recruitment of Pho Repressive Complex (PhoRC) to Polycomb Response Elements. *PLoS Genet.* 10.

Kalb, R., Latwiel, S., Baymaz, H.I., Jansen, P.W.T.C., Müller, C.W., Vermeulen, M., and Müller, J. (2014). Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat. Struct. Mol. Biol.* 21, 569–571.

Kaneko, S., Son, J., Shen, S.S., Reinberg, D., and Bonasio, R. (2013). PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nat. Struct. Mol. Biol.* 20, 1258–1264.

Kaneko, S., Bonasio, R., Saldaña-Meyer, R., Yoshida, T., Son, J., Nishino, K., Umezawa, A., and Reinberg, D. (2014). Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. *Mol. Cell* 53, 290–300.

Kang, S.I., Chang, W.-J., Cho, S.-G., and Kim, I.Y. (2003). Modification of promyelocytic leukemia zinc finger protein (PLZF) by SUMO-1 conjugation regulates its transcriptional repressor activity. *J. Biol. Chem.* 278, 51479–51483.

- Kang, S.I., Choi, H.W., and Kim, I.Y. (2008). Redox-mediated modification of PLZF by SUMO-1 and ubiquitin. *Biochem. Biophys. Res. Commun.* 369, 1209–1214.
- Kassis, J.A., and Brown, J.L. (2013). Polycomb Group Response Elements in *Drosophila* and Vertebrates. *Adv. Genet.* 81, 83–118.
- Kikugawa, T., Kinugasa, Y., Shiraishi, K., Nanba, D., Nakashiro, K., Tanji, N., Yokoyama, M., and Higashiyama, S. (2006). PLZF regulates Pbx1 transcription and Pbx1-HoxC8 complex leads to androgen-independent prostate cancer proliferation. *The Prostate* 66, 1092–1099.
- Kim, E., Kim, M., Woo, D.-H., Shin, Y., Shin, J., Chang, N., Oh, Y.T., Kim, H., Rhee, J., Nakano, I., et al. (2013a). Phosphorylation of EZH2 activates STAT3 signaling via STAT3 methylation and promotes tumorigenicity of glioblastoma stem-like cells. *Cancer Cell* 23, 839–852.
- Kim, W., Bird, G.H., Neff, T., Guo, G., Kerenyi, M.A., Walensky, L.D., and Orkin, S.H. (2013b). Targeted disruption of the EZH2-EED complex inhibits EZH2-dependent cancer. *Nat. Chem. Biol.* 9, 643–650.
- Klauke, K., Radulović, V., Broekhuis, M., Weersing, E., Zwart, E., Olthof, S., Ritsema, M., Bruggeman, S., Wu, X., Helin, K., et al. (2013). Polycomb Cbx family members mediate the balance between haematopoietic stem cell self-renewal and differentiation. *Nat. Cell Biol.* 15, 353–362.
- Kondo, Y., Shen, L., Suzuki, S., Kurokawa, T., Masuko, K., Tanaka, Y., Kato, H., Mizuno, Y., Yokoe, M., Sugauchi, F., et al. (2007). Alterations of DNA methylation and histone modifications contribute to gene silencing in hepatocellular carcinomas. *Hepatol. Res. Off. J. Jpn. Soc. Hepatol.* 37, 974–983.
- Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., et al. (2008). Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains. *PLoS Genet* 4, e1000242.
- LaJeunesse, D., and Shearn, A. (1996). E(z): a polycomb group gene or a trithorax group gene? *Development* 122, 2189–2197.
- Landeira, D., Sauer, S., Poot, R., Dvorkina, M., Mazzarella, L., Jørgensen, H.F., Pereira, C.F., Leleu, M., Piccolo, F.M., Spivakov, M., et al. (2010). Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. *Nat. Cell Biol.* 12, 618–624.
- Lau, P.N.I., and Cheung, P. (2011). Histone code pathway involving H3 S28 phosphorylation and K27 acetylation activates transcription and antagonizes polycomb silencing. *Proc. Natl. Acad. Sci. U. S. A.* 108, 2801–2806.
- Lee, J.M., Lee, J.S., Kim, H., Kim, K., Park, H., Kim, J.-Y., Lee, S.H., Kim, I.S., Kim, J., Lee, M., et al. (2012). EZH2 generates a methyl degron that is recognized by the DCAF1/DDB1/CUL4 E3 ubiquitin ligase complex. *Mol. Cell* 48, 572–586.

- Lee, S.T., Li, Z., Wu, Z., Aau, M., Guan, P., Karuturi, R.K.M., Liou, Y.C., and Yu, Q. (2011). Context-specific regulation of NF- κ B target gene expression by EZH2 in breast cancers. *Mol. Cell* 43, 798–810.
- Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., et al. (2006). Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells. *Cell* 125, 301–313.
- Lemercier, C., Brocard, M.-P., Puvion-Dutilleul, F., Kao, H.-Y., Albagli, O., and Khochbin, S. (2002). Class II Histone Deacetylases Are Directly Recruited by BCL6 Transcriptional Repressor. *J. Biol. Chem.* 277, 22045–22052.
- Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* 276, 565–570.
- Li, B.E., and Ernst, P. (2014). Two decades of leukemia oncoprotein epistasis: the MLL1 paradigm for epigenetic deregulation in leukemia. *Exp. Hematol.* 42, 995–1012.
- Li, J.Y., English, M.A., Ball, H.J., Yeyati, P.L., Waxman, S., and Licht, J.D. (1997). Sequence-specific DNA binding and transcriptional regulation by the promyelocytic leukemia zinc finger protein. *J. Biol. Chem.* 272, 22447–22455.
- Li, X., Gonzalez, M.E., Toy, K., Filzen, T., Merajver, S.D., and Kleer, C.G. (2009). Targeted overexpression of EZH2 in the mammary gland disrupts ductal morphogenesis and causes epithelial hyperplasia. *Am. J. Pathol.* 175, 1246–1254.
- Lin, R.J., Nagy, L., Inoue, S., Shao, W., Miller, W.H., and Evans, R.M. (1998). Role of the histone deacetylase complex in acute promyelocytic leukaemia. *Nature* 391, 811–814.
- Liu, B., Liu, Y.-F., Du, Y.-R., Mardaryev, A.N., Yang, W., Chen, H., Xu, Z.-M., Xu, C.-Q., Zhang, X.-R., Botchkarev, V.A., et al. (2013). Cbx4 regulates the proliferation of thymic epithelial cells and thymus function. *Development* 140, 780–788.
- Lugt, N.M. van der, Domen, J., Linders, K., Roon, M. van, Robanus-Maandag, E., Riele, H. te, Valk, M. van der, Deschamps, J., Sofroniew, M., and Lohuizen, M. van (1994). Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene. *Genes Dev.* 8, 757–769.
- Lynch, M.D., Smith, A.J.H., De Gobbi, M., Flenley, M., Hughes, J.R., Vernimmen, D., Ayyub, H., Sharpe, J.A., Sloane-Stanley, J.A., Sutherland, L., et al. (2012). An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. *EMBO J.* 31, 317–329.
- Mallen-St. Clair, J., Soydaner-Azeloglu, R., Lee, K.E., Taylor, L., Livanos, A., Pylayeva-Gupta, Y., Miller, G., Margueron, R., Reinberg, D., and Bar-Sagi, D. (2012). EZH2 couples pancreatic regeneration to neoplastic progression. *Genes Dev.* 26, 439–444.
- Mansour, M.R., Abraham, B.J., Anders, L., Berezovskaya, A., Gutierrez, A., Durbin, A.D., Etchin, J., Lawton, L., Sallan, S.E., Silverman, L.B., et al. (2014). An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* 346, 1373–1377.

Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C.L., Dynlacht, B.D., and Reinberg, D. (2008). Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol. Cell* 32, 503–518.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., Voigt, P., Martin, S., Taylor, W.R., De Marco, V., et al. (2009). Role of the polycomb protein Eed in the propagation of repressive histone marks. *Nature* 461, 762–767.

