UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE DE STRASBOURG

IGBMC, UMR 7104, U 964

THÈSE

présentée par :

Ujjwal Kumar BANERJEE

soutenue le : **18 décembre 2017**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Aspects moléculaires et cellulaires de la biologie

3-D Genome organization of DNA damage repair

THÈSE dirigée par :

Dr SOUTOGLOU Evi	DR, IGBMC, université de Strasbourg

RAPPORTEURS :

Dr LEGUBE Gaëlle	DR, LBCMCP-UMR5088, université de Toulouse
Dr HUERTAS-SÁNCHEZ Pablo	Professor, CABIMER, University of Seville, Spain

---

AUTRES MEMBRES DU JURY :

Dr SEXTON Thomas	CR, IGBMC, Illkirch, France

Acknowledgements

I would like to express my sincere gratitude to my thesis supervisor Dr. Evi Soutoglou for offering me the precious opportunity to do my PhD in her laboratory and giving me this project. Her leadership skills, enthusiasm, kindness and encouragement will be forever admired and never forgotten. Thank you for having faith in me and giving me a chance. Mere words will not suffice to express my feelings and I hope my simple “Thank you” carries with it all the weight of my gratitude and thankfulness.

I would also like to express my gratitude to the members of the jury Dr. Gaëlle Legube, Dr. Pablo Huertas and Dr. Thomas Sexton for taking the time to evaluate my work. I would like to thank Dr. Gaëlle Legube for her feedback and suggestions during my mid-thesis evaluation. A special thanks to Dr. Sexton for his ideas and conversations which were always encouraging.

I owe an enormous debt of gratitude to all the current and past members of my lab Katerina, Alkmini, Duygu, Audrey, Celine, Ana, Ioanna, Indrajeet, Mixalis and Lyuba. Thank you all for bearing with my antics, my experiments with cooking, my offkey renditions of French songs and my debates on nearly every topic imaginable. Katerina, your motivation and encouragement has been instrumental in keeping me going through some of the toughest times during my PhD, and I can never repay your kindness. Thank you. Thank you very much Duygu, for listening to my weird ideas and for being an interpreter when I had to get things done in French. Thank you Alkmini, for all the discussions on science, food, T.V. series and pop-culture. Being an only child, I never knew what it was like to have an older sister until I met you. Audrey and Celine, thank you for being so patient with my strange Indian accent, French pronunciation and teaching me a smattering of French. Grand merci! Thank you Indrajeet for all your feedback on my writing and help in the 53BP1 project. Thank you Ioanna for your support and motivation. I wish you Mixalis, the very best of luck in your project. Your “Reparito” will always be a hit record in my playlist!!! Thank you Lyuba for being a good sparring partner in our debates. I thank everyone of you for your support and feedback on scientific and non-scientific matters. The winds of time will take us all our separate ways, but for me, all of you shall forever be family.

I thank my friends Sanjay Chahar and Atish Mukherjee and their families for being motivating and treating me as a younger brother. You have seen me at my worst and brought me out of the doldrums. Your unique sense of humour and ‘lemons to lemonade’ outlook on life taught me to look at the brighter and lighter side of things, no matter what the situation.

Thank you very much Pooja Mukherjee for being a friend and confidante. You have been the lighthouse which brought me ashore through the choppy waters of PhD.

Thank you Mohammed Ashick Saleem, for all your help with bioinformatics and understanding the intricacies of coding. A lot of the analysis of this project would not have been possible without your feedback. Thank you very much for your help and support.

I would like to thank my friends Aditya, Adam, Federica, Federico, Samantha, Olivier, Ioannis and everyone else for helping me throughout my time here at the IGBMC. Your kind words of encouragement kept me going. My life is richer for having known you.

I would like to thank Bernard Jost and Romain Kaiser at the Next-Generation sequencing and Tao Ye at the Bioinformatics platform for all their help in the sequencing and data analysis. Thank you very much for all the analysis and bearing with my questions and analysis requests.

I would also like to thank Claudine Ebel and Muriel Phillips at the Flow cytometry platform for all the help in cell sorting and data analysis. Thank you for being so kind and patient when I was running late with my experiments.

I would like to thank La Ligue Contre le Cancer for providing financial support during the course of my PhD.

Last and definitely not the least, I would like to thank my parents and my Nani for being my pillars of support and backing me up in all my endeavours. Everything that I am, is because of your blessings and good wishes.

Ujjwal Kumar Banerjee

Contents

List of figures

List of abbreviations

53BP1	p53-Binding Protein 1
53BP1-NB	53BP1 nuclear bodies
6,4-PP	6-4 pyrimidone photoproducts
AID	Activation-Induced cytidine Deaminase
Alt-EJ	Alternative End-Joining
AP	Apurinic/apyrimidinic
ATM	Ataxia - Telangiectasia Mutated
ATR	Ataxia Telangiectasia and Rad3 related
ATRIP	ATR Interacting Protein
BLM	Bloom helicase
BaP	Benzo[a]pyrene
BER	Base excision repair
BLESS	Break labelling, enrichment on streptavidin and next generation sequencing
BPDE	Benzo[a]pyrene diol epoxide
BRCA1	Breast Cancer 1
BRCA2	Breast Cancer 2
BRCT	BRCA1 C Terminus
CDK	Cyclin-Dependent-Kinase
CHD	Chromodomain-Helicase-DNA binding
ChIP-Seq	Chromatin immunoprecipitation and sequencing
Chk1	Checkpoint kinase 1
Chk2	Checkpoint kinase 2
cNHEJ	classical NHEJ
CPD	Cyclobutene pyrimidine dimers
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DDR	DNA Damage Response
D-loop	Displacement-loop
DNAPK	DNA-dependent Protein Kinase
DNAPKcs	DNAPK catalytic subunit
DSB	Double Strand Break
DUBs	Deubiquitinating enzymes
ERCC1	Excision repair cross-complementing 1
EXD2	Exonuclease 3’-5’ domain-containing protein 2
EXO1	Exonuclease 1
FHA	Forkhead associated
G0 phase	Gap 0 phase
G1 phase	Gap 1 phase
G2 phase	Gap 2 phase
GCN5	General Control Of Amino-Acid Synthesis 5
HC	Heterochromatin
HDAC	histone deacetylase
HJ	Holliday-Junction
HR	Homologous recombination
Ig	Immunoglobulins
IRIF	Ionizing Radiation-Induced Foci

ISWI	Imitation Switch
KAP1	KRAB-associated protein-1
LEDGF	Lens epithelium-derived growth factor
MACS	Model based analysis for ChIP-Seq
MDC1	Mediator of DNA damage checkpoint protein 1
MMSET	Multiple Myeloma SET domain
MRN	Mre11-RAD50-NBS1
NBS1	Nijmegen Breakage Syndrome 1
NCS	Neocarzinostatin
NER	Nucleotide excision repair
NHEJ	Non-Homologous End Joining
PALB2	Partner and localizer of BRCA2
PARP	poly-ADP-ribosyl polymerases
PAXX	Paralog of XRCC4 and XLF
PNKP	Polynucleotide Kinase 3' Phosphatase
PP	Protein Phosphatase
PTM	Post-Translational Modification
RECQ5	RecQ Like helicase 5
RIF1	RAP1-interacting factor 1
RNF8	Ring Finger protein 8
RNF20	Ring Finger Protein 20
RNF168	Ring Finger Protein 168
RNF169	Ring Finger Protein 169
ROS	Reactive Oxygen Species
RPA	Replication Protein A
RPKM	Reads Per Kilobase Million
RSC	Chromatin structure remodeling
SMARCAD1	SWI/SNF-related, matrix-associated actin-dependent regulator of chromatin, subfamily a, containing DEAD/H box 1
SMC	Structural Maintenance of Chromosome
ssDNA	single-stranded DNA
SWI/SNF	Switching defective/sucrose non-fermenting
TADs	Topologically Associated Domains
TIP60	Tat interactive protein 60 kDa
TopBP1	Topoisomerase II Binding Protein 1
TSS	Transcriptional Start Site
UBC13	Ubiquitin-Conjugating Enzyme 13
USP3	Ubiquitin specific peptidase 3
USP16	Ubiquitin specific peptidase 16
UV	Ultraviolet
WSTF	Williams syndrome transcription factor
XLF	XRCC4-like factor
XRCC1	X-ray repair cross-complementing protein 1
XRCC4	X-ray repair cross-complementing protein 4

Thesis summary

Our genome is constantly under attack by endogenous and exogenous factors which challenge its integrity. These can result in deaminations, spontaneous depurinations, oxidative lesions and single- and double-strand breaks (SSBs and DSBs). Out of these damages, DSBs are among the most deleterious since they may lead to loss of genetic information, translocations and cell death. To address these problems, cells have developed a highly conserved mechanism which not only senses DNA damage and its subsequent repair, but also coordinates repair with other cellular processes like, transcription status, cell cycle progression and apoptosis. There are two major pathways for DNA damage repair: Homologous recombination (HR) and non-homologous end joining (NHEJ) depending on the cell cycle state. HR is a high-fidelity mechanism since it uses the homologous chromosome as a template for carrying out repair but only occurs in proliferating cells during S/G2. On the other hand, NHEJ occurs regardless of the cell cycle state and does not need a template but can be an error prone repair mechanism.

Regardless of whether a DSB is generated by endogenous or exogenous sources, the subsequent repair mechanism occurs by a three-step signaling cascade which comprises of sensor proteins, adaptor proteins and effector proteins. The signaling cascade begins with sensing of the lesion by the MRN (Mre11/Rad51/Nbs1) complex which in turn recruits and activates ATM, a key signaling molecule. ATM further phosphorylates the histone variant H2AX to produce γ-H2AX which forms the hallmark for DNA damage. Another crucial player in the repair mechanism is the tumour-suppressor 53 binding protein 1 (p53BP1/53BP1). This protein rapidly associates with DSBs and other DNA damage repair proteins. It is an adaptor protein which is required for processing the DNA damage response and as a platform for the recruitment of other factors. It has also been shown to have a role in the selection of the DSB repair pathway. In addition to damage induced foci, 53BP1 can also be present in undamaged cells due to replication stress but remains limited to the G1 phase of the cell cycle. Some genomic regions are difficult to replicate because of a scarcity of replication origins. These regions, termed as fragile regions, may not complete replication during a given S phase. Recent research has shown the importance of fragile site metabolism for genome integrity and cancer development. Fragile sites are especially challenging for a cell as these unreplicated regions do not classify as a damage to the DNA and hence may not trigger a DNA damage response. Although evidence of the presence of 53BP1 bodies localized around fragile sites exist, the main problem for the analysis of these bodies remains the lack of a suitable antibody against 53BP1 which would facilitate easy analysis of its interaction partners and genomic distribution. The question remains regarding the continued presence of fragile sites with an increased propensity for replication stress induced DNA damage and their gradual elimination from the genome through evolution.

All the above-mentioned repair processes happen in the context of a highly organized and compartmentalized chromatin. The chromatin in the nucleus is not a uniform macromolecule but comprised of different domains characterized by complex DNA structures and histone modifications. These domains are not only arranged linearly along the length of the genome, but also spatially inside the nucleus and can change during cell differentiation. This

organization leads to the formation of 'Topologically Associated Domains' or TADs that fold into discrete structures iteratively throughout the nucleus. This 3D organization confines the chromosomes into distinct territories called chromosome territories. Chromatin is also associated with different nuclear compartments that further define its compaction and function. Based on histone marks, the chromatin is also divided into open transcriptionally active compartment (euchromatin) and a compacted and transcriptionally inactive compartment (heterochromatin). Heterochromatin is mostly composed of repetitive elements which are packaged together to prevent illegitimate recombination events and to maintain genomic stability. This spatial organization is crucial for maintaining the 3 R's of DNA, namely, replication, repair and recombination.

Work from our group as well as from other groups has shown that non-random global genome organization is a key factor in maintaining genomic integrity. Our lab has recently shown that DSBs in pericentromeric chromatin are treated differently from DSBs in centromeric chromatin. Also, it was shown earlier that the presence of heterochromatin at Lamin-associated Domains (LADs) delays DNA damage response and impairs HR (Lemaitre et al. 2014). It has also been shown that transcriptionally active chromatin regions (H3K36me3) repair DSBs in their vicinity by HR. Although these studies have set the proof of concept linking DNA repair to higher-order chromatin structure, there are still many fundamental questions in this area. We do not know what determines the specificity of DNA repair outcome in lesions occurring in different nuclear compartments. Moreover, it is not clear whether there is an inherent hierarchy in repairing different damaged genomic sequences. In other words, it is not known whether there are chromatin or topological domains that are repaired first and others where breaks persist.

The goal of my first project was to understand the influence of 3D genome organization on DNA repair. I followed the kinetics of DSB repair to map in high resolution the locations of DSBs as they get repaired. I also wanted to determine whether the efficiency in DNA repair across the genome correlated with previously defined topological maps of distinct chromatin domains and ChIP-Seq profiles of histone marks. The specific questions I wanted to address are:

i. Are certain genomic regions more susceptible to damage?
ii. Is it easier or more difficult to repair certain regions?
iii. What is the role of 3D organization in DNA repair pathway choice?

To follow the DNA repair kinetics, I induced DNA damage in mES J1 cells -for which high-resolution 3D-topological maps of distinct chromatin domains are available- by treating them with the radiomimetic drug neocarzinostatin (NCS) and allowed DNA repair for 0, 30minutes, 90minutes, 3 and 6 hours post treatment. Subsequently I mapped the position of the breaks at these timepoints using *in situ* Breaks Labeling Enrichment on Streptavidin and next generation Sequencing (BLESS). BLESS is a genome-wide method developed to identify DNA breaks at single nucleotide resolution. I have applied this method to map the DNA breaks induced immediately after the NCS treatment as it has been previously described (Crosetto et al. 2013) and followed the efficiency of repair over time. As a complementary approach to mapping

DSBs, I have performed a ChIP-Seq using antibodies for γ-H2AX, a factor for sensing DNA damage.

As a complementary strategy to inflict DSBs, I tried to establish an mES cell line containing the *Asi*SI restriction enzyme, an 8 base-pair cutter that produces DSBs homogeneously in the genome, and generate DSB maps by BLESS and ChIP-Seq as before. Using both approaches (BLESS and ChIP-Seq) in two different experimental systems (NCS, site specific endonucleases), I wanted to interrogate the genome and hence validate the robustness with which DSBs arising in different chromatin domains are repaired and correlate this efficiency with the 3D genome organization of these domains.

Results from BLESS and ChIP-Seq have suggested that there are fragile regions non-randomly distributed in the mouse genome which could be classified into distinct categories based on their occurrence and repair profiles. Constitutive breaks or those which are present in the chromatin of untreated cells could also be observed in the data-set from ChIP-Seq for γH2AX. Induced breaks were those which were produced because of the treatment with Neocarzinostatin. It was observed that in untreated cells, constitutive damage occurred in active chromatin (H3K4me3, H3K36me3). This was also corroborated by analysis of Lamin-associated domains (LADs) which showed a lower incidence of damage, suggesting that constitutive damage is present preferentially in active chromatin. Interestingly, it was also observed that a higher proportion of the constitutive break regions are present downstream of transcription start sites. This correlates with the findings that the DNA damage response signaling mechanism is also essential for transcriptional elongation. It was seen that the repair starts much further away from a transcription start site and then moves towards the start site. After peak annotation of the ChIP-Seq data, it was seen that after induction of damage, a higher percentage of damage persistence was observed in the intergenic and intronic regions.

For my second project, I wanted to elucidate the genomic footprint of 53BP1 in asynchronous cells and G1 arrested cells to identify other 'hotspots' of 53BP1 body accumulation apart from annotated fragile sites. Multiple studies have shown that DNA replication stress is a major contributor of genome instability which is a precursor for cancer and many other diseases. Over the years, to gain a better understanding of genome stability during replication, advancements have been made in deciphering the role of these 53BP1 bodies. They have emerged as having a substantial role in maintaining genome stability during replication as a shielding factor. Despite the importance of 53BP1 bodies at fragile sites, only a few fragile sites shielded by these bodies have been described. The precise mechanism for the formation of these bodies, yet remains elusive, so this project aims to uncover the functional significance of 53BP1 bodies in cell cycle progression and completion. Because of a lack of high specificity antibodies against 53BP1, it was decided to generate a system which uses Bio-ID technology to identify novel genomic loci and biochemical composition of 53BP1 bodies. Briefly, the system contains the minimal functional region of 53BP1 called Tudor, labelled with a biotin tag (BioTag) and an mCherry molecule. This system enables us to use the BioTag to perform immunoprecipitation of 53BP1 for protein interaction studies and genome wide sequencing and can also be used for microscopy because of the included mCherry. Using this system

should let us systematically identify and characterize novel fragile genomic loci shielded by 53BP1 bodies.

To identify genomic loci shielded by 53BP1 bodies, previous studies have employed ChIP followed by qPCR using primers specific against specific fragile sites. Despite these results, a genome wide approach covering the entire genome to identify novel fragile sites has not been attempted yet. Results from my second project have shown that the Bio-ID system is suitable for chromatin immunoprecipitation as evidenced by the successful pulldown of annotated fragile sites. The technique is now being used to study the distribution of 53BP1 nuclear bodies in G1 arrested cells. The data is currently under bioinformatic analysis and will be further elaborated in the thesis. This data would enable us to further understand the significance of 53BP1 during cell division.

Thesis summary in French

Notre génome est constamment attaqué par des facteurs endogènes et exogènes qui menacent son intégrité. Ceux-ci peuvent entraîner des désaminations, des dépurations spontanées, des lésions oxydatives et des cassures simple et double brins (CSB et CDBs). Parmi ces dommages, les CDB font partie des plus nuisibles car elles peuvent entraîner la perte d'information génétique, des translocations chromosomiques et la mort cellulaire. Pour répondre à ces problèmes, les cellules ont développé un mécanisme hautement conservé qui non seulement détecte les dégâts de l'ADN et conduit à leur réparation, mais coordonne également cette réparation avec d'autres processus cellulaires tels que l'état de transcription, la progression du cycle cellulaire et l'apoptose. Il existe deux voies principales de réparation des dommages à l'ADN, en fonction de l'état du cycle cellulaire : la recombinaison homologue (HR) et la jonction d'extrémités non homologues (NHEJ). HR est un mécanisme hautement fidèle car il se sert du chromosome homologue comme matrice pour effectuer la réparation, mais ne se produit que dans les cellules proliférantes pendant les phases S/G2 du cycle cellulaire. NHEJ, lui, se produit indépendamment de l'état du cycle cellulaire et ne requière pas une matrice pour la réparation, mais peut être un mécanisme sujet à l'erreur.

Indépendamment du fait qu'une CDB soit générée par des sources endogènes ou exogènes, le mécanisme de réparation subséquent se produit par une cascade de signalisation à trois étapes qui comprend des protéines de détection, des protéines adaptatrices et des protéines effectrices. La cascade de signalisation commence par la détection de la lésion par le complexe MRN (Mre11/Rad51/Nbs1) qui, à son tour, recrute et active ATM, une molécule de signalisation clé. ATM phosphoryle ensuite le variant d'histone H2AX pour produire γ-H2AX qui constitue un marqueur caractéristique des dommages à l'ADN. Un autre facteur crucial dans le mécanisme de réparation est le tumor-suppressor 53 binding protein 1 (p53BP1/53BP1). Cette protéine s'associe rapidement avec les CDBs et d'autres protéines de réparation des dommages à l'ADN. C'est une protéine adaptatrice qui est nécessaire au cours de la réponse aux dommages et sert de plate-forme pour le recrutement d'autres facteurs. Il a également été démontré qu'il joue un rôle dans la sélection de la voie de réparation des CDBs. En plus de sa localisation aux foyers induits par les dommages, 53BP1 peut également être présent dans des cellules non endommagées en raison du stress de réplication, mais sa présence reste dans ce cas limité à la phase G1 du cycle cellulaire. Certaines régions génomiques sont difficiles à répliquer en raison d'une rareté d'origines de réplication. Ces régions, appelées sites fragiles, peuvent ne pas se répliquer complètement au cours d'une phase S donnée. De récentes recherches ont montré l'importance du métabolisme des sites fragiles dans l'intégrité du génome et le développement du cancer. Les sites fragiles représentent un défi pour une cellule puisque ces régions non répliquées ne sont pas reconnues comme des dommages à l'ADN et ne peuvent donc pas déclencher la voie de réponse aux dommages. Bien que la preuve de la présence de structures 53BP1 autour des sites fragiles existe, le principal problème pour l'analyse de ces structures est l'absence d'un anticorps anti 53BP1 approprié qui faciliterait l'analyse de ses partenaires d'interaction et de sa distribution génomique. La question reste ouverte puisque ces sites fragiles sont continuellement présents avec une propension accrue de dommages à l'ADN

induite par le stress de réplication et qu'ils sont progressivement éliminés du génome par évolution.

Tous les processus de réparation précédemment mentionnés se déroulent dans le cadre d'une chromatine hautement organisée et compartimentée. La chromatine n'est pas une macromolécule uniforme dans le noyau mais se compose de différents domaines caractérisés par des structures d'ADN complexes et des modifications d'histones. Ces domaines ne sont pas seulement disposés linéairement sur la longueur du génome, ils ont également une organisation spatiale à l'intérieur du noyau qui peut changer pendant la différenciation cellulaire. Cette organisation conduit à la formation de « Domaines Topologiquement Associés » ou TADS qui se replient en structures discrètes de manière répétitive à travers le noyau. Cette organisation 3D confine les chromosomes en des territoires distincts appelés territoires chromosomiques. La chromatine est également associée à différents compartiments nucléaires qui définissent davantage sa compaction et sa fonction. Sur la base de marques d'histone, la chromatine est également divisée en un compartiment ouvert transcriptionnellement actif (euchromatine) et un compartiment compacté et transcriptionnellement inactif (hétérochromatine). L'hétérochromatine est principalement composé d'éléments répétitifs qui sont empaquetés ensemble afin de prévenir les événements de recombinaison illégitime et de maintenir la stabilité génomique. Cette organisation spatiale est cruciale pour le maintien des 3 R d'ADN, à savoir la réplication, la réparation et la recombinaison.

Les travaux de notre groupe ainsi que d'autres groupes ont montré que l'organisation globale non aléatoire du génome est un facteur clé dans le maintien de l'intégrité génomique. Notre laboratoire a récemment mis en évidence que les CDBs dans la chromatine péricentromérique sont traités différemment des CDBs dans la chromatine centromérique. De plus, il a été montré précédemment que la présence d'hétérochromatine dans les LADs (Lamin-associated domains) retarde la réponse aux dommages de l'ADN et porte atteinte à HR. Il a également été montré que les CDBs dans les régions de chromatine transcriptionnellement actives (H3K36me3) sont réparées par HR. Bien que ces études aient établi la preuve du concept reliant la répartition de l'ADN à la structure de la chromatine, il existe encore de nombreuses questions fondamentales dans ce domaine. Nous ne savons pas ce qui détermine la spécificité de réparation de l'ADN dans les lésions se produisant dans différents compartiments nucléaires. Par ailleurs, il n'est pas clair s'il existe une hiérarchie inhérente à la réparation de différentes séquences génomiques endommagées. En d'autres termes, nous ne savons pas s'il existe des domaines de la chromatine ou des domaines topologiques qui sont réparés en premier et d'autres où les cassures persistent.

L'objectif de mon premier projet était de comprendre l'influence de l'organisation du génome 3D sur la réparation de l'ADN. J'ai suivi la cinétique de réparation des CDBs afin de cartographier en haute résolution les emplacements des CDBs à mesure qu'elles sont réparées. Je voulais aussi déterminer si l'efficacité de réparation de l'ADN à travers le génome corrèle avec les cartes topologiques de domaines de chromatine distincts précédemment définies et les profils de CHIP-Seq des marques d'histones. Les questions spécifiques que j'ai voulu aborder étaient les suivantes :

i. Certaines régions génomiques sont-elles plus sensibles aux dommages ?

ii. Certaines régions sont-elles plus faciles ou plus difficile à réparer ?

iii. Quel est le rôle de l'organisation 3D dans le choix de la voie de réparation de l'ADN ?

Pour suivre la cinétique de réparation de l'ADN, j'ai induit des dommages à l'ADN dans des cellules mES J1 - pour lesquelles des cartes 3D haute résolution des domaines topologiques distincts de la chromatine sont disponibles - en les traitant avec la drogue radiomimétique Néocarzinostatine (NCS). J'ai ensuite permis la réparation de l'ADN pendant 30 minutes, 90 minutes, 3 et 6 heures après traitement. Puis, j'ai cartographié la position des cassures à ces temps grâce à la technique de BLESS (*in situ* Breaks Labeling Enrichment on Streptavidin and next generation Sequencing). BLESS est une méthode à l'échelle du génome développée pour identifier les cassures d'ADN avec une résolution nucléotidique. J'ai appliqué cette méthode afin de cartographier les cassures d'ADN induites immédiatement après le traitement NCS, comme décrit précédemment dans la littérature (Crosetto et al. 2013), et suivi l'efficacité de la réparation au cours du temps. Comme approche complémentaire à cette cartographie des CDBs, j'ai réalisé un ChIP-Seq en utilisant des anticorps pour γ-H2AX, marqueur des dommages de l'ADN.

En tant que stratégie complémentaire pour induire des CDBs, j'ai tenté d'établir une lignée cellulaire mES contenant l'enzyme de restriction AsiSI qui produit des CDBs de manière homogène dans le génome, et cartographier ces cassures également par BLESS et ChIP-Seq. En utilisant les deux approches (BLESS et ChIP-Seq) dans deux systèmes expérimentaux différents (NCS, endonucléases site-spécifiques), je voulais interroger le génome, déterminer la robustesse avec laquelle les CDBs qui ont lieu dans les différents domaines de la chromatine sont réparés, et établir une corrélation entre cette efficacité de réparation et l'organisation génomique 3D de ces domaines.

Les résultats de BLESS et de ChIP-Seq ont suggéré qu'il existe des régions fragiles réparties de manière non aléatoire dans le génome de la souris, qui pourraient être classées en catégories distinctes en fonction de leur apparition et de leurs profils de réparation. Les cassures constitutives ou celles qui sont présentes dans la chromatine de cellules non traitées peuvent également être observées grâce aux données de ChIP-Seq pour γH2AX. Les cassures induites étaient celles qui ont été produites par le traitement Neocarzinostatine. Nous avons observé que dans les cellules non traitées, les dommages constitutifs ont lieu dans la chromatine active (H3K4me3, H3K36me3). Cela a également été confirmé par l'analyse des domaines associés à la lamine (LADs) qui a montré une incidence de dommages plus faible dans ces régions, ce qui suggère que les dommages constitutifs sont présents préférentiellement dans la chromatine active. De manière intéressante, nous avons aussi observé qu'une proportion plus élevée de régions avec cassures constitutives sont présentes en aval de sites d'initiation de la transcription. Ceci est en corrélation avec les découvertes montrant que la voie de signalisation de réponse aux dommages à l'ADN est également essentielle pour l'élongation de la transcription. Il a été constaté que la réparation démarre loin du site d'initiation de la transcription, puis se déplace vers le début du site. Après l'annotation des données de ChIP-Seq, il a été constaté qu'après induction de dommages, un pourcentage plus élevé de persistance des dommages était observé dans les régions intergéniques et introniques.

Pour mon deuxième projet, je voulais élucider l'empreinte génomique de 53BP1 dans des cellules asynchrones et des cellules bloquées en G1 afin d'identifier d'autres « points chauds » d'accumulation de structures 53BP1 en dehors des sites fragiles annotés. De multiples études ont montré que le stress de réplication de l'ADN est un facteur majeur de l'instabilité du génome, qui est un précurseur du cancer et de nombreuses autres maladies. Au fil des années, afin de mieux comprendre la stabilité du génome au cours de la réplication, des progrès ont été réalisés pour déchiffrer le rôle de ces structures 53BP1. Ils ont émergé comme ayant un rôle important dans le maintien de la stabilité du génome pendant la réplication en tant que facteur de protection. Malgré l'importance des structures 53BP1 sur les sites fragiles, très peu de sites fragiles protégés par ces structures ont été décrits. Le mécanisme précis de formation de ces structures reste mal compris, et ce projet vise donc à découvrir la signification fonctionnelle des structures 53BP1 dans la progression et l'achèvement du cycle cellulaire. En raison de l'absence d'anticorps de spécificité élevée contre 53BP1, nous avons décidé de générer un système qui utilise la technologie Bio-ID pour identifier de nouveaux loci génomiques et la composition biochimique des structures 53BP1. En bref, le système contient la région fonctionnelle minimale de 53BP1 appelée Tudor, marquée avec une étiquette de biotine (BioTag) et une molécule mCherry. Le BioTag peut être utilisé pour réaliser une immunoprécipitation de 53BP1 permettant ainsi l'étude de ses interactions avec d'autres protéines et le séquençage du génome pour identifier les cibles génomiques de 53BP1 ; mCherry peut être utilisé pour la microscopie. Ce système devrait nous permettre d'identifier et de caractériser systématiquement de nouveaux sites fragiles protégés par les structures 53BP1.

Pour identifier les loci génomiques protégés par les structures 53BP1, des études antérieures avait employé la méthode de ChIP suivi de qPCR en utilisant des amorces spécifiques contre des sites fragiles spécifiques. Mais une approche couvrant l'ensemble du génome pour identifier de nouveaux sites n'avait pas encore été tentée. Les résultats de mon deuxième projet ont montré que le système Bio-ID est adapté à l'immunoprécipitation de la chromatine, comme en témoigne la précipitation et l'identification réussie des sites fragiles annotés. La technique est maintenant utilisée pour étudier la répartition des structures 53BP1 dans les noyaux de cellules arrêtées en G1. Les données sont actuellement soumises à une analyse bioinformatique et seront plus développées dans la thèse. Ces données nous permettraient de mieux comprendre le rôle de 53BP1 pendant la division cellulaire.

INTRODUCTION

1.1 DNA Damage

Assaults on the integrity of the genome, through endogenous and exogenous factors, are a constant occurrence throughout the lifetime of a cell. To protect the integrity of the genetic information, cells have developed sophisticated networks to sense and repair these genotoxic insults. The purpose of such networks is to enable the cell to halt division (cell cycle arrest) and undergo DNA repair or programmed cell death (apoptosis) (Zhang et al., 2004). All DNA repair systems, cell cycle checkpoints and additional pathways whose activity changes upon DNA damage are collectively known as the DNA damage response (DDR) (Derks et al 2014). Improper repair of damaged DNA leads to genomic instability, senescence and apoptosis (Jackson and Bartek, 2009). Defects in an organism's capacity to repair damage to its DNA are associated with a predisposition to immunodeficiency, neurological disorders and cancer (Subba Rao, 2007; Thoms et al., 2007), highlighting the importance of repair pathways for preserving genome integrity. The type of DNA damage created depends on the insult, but may include altered bases, single- and double-strand breaks (DSB), inter-strand crosslinks and bulky DNA adducts (Norbury and Hickson, 2001).

1.1.1 Causes of DNA damage

As with any ordered structure, DNA is subject to error, modification and deterioration. The dynamic chemical nature of the cellular environment and frequent *in vivo* processing provide an endogenous baseline of DNA damage that is compounded by exposure to sources of environmental stress. It is estimated that there are as many as 10^5 spontaneous nucleotide damage events per day in each cell (Lindahl, 1993). The major causes of DNA damage can be broadly classified into four categories. The first is caused by the inherent instability of the DNA molecule which decays over time, predominantly through spontaneous hydrolysis creating abasic sites and deamination products (Lindahl, 1993). Secondly, metabolism produces highly reactive byproducts such as reactive oxygen species (ROS) which create diverse types of DNA damage (de Bont, 2004). Thirdly, damage may also occur due to deletion or misincorporation of DNA bases during replication or DNA repair. These three naturally occurring sources are together classified as endogenous sources of damage. Lastly, exogenous agents, commonly referred to as carcinogens, such as ionizing radiation, benzene and tobacco products are also responsible for DNA damage. Apart from these causes, the cell might also produce breaks during the

creation of immune system repertoire and meiotic recombination. The causes of DNA damage are further discussed below:

Endogenous sources of damage:

a. Hydrolysis

   Hydrolytic DNA damage involves deamination or the total removal of individual bases. The chemical bond between a DNA base and its respective deoxyribose, although relatively stable, is subject to chance cleavage by a water molecule through spontaneous hydrolysis. Loss of DNA bases, known as AP (apurinic/apyrimidinic) sites, can be potentially lethal to the cell, as they act to block the progress of DNA replication, but are efficiently repaired by the base excision repair (BER) pathway. In mammalian cells, it is estimated that depurination occurs at a rate of about 10,000 purine bases lost per cell generation (Lindahl, 1972). The rate of depyrimidination is considerably slower, resulting in the loss of about 500 pyrimidine bases per cell generation.

b. Metabolism by-products

   Cellular metabolism, including oxidative respiration and lipid peroxidation, produces reactive oxygen species (ROS) such as superoxide anions, hydroxyl radicals, and hydrogen peroxide. These ROS generate multiple kinds of DNA damage including oxidized bases and double strand breaks (DSBs). Upon exposure to ROS, guanine is oxidized into 8-oxo-7,8-dihydroguanine (8-oxo-Gua) which can cause transversions such as A·T→T·A or G·C→T·A, accumulation of which can lead to detrimental consequences.

c. Replication stress

   Replication stress is defined as slowing and stalling of the replication fork progression, which can lead to replication fork collapse and DNA breaks. This can induce genomic instability and possible cell death. Replication stress can be caused by both external agents and internal factors, which damage the DNA template or inhibit replication proteins. Replication stress can also be due to abnormal firing of replication origins. In human cells, replication can start from thousands of defined sites along the chromosome at the so-called replication origins, and the newly synthesized DNA forms bidirectional replication forks. The origins are activated at separate times during S-phase with different origins firing in a reproducible temporal sequence throughout S phase. However, only a fraction of the origins fire

under normal S-phase with the rest serving as backup under DNA damage and replication stress conditions (Zeman and Cimprich, 2014). Overexpression of oncogenes like Ras, Myc and Cyclin-E (Di Micco et al., 2006), causes an abnormal increase in replication origin firing. The increased replication initiation leads to nucleotide deficit and stalled forks. The stalled forks may continue to unwind the DNA helix, creating single-stranded DNA (ssDNA). This leads to induction of the intra-S checkpoint that halts the cell cycle and suppresses the origin firing. However, if the cell does not manage to restart replication, the fork will eventually collapse, creating DNA breakage (Magdalou et al., 2014). Collisions between the replication and transcription machineries may further contribute to replication stress. During S-phase, replication and transcription operate on the same DNA template which can lead to collisions that cause replication stalling.

d. Programmed DNA breaks

Other than the above-mentioned damages to DNA which arise because of other processes in the cell, DNA lesions are sometimes produced in cells during cell development. Pachytene spermatocytes have DSBs which are produced by an enzyme called Spo11, a topoisomerase II-like enzyme (Zickler and Kleckner, 1999). Spo11 creates DSBs to generate cross-overs between homologues during meiotic prophase I. These events are resolved by the high-fidelity repair mechanism homologous recombination (HR) to minimize alterations to the germ line genome. DSBs are also produced during the development of immune cells. V(D)J recombination, a process mediated by the enzymes RAG1/2 and nuclease Artemis, occurs specifically in early B and T-cells for generating antigen receptor diversity. Furthermore, in B-cells that have completed V(D)J recombination, a second set of DNA lesions is produced, by a process called Class Switch Recombination, to change their immunoglobulin heavy chain gene from producing IgM to IgG, IgA, or IgE by changing the Igμ to Igγ, Igα or Igε gene respectively. This process is regulated by a B-cell specific cytidine deaminase called activation-induced deaminase (AID) which only acts to convert cytosine to uracil within regions of single-stranded DNA (ssDNA). Once this change in the switch region has been made, uracil glycosylase converts these to abasic sites and nicks are produced at these sites. Participation of other enzymes, such as Exo1, assist in converting the nicks into large overlapping gaps, resulting in DSBs.

Exogenous sources of damage:

a. Ionizing radiation

   Proteins, cell membranes and DNA are targets of radiation exposure, and effects on these structures can be produced by direct damage as well as free radicals, including ROS. DNA damage is caused by the ionizations/excitations of DNA or the surrounding material, mostly water within a radical-diffusion distance of 4 nm from the DNA in the cellular environment (Goodhead 1999). Severity of lesions depends on the energy deposition in time and space of radiation. Previous publications allow the rough estimation of the yields of DNA damage in mammalian cells caused by low dose of radiation: for each diploid cell, 1 Gray induces 1000 simple-strand breaks, 10000 base damage, 150 protein-DNA cross-links, 70 bulky lesions (i.e., clusters of base damage), and 40 double-strand breaks (DSB) (Ward, 1988).

b. Ultraviolet (UV) radiation

   Solar UV radiation is divided into UVC (200-280nm), UVB (280-315nm) and UVA (315-400nm). Short-wavelength UVC-radiation and UVB wavelengths less than 295nm are blocked by the atmosphere (Freeman et al., 1989). Thus, UV radiation reaching the earth's surface is composed of approximately 95% long-wavelength UVA radiation and 5% UVB radiation. UVB radiation is directly absorbed by the DNA bases, inducing the formation of DNA lesions (Piette et al., 1986). The primary UVB induced lesions are cyclobutane pyrimidine dimers (CPDs), formed by the covalent attachment of two adjacent pyrimidines, and pyrimidine 6-4 pyrimidone photoproducts (6,4-PPs), formed by a covalent linkage between a pyrimidine and the adjacent pyrimidine. Both CPDs and 6,4-PPs are mutagenic, with CPDs considered the primary mutagenic lesion, due to the efficient repair of 6,4-PPs (You et al., 2001). Error-prone replication of post UV induced CPDs results in the generation of C-T and CC-TT transition mutations, considered 'UV signature mutations' (Wikonkal and Brash,1999).

c. Genotoxic chemicals and DNA damaging drugs

   Cells often are exposed to genotoxic agents present in the environment and because of professional or personal practices. Many compounds do not produce DSBs directly, but other lesions that can be transformed into DSBs once the cells try to go through DNA replication.

A common example of a genotoxic compound to which humans may be exposed is benzo[a]pyrene (BaP), which is produced through incomplete combustion of any organic compound and can be a result of forest fires, diesel engines and cigarette smoking. Cigarette smoking directly delivers the BaP to lung tissue where it forms the bulky DNA adduct, benzo[a]pyrene diol epoxide (BPDE)-deoxyguanosine, with guanine base pairs. High frequency of such insults along with inefficient repair of this damage by nucleotide excision repair (NER) mechanism, may cause lung cancer. Recent research has shown that repair of BPDE damage occurs more frequently when the BPDE-burdened guanine is next to a cytosine rather than a thymine or adenine (Li et al, 2017).

The DNA damaging properties of chemicals have also been exploited in chemotherapy. Most of the commonly used chemotherapies exert their cytotoxic effects through either inducing DNA damage or interfering with the repair pathways or a combination of both. These chemicals work through different mechanisms of actions like alkylation, cross-linking, inhibition of the DNA synthesis, induction of SSBs and DSBs by free radicals, interaction with topoisomerases and as antimicrotubule agents (Lichtman, 2008). Some of these drugs have been described below:

i. Alkylating agents such as carmustine, cyclophosphamide and melphalan, damage DNA through the chemical transfer of one or more alkyl groups to DNA deoxyribose or phosphate groups leading to either strand scission, or to the production of DNA bases that alter Watson-Crick base pairing preferences.

ii. Antimetabolites such as methotrexate, pemetrexed and cytosine arabinoside, generate replication-associated DNA damage by compromising the synthesis of nucleotides needed for DNA replication fork progression.

iii. Topoisomerase inhibitors such as camptothecin, topotecan, etoposide and doxorubicin covalently bond and thus inhibit the catalytic action of topoisomerase I or topoisomerase II, which are enzymes that regulate DNA topology by breaking and rejoining the phosphodiester backbone of DNA during DNA replication, transcription, recombination and repair.

iv. Antimicrotubule agents such as taxanes, vinblastine and vincristine interfere with the synthesis or degradation of microtubules required for the proper segregation of chromosomes during mitosis.

An important class of these damage inducing drugs is radiomimetic drugs which produce effects like ionizing radiation. Bleomycin and Neocarzinostatin (NCS) are antibiotics which produce sequence specific breaks by a free radical mechanism. While bleomycin has been used as an antitumor agent clinically, NCS has become ubiquitous in research laboratories to induce and study DSBs and repair mechanisms. It comprises of a carrier apoprotein and an unstable hydrophobic chromophore which is the active component. The chromophore forms a highly reactive biradical species which can induce double-stranded damage by simultaneously abstracting hydrogen from C-5' of deoxyribose in one DNA strand, and from C-1' or C-4' in the opposite strand. NCS thus induces clean DSBs, as well as abasic sites with closely opposed strand breaks, which rapidly merge to yield additional DSBs (Smith et al, 1994; Goldberg, 1991).