Martin, C., Cao, R., and Zhang, Y. (2006). Substrate Preferences of the EZH2 Histone Methyltransferase Complex. *J. Biol. Chem.* 281, 8365–8370.

Mathew, R., Seiler, M.P., Scanlon, S.T., Mao, A., Constantinides, M.G., Bertozzi-Villa, C., Singer, J.D., and Bendelac, A. (2012). BTB-ZF factors recruit the E3 ligase cullin 3 to regulate lymphoid effector programs. *Nature* 491, 618–621.

McConnell, M.J., Durand, L., Langley, E., Coste-Sarguet, L., Zelent, A., Chomienne, C., Kouzarides, T., Licht, J.D., and Guidez, F. (2015). Post transcriptional control of the epigenetic stem cell regulator PLZF by sirtuin and HDAC deacetylases. *Epigenetics Chromatin* 8.

Melnick, A., Ahmad, K.F., Arai, S., Polinger, A., Ball, H., Borden, K.L., Carlile, G.W., Prive, G.G., and Licht, J.D. (2000). In-depth mutational analysis of the promyelocytic leukemia zinc finger BTB/POZ domain reveals motifs and residues required for biological and transcriptional functions. *Mol. Cell. Biol.* 20, 6550–6567.

Mendenhall, E.M., Koche, R.P., Truong, T., Zhou, V.W., Issac, B., Chi, A.S., Ku, M., and Bernstein, B.E. (2010). GC-Rich Sequence Elements Recruit PRC2 in Mammalian ES Cells. *PLoS Genet* 6, e1001244.

Mito, Y., Henikoff, J.G., and Henikoff, S. (2007). Histone Replacement Marks the Boundaries of cis-Regulatory Domains. *Science* 315, 1408–1411.

Morey, L., Pascual, G., Cozzuto, L., Roma, G., Wutz, A., Benitah, S.A., and Di Croce, L. (2012). Nonoverlapping Functions of the Polycomb Group Cbx Family of Proteins in Embryonic Stem Cells. *Cell Stem Cell* 10, 47–62.

Morey, L., Aloia, L., Cozzuto, L., Benitah, S.A., and Di Croce, L. (2013). RYBP and Cbx7 Define Specific Biological Functions of Polycomb Complexes in Mouse Embryonic Stem Cells. *Cell Rep.* 3, 60–69.

Morin, R.D., Johnson, N.A., Severson, T.M., Mungall, A.J., An, J., Goya, R., Paul, J.E., Boyle, M., Woolcock, B.W., Kuchenbauer, F., et al. (2010). Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat. Genet.* 42, 181–185.

Mousavi, K., Zare, H., Wang, A.H., and Sartorelli, V. (2012). Polycomb protein Ezh1 promotes RNA polymerase II elongation. *Mol. Cell* 45, 255–262.

Mozzetta, C., Pontis, J., Fritsch, L., Robin, P., Portoso, M., Proux, C., Margueron, R., and Ait-Si-Ali, S. (2014). The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-mediated gene silencing. *Mol. Cell* 53, 277–289.

- Musselman, C.A., Lalonde, M.-E., Côté, J., and Kutateladze, T.G. (2012). Perceiving the epigenetic landscape through histone readers. *Nat. Struct. Mol. Biol.* *19*, 1218–1227.
- 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., et al. (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev. Cell* *7*, 663–676.
- Neff, T., Sinha, A.U., Kluk, M.J., Zhu, N., Khattab, M.H., Stein, L., Xie, H., Orkin, S.H., and Armstrong, S.A. (2012). Polycomb repressive complex 2 is required for MLL-AF9 leukemia. *Proc. Natl. Acad. Sci. U. S. A.* *109*, 5028–5033.
- Nègre, N., Hennetin, J., Sun, L.V., Lavrov, S., Bellis, M., White, K.P., and Cavalli, G. (2006). Chromosomal distribution of PcG proteins during *Drosophila* development. *PLoS Biol.* *4*, e170.
- Nekrasov, M., Wild, B., and Müller, J. (2005). Nucleosome binding and histone methyltransferase activity of *Drosophila* PRC2. *EMBO Rep.* *6*, 348–353.
- Nikoloski, G., Langemeijer, S.M.C., Kuiper, R.P., Knops, R., Massop, M., Tönnissen, E.R.L.T.M., van der Heijden, A., Scheele, T.N., Vandenberghe, P., de Witte, T., et al. (2010). Somatic mutations of the histone methyltransferase gene *EZH2* in myelodysplastic syndromes. *Nat. Genet.* *42*, 665–667.
- Ntziachristos, P., Tsirigos, A., Vlierberghe, P.V., Nedjic, J., Trimarchi, T., Flaherty, M.S., Ferres-Marco, D., da Ros, V., Tang, Z., Siegle, J., et al. (2012). Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. *Nat. Med.* *18*, 298–303.
- O’Carroll, D., Erhardt, S., Pagani, M., Barton, S.C., Surani, M.A., and Jenuwein, T. (2001). The Polycomb-Group Gene *Ezh2* Is Required for Early Mouse Development. *Mol. Cell. Biol.* *21*, 4330–4336.
- Ozato, K. (2009). PLZF outreach: a finger in interferon’s pie. *Immunity* *30*, 757–758.
- Park, I., Qian, D., Kiel, M., Becker, M.W., Pihalja, M., Weissman, I.L., Morrison, S.J., and Clarke, M.F. (2003). Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature* *423*, 302–305.
- Pasini, D., Bracken, A.P., Jensen, M.R., Denchi, E.L., and Helin, K. (2004). Suz12 is essential for mouse development and for *EZH2* histone methyltransferase activity. *EMBO J.* *23*, 4061–4071.
- Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M., and Helin, K. (2007). The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol. Cell. Biol.* *27*, 3769–3779.
- Pasini, D., Malatesta, M., Jung, H.R., Walfridsson, J., Willer, A., Olsson, L., Skotte, J., Wutz, A., Porse, B., Jensen, O.N., et al. (2010a). Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. *Nucleic Acids Res.* *38*, 4958–4969.

- Pasini, D., Cloos, P.A.C., Walfridsson, J., Olsson, L., Bukowski, J.-P., Johansen, J.V., Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. (2010b). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* *464*, 306–310.
- Pirity, M.K., Locker, J., and Schreiber-Agus, N. (2005). Rybp/DEDAF Is Required for Early Postimplantation and for Central Nervous System Development. *Mol. Cell. Biol.* *25*, 7193–7202.
- Pobezinsky, L.A., Etzensperger, R., Jeurling, S., Alag, A., Kadakia, T., McCaughtry, T.M., Kimura, M.Y., Sharrow, S.O., Guintier, T.I., Feigenbaum, L., et al. (2015). Let-7 microRNAs target the lineage-specific transcription factor PLZF to regulate terminal NKT cell differentiation and effector function. *Nat. Immunol.* *16*, 517–524.
- Puda, A., Milosevic, J.D., Berg, T., Klampfl, T., Harutyunyan, A.S., Gisslinger, B., Rumi, E., Pietra, D., Malcovati, L., Elena, C., et al. (2012). Frequent deletions of JARID2 in leukemic transformation of chronic myeloid malignancies. *Am. J. Hematol.* *87*, 245–250.
- Qin, J., Whyte, W.A., Anderssen, E., Apostolou, E., Chen, H.-H., Akbarian, S., Bronson, R.T., Hochedlinger, K., Ramaswamy, S., Young, R.A., et al. (2012). The Polycomb Group Protein L3mbtl2 Assembles an Atypical PRC1-Family Complex that Is Essential in Pluripotent Stem Cells and Early Development. *Cell Stem Cell* *11*, 319–332.
- Raaphorst, F.M., van Kemenade, F.J., Blokzijl, T., Fieret, E., Hamer, K.M., Satijn, D.P., Otte, A.P., and Meijer, C.J. (2000). Coexpression of BMI-1 and EZH2 polycomb group genes in Reed-Sternberg cells of Hodgkin's disease. *Am. J. Pathol.* *157*, 709–715.
- Rao, Z.-Y., Cai, M.-Y., Yang, G.-F., He, L.-R., Mai, S.-J., Hua, W.-F., Liao, Y.-J., Deng, H.-X., Chen, Y.-C., Guan, X.-Y., et al. (2010). EZH2 supports ovarian carcinoma cell invasion and/or metastasis via regulation of TGF-beta1 and is a predictor of outcome in ovarian carcinoma patients. *Carcinogenesis* *31*, 1576–1583.
- Reid, A., Gould, A., Brand, N., Cook, M., Strutt, P., Li, J., Licht, J., Waxman, S., Krumlauf, R., and Zelent, A. (1995). Leukemia translocation gene, PLZF, is expressed with a speckled nuclear pattern in early hematopoietic progenitors. *Blood* *86*, 4544–4552.
- Reynolds, N., Salmon-Divon, M., Dvinge, H., Hynes-Allen, A., Balasooriya, G., Leaford, D., Behrens, A., Bertone, P., and Hendrich, B. (2012). NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. *EMBO J.* *31*, 593–605.
- Rho, S.B., Chung, B.M., and Lee, J.-H. (2007). TIMP-1 regulates cell proliferation by interacting with the ninth zinc finger domain of PLZF. *J. Cell. Biochem.* *101*, 57–67.
- Riising, E.M., Comet, I., Leblanc, B., Wu, X., Johansen, J.V., and Helin, K. (2014). Gene Silencing Triggers Polycomb Repressive Complex 2 Recruitment to CpG Islands Genome Wide. *Mol. Cell* *55*, 347–360.
- Ringrose, L., Rehmsmeier, M., Dura, J.-M., and Paro, R. (2003). Genome-Wide Prediction of Polycomb/Trithorax Response Elements in *Drosophila melanogaster*. *Dev. Cell* *5*, 759–771.

- Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311–1323.
- da Rocha, S.T., Boeva, V., Escamilla-Del-Arenal, M., Ancelin, K., Granier, C., Matias, N.R., Sanulli, S., Chow, J., Schulz, E., Picard, C., et al. (2014). Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. *Mol. Cell* 53, 301–316.
- Sadler, A.J., Rossello, F.J., Yu, L., Deane, J.A., Yuan, X., Wang, D., Irving, A.T., Kaparakis-Liaskos, M., Gantier, M.P., Ying, H., et al. (2015). BTB-ZF transcriptional regulator PLZF modifies chromatin to restrain inflammatory signaling programs. *Proc. Natl. Acad. Sci.* 112, 1535–1540.
- Sanda, T., Lawton, L.N., Barrasa, M.I., Fan, Z.P., Kohlhammer, H., Gutierrez, A., Ma, W., Tatarek, J., Ahn, Y., Kelliher, M.A., et al. (2012). Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. *Cancer Cell* 22, 209–221.
- Sanulli, S., Justin, N., Teissandier, A., Ancelin, K., Portoso, M., Caron, M., Michaud, A., Lombard, B., da Rocha, S.T., Offer, J., et al. (2015). Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation. *Mol. Cell* 57, 769–783.
- Sasaki, D., Imaizumi, Y., Hasegawa, H., Osaka, A., Tsukasaki, K., Choi, Y.L., Mano, H., Marquez, V.E., Hayashi, T., Yanagihara, K., et al. (2011). Overexpression of Enhancer of zeste homolog 2 with trimethylation of lysine 27 on histone H3 in adult T-cell leukemia/lymphoma as a target for epigenetic therapy. *Haematologica* 96, 712–719.
- Sashida, G., Harada, H., Matsui, H., Oshima, M., Yui, M., Harada, Y., Tanaka, S., Mochizuki-Kashio, M., Wang, C., Saraya, A., et al. (2014). Ezh2 loss promotes development of myelodysplastic syndrome but attenuates its predisposition to leukaemic transformation. *Nat. Commun.* 5, 4177.
- Saurin, A.J., Shao, Z., Erdjument-Bromage, H., Tempst, P., and Kingston, R.E. (2001). A Drosophila Polycomb group complex includes Zeste and dTAFII proteins. *Nature* 412, 655–660.
- Savage, A.K., Constantinides, M.G., Han, J., Picard, D., Martin, E., Li, B., Lantz, O., and Bendelac, A. (2008). The transcription factor PLZF (Zbtb16) directs the effector program of the NKT cell lineage. *Immunity* 29, 391–403.
- Schwartz, Y.B., and Pirrotta, V. (2013). A new world of Polycombs: unexpected partnerships and emerging functions. *Nat. Rev. Genet.* 14, 853–864.
- Schwartz, Y.B., Kahn, T.G., Nix, D.A., Li, X.-Y., Bourgon, R., Biggin, M., and Pirrotta, V. (2006). Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nat. Genet.* 38, 700–705.
- Score, J., Hidalgo-Curtis, C., Jones, A.V., Winkelmann, N., Skinner, A., Ward, D., Zoi, K., Ernst, T., Stegelmann, F., Döhner, K., et al. (2012). Inactivation of polycomb repressive complex 2 components in myeloproliferative and myelodysplastic/myeloproliferative neoplasms. *Blood* 119, 1208–1213.

Scott, C.L., Gil, J., Hernando, E., Teruya-Feldstein, J., Narita, M., Martínez, D., Visakorpi, T., Mu, D., Cordon-Cardo, C., Peters, G., et al. (2007). Role of the chromobox protein CBX7 in lymphomagenesis. *Proc. Natl. Acad. Sci. U. S. A.* *104*, 5389–5394.

Shaknovich, R., Yeyati, P.L., Ivins, S., Melnick, A., Lempert, C., Waxman, S., Zelent, A., and Licht, J.D. (1998). The promyelocytic leukemia zinc finger protein affects myeloid cell growth, differentiation, and apoptosis. *Mol. Cell. Biol.* *18*, 5533–5545.

Shi, B., Liang, J., Yang, X., Wang, Y., Zhao, Y., Wu, H., Sun, L., Zhang, Y., Chen, Y., Li, R., et al. (2007). Integration of Estrogen and Wnt Signaling Circuits by the Polycomb Group Protein EZH2 in Breast Cancer Cells. *Mol. Cell. Biol.* *27*, 5105–5119.

Shi, J., Wang, E., Zuber, J., Rappaport, A., Taylor, M., Johns, C., Lowe, S.W., and Vakoc, C.R. (2013a). The Polycomb complex PRC2 supports aberrant self-renewal in a mouse model of MLL-AF9;Nras(G12D) acute myeloid leukemia. *Oncogene* *32*, 930–938.

Shi, J., Whyte, W.A., Zepeda-Mendoza, C.J., Milazzo, J.P., Shen, C., Roe, J.-S., Minder, J.L., Mercan, F., Wang, E., Eckersley-Maslin, M.A., et al. (2013b). Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. *Genes Dev.* *27*, 2648–2662.

Simon, J., Chiang, A., Bender, W., Shimell, M.J., and O'Connor, M. (1993). Elements of the *Drosophila* bithorax complex that mediate repression by Polycomb group products. *Dev. Biol.* *158*, 131–144.

Sing, A., Pannell, D., Karaiskakis, A., Sturgeon, K., Djabali, M., Ellis, J., Lipshitz, H.D., and Cordes, S.P. (2009). A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. *Cell* *138*, 885–897.

Smith, L.-L., Yeung, J., Zeisig, B.B., Popov, N., Huijbers, I., Barnes, J., Wilson, A.J., Taskesen, E., Delwel, R., Gil, J., et al. (2011). Functional crosstalk between Bmi1 and MLL/Hoxa9 axis in establishment of normal hematopoietic and leukemic stem cells. *Cell Stem Cell* *8*, 649–662.

Sneeringer, C.J., Scott, M.P., Kuntz, K.W., Knutson, S.K., Pollock, R.M., Richon, V.M., and Copeland, R.A. (2010). Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. *Proc. Natl. Acad. Sci. U. S. A.* *107*, 20980–20985.