1.2 Double Strand Breaks

Among the diverse types of DNA lesions, DSBs are the most potent. They are the major cause of cell sensitivity after radiation and 1 Gray of radiation is believed to cause 40 DSBs (Ward, 1988; Lomax et al., 2013). Besides radiation, different chemotherapeutic drugs also generate DSBs either directly or from lesions during DNA replication. DSBs are also produced endogenously during DNA replication due to damage in the template strand. Endogenous site-specific DSBs are also generated during certain cellular processes such as V(D)J recombination, class-switch recombination and meiosis (Lieber, 2010). DSBs occur when both strands of the DNA double helix are broken. They pose a threat to the genomic integrity and cell survival as any unrepaired DSBs may result in growth arrest, cell death, mutation and faulty repair can lead to chromosomal aberrations and translocations which are a precursor for cancer (Jeggo and Löbrich, 2007).

1.3 DNA damage response (DDR)

The DNA damage response (DDR) is a signal transduction pathway that senses DNA damage and orchestrates a highly structured response to rectify the damage. The first step involves detection, by sensor proteins, of DNA damage or chromatin alterations

that have occurred after damage induction. The signal is then transmitted to transducer proteins that amplify the damage signal. Ultimately, the signals are passed to the effectors which are involved in specific pathways to carry out the repair (Zhou and Elledge, 2000). This multi-step cascade of DDR results in a variety of cellular responses including cell-cycle arrest, induction of stress response genes, DNA repair and cell death (Fig. 1) (Ciccia & Elledge 2010).

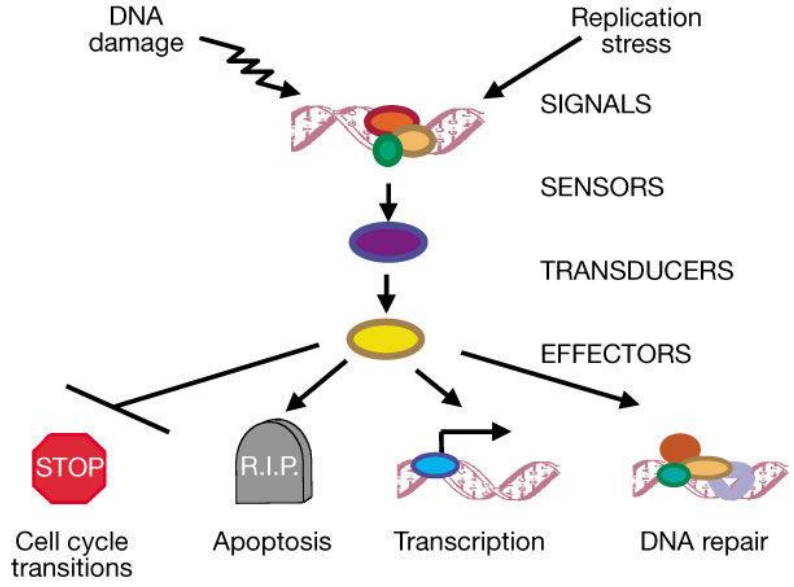


Figure 1. ***General outline of the DNA damage response pathway****. Arrowheads represent activating events and perpendicular arrows represent inhibitory events. Cell cycle arrest is depicted with a stop sign and apoptosis with a tombstone. DNA helix with an arrow represents damaged-induced transcription, while DNA helix with oval shaped subunits represents damage-induced repair. The network has been represented as a linear pathway consisting of signals, sensors, transducers and effectors. (Taken from Zhou and Elledge, 2000).*

1.3.1 DNA damage sensing

Distinct sensor proteins have evolved to recognize and bind to aberrant DNA structures at the sites of DNA damage to recruit and activate transducer enzymes. The MRN complex (Mre11/Rad50/Nbs1) and the Ku70/80 heterodimer bind DNA ends present at DSBs (Blier et al., 1993; de Jager et al., 2001; Mimori and Hardin, 1986). Single-strand DNA formed at various lesion types and stalled/broken replication forks are coated by the heterotrimeric complex Replication Protein A (RPA) (Wold, 1997). The poly(ADP-ribose) polymerase (PARP1) senses both single- and double-strand breaks (D'Amours et al., 1999), while the heterodimers MSH2/MSH6 and MSH2/MSH3 are specific for the identification of mismatched and improperly inserted or deleted nucleotides (Germann et al., 2010).

Many other proteins, including BRCA1, tumor suppressor p53-binding protein 1 (53BP1) and mediator of DNA damage checkpoint protein 1 (MDC1) (DiTullio et al.,

2002; Bekker-Jensen et al., 2006; Schultz et al., 2000), are also rapidly recruited to the sites of damage. Agglomerations of such proteins at sites of damage or altered chromatin states, in the form of complexes, are referred to as foci (Lisby and Rothstein, 2005).

The MRN complex, RPA and Ku70/80 heterodimer activate the transducer enzymes ATM (ataxia telangiectasia mutated), ATR (ATM and Rad3 related), and DNA-PKcs (DNA-dependent protein kinase, catalytic subunit), respectively (Lieber, 2010; Smith et al., 2010, Goodarzi and Jeggo, 2013).

1.3.1.1 Sensing by MRN complex

The meiotic recombination 11 nuclease (Mre11) has an affinity for DNA ends which is further stimulated by Rad50 and causes the Mre11-Rad50 dimer to bind to the free ends of DNA using the two C-terminal DNA-binding domains of Mre11 (de Jager et al., 2001; Paull and Gellert,1999). A direct interaction with the C-terminal region of Nbs1 facilitates further ATM recruitment and activation (Assenmacher and Hopfner, 2004; Falck et al., 2005). The MRN complex aids the recruitment and activation of ATM at sites of DNA damage (Carson et al., 2003; D'Amours and Jackson, 2002; Falck et al., 2005; Lee and Paull, 2005; Uziel et al., 2003). ATM phosphorylates H2AX to generate γH2AX. This is one of the earliest signaling events in DNA damage response (Rogakou et al., 1999) and serves as a hallmark for DNA damage. (The role of γH2AX in the DDR has been discussed in further detail in section 1.3.3.) γH2AX creates a further focus for the attachment of MDC1 (Stucki and Jackson, 2006; Xie et al., 2007); subsequently recruiting E3 ubiquitin-protein ligase RING finger protein 8 (RNF8) as part of a ubiquitin signaling cascade (Wu et al., 2011). RNF8 ubiquitinylates H2AX creating a focus for the stable attachment of oligomerised 53BP1 (Kolas et al., 2007; Mailand et al., 2007). 53BP1 in turn stimulates ATM activity through the MRN complex resulting in further amplification of the ATM signal (Lee and Paull, 2007). The MRN complex is involved not only in DSB sensing and signaling with ATM but also in DSB repair, maintenance of the S-phase checkpoint and telomere stability (Lavin, 2004; Shiloh, 2001). It has also been shown that the sensor protein PARP1 partially controls ATM activation and mediates the initial recruitment of Mre11 and Nbs1 to DNA damage, which occurs within the first minute after damage (Haince et al., 2007; Haince et al., 2008).

1.3.1.2 Sensing by PARPs

The poly (ADP-ribose) polymerase (PARP) proteins PARP1 and PARP2 can sense both single- and double-stranded DNA damage, whereas PARP3 predominantly responds to DSBs (Beck et al., 2014). Upon their activation as a response to DNA damage, PARPs catalyze the synthesis of poly (ADP)-ribose (PAR) chains on themselves and other proteins. By this mechanism, PARPs promote the recruitment of the X-ray repair cross-complementing protein 1 (XRCC1) and the DNA ligase 3 (LIG3) (Weinfeld et al., 1997; El-Khamisy et al., 2003; Mortusewicz et al., 2006). It has also been shown that PARP1 is involved in HR repair pathways of hydroxyurea (HU)-mediated collapsed replication forks by recruiting Mre11 and initiating end resection activity (Bryant et al., 2009; Yang et al., 2004). On the other hand, PARP3 drives repair towards the NHEJ repair pathway by preventing excessive end-resection mediated by Mre11 and repair by HR (Beck et al., 2014). Recently, PARP1 was also shown to be responsible for recruiting the chromatin remodeler CHD2 which promotes the assembly of NHEJ repair complexes at DNA break sites (Luijsterburg et al., 2016).

1.3.1.3 Sensing by Ku70/80

Ku is extremely abundant in human cells with estimates at 400 000 molecules per cell and it binds strongly to DNA DSBs (Lieber, 2010). Ku is localized primarily in the nucleus. Ku exists as a heterodimer of its subunits Ku70 and Ku80 and through the dimerization domains, creates a ring structure lined with positive charges which bind free DNA ends in a sequence independent manner (Fell and Schild-Poulter, 2015). An early consequence of this Ku-DNA complex is to keep the DNA ends near each other before repair and to stop them from drifting apart. The longer carboxy-terminal of Ku80 enables it to bind to DNA-PKcs which primes the cell for initiating repair by Non-Homologous End Joining (NHEJ). This activation causes the recruitment of nucelases, polymerases and ligases in any order. This flexibility gives rise to a wide variety of outcomes from identical starting ends (Downs and Jackson, 2004; Lieber, 2010).

1.3.1.4 Sensing by RPA

The heterotrimeric complex of Replication Protein A (RPA) is composed of 70, 32 and 14 kDa subunits. This complex has a very high affinity with single stranded DNA (ssDNA) through multiple oligonucleotide/oligosaccharide binding fold domains (Marechal and Zou, 2015; Kim et al., 1994). It was initially thought to have a role in replication by enhancing the assembly and recruitment of DNA polymerases α, δ, and ε (Dornreiter et al., 1992; Yuzhakov et al., 1999). RPA protein coats ssDNA overhangs of DSBs with a definite 5'→3' polarity (Kolpashchikov et al., 2001). This RPA-ssDNA complex causes activation and localization of the ataxia telangiectasia and RAD3-related (ATR) kinase and its co-activator, ATR-interacting protein (ATRIP) to the ssDNA (Cimprich and Cortez, 2008). Activated ATR-ATRIP complexes contribute to various processes including cell-cycle checkpoints and stabilization of stalled replication forks.

1.3.2 Signal propagation

The initial sensing of DNA damage is accompanied by a massive buildup of multiprotein foci at the damage site and the induction of post translational modifications. The DDR signal is propagated by the phosphatidylinositol-3-kinase (PIKK) family which includes ATM, ATR and DNA-PKcs. Upon activation, DDR transducer enzymes go on to modify downstream effector proteins through post-translational modifications, including phosphorylation, ubiquitylation, sumoylation, PARylation, etc, to disseminate the damage signal and further recruit repair factors (Smith et al., 2010; Huang and D'Andrea, 2006; Branzei et al., 2008; Galanty et al.,2009; Morris et al., 2009; D'Amours et al., 1999). The signal propagation occurs both as a systematic recruitment of other factors and positive feedback loops which tend to amplify the signal.

1.3.2.1 ATM signaling

The serine/threonine protein kinase ATM is a 350 kDa nuclear protein kinase, which exists as an inactive homodimer that rapidly dissociates in response to changes in the highly organized structure of DNA (i.e. chromatin alterations) and autophosphorylates at serine-1981 following low-dose ionizing radiation (Bakkenist and Kastan, 2003). It acts as a chief mobilizer for the cellular response to DNA damage, especially in the case of DSBs (for review (Ziv and Shiloh,

2013)). Mutations in the ATM gene cause ataxia telangiectasia, an autosomal recessive neurodegenerative disorder. This disorder involves a marked defect in responding to DSB repair. ATM mutants also present chromosomal instability, translocations involving the T cell receptor locus, telomere–telomere fusions, reduced telomere length and increased rates of intrachromosomal recombination (Enoch and Norbury, 1995; Kirsch, 1994). An important substrate of ATM is the histone variant H2AX which represents 2%-25% of the total H2A pool (Rogakou, et al., 1998). Phosphorylation of H2AX at Ser139 gives rise to γH2AX. γH2AX is used as a biomarker for the presence of DSBs. Another important interactor of ATM is MDC1 which is anchored to DSB sites through interactions with γH2AX and binds ATM in parallel. This allows ATM to phosphorylate more H2AX molecules which gives rise to a positive feedback loop and causes the rapid acceleration of the DSB-associated focus. Notably, ChIP-chip studies have shown that unlike γH2AX which spreads more than a megabase around the region flanking a DSB, activated ATM exhibits a more restrictive pattern around the region flanking a DSB spreading upto 10kb away from the break site (Caron et al., 2015). In addition to the phosphorylation of several target molecules which regulate DDR, ATM also modulates the action of several other protein kinases which further initiate their own signaling pathways leading to complex signaling cascades. The best documented example is Checkpoint kinase 2 (Chk2), which is phosphorylated at residue T62 by ATM following DSB formation (Chaturvedi et al., 1999; Matsuoka et al., 1998, 2000). ATM-mediated pathways have also been shown to be involved in chromatin relaxation (Goodarzi et al., 2008; Polo et al., 2010; Ziv et al., 2006), nucleosome remodeling (Goodarzi et al., 2011) and activation of p53 (Turenne et al., 2001). Activation of Chk2 and p53 have been shown to block the cell cycle in G2/M (Matsuoka et al., 2000; Saito et al 2002) (Fig. 2).

1.3.2.2 ATR signaling

ATR is recruited to ssDNA structures coated with the single-strand binding heterotrimer RPA. ATR forms a heterodimer with ATRIP which is thought to be essential for its activity since ATR and ATRIP mutants have similar phenotypes (Cimprich and Cortez, 2008). ATRIP binds the large subunit of RPA to facilitate ATR localization to ssDNA structures (Ball et al., 2007; Cortez et al., 2001; Zou

and Elledge, 2003), and colocalization of ATR-ATRIP with the independently recruited TopBP1. This promotes activation of ATR kinase activity through stimulatory binding between ATR-ATRIP and TopBP1 (Kumagai et al., 2006; Mordes et al., 2008). TopBP1 accumulates on RPA-coated ssDNA through the coordinated actions of the RAD17-RFC2/4 damage-specific clamp loader and the RAD9-RAD1-HUS1 (9-1-1) complex. First, RAD17-RFC2/4 loads on to RPA-coated ssDNA structures, preferentially those with 5'-recessed ssDNA-dsDNA junctions (Ellison and Stillman, 2003), and facilitates docking of the 9-1-1 complex (Bermudez et al., 2003). 9-1-1 then recruits TopBP1 which causes a further increase in signal spreading. The best-characterized ATR effector, Chk1 (Zhao and Piwnica-Worms, 2001), is activated through ATR-mediated phosphorylation at residues S317 and S345 (Liu et al., 2000). The ATR-mediated Chk1 pathway plays a significant role in preventing cells from entering mitosis with unreplicated or damaged DNA (Brown and Baltimore, 2003) (Fig. 2).

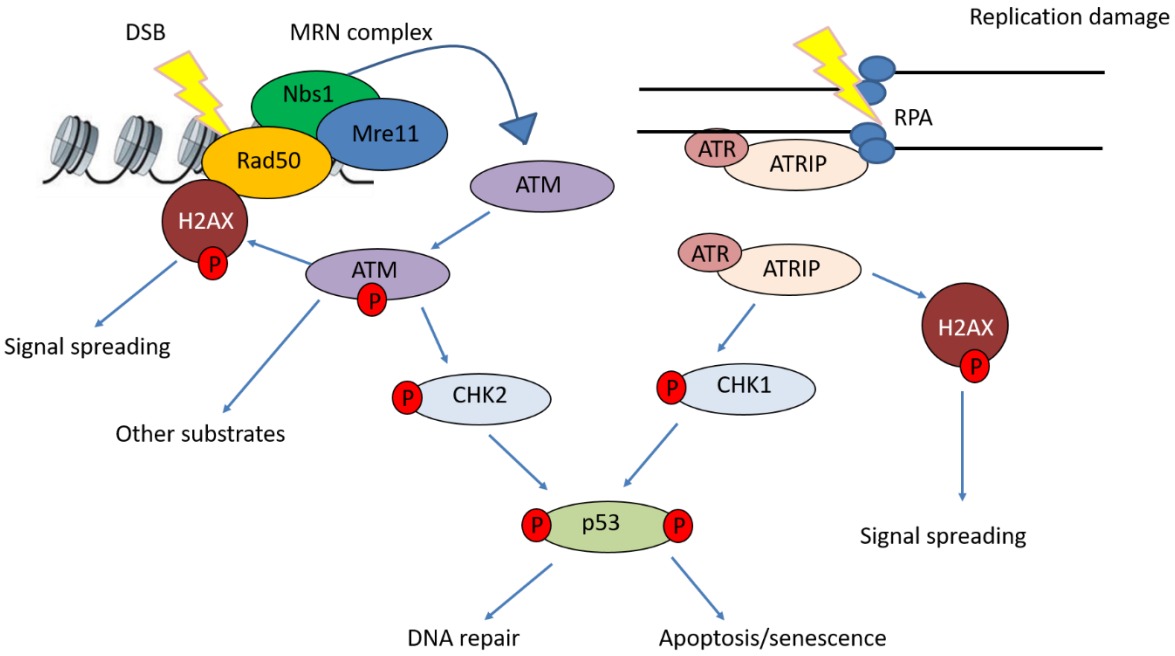


Figure 2. ***ATM and ATR as apex kinases in the DNA damage response signaling cascade****. On the left side, DSB is recognized by the* MRN *complex which results in the recruitment of* ATM*. Autophosphorylation of* ATM *leads to the phosphorylation of* H2AX *at Ser139 (*γH2AX*) as well as other downstream substrates. The right side shows damage at a replication fork due to replication stress. The site is bound by the ssDNA binding* RPA *complex which recruits the* ATR-ATRIP *complex.* ATR-ATRIP *complex activates* γH2AX*.* CHK1 *and* CHK2 *are downstream targets of* ATR *and* ATM *respectively.* CHK1 *and* CHK2 *transmit the signal to effector proteins such as* p53*.*

1.3.2.3 DNA-PKcs signaling

The catalytic subunit of DNA-dependent protein kinase (DNA-PKcs) is a serine theronine protein kinase of the PIKK family and is the largest protein kinase in biology and the only one which is activated by binding to DNA (Lieber, 2010). It has key roles in DNA damage repair through NHEJ besides having roles in V(D)J recombination of immunoglobulin genes and T-cell receptor genes, and telomere length maintenance. The damage sensing is initiated by Ku70/80 heterodimer which encircles the DNA with its ring-like structure and initiates repair. Ku70/80 interacts with the DNA-dependent protein kinase catalytic subunit (DNA-PKcs) forming a complex, DNA-dependent protein (DNA-PK). DNA-PKcs tethers to the damaged DNA ends and forms a synaptic complex that keeps the two DNA ends together. The interaction with Ku enhances DNA-PKcs kinase activity, continuing the repair process, and starts the end processing of DNA ends. DNA-PKcs can also be phosphorylated by ATR at T2609, facilitating its activation (Yajima et al., 2006). Through this activation, DNA-PKcs assists in the amplification of ATR checkpoint signaling (Shimura eta al., 2007) by interacting with Chk1 and promoting the stabilisation of the Chk1-Claspin complex (Lin et al., 2014). It can also act in concert with ATR to phosphorylate RPA32 (Liu et al., 2012) which further promotes phosphorylation of RPA32 at S4, S8 and T21 by DNA-PKcs (Block et al., 2004; Liu et al., 2012; Niu et al., 1997). This hyperphosphorylation of RPA32 stimulates the recruitment of Rad51 (Shi et al., 2010) and PALB2 (Murphy et al., 2014). There is also evidence of a complementary role with the ATM signaling pathway to phosphorylate H2AX resulting in spreading of the DDR signal (Caron et al., 2015; Stiff et al., 2004).

1.3.2.4 Ubiquitylation mediated signaling

Ubiquitylation has been shown to regulate almost all DNA repair pathways, especially in early signaling following DSBs where BRCA1, 53BP1 and RAD51 are recruited to damaged sites in a ubiquitin-mediated manner. The conjugation of ubiquitin requires an enzymatic process, which is divided into three steps: E1-(activating), E2-(conjugating) and E3-(ligase) enzymes (Ciechanover et al., 1982; Hershko et al., 1983). The E3 ligase RNF8 forms a critical link between phosphorylation and ubiquitylation events in the DDR (Kolas et al., 2007; Huen et al., 2007; Mailand et al., 2007). RNF8 contains both a RING domain, which is

important for E2 interaction, and a Forkhead associated (FHA) domain that binds phospho-threonine residues, thus facilitating interaction with substrates phosphorylated by the PIKKs. The FHA domain of RNF8 mediates its interaction with phosphorylated MDC1 (Kolas et al., 2007; Huen et al., 2007; Mailand et al., 2007). The E2 ligase ubiquitin-conjugating enzyme 13 (UBC13) interacts with RNF8's RING domain, while ATM-phosphorylated HECT and the E3 ubiquitin protein ligase 2 (HERC2), another E3 ligase, interacts with the FHA domain of RNF8. HERC2 is believed to stabilize the interaction between RNF8 and UBC13 and maintain RNF8 levels at DNA lesions (Bekker-Jensen et al., 2010). The E3 ligase RNF168 is also critical for repair of DSBs and is recruited to DNA breaks in a RNF8 ubiquitylation-dependent manner. RNF168 catalyzes the ubiquitylation of histone H2A and H2AX (Mattrioli et al., 2012), which is important for the recruitment of 53BP1, receptor-associated protein 80 (RAP80) and BRCA1 (Doil et al., 2009; Stewart et al., 2009; Gatti et al., 2015, Thorslund et al., 2015). RNF8 can also interact with the E2 ubiquitin-conjugating enzyme UBCH8, which promotes polyubuiquitylation of Ku80 (Feng et al., 2012).

The ubiquitin-mediated signaling response is attenuated by the action of deubiquitylating enzymes (DUBs). Several DUBs function to limit RNF8 and RNF168-mediated ubiquitylation. The ubiquitin-specific proteases 3 and 16 (USP3 and USP16) both catalyze the disassembly of RNF8 and RNF168-generated ubiquitin chains (Nicassio et al., 2007; Joo et al., 2007). The E3 ubiquitin ligase paralogous to RNF168 called RNF169 acts as a negative regulator of RNF168-catalyzed ubiquitylation products by competing with 53BP1 and BRCA1 for binding the ubiquitin chains eventually limiting the magnitude of their recruitment to DNA lesions (Poulson et al., 2012).

1.3.3 Role of γH2AX in signal amplification

Upon DNA double-strand break (DSB) induction by ionizing radiation (IR), hundreds of molecules of multiple DNA damage response (DDR) protein species accumulate at DNA DSB sites forming foci known as ionizing radiation induced foci (IRIF) (Bonner et al., 2008; Fernandez-Capetillo et al., 2003; Paull et all., 2000; Rogakou et al., 1999). Phosphorylation of the H2A variant, H2AX, to generate γH2AX is considered the hallmark of DNA damage. H2AX is one of the most conserved H2A variants and is unique amongst other DDR factors since it is already prepositioned on nucleosomes

rather than being recruited to chromatin as a result of damage. Its levels vary from 2 to 20% of the total H2A pool (i.e. in some cells, 1 in 2.5 nucleosomes contain H2AX, whereas in other cells as few as 1 in 25 nucleosomes may contain H2AX), depending on the cell line or tissue examined (Rogakou et al., 1998). Reasons for diverse levels of H2AX might be because H2AX is synthesized in both replication-dependent and replication-independent manners. H2AX is phosphorylated on serine 139 by activated ATM or DNA-PK (Stiff et al., 2004; Savic et al., 2009) and these overlapping pathways may signify the importance of this event. For the phosphorylation of H2AX at Ser 139, it needs to be initially methylated. Dimethylation of H2AX at lysine 134 is carried out by the suppressor of variegation 3-9 homologue 2 (SUV39H2), which is a SET containing methyltransferase. Loss of methylated H2AX has revealed substantially lower levels of γH2AX formation and expression and also leads to reduced affinity between the PIKKs and H2AX. Therefore, ATM and ATR at least need to bind to H2AX K134 to be able to mediate the phosphorylation at S139 for γH2AX formation (Sone et al., 2014). H2AX can also be phosphorylated on tyrosine 142 by the Williams syndrome transcription factor (WSTF) remodeling factor kinase, which has been shown to be involved in chromatin remodeling (Xiao et al., 2009; Cook et al., 2009). γH2AX is eventually dephosphorylated when the lesion has been repaired and several phosphatases such as PP1, PP2A, PP4 and WIP1 are involved in this process (Nazarov et al., 2003; Chowdhury et al., 2005; Chowdhury et al., 2008; Macurek et al., 2010). However, a portion of the phosphorylated H2AX is also evicted from the nucleosome by histone exchange (Svetlova et al., 2010).

Signal amplification occurs by spreading through the action of MDC1 binding to γH2AX (Stucki et al., 2005). MDC1 in turn recruits the MRN complex (MRE11–RAD50–NBS1) (Lukas et al., 2004) and the MRN complex further activates ATM (Uziel et al., 2003). This generates a positive feedback loop to drive spreading of the phosphorylation signal away from the break. This phosphorylation, which occurs minutes after the induction of damage, is not only localized to the site of the DSB but extends along the length of the chromatin on either side of the DSB (Nakamura et al., 2010) (Fig. 3). To explain the spreading of γH2AX, high resolution 4Pi microscopy was used to show that non-phosphorylated H2AX is localized in clusters within the nucleus. Upon damage induction, H2AX becomes phosphorylated to γH2AX and the spreading occurs until there is a gap in the H2AX substrate at which point the γH2AX spreading ceases (Bewersdorf et al., 2006). To further elaborate on the spreading of

γH2AX, a genome-wide ChIP-chip on the Damage-inducible via AsiSI (DiVA) system, showed that γH2AX was also limited to its domain around a DSB by the tethering molecule cohesin (Caron et al., 2012). Investigations using the DIvA system have also shown that the selection of break repair pathways is regulated by the transcriptional status of the break site (Aymard et al., 2014).

As H2AX knockout cells present only a moderate sensitivity to ionizing radiation but fail to recruit damage repair proteins like Nbs1 and BRCA1 to the repair foci (Celeste et al., 2002), the key role of γH2AX might not be the direct recruitment of repair factors but keeping the repair factors in the vicinity of the damage site thereby preventing the movement of damaged ends away from each other (Bassing and Alt, 2004; Celeste et al., 2003). Since the phosphorylation of H2AX occurs in the context of chromatin which undergoes decondensation around the DSB, one suggested role of γH2AX is the retention of chromatin remodeling factors around the damage site to demarcate a neighborhood for efficient repair.

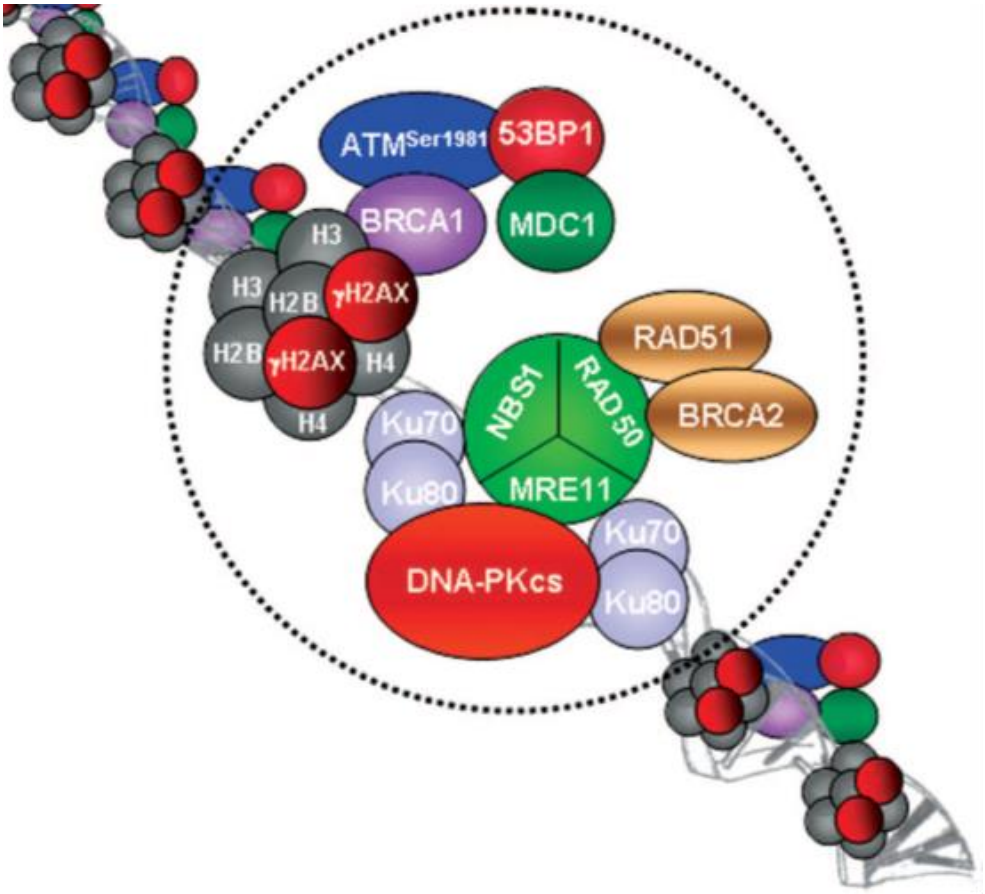


Figure 3. ***Recruitment of DDR proteins at DSB.*** *Region around a DSB expanded to show histone octamer containing* γH2AX *which can spread around the DSB. At the DSB,* DNA-PK *(recruited by* Ku70-Ku80 *heterodimers which initiate NHEJ),* Rad51-BRCA2 *(if cell is in S or G2 phase to initiate HR).* γH2AX *mediates the recruitment of DSB recognition and repair factors including* MDC1, BRCA1, 53BP1 *and* ATM. *(Adapted from Bhogal et., 2009).*

1.3.4 Redundancy in DNA damage response

Despite the activation of different signaling proteins and recruitment of diverse repair factors dependent on the type and location of the break and cell-cycle phase of the cell, there exists an inherent redundancy between these dynamic processes. The PIKK family kinases are distinct in their response to the type of DNA lesions, activation signals and kinetics but can share common downstream substrates (Culligan et al., 2006; Falck et al., 2005; Helt et al., 2005). For example, ATM is primarily activated in response to DSBs whereas ATR responds to ssDNA coated with RPA. It has been shown that these lesions can be sensed interchangeably indicating redundancy between the kinases (Cimprich and Cortez, 2008; Reaper et al., 2011). ATR can respond to DSBs, although to a lesser extent than ATM, and both ATM and ATR have common phosphorylation targets (Bartek and Lukas, 2003; Brown and Baltimore, 2003; Shiloh, 2003). ATM deficiency can be partially rescued by an overexpression of ATR and ATR is indeed upregulated in cells that lack a functional ATM (Cliby et al., 1998; Khanna and Jackson, 2001) indicating a redundant back-up role of ATR in the absence of ATM. The converse however does not occur, as indicated by the embryonic lethality of ATR knockouts and genotoxic sensitivity of ATR hypomorphic mutant cell lines (Brown and Baltimore, 2003; Gamper et al., 2013; Lewis et al., 2005; Nghiem et al., 2001; Wright et al., 1998).

1.3.5 Cellular events initiated by DNA damage response

Various forms of DNA damage encountered by the cell trigger different responses and outcomes. One such response is the utilization of 'checkpoints', which are surveillance and response pathways that halt the progression of the cell through cell cycle stages in response to DNA damage to prevent the conversion of lesions into permanent mutations. In some cases, cells with damaged DNA skip DNA repair and carry out programmed cell death (apoptosis) or remove themselves from the pool of proliferative cells by entering a state of permanent cell cycle withdrawal (senescence) (Kastan and Bartek, 2004). In addition to these outcomes, there also exist numerous repair pathways which regulate the repair of damage based on the nature of damage and the cell cycle stage in which the damage occurred.

1.3.5.1 Cell cycle arrest

The sequential series of events that give rise to two daughter cells is called the cell cycle. This process is divided into three stages: gap (G)1, synthesis (S), G2 and mitosis (M). During the G1 phase, the cell experiences growth and is characterized by the synthesis of new cellular organelles resulting in high protein synthesis rates and increased metabolic rates (Boye et al., 2009). External signals during this phase instruct the cell whether to prepare for DNA synthesis during the S phase. In the S phase, the DNA is duplicated for subsequent chromosome segregation and cell division. During the G2 phase, the cell prepares for cell division which occurs in M phase. Segregation of the newly duplicated chromosomes and ultimately cytokinesis happens during the M phase to give rise to two new daughter cells. Apart from these phases, another phase, G0, exists during which the cell neither divides nor makes any preparation to divide (Santoro and Blandino, 2010).

Cell cycle checkpoints are present to ensure fidelity of the cell division process and limit the number of mutations that are passed into the next generation of daughter cells. Three major cell cycle checkpoints exist, one for each of the transition steps between phases of the cell cycle. The cell can reversibly arrest the cell cycle progression in response to DNA damage to allow time for repair. Different damage response pathways are activated in different cell cycle phases, and inactivation of a pathway has a phase dependent outcome. Cyclins and cyclin dependent kinases (Cdks) regulate cell cycle transitions. Cdks require cyclin binding for their activity and substrate selectivity, and are regulated by activating and inhibiting phosphorylations (Shaltiel et al., 2015). In response to DNA damage, the activity of cyclin-Cdk complexes can initiate cell cycle arrest in the G1 or G2 phase, or slow down regulation in the S-phase.

i. G1/S checkpoint:

ATM and Chk2 are activated around damaged sites in response to damage in G1 and contribute to the response (Matsuoka et al., 1998). ATM and Chk2 stabilize p53, thereby stimulating transcriptional targets, such as Cdk inhibitor protein p21 (Harper et al., 1993; Harper et al., 1995). Cell cycle progression is hindered by p21 which accumulates and binds to cyclin-Cdk complexes. Loss of p53 or p21 causes a complete loss of the G1 checkpoint (Deng et al., 1995). Activation of the p38 MAPK family degrades cyclin D and hinders the activity of Cdk2 hence preventing entry into the S-phase.

ii. Intra-S checkpoint:

DNA damage in the early S-phase is particularly harmful since damage can interfere with replication fork progression and damage to single bases can result in mutational changes. The intra-S checkpoint is triggered by ATR and Chk1 (Paulsen and Cimprich, 2007). Cdk2 phosphorylates Exo1 and CtIP, which carry out extensive resection of DSB ends. This also promotes HR causing additional activation of ATR. Cdk2-dependent phosphorylation of ATRIP and Chk1 further restrict full activation of ATR and Chk1 to the S and G2 phases.

iii. G2/M checkpoint:

G2/M cell cycle arrests are regulated by Chk1 and Chk2 which are under the control of ATR and ATM respectively. Downstream accumulation of p21 during S-phase is inhibited by the PCNA-associated CRL4 ubiquitin ligase. In G2, Wee1 remains crucial for checkpoint control. Wee1 kinase, which becomes expressed in S-phase, targets Cdc25A for degradation restricting further Cdk activation. ATR and Chk1 signaling are essential for checkpoint maintenance in G2, and ATM and Chk2 are needed to control the arrest. However, the G2 arrest does not depend on ATM, p53 and p21.

During mitosis, DNA repair and checkpoint signaling are inhibited. However, DSBs formed during mitosis still activate ATM but further downstream signaling does not occur. Phosphorylation of XRCC4 by mitotic polo-like kinase 1(Plk1) and Cdk1 suppresses NHEJ (Shaltiel et al., 2015).

1.3.5.2 Limitations of cell cycle checkpoints

Despite the safeguards in place to correct DNA damage, cell cycle checkpoints have significant limitations (Deckbar et al., 2011). There are differences in the cell's ability to sense damage between early and late G1 phase which affect the cell´s ability to undergo checkpoint arrest. The late G1/S-phase checkpoint is less sensitive and does not always completely respond to DNA damage. All breaks are not repaired before moving on to the next step and the fraction of damaged cells undergoing arrest at this stage seems to decrease compared to damage arrested cells in early G1. Due to a limitation in the G2 checkpoint, cells where the number of DSBs falls below a threshold (10-20 DSBs), are released from G2 arrest (Deckbar et al., 2007). These cells can then proceed to mitosis with unrepaired DSBs which elevates the risk for

chromosomal misalignments leading to cell death or persistent DNA damage in the following G1 (Linke et al., 1997). When such cells carrying persistent DNA damage enter the G1 phase of the next cell cycle, arrest is initiated in a p53-dependent manner. This limitation explains the prominent role played by p53 in the regulation of DNA damage response over multiple cell cycles.

1.3.5.3 Senescence

The irreversible growth arrest of the cell in G1 but not the arrest of cellular or metabolic activities is termed as senescence. Senescence can be activated by loss of replicative ability due to telomere shortening (Replicative senescence) or stress induced senescence. Both types are characterized by altered cellular morphology, often becoming large, irregular and flat with an increased cytoplasmic size and protein content, altered gene expression and activation of the DDR, tumor suppressors and increased accumulation of γ-H2AX foci (Harley et al., 1990; Serrano et al., 1997; De Cecco et al., 2011; Sedelnikova et al., 2004; Schmitt et al., 2007).

During telomere shortening, only a few of the shortest telomeres of the cell are crucial to initiate DDR (Hemann et al., 2001). Telomere dysfunction, whether by shortening or through general damage, initiates senescence through the activation of DDR. This leads to p53 mediated (Di Fagagna et al., 2003; Herbig et al., 2004) or a p16 mediated (Smogorzewska and Lange, 2002) cell cycle arrest.

1.3.5.4 Apoptosis

Apoptosis or programmed cell death is an evolutionary conserved mechanism that normally occurs during embryo development and in response to cellular stress (Rich et al., 2010). It can also occur when DNA repair mechanisms are unable to repair DNA damage. Upon damage induction, ATM induces the phosphorylation of p53 (Shieh et al., 2000). This activation of p53 results in transcriptional regulation of both anti and pro apoptotic genes with dual roles in promoting repair and initiating apoptosis based on the severity of the DNA lesion (Roos and Kaina, 2013). ATM is also responsible for the activation of the nuclear transcription factor nuclear factor KappaB (NF-κB) an anti-apoptotic mediator (Piret et al., 1999).

1.4 DNA Damage Repair mechanisms

After the detection of lesions in the genome through the damage sensing apparatus, the cell employs a variety of repair mechanisms to repair the damage. The cell repair mechanisms include the excision repair pathways: base excision repair (BER) and nucleotide excision repair (NER); mismatch repair (MMR); and DSB repair pathways: homologous recombination repair (HR) and non-homologous end joining (NHEJ). Even though these repair pathways target specific types of damage, sometimes cells can use a combination of these pathways to repair complex forms of damage.

1.4.1 Base Excision repair

BER is a repair mechanism focused mainly against endogenous DNA damage such as spontaneous degradation of DNA or due to ROS. Bases that have undergone chemical modifications such as methylation, oxidation and deamination are removed through BER. Damaged DNA from one DNA strand is removed and the resultant gap is filled in using the complementary strand as a template, which makes BER a 'cut and patch' type of repair system.

Specifically, glycosylases recognize the damaged bases and catalyze the hydrolysis of the N-glycosidic bond between the damaged base and sugar, resulting in either a single apurinic/apyrimidinic (AP) site or a series of sites. This AP site is processed by an AP endonuclease that leads to cleavage of the phosphodiester bond either 5' or 3' to the AP site (Maynard et al., 2009). A single nucleotide is filled in by DNA polymerase β and the strand is repaired by DNA ligase III (Lindahl and Wood, 1999) while DNA polymerase β, δ and ε fill in longer gaps and DNA ligase I finishes repair (Mol et al., 2000, Hitomi et al., 2007, Rastogi et al., 2010) (Fig. 4).

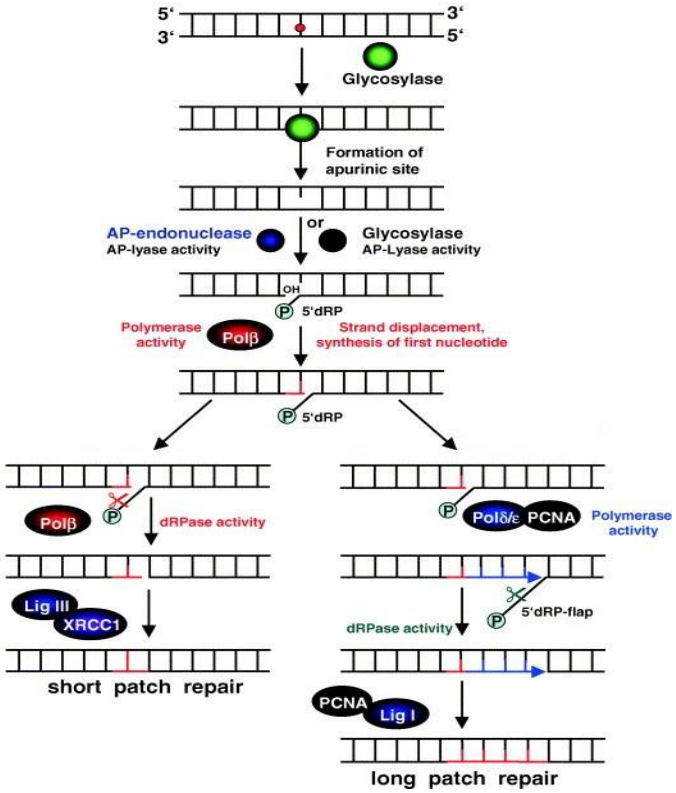


Figure 4. ***Mechanism of base excision repair (BER).*** *Recognition of the DNA lesion occurs by DNA glycosylase which removes the damaged base by hydrolyzing the N-glycosidic bond. The remaining AP site is processed by APE. Depending on the cleavability of the resulting 5'dRP by Polβ, repair is performed via the short or long patch BER pathway. (Adapted from Christmann et al., 2003)*

1.4.2 Nucleotide Excision repair

NER is a versatile repair pathway that removes bulky helix distorting DNA adducts arising from damage caused by UV radiation and alkylating agents. NER is composed of two pathways: transcription coupled repair, where actively transcribed strands of genes are given preference for repair; and global genome repair, where both the non-transcribed strands of a transcribed gene, as well as non-transcribed regions of the genome, are repaired (Lindahl and Wood, 1999). In global genome repair–NER, DNA lesions are recognized by the protein complex XPC-HR23B that initiates repair events. Transcription-coupled repair-NER pathway is activated when damaged DNA blocks RNA polymerase II from the transcribed strand of active genes (de Boer and Hoeijmakers, 2000). The subsequent steps in both global genome repair-NER and transcription-coupled repair-NER pathways involve the helicases XPB and XPD, which are part of the multi-subunit transcription factor TFIIH. RPA coats the undamaged strand and protects single-stranded DNA, while the endonucleases XPG and ERCC1/XPF act as molecular scissors to cut the 3' and 5' ends of the damaged

strand, respectively. Following removal of the damaged strand, the newly formed gap is filled in by either DNA polymerase ε or δ, and the resulting nick is closed by DNA ligase I (Benhamou and Sarasin, 2000; Gillet and Scharer, 2006; Wood, 2010) (Fig. 5).