Sobieszczuk, D.F., Poliakov, A., Xu, Q., and Wilkinson, D.G. (2010). A feedback loop mediated by degradation of an inhibitor is required to initiate neuronal differentiation. *Genes Dev.* *24*, 206–218.

Sparmann, A., and van Lohuizen, M. (2006). Polycomb silencers control cell fate, development and cancer. *Nat. Rev. Cancer* *6*, 846–856.

Spicuglia, S., Vincent-Fabert, C., Benoukraf, T., Tibéri, G., Saurin, A.J., Zacarias-Cabeza, J., Grimwade, D., Mills, K., Calmels, B., Bertucci, F., et al. (2011). Characterisation of Genome-Wide PLZF/RARA Target Genes. *PLoS ONE* *6*, e24176.

Steffen, P.A., and Ringrose, L. (2014). What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. *Nat. Rev. Mol. Cell Biol.* *15*, 340–356.

- Stojic, L., Jasencakova, Z., Prezioso, C., Stützer, A., Bodega, B., Pasini, D., Klingberg, R., Mozzetta, C., Margueron, R., Puri, P.L., et al. (2011). Chromatin regulated interchange between polycomb repressive complex 2 (PRC2)-Ezh2 and PRC2-Ezh1 complexes controls myogenin activation in skeletal muscle cells. *Epigenetics Chromatin* 4, 16.
- Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* 403, 41–45.
- Strutt, H. (1997). Co-localization of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression. *EMBO J.* 16, 3621–3632.
- Su, I.-H., Basavaraj, A., Krutchinsky, A.N., Hobert, O., Ullrich, A., Chait, B.T., and Tarakhovsky, A. (2003). Ezh2 controls B cell development through histone H3 methylation and Igh rearrangement. *Nat. Immunol.* 4, 124–131.
- Suliman, B.A., Xu, D., and Williams, B.R.G. (2012a). The promyelocytic leukemia zinc finger protein: two decades of molecular oncology. *Front. Oncol.* 2, 74.
- Suliman, B.A., Xu, D., and Williams, B.R.G. (2012b). The Promyelocytic Leukemia Zinc Finger Protein: Two Decades of Molecular Oncology. *Front. Oncol.* 2.
- Suzuki, M., Mizutani-Koseki, Y., Fujimura, Y., Miyagishima, H., Kaneko, T., Takada, Y., Akasaka, T., Tanzawa, H., Takihara, Y., Nakano, M., et al. (2002). Involvement of the Polycomb-group gene Ring1B in the specification of the anterior-posterior axis in mice. *Development* 129, 4171–4183.
- Szalkowski, A.M., and Schmid, C.D. (2011). Rapid innovation in ChIP-seq peak-calling algorithms is outdistancing benchmarking efforts. *Brief. Bioinform.* 12, 626–633.
- Takahashi, S., and Licht, J.D. (2002). The human promyelocytic leukemia zinc finger gene is regulated by the Evi-1 oncoprotein and a novel guanine-rich site binding protein. *Leukemia* 16, 1755–1762.
- Tan, J., Jones, M., Koseki, H., Nakayama, M., Muntean, A.G., Maillard, I., and Hess, J.L. (2011). CBX8, a polycomb group protein, is essential for MLL-AF9-induced leukemogenesis. *Cancer Cell* 20, 563–575.
- Taniguchi, H., Jacinto, F.V., Villanueva, A., Fernandez, A.F., Yamamoto, H., Carmona, F.J., Puertas, S., Marquez, V.E., Shinomura, Y., Imai, K., et al. (2012). Silencing of Kruppel-like factor 2 by the histone methyltransferase EZH2 in human cancer. *Oncogene* 31, 1988–1994.
- Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H., et al. (2012). RYBP-PRC1 Complexes Mediate H2A Ubiquitylation at Polycomb Target Sites Independently of PRC2 and H3K27me3. *Cell* 148, 664–678.
- Thirkettle, H.J., Mills, I.G., Whitaker, H.C., and Neal, D.E. (2009). Nuclear LYRIC/AEG-1 interacts with PLZF and relieves PLZF-mediated repression. *Oncogene* 28, 3663–3670.

- Tie, F., Banerjee, R., Stratton, C.A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., Diaz, M.O., Scacheri, P.C., and Harte, P.J. (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes *Drosophila* Polycomb silencing. *Development* *136*, 3131–3141.
- Tie, F., Banerjee, R., Saiakhova, A.R., Howard, B., Monteith, K.E., Scacheri, P.C., Cosgrove, M.S., and Harte, P.J. (2014). Trithorax monomethylates histone H3K4 and interacts directly with CBP to promote H3K27 acetylation and antagonize Polycomb silencing. *Development* *141*, 1129–1139.
- Tien, C.-L., Jones, A., Wang, H., Gerigk, M., Nozell, S., and Chang, C. (2015). Snail2/Slug cooperates with Polycomb repressive complex 2 (PRC2) to regulate neural crest development. *Dev. Camb. Engl.* *142*, 722–731.
- Tolhuis, B., de Wit, E., Muijters, I., Teunissen, H., Talhout, W., van Steensel, B., and van Lohuizen, M. (2006). Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in *Drosophila melanogaster*. *Nat. Genet.* *38*, 694–699.
- Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science* *329*, 689–693.
- Turner, S.A., and Bracken, A.P. (2013). A “Complex” Issue: Deciphering the Role of Variant PRC1 in ESCs. *Cell Stem Cell* *12*, 145–146.
- Ueda, T., Sanada, M., Matsui, H., Yamasaki, N., Honda, Z.-I., Shih, L.-Y., Mori, H., Inaba, T., Ogawa, S., and Honda, H. (2012). EED mutants impair polycomb repressive complex 2 in myelodysplastic syndrome and related neoplasms. *Leukemia* *26*, 2557–2560.
- Uhlén, M., Björling, E., Agaton, C., Szigartyo, C.A.-K., Amini, B., Andersen, E., Andersson, A.-C., Angelidou, P., Asplund, A., Asplund, C., et al. (2005). A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell. Proteomics MCP* *4*, 1920–1932.
- Visser, H.P., Gunster, M.J., Kluin-Nelemans, H.C., Manders, E.M., Raaphorst, F.M., Meijer, C.J., Willemze, R., and Otte, A.P. (2001). The Polycomb group protein EZH2 is upregulated in proliferating, cultured human mantle cell lymphoma. *Br. J. Haematol.* *112*, 950–958.
- Vizán, P., Beringer, M., Ballaré, C., and Di Croce, L. (2015). Role of PRC2-associated factors in stem cells and disease. *FEBS J.* *282*, 1723–1735.
- Voigt, P., Tee, W.-W., and Reinberg, D. (2013). A double take on bivalent promoters. *Genes Dev.* *27*, 1318–1338.
- Voncken, J.W., Roelen, B.A.J., Roefs, M., de Vries, S., Verhoeven, E., Marino, S., Deschamps, J., and van Lohuizen, M. (2003). Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. *Proc. Natl. Acad. Sci. U. S. A.* *100*, 2468–2473.
- Waddington, C.H. (2012). The Epigenotype. *Int. J. Epidemiol.* *41*, 10–13.
- Walker, J.A., Barlow, J.L., and McKenzie, A.N.J. (2013). Innate lymphoid cells--how did we miss them? *Nat. Rev. Immunol.* *13*, 75–87.
- Wang, L., Brown, J.L., Cao, R., Zhang, Y., Kassiss, J.A., and Jones, R.S. (2004). Hierarchical Recruitment of Polycomb Group Silencing Complexes. *Mol. Cell* *14*, 637–646.

Wang, L., Jin, Q., Lee, J.-E., Su, I. -hsi., and Ge, K. (2010). Histone H3K27 methyltransferase Ezh2 represses Wnt genes to facilitate adipogenesis. *Proc. Natl. Acad. Sci. U. S. A.* *107*, 7317–7322.

Wang, X., Wang, L., Guo, S., Bao, Y., Ma, Y., Yan, F., Xu, K., Xu, Z., Jin, L., Lu, D., et al. (2013). Hypermethylation reduces expression of tumor-suppressor PLZF and regulates proliferation and apoptosis in non-small-cell lung cancers. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* *27*, 4194–4203.

Warrell, R.P., de Thé, H., Wang, Z.Y., and Degos, L. (1993). Acute promyelocytic leukemia. *N. Engl. J. Med.* *329*, 177–189.

Wen, J., Huang, S., Pack, S.D., Yu, X., Brandt, S.J., and Noguchi, C.T. (2005). Tal1/SCL binding to pericentromeric DNA represses transcription. *J. Biol. Chem.* *280*, 12956–12966.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell* *153*, 307–319.

Woo, C.J., Kharchenko, P.V., Daheron, L., Park, P.J., and Kingston, R.E. (2010). A region of the human HOXD cluster that confers polycomb-group responsiveness. *Cell* *140*, 99–110.

Xiao, G.-Q., Unger, P., Yang, Q., Kinoshita, Y., Singh, K., McMahon, L., Nastiuk, K., Sha, K., Krolewski, J., and Burstein, D. (2015). Loss of PLZF expression in prostate cancer by immunohistochemistry correlates with tumor aggressiveness and metastasis. *PloS One* *10*, e0121318.

Xie, H., Xu, J., Hsu, J.H., Nguyen, M., Fujiwara, Y., Peng, C., and Orkin, S.H. (2014). Polycomb repressive complex 2 regulates normal hematopoietic stem cell function in a developmental-stage-specific manner. *Cell Stem Cell* *14*, 68–80.

Xu, C., Bian, C., Yang, W., Galka, M., Ouyang, H., Chen, C., Qiu, W., Liu, H., Jones, A.E., MacKenzie, F., et al. (2010). Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). *Proc. Natl. Acad. Sci. U. S. A.* *107*, 19266–19271.

Xu, D., Holko, M., Sadler, A.J., Scott, B., Higashiyama, S., Berkofsky-Fessler, W., McConnell, M.J., Pandolfi, P.P., Licht, J.D., and Williams, B.R.G. (2009a). Promyelocytic leukemia zinc finger protein regulates interferon-mediated innate immunity. *Immunity* *30*, 802–816.

Xu, D., Holko, M., Sadler, A.J., Scott, B., Higashiyama, S., Berkofsky-Fessler, W., McConnell, M.J., Pandolfi, P.P., Licht, J.D., and Williams, B.R.G. (2009b). Promyelocytic Leukemia Zinc Finger Protein Regulates Interferon-Mediated Innate Immunity. *Immunity* *30*, 802–816.

Xu, F., Li, X., Wu, L., Zhang, Q., Yang, R., Yang, Y., Zhang, Z., He, Q., and Chang, C. (2011). Overexpression of the EZH2, RING1 and BMI1 genes is common in myelodysplastic syndromes: relation to adverse epigenetic alteration and poor prognostic scoring. *Ann. Hematol.* *90*, 643–653.

Xu, J., Shao, Z., Li, D., Xie, H., Kim, W., Huang, J., Taylor, J.E., Pinello, L., Glass, K., Jaffe, J.D., et al. (2015). Developmental Control of Polycomb Subunit Composition by GATA Factors Mediates a Switch to Non-Canonical Functions. *Mol. Cell* 57, 304–316.

Xu, K., Wu, Z.J., Groner, A.C., He, H.H., Cai, C., Lis, R.T., Wu, X., Stack, E.C., Loda, M., Liu, T., et al. (2012). EZH2 Oncogenic Activity in Castration-Resistant Prostate Cancer Cells Is Polycomb-Independent. *Science* 338, 1465–1469.

Yeyati, P.L., Shaknovich, R., Boterashvili, S., Li, J., Ball, H.J., Waxman, S., Nason-Burchenal, K., Dmitrovsky, E., Zelent, A., and Licht, J.D. (1999). Leukemia translocation protein PLZF inhibits cell growth and expression of cyclin A. *Oncogene* 18, 925–934.

Yoshida, C., Tokumasu, F., Hohmura, K.I., Bungert, J., Hayashi, N., Nagasawa, T., Engel, J.D., Yamamoto, M., Takeyasu, K., and Igarashi, K. (1999). Long range interaction of cis-DNA elements mediated by architectural transcription factor Bach1. *Genes Cells Devoted Mol. Cell. Mech.* 4, 643–655.

Yu, M., Mazor, T., Huang, H., Huang, H.-T., Kathrein, K.L., Woo, A.J., Chouinard, C.R., Labadorf, A., Akie, T.E., Moran, T.B., et al. (2012). Direct Recruitment of Polycomb Repressive Complex 1 (PRC1) to Chromatin by Core Binding Transcription Factors. *Mol. Cell* 45, 330–343.

Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011). H3K36 Methylation Antagonizes PRC2-mediated H3K27 Methylation. *J. Biol. Chem.* 286, 7983–7989.

Zhang, J., Ding, L., Holmfeldt, L., Wu, G., Heatley, S.L., Payne-Turner, D., Easton, J., Chen, X., Wang, J., Rusch, M., et al. (2012). The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* 481, 157–163.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.-J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X-chromosome. *Science* 322, 750–756.

Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E., Borowsky, M., and Lee, J.T. (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol. Cell* 40, 939–953.

Zollman, S., Godt, D., Privé, G.G., Couderc, J.L., and Laski, F.A. (1994). The BTB domain, found primarily in zinc finger proteins, defines an evolutionarily conserved family that includes several developmentally regulated genes in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 91, 10717–10721.

Annexe

Annexe 1 : Matériel et méthodes détaillé

Cell culture and transfection methods.

Human myeloblastic cell line KG1 was maintained at exponential growth in RPMI supplemented with 10% foetal calf serum and 1% sodium pyruvate. 293T cells were maintained at exponential growth in DMEM supplemented with 10% foetal calf serum and 1% sodium pyruvate.

293T cells (3×10^6) were seeded previous day into a 10 cm dish and transfected using Lipofectamine 2000 (Life Technologies) according to the manufacturer's instructions. Cells were transfected with 1µg of WT or mutant PLZF and with 5µg of EZH2 or empty plamid. Cells were harvested 24 h later and lysed.

Nuclear extracts.

KG1 cells were harvested by centrifugation and washed twice with ice-cold 1x PBS. After centrifugation at 4°C for 10 minutes, the cell pellet was resuspended and incubated 10 minutes in hypotonic buffer (10 mM Hepes, pH 7,9, 1,5 mM MgCl₂, 10 mM KCl, 0,5 mM DTT and 0,5 mM PMSF). The suspension was dounced thirty to forty times in a Dounce homogenizer and centrifuged 15 minutes at 4°C. The nuclear pellet was resuspended in one volume of low salt buffer (20 mM Hepes, pH 7,9, 25% glycerol, 1,5 mM MgCl₂, 20 mM KCl, 0,2 mM EDTA, 0,5mM DTT and 0,5 mM PMSF) and one volume of high salt buffer (20 mM Hepes, pH 7,9, 25% glycerol, 1,5 mM MgCl₂, 1,2 M KCl, 0,2 mM EDTA, 0,5mM DTT and 0,5 mM PMSF) and left for 2 hours at 4°C. After centrifugation at 4°C and 13.000 rpm, nuclear extracts were kept frozen at -80°C.

293T cells were harvested and washed twice with ice-cold 1x PBS. Pellets were resuspended in buffer A (10mM Hepes pH 7,9, 1,5mM MgCl₂, 10 mM KCl and 0,5mM DTT,) and incubated 10 minutes at 4°C. After centrifugation pellets were resuspended in A* buffer (buffer A plus 0,2% NP40, 0,5 mM PMSF and protease inhibitors EDTA free, Roche) and incubated 2 minutes at 4°C. Cytoplasmic fraction was discarded and nuclear pellet was resuspended and incubated in buffer B (20 mM Hepes pH 7,9, 0,2 mM EDTA pH8, 420 mM NaCl, 25% glycerol, 0,5 mM DTT, 0,5 mM PMSF and protease inhibitors EDTA free). After centrifugation at 4°C and 13.000 rpm, nuclear extracts were kept frozen at -80°C.