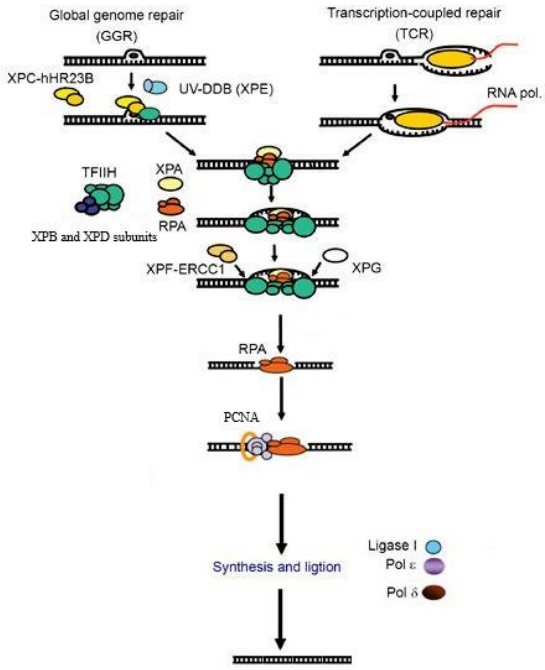


Figure 5. Two subpathways of mammalian NER. Detection of helix distortion by XPC-HR23B *in global genome-NER. Transcription coupled-NER is triggered by DNA damage-mediated blockage of* RNA polymeraseII (RNAPolII). *Sub-pathways converge followed by helix opening and damage verification by subunits of* TFIIH *and* RPA. ERCC1-XPF *and* XPG *structure-specific endonucleases incise the damaged strand. The single strand gap is filled by* DNA polymerase, *and sealed by* DNA ligase III-XRCC1. *(Adapted from Fousteri and Mullenders, 2007).*

1.4.3 Mismatch repair

MMR targets insertion/deletion loops and base-base mismatches which arise due to base misincorporation and DNA polymerase slippage during DNA synthesis and recombination (Li, 2008). Loss of MMR activity causes an increase in genome-wide point mutations and problems with replication of repeat rich microsatellite DNA regions leading to microsatellite instability (Hegan et al., 2006). It also plays a role in regulating the G2/M cell cycle checkpoint and apoptosis in response to DNA-damaging agents (Hardman et al., 2001; Wei-feng et al., 2006).

The E. coli MMR genes MutS and MutL have remained conserved throughout evolution (Acharya et al., 2003). Five MutS homologues (MSH) have been identified in humans that function as two major heterodimers: MutSα (MSH2 and MSH6 heterodimer) and MutSβ (MSH2 and MSH3 heterodimer). MutSα is the most

abundant and exhibits a high affinity for single base mispairing and single strand loops resulting from the insertion or deletion of one nucleotide. MutSβ functions in the repair of 2-4 insertion/deletion loop mispairs (Fishel and Wilson, 1997, Buermeyer et al., 1999, Jun et al., 2006). The MutL homologues (MLH) consist of MLH1, MLH3 and post-meiotic segregation (PMS) PMS1 or PMS2. The recognition is done by MutSα and the MutLα heterodimer (comprising of MLH1 and PMS2). PCNA and exonucleases (EXO1) interact with MutSα/MutLα homologs to excise the mismatched base. Polymerase δ then fills the gap and the nick is sealed by DNA ligase 1 (Jascur and Boland, 2006; Jiricny, 2006; Bak et al., 2014) (Fig. 6).

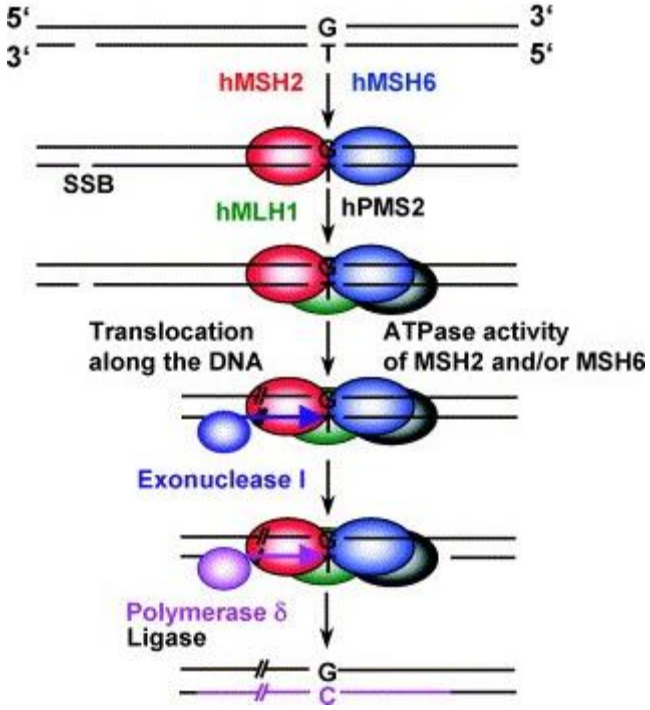


*Figure 6.**Model for Mismatch Repair**. Mismatched DNA bases are recognized by* MutSα *heterodimer (*MSH2/MSH6*). The* MLHl/PMS2 *heterodimer (*MutLα*) recruits and binds to* MutSα*.* PCNA *and exonucleases (*EXO1*) interact with* MutSα/MLH *homologs to excise the mismatched base. The resulting ssDNA is stabilized by* RPA *and filled by DNA polymerases in the presence of* PCNA*. The gap is sealed by* DNA ligase I*. (Adapted from Christmann et al., 2003)*

1.4.4 Homologous recombination

HR uses the sequence homology found in a sister chromatid or homologous chromosome to be used as a template to facilitate repair. Since this repair mechanism uses a template for strand synthesis, it is regarded as an error-free repair mechanism (Thompson 2012). It usually occurs during the S and G2 phases of the cell cycle. As mentioned previously, one of the earliest processes in HR is the recruitment of the MRN complex to the break site. This causes the activation of ATM which leads to the phosphorylation of downstream targets involved in cell cycle regulation. A direct

consequence of ATM activation is the resection of DSBs to produce 3' single-stranded DNA tracts. This resection is done by Mre11 which nicks the target strand up to 300 bases from the break site, and resects it in the 3' to 5' direction (Garcia et al., 2011). Simultaneously, exonuclease1 (Exo1) digests the strand in a 5' to 3' direction away from the DSB site (Tomimatsu et al., 2012). Mre11 endonuclease activity is stimulated by interaction with the C-terminal binding protein (CtBP) interacting protein (CtIP) (Sartori et al., 2007; Huertas and Jackson, 2009; Ramírez-Lugo et al., 2011). CtIP also interacts with BRCA1 to exclude the TopBP1-interacting protein, RIF1, a component of the non-homologous end-joining pathway which limits MRN resection (Chapman et al., 2013; Di Virgilio et al., 2013; Escribano-Díaz et al., 2013; Peterson et al., 2013; Xu et al., 2010a; Zimmermann et al., 2013). However, recently it has been shown that the interaction of CtIP with BRCA1 is dispensable for resection initiation in HR but is involved in regulating resection speed and efficiency (Cruz-Garcia et al., 2014). EXD2 has also been shown to promote end resection and HR in G2. EXD2 has a 3'-5' polarity and interacts with CtIP. It has been postulated that EXD2 promotes the generation of ssDNA downstream of Mre11 endonuclease where it cooperates with Mre11 exonuclease (Broderick et al., 2016).

An important requirement for end-resection is the unwinding of the DNA helix upstream of the DSB site. This is done by the RecQ helicase BLM which is recruited to damage sites by a direct interaction with the MRN complex, where it is phosphorylated by ATM kinase (Ababou et al., 2000). This recruitment is enhanced by the human heterogeneous nuclear ribonucleoprotein U-like (hnRNPUL) proteins 1 and 2, both of which have been shown to bind the Nbs1 C-terminus and to promote efficient end-resection (Polo et al., 2012). BLM forms two separate end-resection complexes, with Exo1 and Dna2, through which it stimulates their exonuclease activity (Cejka et al., 2010; Gravel et al., 2008; Nimonkar et al., 2011). RPA also appears to play a key role in end resection by stimulating the exonuclease efficiency of Exo1 and Dna2 (Yan et al., 2011). The binding of RPA to the exposed 3' DNA strand prevents the formation of inappropriate hairpin secondary structures which may present as targets for Mre11-mediated degradation (Chen et al., 2013). It is also essential for the recruitment of downstream repair proteins. Rad51 recombinase is an important RPA-interacting protein which, following the eviction of RPA, rapidly coats the exposed ssDNA strand (Sung et al., 2003) (Fig. 7). The eviction is initiated as a result of Rad51 capturing transient RPA that has been removed from the ssDNA

(Fanning et al., 2006). Rad52 has also been implicated in RPA-Rad51 exchange as it binds to the RPA34 and RPA70 subunits in response to prior RPA phosphorylation and stimulates RPA displacement (Sugiyama and Kowalczykowski, 2002; Jackson et al., 2002; Deng et al., 2009). The removal of RPA is also promoted by BRCA2 which promotes RPA eviction through competitive ssDNA interaction (Jensen et al., 2010; Liu et al., 2010; Yang et al., 2002). BRCA2 interacts with Rad51 through a series of eight 35 amino acid repeats, known as the BRC repeats (Bignell et al., 1997; Bork et al., 1996; Galkin et al., 2005). Later it was shown that these BRC repeats fall into two distinct classes: BRC1-4 which bind free Rad51; and BRC5-8 which interact with ssDNA-bound Rad51 (Carreira and Kowalczykowski, 2011). Formation of the Rad51 nucleoprotein filament is followed by sister chromatid capture and homology search (San Filippo et al., 2008). This is made possible by the ability of Rad51 to hold two DNA molecules in close proximity by the formation of a dsDNA binding groove on the exterior of the filament (Sung et al., 2003). Homology search by Rad51 is promoted by a direct interaction with Rad54 which translocates along dsDNA and allows the checking of the sister chromatid (Mazin et al., 2010). After a homologous region has been found, Rad54 promotes strand invasion by Rad51 where the invading filament base-pairs with the homologous template leading to the formation of a D-loop (Cejka et al., 2010; Mazon et al., 2010; Goodarzi and Jeggo, 2013). RPA binds and stabilizes the displaced strand Eggler et al., 2002). The RECQ5 helicase protects the invading strand by preventing the formation of new Rad51 filaments (Paliwal et al., 2013). Rad54 then promotes dissociation of Rad51 from the filament, leaving a 3' end for polymerase-mediated extension (Kiianitsa et al., 2006; Li and Heyer, 2009; Solinger et al., 2002). Synthesis is performed by DNA pol δ (Lydeard et al., 2007; Maloisel et al., 2008), in association with the replicative PCNA clamp (Li et al., 2009).

Rad52 is responsible for second-end capture of the extended invading strand by allowing the annealing of this invading strand with the RPA-coated complementary sequence on the second end of the DSB (Nimonkar et al., 2009). This leads to the formation of an entangled structure known as a double Holliday junction (Bachrati and Hickson, 2009). These Holliday junctions are processed by either junction dissolution or reolution to release two intact dsDNA strands. Dissolution is mediated by BLM in an ATP-dependent mechanism. The strands are then cut, unwound and reannealed by Topoiomerase IIIα, producing two non-crossover dsDNA molecules

(Swuec and Costa, 2014). As dissolution results in the formation of two accurate copies of the two original dsDNA strands, it is the favored method for Holliday junction processing in somatic cells. Alternately, Holliday junction processing by resolution involves nucleolytic cleavage by Yen1 and Mus81/Mms4 at each branch point of the junction (Blanco et al., 2010; Chan and West, 2015; Ho et al., 2010) which may result in the formation of crossover and non-crossover products depending on the direction of junction cleavage (Lee et al., 2015) (Fig. 7).

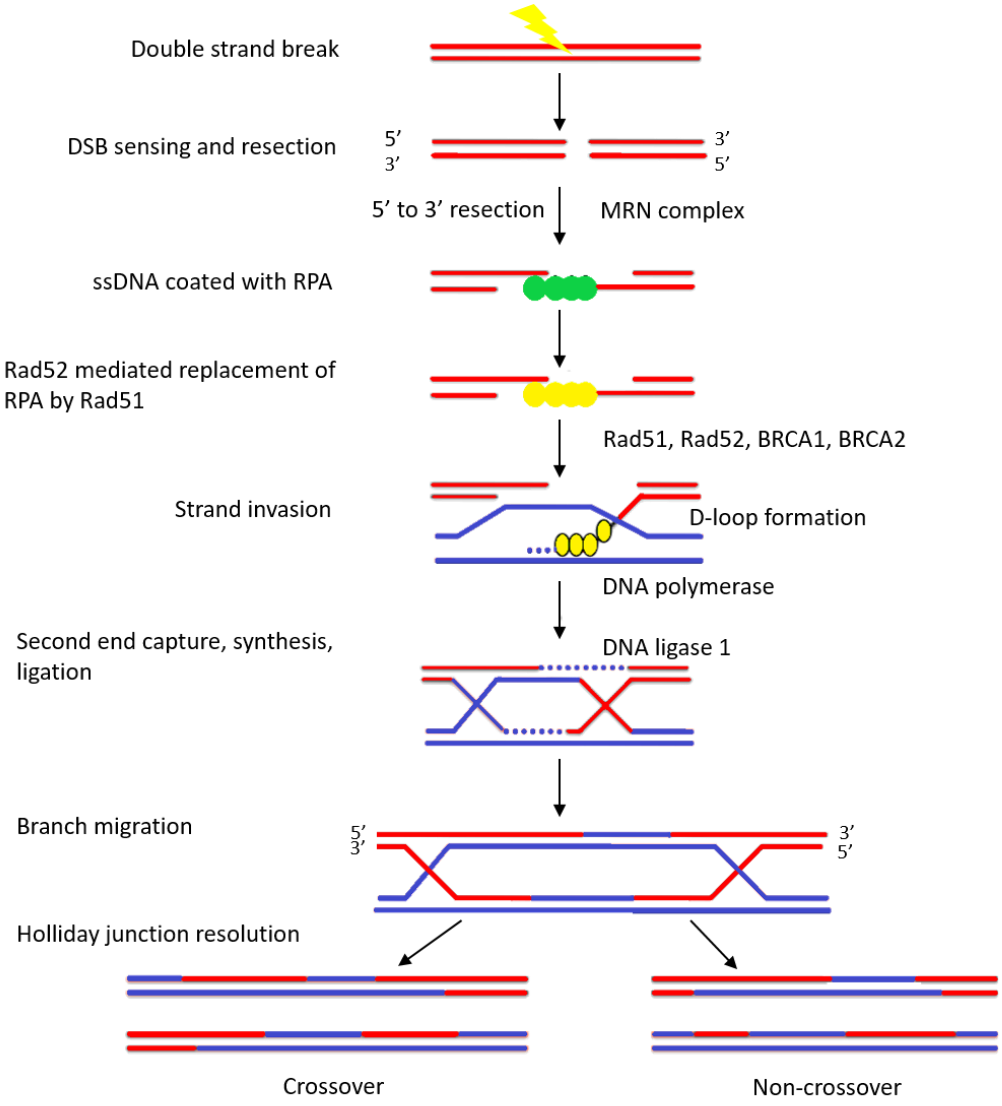


Figure 7. ***Homologous recombination repair (HR) pathway.*** *DSBs are recognized and bound by the MRN complex. MRN complex recruits other repair complexes, such as the ATM complex. RPA binds to the ssDNA and is then replaced by Rad51. Rad51 interacts with a complex of BRCA1, BRCA2 and Rad52, which perform homology search on the sister chromatid, leading to the formation of heteroduplex molecules or a displacement loop (D-loop) that matures into Holliday junctions (HJs). Then, the HJs are resolved by restriction endonucleases. Finally, the DNA ends are re-joined together by DNA ligase I. (modified from Misteli et al., 2009).*

1.4.5 Non-homologous end joining

Non-homologous end joining is so named because unlike HR, where the sister chromatid is used as a template to repair the damage, the free ends of the DNA at the site of the break are directly ligated. This makes the repair process error prone (Lieber and Wilson, 2010; Valerie and Povirk, 2003).

1.4.5.1 Classical Non-homologous end joining

In cNHEJ, DSB ends are simply religated without needing a homologous template for synthesis. This pathway is fast and active throughout the cell cycle but is also error-prone. Approximately 80% of the DSBs occurring in both G1 and G2 are repaired by c-NHEJ (Beucher et al. 2009). Depending on the nature of the lesion, limited end processing may be needed, which may lead to the loss of a few base-pairs hence increasing the chances of erroneous repair (Davis and Chen 2013).

As mentioned previously, the break detection is initiated by the Ku heterodimer which protects the DNA ends from nucleolytic degradation and recruits signaling molecules such as DNA-PKcs to tether the broken ends together. DNA-PKcs auto-phosphorylates and phosphorylates several molecules such as Artemis, polynucleotide kinase 3' phosphatase (PNKP), XRCC4, XLF4 and DNA Ligase IV which are responsible for end-processing and ligation. After autophosphorylation, DNAPKcs loses its kinase activity and dissociates from the break ends (Chan et al., 2002).

Depending on the free ends of the DNA at the break site, simple ligation or end-processing occurs. For example, the presence of a 3' damaged sugar in the form of a phosphoglycolate or a 5'-OH prevents direct ligation and thus needs end processing to produce ligatable break ends (Schipler and Iliakis 2013). End-processing is done by PNKP and Artemis. PNKP generates 5'-P ends and removes 3'-P ends because of its kinase and phosphatase activities respectively (Davis and Chen 2013; Goodarzi and Jeggo 2013). Artemis has autophosphorylated DNAPKcs-dependent endonuclease activity at DNA hairpins and ssDNA overhangs (Goodarzi et al., 2006). It has also been shown that Artemis has DNAPKcs independent 5'-exonuclease activity in vitro (Li et al., 2014). Recently, Artemis has also been shown to be involved in completing resection in DSBs generated during G1 cell cycle (Biehs et al., 2017).

Ligation of DSBs is the last step of cNHEJ. Ligation is carried out by DNA ligase IV which is stabilized by X-ray cross complementing protein 4 (XRCC4) and stimulated by XRCC4-like factor (XLF) (Grawunder et al., 1997; Andres et al., 2007; Riballo et

al., 2009). The newly identified paralog of XRCC4 and XLF (PAXX) interacts directly with Ku and promotes Ku-dependent ligation (Ochi et al. 2015) (Fig. 8). Because of its ring-like structure, Ku becomes trapped on the DNA strands after ligation has been completed and is eventually degraded by the proteasome after polyubiquitination (Davis and Chen, 2013).