Co-immunoprecipitation (Co-IP).

Co-IP of endogenously expressed proteins was performed using KG1 cells. Nuclear extracts were diluted in hypotonic buffer and incubated 3 hours with 2 µg of anti-PLZF (Santa Cruz Biotechnology, H300) or anti-EZH2 (Active Motif, pAb, cat 39901 or AC22 antibody) antibody in the presence of Protein G beads using (Dynabeads, Life Technologies) and resulting complexes were washed, denatured and eluted according to the manufacturer's instructions.

Nuclear extracts of 293T cells were IP with the ProFound c-Myc tag IP/coIP kit (No 23 622) following the manufacturer's instructions.

Yeast two-hybrid system.

The baits and preys were cloned into modified vectors (pGBT9 and pACT2 respectively, Clontech) using the Gateway technology (Invitrogen™).

Baits and preys were then cotransformed in AH109 yeast strain (MATa, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2::GAL1UAS-GAL1TATAHIS3, GAL2UAS-GAL2TATA-ADE2, URA3::MEL1 UASMEL1TATA-lacZ, MEL1) by the Lithium acetate method.

Following transformation, yeast were plated onto Synthetic Complete (SC) medium minus Leucine (-L), minus Tryptophane (-W) and incubated at 30°C for 4 to 5 days. Yeast clones were then transferred in liquid SC-WL for 3 days at 30°C with agitation to normalize the yeast cell concentration used for the phenotypic assay. The cells were then or directly spotted on nitrocellulose membrane on YPD plate for BGal assay or diluted 1/20 in water and spotted onto selective medium (-WHL) for phenotypic assay.

Plasmids.

The expression vectors for PLZF and PLZF mutants were previously described in (Boukarabila et al., 2009). The expression vector for EZH2 was obtained courtesy of Dr Djabali.

Knockdown.

EZH2 and PLZF knockdown were achieved using a doxycycline-induced short hairpin RNA (sh-RNA)-targeting EZH2 or PLZF (pTRIPZ-EZH2, openbiosystem # V2THS_63066, pTRIPZ-PLZF, openbiosystem # V2LHS_72096). A non-silencing sh-RNA (pTRIPZ-NS, openbiosystem # RHS4743) was used as control. Viral particules were produced by the AniRA platform (Lyon). KG1 cells were infected by spinoculation and cells containing the

pTRIPZ were selected on puromycin (1,75µg/mL). Knockdown of PLZF and EZH2 was obtained by the addition of doxycycline (2µg/ml) to the cells during 7 to 10 days.

Western blotting

WB were performed as previously described (Boukarabila et al., 2009). Briefly, nuclear extracts were separated on polyacrylamide gels, blotted on nitrocellulose membrane and incubated with different primary antibodies (see list 1).

Chromatin Immunoprecipitation (ChIP).

KG1 cells were fixed in 1% formaldehyde for 8 min at room temperature with shaking. The reaction was quenched by 125mM glycine for 5 min at room temperature. Cells were washed with ice-cold 1 x PBS and resuspended in cell lysis buffer (10 mM Tris-HCl pH 8, 10 mM EDTA, 0,5mM EGTA, 0,25% triton, 0.5mM PMSF and 1X protease inhibitor cocktail) and incubated on ice for 5 min to separate nuclei from cytoplasm. The cells were centrifuged at 6000 rpm for 1 min and nuclei were re-suspended in 1 ml nuclei lysis buffer (250mM NaCl, 50mM Tris-HCl pH 8, 1mM EDTA, 0,5mM EGTA, 0.5mM PMSF and 1X protease inhibitor cocktail) and incubated 15 min at 4°C. Nuclear pellets were resuspended in SDS buffer for the chromatin shearing (1% SDS, 10 mM Tris HCl pH8, 1 mM EDTA, 0.5 mM EGTA, 0.5mM PMSF and 1X protease inhibitor cocktail). DNA was sonicated to obtain 100-500 base pair fragments (ultrasonic processor, Fischer Scientific, ref 75041, 40% amplitude, 30 s on and 30 s off for 20 minutes). The chromatin was diluted 1:7 with ChIP dilution buffer (0.6% Triton, 0.06% NaDOC, 150 mM NaCl, 12 mM Tris.HCl pH8, 1 mM EDTA, 0,5mM EGTA, 0.5mM PMSF and 1X protease inhibitor cocktail) and was pre-cleared with 80 µl protein G sepharose beads (GE healthcare) for 2 h at 4°C with rotation before incubation with the respective immunoprecipitating antibody (see list 2) at 4°C overnight with shaking. Immunoprecipitated complexes were isolated by incubation with 100 µl of protein G sepharose for 3 h at 4°C, washed extensively with six different buffers (wash 1 : 1% Triton-X, 0.1% sodium deoxycholate, 10 mM Tris-Cl pH 8, 150 mM NaCl ; wash 2 : 0.5% NP40, 0.5% sodium deoxycholate, 0.5% Triton-X, 10 mM Tris-Cl pH 8, 150 mM NaCl ; wash 3 : 0.7% Triton, 0.1% sodium deoxycholate, 250 mM NaCl, 10 mM Tris.HCl pH8 ; wash 4 : 0.5% NP40, 0.5% sodium deoxycholate, 250 mM LiCl, 20 mM Tris.HCl pH8, 1 mM EDTA ; wash 5 : 0.1% NP40, 150 mM NaCl, 20 mM Tris.HCl pH8, 1 mM EDTA ; wash 6 : 20 mM Tris.HCl pH8, 1 mM EDTA). The third Wash was omitted for PLZF ChIP. Washes were performed for 5 minutes at 4°C with rotation and beads were collected by centrifugation at 3000 rpm, at 4°C

for 1 min. For sequencing, DNA was eluted with the I-Pure kit (Diagenode) following the manufacturer's instructions. For qPCR, DNA was purified using chelex-100 (Bio-Rad Laboratories). In brief, 100 µl of a 10% chelex-100 resin solution was added to the tubes containing sepharose beads and the tubes were boiled at 95°C for 10 min. RNase A (Sigma, R6513) was added during 10 minutes at RT before the proteinase K (Diagenode) treatment 20 minutes at 65°C. The samples were centrifuged (13 000 rpm, 1 min) to pellet the sepharose beads and resin after heat inactivation of proteinase K at 100°C during 10 min. DNA is recovered using a QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's instructions.

RNA isolation and quantitative PCR

Total RNA was extracted using the RNeasy Mini kit according to the manufacturer's protocol (Qiagen). Reverse transcription was performed with 500 ng of the total RNA using the Transcriptor High Fidelity cDNA Synthesis Kit (Roche) according to the manufacturer's protocol.

Quantitative (q)-PCR analysis

Quantification of ChIPed DNA or cDNA was performed by quantitative real-time PCR. DNA target sequences were selected based on our ChIP-Seq analyses in KG1 cells and primers were designed for amplifying the DNA location of interest with Primer-BLAST. Real-time PCR was performed using Power SYBR Green PCR Master Mix (Applied Biosystems) according to the manufacturer's instructions with the 7500 Fast Real Time PCR system (Life Technologies). For ChIPed DNA, IgG control "cycle over the threshold" Ct values were subtracted to Input or IP Ct values and converted into bound value by $2^{-(IP\ Ct\ or\ input\ Ct - IgG\ IP\ Ct)}$. Data are expressed as % of bound /input. Primers used are given in table 1.

For cDNA, relative expression levels were determined by the 2 delta-delta Ct method, using HPRT or PBG-D as reference genes. Experiments were performed with two individual clones treated or not with the doxycycline. Technical triplicates are presented as mean values $\pm$ SD. Primers used are listed in table 2.