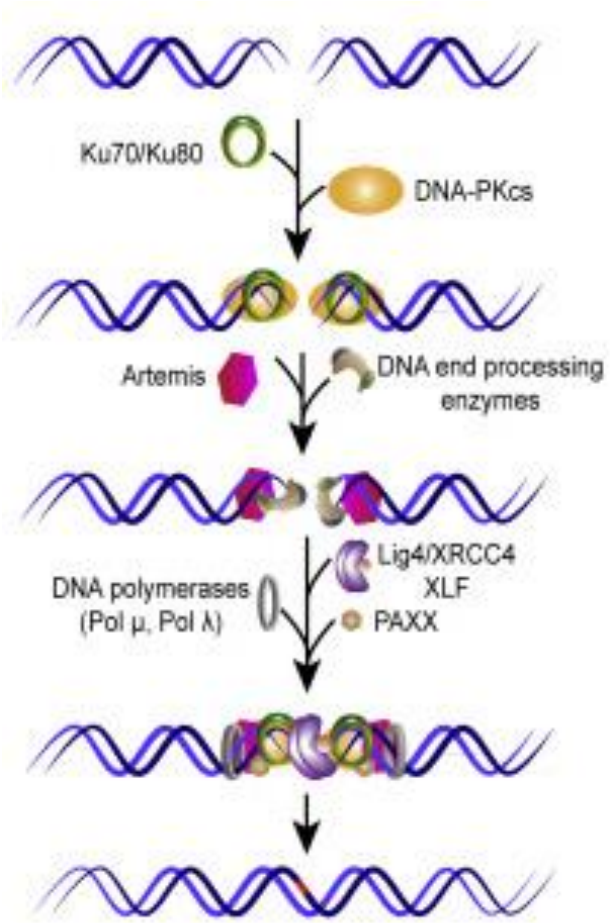


Figure 8. ***Model for DSB repair by c-NHEJ****. The DSB ends are bound by* Ku70/80*, which recruits* DNA-PKcs*, forming* DNA-PK*. After end processing by* Artemis *and other end processing enzymes, the DNA ends are ligated by* Lig4, XRCC4, and XLF *(modified from Iliakis et al. 2015).*

1.4.5.2 Alternate End Joining

In cells lacking the core component of cNHEJ pathway (Ku), a slower alternate end joining pathway is activated (Mansour et al. 2013). This pathway requires PARP1 (Wang et al. 2006) which has high affinity for SSBs and DSBs. Alt-EJ involves DSB end resection at the break site mediated by CtIP and Mre11. This can result in deletions of up to 100 base pairs which makes Alt-EJ a highly mutagenic repair

mechanism (Mansour et al. 2010). Ligase I/III and X-ray cross complementing protein 1 (XRCC1) mediate the ligation step (Iliakis et al., 2015) (Fig. 9).

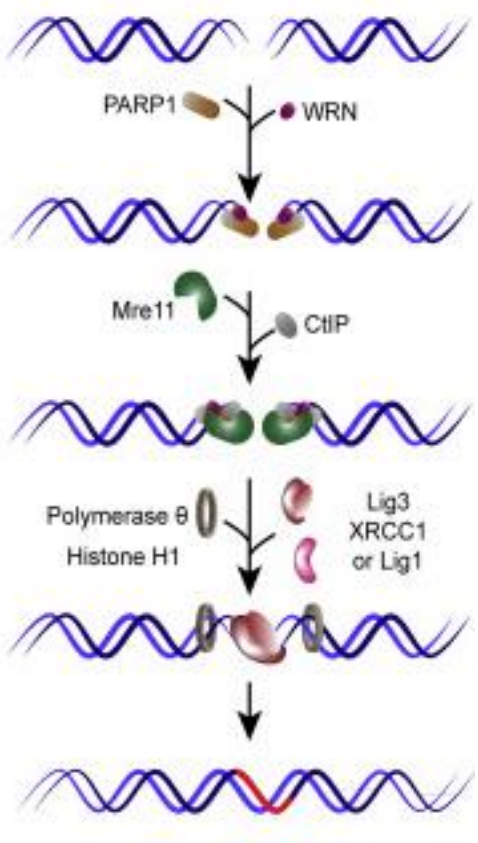


Figure 9. ***Model for DSB repair by alt-NHEJ***. PARP1 *is rapidly recruited to the DSB ends. After* CtIP- *and* Mre11-*dependent end resection, DNA ends are ligated by* Lig1, Lig3, *and* XRCC1 *(modified from Iliakis et al. 2015).*

The two DSB repair pathways have different mechanisms for break repair which affect the kinetics. Monitoring the kinetics of repair has revealed that repair processes have a fast and slow component. The fast component is believed to represent repair by NHEJ while the slow component is dependent on the phase of the cell cycle which in G1 is represented by A-EJ and by HR in G2. Both mechanisms involved in the slow component of repair have an end resection step which is dependent on Artemis and ATM which could indicate repair of complex DSBs (DiBiase et al., 2000, Riballo et al., 2004, Beucher et al., 2009, Biehs et al., 2017). Given the compact nature of heterochromatin, it is less accessible to the repair machinery and so is repaired slower than euchromatin (Cowell et al., 2007, Goodarzi et al., 2008, Natale et al., 2017).

1.5 Role of chromatin organization in DNA damage repair

All DNA metabolism activities such as transcription, replication and repair happen in the context of a highly organized chromatin. Chromatin shows a dynamic response to DNA damage to facilitate repair. The currently favored "prime-repair-restore" model suggests that chromatin modifications give access to the damage site, prime the environment for repair and restore the chromatin structure after the DNA repair has been accomplished (Misteli and Soutoglou, 2009; Soria et al., 2012).

1.5.1 Chromatin organization

Eukaryotic cells store and organize their DNA in the form of a highly compacted and structured chromatin. The nucleosome forms the basic repeating structural unit of chromatin and appears as "beads on a string" under an electron microscope (Olins and Olins, 1974). It is comprised of 147 base pairs of DNA wrapped around a histone octamer. The histone octamer consists of 2 copies of the histone proteins H2A, H2B, H3 and H4. Nucleosomes are held together by a linker histone H1. This first level of organization compacts the DNA by 5-10-fold (Kornberg, 1974). The nucleosome strings then fold into ~30nm diameter fibers called solenoids (Luger et al., 2012) which further fold into higher order structures (Horn and Peterson 2002; Felsenfeld and Groudine 2003). Maximum chromatin compaction occurs at mitosis, when each DNA molecule is packed into a mitotic chromosome. Rather than intermingling freely, chromosomes occupy distinct territories in the nucleus (Cremer et al., 2000) (Fig. 10). The relative position of individual chromosomes is not random and defined pairings of chromosomes are found in many cell types (Cremer et al., 2001; Dundr and Misteli, 2001; Kuroda et al., 2004; Parada et al., 2004). The position of chromosomes relative to the nuclear envelope depends on the gene density, with gene poor chromosomes located closer to the periphery (Boyle et al., 2001; Cremer et al., 2001; Croft et al., 1999). Regions that are gene-rich require a continuous or frequent access to DNA and hence, have a relatively accessible form of chromatin called euchromatin. Gene-poor or transcriptionally silent regions are maintained in a highly compacted form known as heterochromatin (Kouzarides, 2007). Heterochromatin exists in two forms: constitutive and facultative heterochromatin. Regions which contain a high density of repetitive sequences such as satellite, centromeric and telomeric regions are packaged as constitutive heterochromatin. Chromocenters are prominent structures visible in mouse nuclei and are formed by clustered constitutive

heterochromatin. Facultative heterochromatin is located in developmentally regulated regions where chromatin state can be decondensed due to gene activities and cellular signals (Jost et al., 2012; Oberdoerffer and Sinclair, 2007; Trojer and Reinberg, 2007). Mammalian inactive X (Xi) or the random inactivation of an X chromosome in females is a classic example of facultative heterochromatin (Craig, 2005).

Initial studies of chromatin organization revealed the presence of 1 megabase wide focal DNA replication structures that were spatially stable over several cell cycles and were thought to represent chromatin structures (Nakamura et al., 1986; Nakayasu and Berezney, 1989). It was later shown that these regions contained aggregates of several 100kb domains/loops (Jackson and Pombo, 1998, Ma et al., 1998, Berezney and Wei, 1998, Zink et al., 1999, Cremer et al., 2000). New methods have become available which allow for a more detailed analysis of chromatin organization. Techniques like Chromatin Conformation Capture (3C) and its extension Hi-C allow the analysis of chromatin conformation and determine if two or more stretches of chromatin are in close spatial proximity with each other. From these approaches, it was shown that chromatin folds into distinct modular domains within chromosome territories. These modular domains, referred to as topologically associated domains (TADs), form when a continuous stretch of chromatin folds into a globular structure with a higher preference for internal interactions (Dixon et al., 2012; Lieberman-Aiden et al., 2009; Sexton et al., 2012). The size of these domains varies from 100kb to 10Mb in Drosophila (Sexton et al., 2012) and have a median size of 880kb in mouse embryonic stem cells. Similar chromatin states are found within TADs (i.e either active or repressive) and active genes cluster with other active genes (Lieberman-Aiden et al., 2009; Rao et al., 2014). TAD borders contain CTCF binding sites and are well defined, although they can shift between cell types (Andrey et al., 2013; Noordermeer et al., 2011). This is also true for chromatin states of individual TADs which can vary between cell types (Ciabrelli and Cavalli, 2014). Higher resolution Hi-C maps have shown the presence of previously undetected contacts forming loops of ~185kb (Rao et al., 2014). TADs have been shown to be repsonsible for interactions between enhancers and promoters (Whalen et al., 2016) and in regulating genome replication (Pope et al., 2014).

DamID experients in fruit-fly, human and mouse cells have also revealed genomic regions interacting with the nuclear lamina (Guelen et al., 2008; Peric-Hupkes et al., 2010; Pickersgill et al., 2006). These regions, termed as lamina-associated domains

(LADs) vary in size from 0.1 to 1 Mb and are characterized by lower expression levels, late replication times and are marked by silent chromatin marks with CTCF binding at the borders (van Steensel and Belmont, 2017; Pope et al., 2014; Kind and van Steensel, 2010). The architecture for this feature remains unchanged during development but individual genes that reside in LADs can dissociate from the lamina in response to differentiation signals (Peric-Hupkes et al., 2010).

All levels of nuclear architecture are dynamic but to varying extents. The behavior of a single locus depends upon the local structure, the position inside a TAD, the distance to the lamina and the position within the chromosome territory.

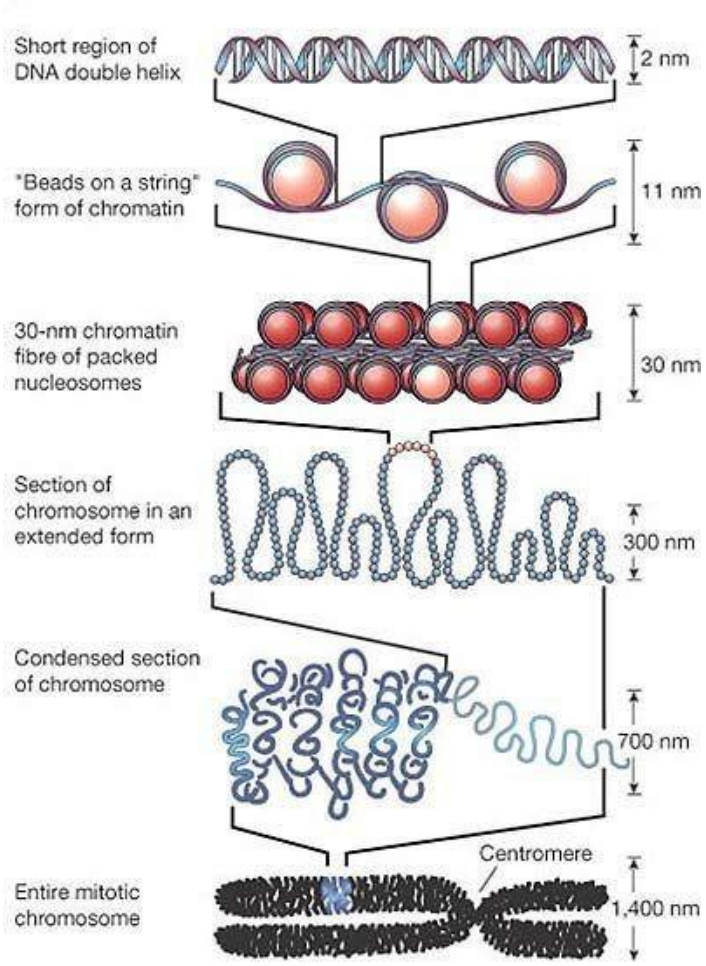


Figure 10. Chromatin organization in the nucleus. Hierarchical organization of DNA from double helix which coils around histone octamers, further compacted in chromatin fibers of packed nucleosomes and finally condensing into chromosome territories and chromosomes. (Adapted from Felsenfeld and Groudine, 2003).

1.5.2 DNA damage repair in the context of chromatin organization

The organization of the chromatin regulates the accessibility of the genome in a cell-lineage specific manner. Some of the major mechanisms that have been developed by

the cell to maintain this organization include ATP-dependent chromatin remodeling, post-translational histone modifications and histone variant incorporation. These mechanisms work together in a network to modify nucleosomes either locally at a specific locus, or globally leading to changes in the global chromatin structure (Smerdon and Thoma, 1990; Wellinger and Thoma, 1997; Thoma, 1999; Peterson and Côté, 2004). Chromatin remodeling and dynamics are essential for DNA damage repair, since the inherently compact nature of chromatin restricts the ability of repair proteins to access the site of damage (Price and D'Andrea., 2013).

1.5.2.1 ATP-dependent chromatin remodeling

ATP-dependent chromatin remodeling involves enzymes that use the energy released from the hydrolysis of ATP to alter the DNA-histone complex causing structural changes to the structure of nucleosomes (Osley et al., 2007) thereby making the nucleosome bound DNA more accessible to interacting proteins (Falbo and Shen, 2006, Farrell et al., 2011). This remodeling can also unfold the chromatin around a damage site to give access to repair factors. Chromatin remodeling is carried out by four identified families of multi-subunit complexes including SWI/SNF, INO80/SWR1, ISWI and CHD complexes. These subunits work by facilitating either nucleosomal sliding, which changes the position of the DNA relative to the nucleosome; nucleosomal displacement, which creates nucleosome-free regions; or adjust the spacing of nucleosomes which changes the position of histones in relation to DNA. These families have distinct roles in subcellular processes like replication, transcription, repair and recombination (Clapier and Cairns, 2009).

As a response to DNA damage, ATP-dependent remodelers are rapidly recruited to DSB sites in order to evict nucleosomes from surrounding chromatin. This allows accessibility to the MRN complex to initiate repair (Shim et al., 2005). The INO80 complex is recruited by γH2AX to favor DDR by chromatin remodeling and further recruitment of DNA repair factors (Kadamb et al., 2013). It has also been implicated in HR repair mechanism along with the SWI/SNF complex (Symington and Gautier, 2011) and CHD4 remodeling enzymes (Papamichos-Chronakis and Peterson, 2012). Studies in yeast have shown that the RSC complex (p400 in mammals), which is a member of the SWR-C family, regulates the incorporation of the H2A.Z variant to DSBs which is required for efficient NHEJ (Shim et al., 2005; Xu et al., 2012). The displacement of H2A-H2B dimers by the SWI/SNF-like remodeler Fun30 (in yeast;

mammalian SMARCAD1) has also been described to facilitate DNA processing by promoting Exo1 and CtIP driven resection (Costelloe et al., 2012; Chen et al., 2012; Eapen et al., 2012).

1.5.2.2 Post-translational histone modifications

The N-and C-terminal histone tails are subject to various post translational modifications (PTMs) such as acetylation, methylation, phosphorylation, ubiquitination, SUMOylation and PARylation among others (Cohen et al., 2011). These histone PTMs play a key role in DNA metabolism and are involved in transcription and DNA repair (Margueron and Reinberg, 2010; Gong et al., 2005).

The most important and well characterized histone PTM in the context of DNA damage repair is the phosphorylation of the histone variant H2AX to produce γH2AX which promotes recruitment of repair proteins to the damage site (See section 1.3.3). In addition to this there are other histone modifications and histone modifying enzymes that play a role in DNA damage response. Histone marks associated with the transcription status have been reported to be involved in controlling DNA repair pathway choice. Genes undergoing active transcription are repaired by HR while repressed genes are repaired by NHEJ. The histone mark H3K36me3, which is associated with gene activation, acts as a platform for recruitment of HR factors during S/G2 phase of the cell cycle (Aymard et al., 2014; Pfister et al., 2014). In accordance with these findings, the chromatin binding protein LEDGF which binds H3K36me3, was shown to promote resection and HR by recruiting CtIP (Daugaard et al., 2012). Mono- and di-methylation of histone H4 (H4K20me1-2) has also been shown to increase after laser-irradiation induced DSBs which facilitates recruitment of 53BP1 to DSB sites (Hartlerode et al., 2012). These histone modifications differ from γH2AX since these are not produced as a result of DNA damage and are constitutively present on euchromatin (H4K20me1/me2) and have been proposed as binding sites for 53BP1 and an early sensing mechanism for DDR (Jacquet et al., 2016). The role of H3K79me remains controversial due to contrary reports regarding its involvement in 53BP1 recruitment (Huyen et al., 2004; FitzGerald et al., 2011). Acetylation of histones is brought about by the HATs Tip60, p300, MOF and Gcn5 which are important in DDR and favor an open chromatin state (Narlikar et al., 2002; Sterner and Berger, 2000). Regulation of DNA damage repair by acetylation needs to be tightly regulated since some histone acetylations interfere in DNA damage

signaling. An example of this is H3K56ac which is associated with DDR. This histone mark does not appear to be responsible for recruitment of repair factors, but seems essential for final DNA repair processes and chromatin assembly after repair (Chen et al., 2008; Munoz-Galvan et al., 2013). However, other studies have also shown that this mark needs to be deacetylated by sirtuin family members and other HDACs (Michishita et al., 2009; Munoz-Galvan et al., 2013; Vempati et al., 2010; Yuan et al., 2009). Acetylation of H4K16 by MOF is essential for γH2AX foci formation and regulates the binding of MDC1, BRCA1 and 53BP1 to γH2AX domains through its chromatin unfolding action (Li et al., 2010; Sharma et al., 2010). RNF20 monoubiquitinates H2B at DSBs and allows local chromatin reorganization so that Rad51, BRCA1 and CtIP can access the DNA (Price and D'Andrea, 2013). Ubiquitination of H2A, H3 and H4 have also been reported to increase in response to UV-irradiation, reducing nucleosomal stability which promotes the recruitment of DNA repair complexes (Zhu and Wani, 2010). PARylation has been shown to have an important role in DDR (Polo and Jackson, 2011). PARylation of H2A, H2B and H3 tails by PARP1 facilitates DNA repair by transcriptional silencing of the regions flanking the damage site which prevents further DNA breakage due to collisions between RNA polymerases and repair factors (Lukas et al., 2011).

1.5.2.3 Histone variant incorporation

Histone variants are parologues of canonical core histones that differ by a few amino acids or the presence of larger modified domains, which confer specific attributes to chromatin structure. They may also possess unique sites which serve as targets for post-translational modifications. These variants replace the canonical histones under specific conditions such as transcription and during repair of damaged chromatin.

The best characterized histone variant in DDR is γH2AX which is a variant of H2A and acts as a marker for DNA damage. The chromatin remodeler p400 deposits H2A.Z, another variant of H2A, at DSB sites (Xu et al., 2012). The histone variant H3.3 is also involved in DNA damage repair (Adam et al., 2013). It plays a role in the nucleosome gap filling process (Schneiderman et al., 2012) and its deposition at UV-induced damage sites is crucial for restarting transcription after the completion of repair (Adam et al., 2013).

1.5.3 DNA damage repair in the context of spatial organization

As the genome is non-randomly organized, with activities such as transcription and replication happening at well-defined loci, the spatial organization also regulates the DDR (Misteli and Soutoglou, 2009). In yeast, the repair is organized in specific repair centers where multiple DSBs can be repaired at the same time (Lisby et al., 2003). Irreparable breaks can also migrate to the nuclear pores where their repair can be facilitated (Nagai et al., 2008; Oza et al., 2009). However, given the size difference between yeast and mammalian nuclei and differences in the nuclear architecture, DSB mobility in yeast is limited (Soutoglou et al., 2007; Soutoglou and Misteli, 2009). DSBs located at the nuclear periphery do not migrate to nuclear pores and instead are repaired by A-EJ (Lemaitre et al., 2014). Recently, the phenomenon of DSB clustering was reported in human cells with breaks in transcriptionally active genes clustering in an MRN complex dependent manner (Aymard et al., 2017). This can be considered as a mechanism to prevent mutations in such areas since these breaks are clustered at G1 and are repaired by HR in post-replicative cells.

An important event during repair is local chromatin decompaction at the break site. Since heterochromatin is composed of highly compacted chromatin, it needs to be made more accessible for repair factors. This decondensation is regulated by the ATM mediated phosphorylation of KAP1 which relaxes the compact state of heterochromatin allowing for its repair (Noon et al., 2010). SET1 was identified as an interactor of KAP1 which helps in its retention to the chromatin and its overexpression increases chromatin compaction and decreases HR (Kalousi et al., 2015). The decompaction and relocation of DSBs at heterochromatin into an environment more amenable for repair is extremely important for proper repair (Chiolo et al., 2011; Tsouroula et al., 2016).

The compaction of the genome also influences the spread of the DDR signal in the nucleus. For example, γH2AX cannot effectively disperse over actively transcribed regions (Iacovoni et al., 2010). Heterochromatin is also refractory to γH2AX formation and when γH2AX is eventually formed on heterochromatin, it remains persistent (Cowell et al., 2007; Goodarzi et al., 2010). The euchromatin-to-heterochromatin trend has been reported (Natale et al., 2017) and has been observed in the work carried out for this project.

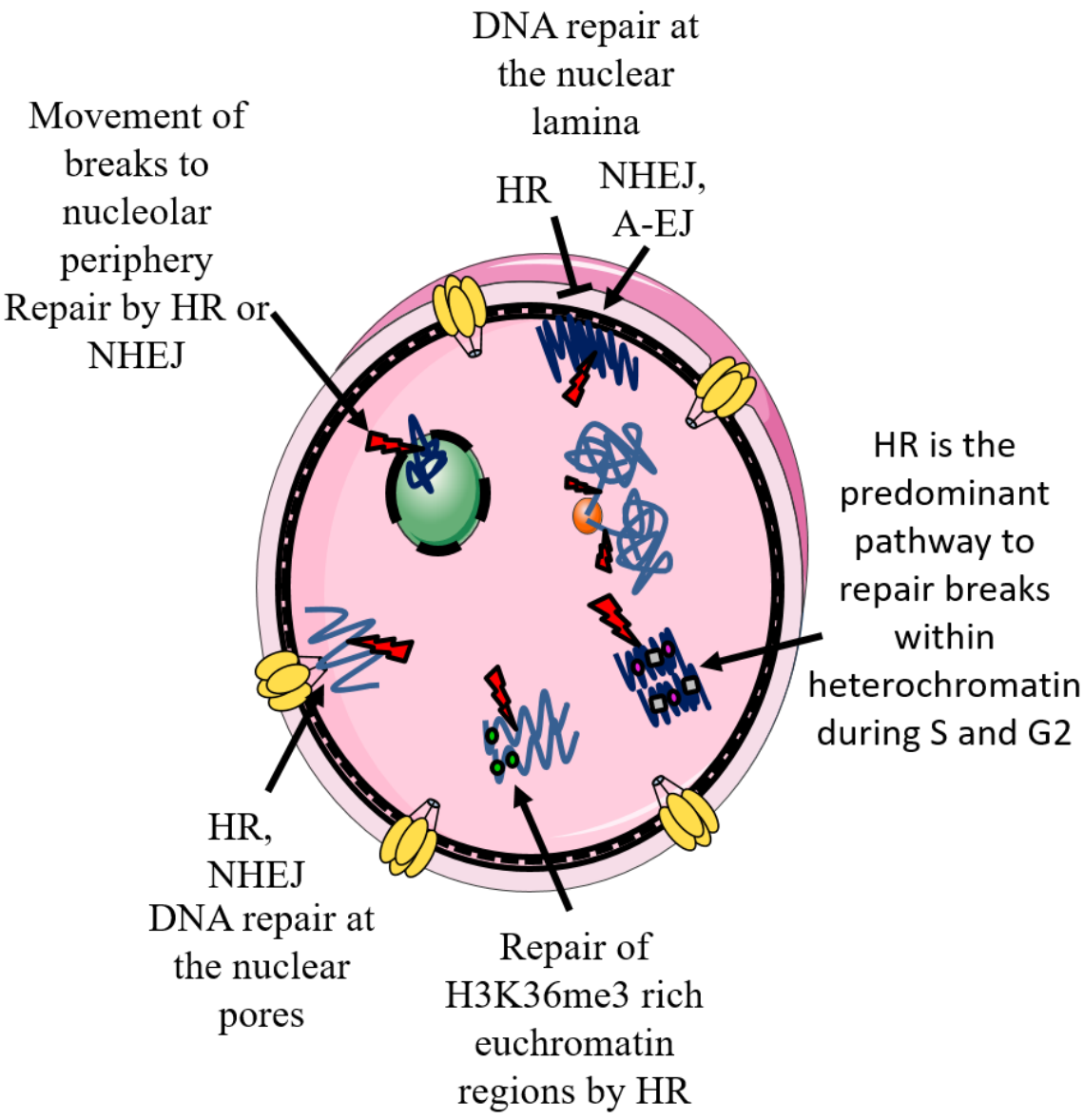


Figure 11. ***Compartmentalization of DNA damage response.*** *DNA damage response is differently organized in different nuclear compartments, cell cycle stages and based on the transcription status. (Adapted from Kalousi and Soutoglou, 2016)*

1.6 Embryonic stem cells

Stem cells are non-specialized cells with the ability to maintain their self-renewal potential while simultaneously undergoing asymmetric cell division to give rise to other specialized cell populations. Stem cells are divided into two broad sub-classes- embryonic stem (ES) cells and adult stem (AS) cells. Since ES cells can give rise to the three germ layers ectoderm, mesoderm and endoderm, they are termed as pluripotent cells (Armstrong, 2012). AS cells on the other hand, can only generate specific subsets of closely related cell types and are hence considered multipotent. For example, hematopoietic cells are AS cells which can generate all types of blood cells.

An important differentiating characteristic between ES cells and AS cells is their proliferative capacity. ES cells proliferate rapidly while AS cells are usually quiescent and proliferate only in response to specific stimuli.

Given their small size and limited number, ES cells are studied in vitro to understand their biology. After their extraction from the inner cell mass of the blastocyst, there are several ways of cultivating ES cells in culture. The most commonly used method is culturing on a layer of inactivated mouse embryonic fibroblasts (termed feeders) with the addition of leukemia inhibiting factor (LIF) which maintains the proliferative capacity of ES cells and prevents spontaneous differentiation, thereby maintaining their 'stemness'. Recent methods for their culture have also demonstrated that inhibitors of differentiation pathways like GSK3β and MEK1/2 also maintain the stemness, allow for feeder-free growth and resemble pluripotency conditions in vivo (Wray et al., 2010, Tamm et al., 2013). But since it is not known whether blocking these pathways would have any effects on the cell cycle or proliferative capacity, all experiments in this work have been carried out with standard ES cell cultivation protocols (Bibel et al., 2007).

1.6.1 DNA damage during embryonic stem cell cycle

Somatic cells spend a relatively large proportion of their cell cycle in the G1 and G2 phases and relatively less time in the S phase. ES cells, on the other hand, spend very little time in the gap phases and spend most of their cell cycle in the S phase (White et al., 2005). This however does not imply a longer S-phase in ES cells (Li et al., 2012). Given the rapid rate of proliferation, the G1-S checkpoint is often reported to be compromised in ES cells. Following DNA damage, ES cells fail to activate G1 checkpoint arrest (Fluckiger et al., 2006; Suvorova et al., 2012; Hong and Stambrook, 2004; Momcilovic et al., 2009). Absence of G1 arrest in ES cells is associated with cytoplasmic suppression of p53 (Aladjem et al., 1998) and Chk2 (Hong and Stambrook, 2004). Absence of a G1 checkpoint may also make ES cells more sensitive to DNA damage since overexpression of Chk2 in ES cells restores the G1 checkpoint and decreases radiation induced apoptosis (Hong and Stambrook, 2004; Hong et al., 2007). The tumour suppressor protein retinoblastoma (RB), which prevents inappropriate G1-S progression and damaged DNA from being replicated, is hyperphosphorylated in ES cells making it inactive (Savatier et al., 1994). Cell cycle regulators like Cdc25a, Cdc6, cyclins etc. - are extremely abundant in ES cells (Tichy

et al., 2012). Due to the short gap phases before and after replication, it may not be possible for ES cells to prevent damaged or partially replicated DNA from entering the next replication cycle. ES cells also have a higher expression of ROS-inactivating enzymes and hence have lower endogenous levels of ROS than differentiated cells (Saretzki et al., 2004; Giachino et al., 2013). The mutation rate in ES cells is also low with a 1000-fold less baseline mutation frequency than MEFs, at certain loci, indicating a very efficient mismatch repair mechanism (Giachino et al., 2013; Hong et al., 2007). The fast replication cycle, in addition to various other endogenous sources of damage, require a very robust DDR during ES cell proliferation since these cells will give rise to all other cell types. An active DDR in ES cells is also required to prevent developmental abnormalities and growth disorders. DDR and especially the presence or absence of DSBs in ES cells has not been exhaustively characterized and some of the reports are contradictory and remain controversial.

There have been reports of increased HR activity in ES cells which allows for gene targeting in mice (Te Riele et al., 1992). This could be due to the suppression of p53 in ES cells (Aladjem et al., 1998). Since p53 also suppresses HR (Mekeel et al., 1997), this may explain the high HR rates in ES cells. A landmark study also shows that p53 inhibits the transcription of Nanog, which is essential for maintaining the stemness (Lin et al., 2005). Reports also show that ES cells are hypersensitive to DNA damage and undergo apoptosis or differentiation to remove the pool of damaged cells (Van Sloun et al., 1999, de Waard et al., 2008). Suppression of p53 may also be associated with preventing cell cycle arrest during embryogenesis where rapid cell proliferation is essential and an increase in HR would be required for restarting stalled replication forks (Shrivastav et al., 2008). Higher levels of Rad51 are also found in ES cells which point towards the prevention of illegitimate HR independent of p53 and protection of stalled replication forks (Domínguez-Bendala et al., 2003, Tichy et al., 2012). Given the differences in Rad51 levels between ES cells and MEFs, no differences in the efficiency of HR have been observed (Tichy et al., 2012). Apart from this, efficient NHEJ has also been shown to contribute to DNA repair efficiency (Fan et al., 2011; Adams et al., 2010; Lan et al., 2012). On the other hand, some studies also report a reduced DNA repair capacity in ES cells and differences between mouse and human cells in their repair efficiencies, with human ES cells more efficient in repairing damage (Banuelos et al., 2008; Wyles et al., 2014).

Endogenous γH2AX foci have been observed in WT ES cells (Saretzki et al., 2008; Banath et al., 2009; Ziegler-Birling et al., 2009; Turinetto et al., 2012). However not all of these reports have explored the reason for the presence of these foci. Reports which do investigate these foci have given chromatin remodeling or alternative structures as the reason for their presence (Banath et al., 2009; Ziegler-Birling et al., 2009). These claims are based on the assumption that γH2AX would arise only in response to DSBs. γH2AX can also arise due to replication stress without the presence of a DSB (Löbrich et al., 2010). In a report, it has been (wrongly) postulated that since the γH2AX foci do not colocalize with 53BP1 foci in ES cells (which is a more specific DSB marker), γH2AX is not an indicator of DDR. It was also concluded that the presence of γH2AX foci is not due to the presence of a DSB in ES cells (Ziegler-Birling et al., 2009). Another study shows higher levels of γH2AX in ES cells compared to MEFs which decreases upon ES cell differentiation. These higher levels of γH2AX have been linked with self-renewal but this conclusion has again been made based on the lack of colocalization of 53BP1 in ES cells indicating an absence of DSBs (Turinetto et al., 2012). It is clear from these studies however, that the γH2AX foci do not arise due to culture conditions since the inner cell mass of the blastocyst also stains positive for γH2AX.

1.7 Review of techniques used to assess DNA damage

Our understanding of DNA damage sensing, the cellular response and subsequent repair mechanisms has improved rapidly over the last two decades. It has become increasingly evident that the packaging of chromatin in the cells regulates DNA damage repair (Kim et al., 2007; Peterson and Almouzni, 2013; Price and D'Andrea, 2013; Jeggo and Downs, 2014). The location of the DSB relative to the nuclear lamina, for example, has been shown to affect the repair by NHEJ and A-EJ (Lemaitre et al., 2014). Recently, it has been shown that repair of DSBs occurring in heterochromatin is cell cycle dependent, with DSBs occurring in G1 being repaired in situ while DSBs occurring during S/G2 relocate to the periphery of heterochromatin (Tsouroula et al., 2016). All of these point towards the fact that the repair kinetics, and ultimately the outcome of repair are intertwined with the chromatin context of the damage site. The methods to identify and follow DNA lesions at specific loci have evolved over the years and have enabled the quantitative and qualitative detection of damage and repair products. Before the use of Next Generation Sequencing (NGS) approaches, the methods to quantify DNA

damage were based on indirect methods such as staining for proteins that are associated with breaks or quantifying the global DNA fragmentation. These techniques had a limited scope as they used proteins as proxies for damage or were not able to give a whole genome view of damage and repair. NGS approaches have helped in identifying "hotspots" for damage and translocations within the genome (Crosetto et al., 2013; Chiarle et al., 2011; Hu et al., 2016). NGS approaches have also elucidated the organization of DNA repair with respect to replication, transcription, chromatin organization and the mobility of DSBs (Barlow et al., 2013; Bunch et al., 2015; Schick et al., 2015; Adar et al., 2016; Hu et al., 2017; Iannelli et al., 2017; Aymard et al., 2017). Techniques can be classified into two categories: for detecting the presence of DNA damage; and techniques for evaluating the repair mechanism. Some of the techniques commonly used to study and quantify DNA damage have been briefly discussed here. The NGS approaches to quantify DNA damage have been elaborated.

1.7.1 Polymerase chain reaction based methods:

Polymerase chain reaction (PCR) is a reliable method for measuring DNA damage at the gene level. The amplification of the product is blocked at the site of the lesion since the polymerase cannot continue beyond this site (Govan et al., 1990). The amount of damage can be estimated by quantifying the difference in the amplification of the products between the damaged and undamaged samples. This technique can detect bulky DNA adducts but does not provide the exact location of the damage within a specific region. It also requires a significant amount of damage in order to give a quantifiable difference between the samples (Karakoula et al., 2003).

Ligation-mediated PCR (LM-PCR) is a variant of this approach and can identify the exact position of damage within a specific region. Specific linkers are ligated to DNA ends which are then amplified by PCR using another specific linker within the region of interest. These amplified fragments are separated on agarose, blotted on nylon membranes and detected with gene-specific probes. This technique has been used to map damage caused by ROS and UV-induced damage (Rodriguez et al., 2000; Pfeifer et al., 1998; Douki et al., 2000). LM-PCR is technically cumbersome and relies on a priori information of the damage site to produce ligatable structures.

Immuno-coupled PCR (IPCR) relies upon the specificity of antibodies to bind to damaged lesions and the subsequent amplification by PCR of the targeted region. This

technique was used to quantify thymine dimers at gene levels and global levels after UV-irradiation (Karakoula et al., 2003).

1.7.2 Electrophoresis based methods:

i. Pulsed field gel electrophoresis (PFGE) was designed to separate large molecular weight yeast chromosomes on an agarose gel subjected to a periodically changing direction of current (Schwartz and Cantor, 1984). This technique was modified to study strand breakage in DNA (Sutherland et al., 1987). This technique was able to detect DNA fragments lower than 100kb but it is not possible to separately identify unique fragments with the same electrophoretic mobility. The high amount of sample needed along with the false positive signals generated from apoptotic and necrotic cells, shearing of DNA during sample preparation and inability to detect clustered DSBs which generate smaller fragments allow for only an approximate estimation of the number of DSBs present in the cell.

ii. Single cell gel electrophoresis (Comet assay)

Single cell gel electrophoresis, popularly referred to as the comet assay, is a sensitive assay to assess DNA damage. It is particularly effective in detecting SSBs and DNA lesions caused by ROS. The cells, which are embedded in agarose plugs before lysis to prevent any extraneous breaks being produced due to handling, are subjected to an electric current that pulls the free ends of DNA away from the central body, resulting in a distinctive comet-like shape. The level of damage is measured by the length of the tail (Singh et al., 1988; Collins, 2004; Olive and Banáth, 2006). The basic comet assay can only detect strand breaks and alkali labile sites while digesting DNA with lesion specific enzymes can convert DNA damage into additional strand breaks (Azqueta et al., 2011). Although this technique is simple and reliable, it lacks in specificity and quantitation.

1.7.3 Fluorescence based methods:

DNA repair proteins such as the Ku heterodimer, γH2AX, 53BP1, etc. have been used as molecular markers for damage and repair and have been analyzed through confocal microscopy and flow cytometry. DNA breakage dependent fluorescence in situ hybridization (DBD-FISH) is a technique which can be used to study DNA breaks in

specific genomic loci. DBD-FISH has been used to identify unstable genomic regions in the genome during the progression of cervical cancer (Cortés-Gutiérrez et al., 2015). Terminal deoxynucleotidyl transferase (TdT) dUTP nick-end labeling or TUNEL assay has also been used to detect SSBs or DSBs and to detect levels of apoptosis by quantifying the levels of DNA fragmentation (Darzynkiewicz et al., 2008). Fluorescence based methods are versatile and reliable although they have some inherent disadvantages such as background autofluorescence and reproducibility of signals.

1.7.4 Next-Generation Sequencing (NGS) based methods

Advances in sequencing technologies coupled with plummeting sequencing costs have made NGS techniques increasingly accessible and accepted for diagnostic purposes and monitoring disease progression in patients. Unlike microarray based techniques, which rely on a priori information of the genome or the genomic regions being studied, NGS can map protein-DNA interaction and DNA-DNA interaction at nucleotide resolution besides giving information about splice variants, post-translational modifications and novel non-coding RNA. NGS approaches serve as an important tool to monitor DDR processes upto the nucleotide level. An increasing number of NGS techniques have been developed to determine the genomic footprint of DDR proteins, the location of damage and correlations with epigenomic markers. Some of the major techniques and their findings have been covered in this section.

1.7.4.1 Chromatin immunoprecipitation and sequencing (ChIP-Seq)

Coupling chromatin immunoprecipitation with high-throughput sequencing allows the interrogation of the complete genome to identify the binding sites of target molecules, histone modifications and chromatin modifying enzymes throughout the genome and correlate it with other genomic features. In the ChIP-Seq protocol, the cells are cross-linked so that they retain their original interaction pattern. After this fixation, the chromatin is fragmented (by sonication or enzymatic reaction) to yield short fragments of chromatin with the target of interest still bound to the chromatin. Antibodies coupled with magnetic beads are used to pull down the fragments which are bound to the target of interest. These fragments are reverse crosslinked and purified for preparing sequencing libraries before being sequenced (Fig. 12).

ChIP-Seq has been extensively used in the field of DNA damage repair to identify hotspots of damage, chromatin status at the site of damage and repair kinetics. This technique has been used to show that DNA damage caused by high salt stress is limited to gene desert regions and further suggested that this non-random accumulation of damage could be the reason for the evolution of genome organization (Dmitrieva et al., 2011). ChIP-Seq was also used to show that DSBs occurring in regions undergoing active transcription are repaired by HR (Aymard et al., 2014). This technique was also used to show that γH2AX accumulated at transcriptionally active sites (Bunch et al., 2015). Very recently, γH2AX ChIP-Seq on human cancer cells has shown that the DNA damage is non-randomly distributed in the nucleus (Natale et al., 2017). The advantages of ChIP-Seq over microarray based ChIP-Chip include the fact that ChIP-Seq has a higher resolution and lower background noise. Also, ChIP-Chip does not allow the total coverage of the genome and excludes repeat rich regions and regions without complete sequence information. An important caveat of ChIP-Seq is that it uses a protein-antibody affinity reaction to do the pulldown and so can be affected by variability of the antibody which needs to be validated every time before the experiment. This protein-based approach also only provides information about the genomic location of the target, and for proteins such as γH2AX, which are known to spread away from the break site, only helps in identifying domains of damage and not the actual break-site.