ChIP-sequencing and data processing.

Before sequencing, ChIPed DNA was quantified using the Qubit fluorometer method (Life Technologies). ChIP-seq experiments were performed on two biological replicates.

Sequencing and mapping.

Libraries for sequencing after chromatin immunoprecipitation were performed according to the manufacturer's protocol (Illumina) either at the Montpellier GenomiX (MGX) or at Harvard sequencing MIT BIO MICRO. Shortly, ChIPed DNA fragments were size selected on an agarose gel to enrich for 350 bp fragments, adapters were ligated and libraries were amplified using polymerase chain reaction (PCR). Size of the amplified fragments was analyzed on a bioanalyzer. Libraries were multiplexed (4 or 5 per lane) and sequencing was performed with a *HiSeq* 2000 next-generation sequencing platform (Illumina) using a 40 bp single-end protocol. Total read counts and unique read counts are presented in table 3.

Sequence data quality in each sample was assessed using FastQC (a quality control tool for high throughput sequence data, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads with Phred quality score less than 30 were filtered out and sample replicates were merged into a single fastq file. Sequences were mapped to the human genome (hg19) using BFAST alignment tool (ref: PMID 19907642) with default parameters.

Mapped sequencing data (BAM files) can be down loaded from the GEO under accession number GSE XXXX (in process).

In addition to our own ChIP-seq data, previously published data sets were retrieved from the **Gene Expression Omnibus** (GEO) database for re-analysis and comparisons. Sequencing datasets were thus retrieved for: human hematopoietic cells enriched within the immature cell surface marker CD34+ (HSC-CD34⁺) H3K27me3 (GSM772951), HSC-CD34⁺ H3K4me3 (GSM773041), (HSC-CD34⁺) H3K27ac. In addition, expression data from the HSC CD34+ cells of healthy donor (GSM308470) and expression data from human untreated KG1 cells (GSM1316691) were extracted.

Peak calling and annotation.

High confidence binding sites were determined through peak calling using HOMER tool (Heinz, 2010 #2532) with the “factor setting” for EZH2, PLZF and PolII and the “histone setting” for histone marks (H3K27me3 and H3K4me3) and the appropriate input as control. False-discovery-rate (FDR) cutoff was set at <0.05 and both overall and local p-value cutoff at 0.01. To increase the accuracy of the identified ChIP enrichment sites, only peaks with ≥ 3 -fold tag count enrichment compared to both input control and local tag density (on region of 10kb around putative peak) were considered for downstream analyses with an additional filter of peak scores ≥ 3 . Moreover, for all ChIP-seq datasets, tag counts were normalized by the total number of mapped tags such that each directory contains 10 million reads.

In order to identify genes that contain enriched regions or peaks, we used HOMER annotate Peaks tool on BED files containing the retained peaks of each sample. We identified therefore all target genes and the distance of every peak to the nearest TSS (Transcription Start Site). Binding sites were at considered the promoters if they were found at +/- 1kb of the TSS. Binding sites were considered at the gene body if they could be detected between 1-kb upstream of the TSS and the Transcription End Site (TES).

Visualization and plotting enrichment profiles on genomic features.

Heatmap visualisation of ChIP enrichment at EZH2 and PLZF-enriched non-redundant transcripts was performed using ngs.plot (version 2.47; ref: <http://www.biomedcentral.com/1471-2164/15/284>).

Plotting of the enrichment profiles were done by ranking the genes using K-means clustering method (here set up at 6 clusters). The average profiles were calculated using the mean read coverage among the gene bodies in each resulted cluster. Quantification of ChIP enrichment for profile plots was performed using the mean (factor binding) or median (histone marks) enrichment signal for each gene. The regions used to calculate enrichment signals were different for EZH2 and for H3K4me3 heatmap figure and correspond to “promoter -proximal regions” (defined as -2000 bp to the TSS) for EZH2 and an H3K4me3 region (first 33% of the gene body) for H3K4me3. The calculated enrichment signal for each gene was then mean smoothed using a 40-gene sliding window across all genes and average profile plots generated in R.

Expression data

To explore gene expression in both KG1 and HSC CD34+ cells, we retrieved publicly available expression data. To determine if a gene was expressed or not (present or absent), we used the “ABS_CALL” values determined by the Affymetrix GeneChip with “Detection Call” algorithm described by (Liu et al., 2002). The corresponding probe set is present (flagged “P”, $p\text{-value} \leq 0,05$), absent (flagged “A”, $p\text{-value} > 0,066$) or marginal (flag “M”, $0,05 \leq p\text{-value} \leq 0.065$).

List 1: Antibodies used for Western Blot

Anti-EZH2 BD Biosciences # 612666 (1/1000), anti-PLZF H300 (Santa Cruz Biotechnology, 1/2000), anti-LSD1 (Active Motif, cat 39186, 1/5000), anti-BMI1 (Abcam, 1.T.21, 1/500), EZH1 (Abcam, ab13665, 1/500), anti-MLL1 (Millipore, N4.4, 1/400), anti-WDR5 (Active

Motif, cat 61485, 1/1000), anti-H3K27me3 (Millipore, cat 07-449, 1/500), anti-H3K4me3 (Diagenode, pAb-003-050, 1/1000), anti-Tubulin (Sigma Aldrich, DM4A, 1/5000), anti-H3 (Active Motif, 39163, 1/10000).

List 2: Antibodies used for ChIP

4 µg of anti-PLZF (H300, Santa Cruz Biotechnology) or 10 µL of anti-EZH2 (Active Motif, pAb, cat 39901) or 2µg of anti-H3K27me3 (Abcam, ab6002) or anti-H3K4me3 (Diagenode, pAb-003-050) or 8µg PolII (N20X, Santa Cruz Biotechnology) or 4µg of H3K36me3 (Abcam, ab9050) were used for ChIP.

Table 1:

a) Primers used for ChIP-qPCR (EZH2 and PLZF ChIP)

Gene	Forward primer	Reverse primer
GAPDH	ACTCACCTGCCCTCAATATC	AGACAGTGTGCCTTTCATTCCAT
SIX4	GCCACGCAGTCACCATTAAG	TAGTATGCAGCAACCGGAGG
MYO6	AAAGTAAATGCGTCGGCGAG	GCCATTCGGTCTTTGGAGGA
NEO1	CACAATTGCCCGCTCACAAC	CAGATACGATCACTCCGAGCC
ASXL2	CATTCCCACAGGGATCGCAA	AGGGAAAGGGAGGTCACTGG
RARA	AGTGAAACCTTCAGCCCAGG	TTTCCAATTCTGCCGTGAC
RARA (K4 peak)	CCGCTGGGAACTTCTGTTCT	GGCTAGTCCGTTGAGCATCC
RUNX2 (EZH2 peak)	TAAGATGGGAGGGCGTGAGA	TAGAATGCCTCCCCCTCGAA
RUNX2 (PLZF peak)	CATCCTGTGCCCCAACTGTA	ATGGGCTGGCTGAGTAAAGG
SPI1	GGGTCATGAGGTCACTTGGG	GGAAGCTGACACAGGCTCTT
CCND3	GACACCAAGCCCAGAGTAA	TTCCTGTCAGTTAACCCGGC
ELMO1	AAGCCTGTCACTCGGTGATG	CTGTGGCTTCGAGTCTGGTT
ID2	CCCACACTAAGCCTGTCGTC	ACGTTTGTAGAGTGAGTTCGTCT
CALY	AAATAGCTCACAGGCGGCTC	GCATCCTCCGCATCCACAT
KLF3	ACTAGCCATTGGAATGGGGC	CAAGCAGCAGAGACCTTCCT
P16	Described in (Kia et al., 2009)	

Table 2: Primers used for cDNA RT-Q-PCR

Gene	Forward primer	Reverse primer
PLZF	TTCATCCAGAGGGAGCTGTT	CCGTACGTCTTCATCCCACT
EZH2	Described in (Shen et al., 2012)	
HOXA9	TGTCCCACGCTTGACACTC	GGGTTATTGGGATCGATGGG
RUNX1	GGTTTCGCAGCGTGGTAAAA	GCACTGTGGGTACGAAGGAA
MYC	CCTTCTCTCCGTCCTCGGAT	CTTGTTCCCTCCTCAGAGTCGC
MEIS1	GCGCAAAGGTACGACGATCT	TACTGATGCGAGTGCAGAGG
DUSP6	CCTGGAAGGTGGCTTCAGTA	CGGTCAAGGTCAGACTCGAT
GFI1	TCCACACTGTCCACACACCT	CTTGAAGCCTGTGTGTTTGC
JAK2	GCTGGTCGGCGTAATCTAAA	TGCCAGATCCCTGTGGATA
RARA	CTGCGAGCATCACAGGACAT	GGGTTGGGGGATGTCTCAAG
SPI1	ACAGGCGTGCAAAATGGAAG	GAAGTCCCAGTAATGGTCGCT
ID2	TGTGGCTGAATAAGCGGTGT	GGTCCATTCAACTTGTCTCTCT
SMAD7	GGGGGCTTTCAGATTCCCAAC	TTGAGCTGTCCGAGGCAAAA
HPRT	GACCTTGATTTATTTTGCATACC	CGAGCAAGACGTTTCAGTCCT
PBG-D	TGGGCAACTGTACCTGACTG	TCGTGGAATGATACGAGCAG

Table 3: Unique and Total read counts for EZH2, PLZF, H3K27me3 and H3K4me3 ChIPs

Samples	Unique Read Counts	Total Read Counts
ChIP_EZH2_shCtrlA (run 386)	15 025 997	14 707 671
ChIP_EZH2_shCtrlA (run 389)	21 203 632	20 636 900
ChIP_EZH2_shCtrlB	31 231 505	30 226 851
ChIP_EZH2_shPLZF_A	27 534 761	26 502 663
ChIP_EZH2_shPLZF_B	37 917 649	36 102 049
ChIP_H3K27me3_shCtrlA	31 099 531	27 207 673
ChIP_H3K27me3_shCtrlB	30 479 242	29 262 442
ChIP_H3K27me3_shPLZF_A	43 641 088	41 595 787
ChIP_H3K27me3_shPLZF_B	30 635 420	29 705 284
ChIP_H3K4me3_shCtrlA	31 767 633	24 346 865
ChIP_H3K4me3_shCtrlB	53 182 303	37 111 248
ChIP_H3K4me3_shPLZF_A	26 670 132	23 519 878
ChIP_H3K4me3_shPLZF_B	35 990 714	30 855 017
ChIP_PLZF_shCtrlA	31 834 483	30 847 029
ChIP_PLZF_shCtrlB	37 576 666	35 482 759
Input_shCtrlB	21 567 388	20 419 119
Input_shPLZF_A	31 203 169	30 216 072

Résumé de la thèse

Les protéines du groupe Polycomb (PcG) sont des facteurs épigénétiques dont le rôle est de maintenir la répression de leurs gènes cibles au niveau de la chromatine via la modification des protéines histones. EZH2 est une protéine clé dans les mécanismes de régulation puisqu'elle catalyse la mise en place de la marque répressive H3K27me3. Dans le cadre de ma thèse, je me suis intéressée au modèle des leucémies aiguës myéloïdes (LAM) dans lesquelles, contrairement à d'autres pathologies myéloïdes, des mutations affectant EZH2 ou des membres Polycomb ne sont retrouvées que très rarement (<1%). Des études ont montré que dans ce type de leucémies, de nombreux gènes cibles d'EZH2 sont dérégulés bien que son activité répressive soit toujours présente, mettant en évidence d'éventuels défauts de recrutement de cette protéine. Parmi les facteurs de transcription susceptibles de réguler l'association des protéines PcG à la chromatine, se trouve PLZF qui est un candidat intéressant. En effet, le laboratoire a mis en évidence une interaction entre PLZF et la protéine Polycomb BMI-1 et a montré que la distribution génomique de PLZF concorde avec celle de certains composants des Polycomb. L'objectif de mon travail de thèse a donc été de déterminer dans quelle mesure PLZF intervient dans le recrutement ou l'activité d'EZH2. J'ai pour cela étudié l'interaction fonctionnelle entre PLZF et EZH2 à la chromatine dans un modèle cellulaire de LAM. Après avoir montré que les deux protéines interagissent *in vivo*, j'ai montré par des expériences de ChIP-seq que PLZF et EZH2 régulent un groupe de gènes communs. De façon intéressante et contrairement aux résultats attendus, les gènes cibles de PLZF sont majoritairement associés à la marque histone activatrice H3K4me3 et à une transcription active. Par ailleurs, nos résultats montrent également que la localisation génomique de PLZF (promoteur versus gene body) influence le niveau d'expression des gènes. Enfin, nous avons montré que PLZF régule les gènes bivalents de façon différente des autres gènes puisque dans ce cas particulier, PLZF a un effet répresseur. Ces résultats apportent donc de nouveaux éléments dans la compréhension des mécanismes de la régulation épigénétique impliquant PLZF et EZH2. Ces travaux m'ont permis de prendre part à deux projets collaboratifs au cours desquels nous avons exploré de façon plus poussée la régulation hématopoïétique normale et leucémique. Le premier projet a permis de mettre en évidence le rôle régulateur de PLZF sur les gènes du cycle cellulaire dans les cellules souches hématopoïétiques lors de l'hématopoïèse normale. Le second projet a révélé un nouveau mécanisme épigénétique de dérégulation de l'oncogène TAL1 dans le cas de patients atteints de leucémies aiguës lymphoïdes T. Dans son ensemble, mon travail a donc permis de mettre l'accent sur des mécanismes épigénétiques originaux qui régulent le système hématopoïétique et qui se retrouvent dérégulés dans les hémopathies malignes.

Abstract

Polycomb group (PcG) proteins are epigenetic factors which play a major role in maintaining epigenetic silencing via histone modifications at the chromatin level. EZH2 is a key regulator that catalyzes the trimethylation of H3K27, which is a repressive mark. During my PhD, I was interested in the acute myeloid leukemia (AML) model in which, unlike other myeloid malignancies, EZH2 or other PcG protein mutations are very rare (<1%). Studies have shown that in this type of leukemia, many of EZH2 target genes are deregulated although its repressive activity is still present highlighting possible EZH2 recruitment defects. Among the transcription factors that regulate the association of PcG proteins to chromatin, the transcription factor PLZF is an interesting candidate. Indeed, the laboratory has demonstrated an interaction between PLZF and the Polycomb protein BMI -1 and showed that the genomic distribution of PLZF is consistent with that of some Polycomb components. The aim of my thesis was therefore to determine in which extent PLZF is involved in the recruitment or activity of EZH2. I studied the functional interaction between PLZF and EZH2 at the chromatin level in a cellular model of human AML. After showing the *in vivo* interaction between the two proteins, ChIP -seq experiments showed that EZH2 and PLZF regulate a group of common genes. Interestingly and contrary to the expected results, PLZF target genes are mostly associated both with the active histone mark H3K4me3 and with an active transcription. Furthermore, our results also show that the genomic localization of PLZF (promoter versus gene body) influences the expression level of genes. Finally, we have shown that PLZF regulates bivalent genes differently from other genes because in this particular case, PLZF had a repressive effect. These results bring new elements for the understanding of epigenetic regulation mechanisms involving PLZF and EZH2. These results were also integrated in two collaborative projects that have allowed us to explore more extensively the regulation of normal and leukemic hematopoiesis. The first project highlighted the PLZF regulatory role on cell cycle genes in hematopoietic stem cells during normal hematopoiesis. The second project revealed a new mechanism of epigenetic dysregulation of the TAL1 oncogene in the case of patients with T acute lymphoblastic leukemia. Overall, my work focused on original epigenetic mechanisms that regulate hematopoietic system or, and that are dysregulated in malignant hematopoiesis.