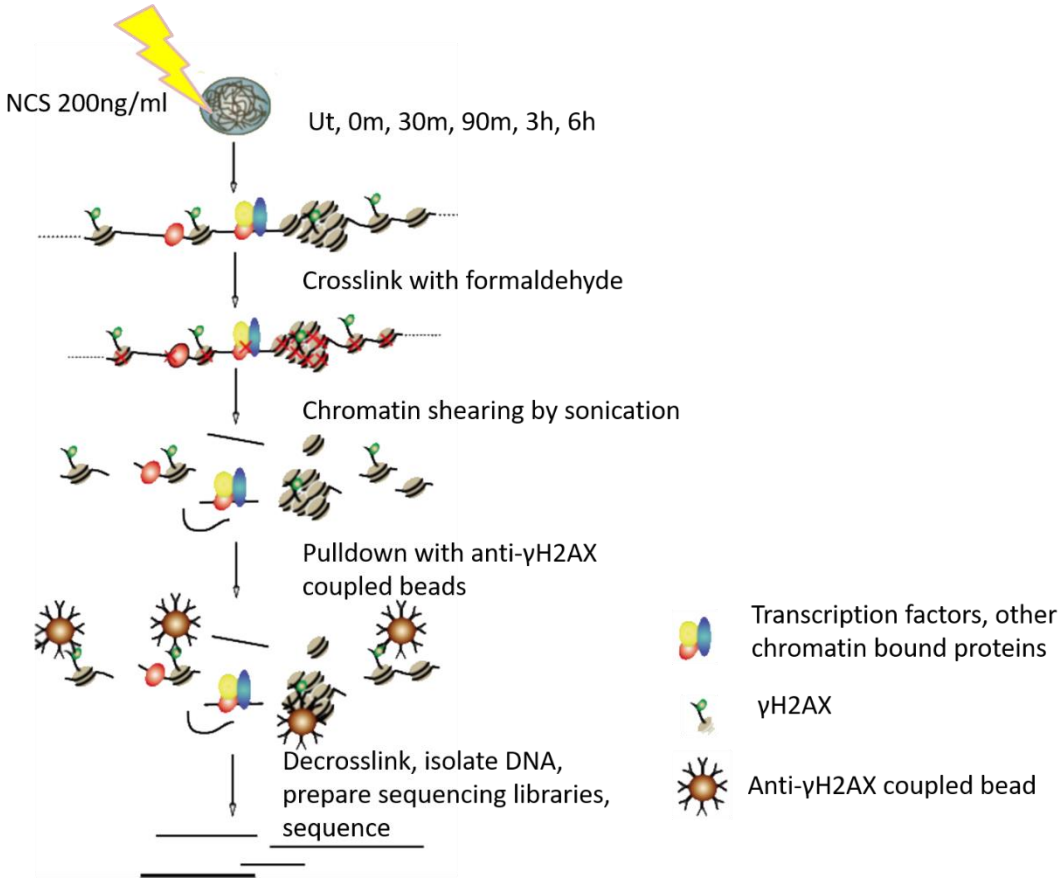


Figure 12. ***Chromatin Immunoprecipitation and sequencing.*** *Ut, 0m, 30m, 90m, 3h, 6h indicate the timepoints at which samples were collected after exposing mouse embryonic stem cells to global genomic DNA damage with 200ng/ml of NCS during experiments for this project. (Ut: untreated cells).*

1.7.4.2 In-situ breaks labeling, enrichment on streptavidin, and next-generation sequencing (BLESS)

As mentioned earlier, ChIP-seq uses proteins as a proxy for damage but does not provide any information of the actual break site. BLESS was developed to precisely identify the break-site without depending on a proxy for the break (Crosetto et al., 2013). The technique relies on a biotin-labeled primer which binds to the break site (endogenously present or induced) thereby labeling the break-site. The chromatin is sheared by enzymatic digestion to give short hairpin molecules which are pulled down (enriched) on streptavidin beads. These hairpin molecules are then circularized by ligating a second unlabeled primer. These circular molecules are digested with restriction enzymes whose restriction sites are uniquely present only on the primers ligated to the DNA. Sequencing libraries are prepared from these fragments before they are sequenced (Fig. 13). Because of the design of the primers which are ligated, it is possible to identify the precise point where a break has occurred in the genome.

This technique was used to identify endogenous breaks and aphidicolin and neocarzinostatin sensitive regions in HeLa cells (Crosetto et al., 2013) and in identifying off-target cutting sites of CRISPR-Cas9 (Ran et al., 2015). It has also been used as a benchmark for judging the efficiency of other break-mapping techniques that have since been developed. An important drawback of this technique is the creation of fixation-induced DNA breaks during sample preparation. Other reported issues with this technique include inefficient linker ligation to the DNA ends, low library complexity due to the presence of linkers in all reads and high background signals (Mitra et al., 2015). During comparisons with other methods to identify DSBs, it was also reported that both ends of the break could occasionally not be detected (Canela et al., 2016).

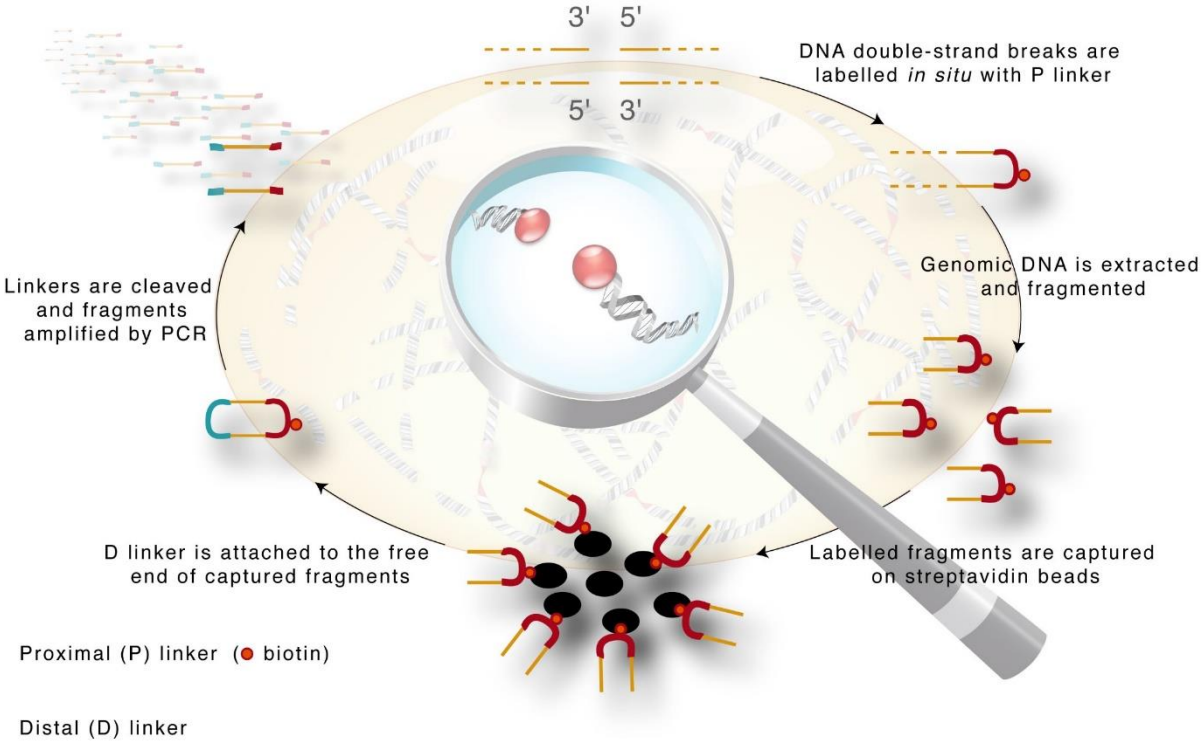


Figure 13. ***BLESS workflow.*** *DSBs are ligated in situ to a proximal linker (red arch) covalently linked to biotin (red circle, genomic DNA is extracted and fragmented; labeled fragments are captured on streptavidin beads (black circles). A distal linker (cyan arch) is then ligated to the free extremity of captured fragments, and fragments are released by linker digestion with I-SceI. Released fragments are amplified by PCR using linker-specific primers and sequenced. (Adapted from Crosetto et al., 2013).*

Other methods that provide high-resolution information about the genomic location of lesions have also been developed. These methods can be classified into two classes based on the information they provide, i.e., mapping the presence of free DSBs (unrepaired); and, mapping faulty repair through translocations or deletions.

1.7.5 Mapping unrepaired DSBs

These techniques detect the presence of free DNA ends and have some common features including ligating labelled primers or nucleotides to the break-site before shearing the genome and enriching for the tagged fragments. These techniques also involve an amplification step to generate libraries suitable for next-generation sequencing.

1.7.5.1 Damaged DNA immunoprecipitation (dDIP) and DNA break immunocapture (DBrIC)

dDIP was the first technique developed for uniquely identifying DNA ends in cells and involved using terminal deoxynucleotidyl transferase (TdT) to add biotinylated dNTP to the DSBs before shearing the genome and pulling down the tagged fragments on streptavidin beads (Leduc et al., 2011) (Fig. 14). This technique detected HO-induced DSBs in yeast and telomeres. An improvement of this technique, called DBrIC, was able to detect a DSB induced by the enzyme I-SceI in at least 2% of the cell population (Gregoire et al., 2016).

1.7.5.2 DSB-Seq

DSB-Seq is similar to dDIP and DBrIC since it also labels the free ends of the DSB. A variant of this technique called SSB-Seq could also specifically detect single-strand breaks (Baranello et al., 2014). These two methods were used to map DNA lesions produced by the topoisomerase II poison etoposide.

1.7.5.3 Break-Seq

This method is similar to the previous methods and was used to generate maps of lesions produced in cells after treatment with the DNA replication inhibitor hydroxyurea (Feng et al., 2011; Hoffman et al., 2015).

Since these techniques relied on the addition of dNTPs to the free ends, an important caveat of these approaches was the initial structure of the break which would affect the number of dNTPs added and hence lead to a biased quantification of breaks.

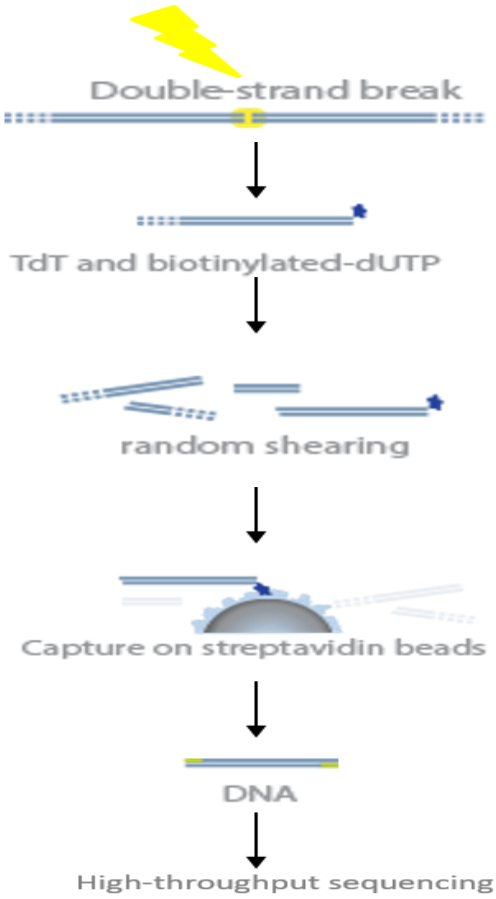


Figure 14. ***Common workflow for dDIP, DBrIC, DSB-Seq and BREAK-Seq.*** *Biotinylated dUTP is added to DSB ends followed by random shearing of the genome. Biotin-labeled fragments are pulled down on streptavidin beads, purified and sequenced. (Adapted from https://www.illumina.com/science/sequencing-method-explorer/kits-and-arrays/dsb-seq.html)*

1.7.5.4 DSBCapture

This technique is similar to BLESS but uses modified biotinylated linkers which permit more efficient tagging of DNA ends (although direct comparisons have not been performed). An important improvement over BLESS addressed the issue of low-complexity library by changing the amplification steps necessary for the library preparation. It has been used to map DSBs produced in the inducible AsiSI system and endogenous DSBs in normal human epidermal keratinocyte cells and outperformed BLESS by detecting 4.5-times more DSBs than BLESS (Lensing et al., 2016).

1.7.5.5 END-Seq

This method, although similar to DSBCapture and BLESS, avoids the creation of breaks during sample preparation by embedding the cells in agarose plugs instead of in situ fixation with formaldehyde. During validation of the technique, it could detect a higher number of reads at cleavage sites of the restriction enzyme AsiSI. It was also reported to be more sensitive for damage detection as it could detect one DSB in 10,000 cells. This technique also mapped RAG cleavage sites in pre-B cells and primary thymocytes undergoing V(D)J recombination (Canela et al., 2016). I have reviewed and compared the efficiency of this technique against other techniques including BLESS (Banerjee and Soutoglou, 2016).

Molecular Cell
Technology Preview

Finding DNA Ends within a Haystack of Chromatin

Ujjwal Banerjee [1, 2, 3, 4] and Evi Soutoglou [1, 2, 3, 4, *]
[1]Institut de Ge´ne´tique et de Biologie Mole´culaire et Cellulaire (IGBMC), 67404 Illkirch, France
[2]Institut National de la Sante´ et de la Recherche Me´dicale (INSERM), U964, 67404 Illkirch, France [3]Centre National de Recherche Scientifique (CNRS), UMR7104, 67404 Illkirch, France
[4]Universite´ de Strasbourg, 67081 Strasbourg, France
*Correspondence: evisou@igbmc.fr
http://dx.doi.org/10.1016/j.molcel.2016.08.012

Identifying DNA fragile sites is crucial to reveal hotspots of genomic rearrangements, yet their precise mapping has been a challenge. A new study in this issue of Molecular Cell (Canela et al., 2016) introduces a highly sensitive and accurate method to detect DNA breaks in vivo that can be adapted to various experimental and clinical settings.

Perhaps the most dangerous form of DNA damage is the DNA double-strand break (DSB), since both strands of the double helix are disrupted. Despite this, DSBs are also necessary intermediates in various physiological reactions, includ-ing antibody diversification and meiotic recombination. Cells have evolved robust DSB detection and repair pathways, and while these pathways are well character-ized, less is known about the chromo-somal location and end structure of DSBs. This makes it difficult to under-stand the origins of the remarkable diver-sity in mutations and structural variation that have been revealed by recent genomic studies of various cancers. In this issue of *Molecular Cell*, Canela et al. (2016) provide a comprehensive method to quantitatively determine this DSB land-scape, thereby having the potential to uncover mechanisms for these genome-wide variations.

The first DSB-detection methods quantified global DNA fragmentation by electrophoresis (e.g., pulsed field gel electrophoresis; Sutherland et al., 1987; Table 1). PCR-based methods, such as ligation-mediated PCR (Rodriguez et al., 2000), were developed to provide a nucle-otide resolution view of the lesion, but this works only if there is a priori information about the location of the damage. One of the first methods used to produce a genome-wide map of DNA damage was ''damage DNA immunoprecipitation'' or dDIP, which involved labeling of damaged DNA sites with biotinylated nucleotides by terminal deoxynucleotidyl transferase fol-lowed by immunoprecipitation and qPCR (Leduc et al., 2011). Concurrent advances in technology and decreased sequencing costs have led to a rise in next-generation sequencing (NGS)-based techniques for analyzing genome organization, and mapping protein-DNA interactions and chromatin modifications on a genome-wide level. Chromatin-immunoprecipitation followed by sequencing (ChIP-seq) or hybridization on microarrays (ChIP-on-chip) can map the genomic locations of proteins in the DDR (DNA damage response) pathway (Szilard et al., 2010; Iacovoni et al., 2010). A caveat of this approach is that DDR factor enrichment is a proxy for sites of DNA damage. Com-pounding this problem, many DDR pro-teins (such as gH2AX) spread over long stretches of DNA, so precise information about the original DNA damage site is lacking. There is also considerable varia-tion in immunoprecipitation efficiency due to variability in antibody affinities.

Consequently, two more approaches were recently developed to measure DNA breaks genome wide: GUIDE-seq (genome-wide unbiased identification of DSBs enabled by sequencing; Tsai et al., 2015) is based on the efficient inte-gration of a blunt oligonucleotide tag fol-lowed by amplification of the tag and sequencing. HTGTS (high-throughput genome-wide translocation sequencing; Chiarle et al., 2011) is based on the detec-tion of a translocation event between a nuclease-induced break and unknown break sites. Because of the repair bias, GUIDE-seq only detects the end if the introduced oligonucleotide is efficiently delivered in cells and ligated at the break site, and HTGTS is most efficient if DSBs are proximal to each other, resulting in a chromosomal translocation. Therefore, both approaches are indirect beause they rely on the repair of the ends from which the sites of DSBs are inferred.

A third method, ''direct in situ break labeling, enrichment on streptavidin and next-generation sequencing'' or BLESS (Crosetto et al., 2013) was developed to be independent of any proxies for DSBs and relies on the ligation of barcoded hairpin linker molecules directly to the break sites within fixed nuclei. Even though this technique has generated maps of aphidicolin and neocarzinostatin-sensitive regions and off-targets of CRISPR-Cas9 (Ran et al., 2015), limitations include high background and low sensitivity.

To address these problems, Canela et al. (2016) established an elegant technique enabling nucleotide resolution mapping of DSBs and end-resection (END-seq). Briefly, live cells are embedded in low-melting agarose plugs and digested with proteinase K and RNase A to remove DNA-bound proteins. Embedding live cells in agarose minimizes some of the technical issues encountered in other methodologies, including spurious gener-ation of DSBs during fixing or mechanical shearing and the use of formaldehyde, which prevents efficient adaptor ligation and possibly alters DNA end structure. After blunting and A-tailing the DNA ends, a proximal biotinylated hairpin adaptor containing a 3^0 T overhang and Illumina's p5 sequence is ligated to the break site. After melting the agarose plug, the DNA is extracted and sheared, followed by a capture of adaptor-ligated fragments on streptavidin beads.

Table 1. A Chronology of Different Methods Utilized to Identify DNA Breaks

Technique	Year	Application	Reference
Pulsed field gel electrophoresis	1987	Designed to separate bacterial and yeast chromosomes; adapted to estimate number of DNA lesions.	Sutherland et al. (1987)
Ligation-mediated PCR (LM-PCR)	2000	Mapping rare single- and double-stranded DNA breaks at single-nucleotide resolution. Works for cells with defined or known damage positions. Can be adapted for microarray and NGS approaches.	Rodriguez et al. (2000)
ChIP-Chip, ChIP-seq	2010	Mapping protein-DNA interactions of DDR proteins. Relies on proteins as a proxy for damage.	Szilard et al. (2010), Iacovoni et al. (2010)
Damage DNA immunoprecipitation (dDIP)	2011	Genome-wide mapping of DNA damage hotspots. Can be adapted for tissue samples and NGS techniques.	Leduc et al. (2011)
High-throughput genome-wide translocation sequencing (HTGTS)	2011	Identification of translocation sites and chromosomal rearrangements.	Chiarle et al. (2011)
In situ break labeling, enrichment on streptavidin and sequencing (BLESS)	2013	Genome-wide detection of DSBs. Independent of DNA end processing and uses unique barcodes to label damage site.	Crosetto et al. (2013)
Genome-wide unbiased identification of DSBs enabled by sequencing (GUIDE-seq)	2015	Identification of targets of CRISPR-Cas9 and associated off-target activity.	Tsai et al. (2015)
End-seq	2016	Genome-wide identification of DSBs. Higher sensitivity and robustness than BLESS; can identify end resection sites.	Canela et al. (2016)

Experimentally, END-seq outperformed BLESS in terms of the number of reads mapped with an average 319-fold in-crease at *AsiSI* restriction enzyme recog-nition sites. The sensitivity of the method is unprecedented: a single DSB was detectable in 10,000 cells lacking a break. As a result, END-seq is able to capture very rare ''off-target'' DSBs generated by the RAG proteins during V(D)J recombina-tion, which can initiate the development of B and T cell lymphomas. Relatedly, off-target cleavage sites for a zinc finger genome-editing nucleases were also detected by END-seq.

END-seq also provides information about how DSB ends are processed, which occurs prior to DSB repair. Resec-tion, or the lack thereof, can influence the utilization of distinct DNA repair path-ways. For example, the progressive 5^0 –3^0 resection of DSB termini is a prerequisite for homologous recombination, but this process has been very difficult to mea-sure. Canela et al. (2016) observed extensive resection in the absence of the non-homologous end-joining proteins Lig4 and 53BP1 when DSBs were pro-duced by *AsiS1* and also observed extensitve resection in ATM-deficient thy-mocytes during V(D)J recombination. It will be of great interest to utilize the method to determine how resection rates and DSB repair varies across different genomic and chromatin environments. This could have important implications for our understanding of cancer etiology and variations observed in genome-edit-ing efficiencies.

END-seq could potentially be used as a reporter to map the position in the genome of other types of non-DSB le-sions (such as nicks) by enzymatically converting them in vitro to DSBs prior to the blunting step in the END-seq protocol. More generally, DSBs could theoretically be used to map the location in the genome of any DNA-bound or chro-matin-associated protein. For instance, a protein of interest fused to a nuclease would introduce DSBs wherever the pro-tein binds. END-seq mapping of these DSBs would infer the protein's location. This is reminiscent of the ''ChEC (chro-matin endogenous cleavage)'' method developed by Laemmli and colleagues (Schmid et al., 2004). ChEC coupled to END-seq would be particularly useful in cases in which conventional ChIP is un-feasible due to antibody limitations or other technical drawbacks.

Because the method is readily appli-cable to detection of DSBs in vivo, as Canela et al. (2016) have demonstrated with primary thymocytes, END-seq can be adapted to patient tissue samples. Future applications might include identi-fying novel fragile sites in cancer cells, profiling the DSB repertoire of newly developed chemotherapeutic agents, and determining off-target activities of genome-editing enzymes such as ZFNs, TALENs, and CRISPR/Cas9. Several clin-ical trials are underway using CRISPR-Cas9-modified cells for the treatment of different cancers. DNA-edited or ex-vivo-corrected cells will be introduced back into the patient, and END-seq could be readily deployed as a specifity moni-toring tool to ensure that the intended re-gions are corrected before proceeding further. The coupling of ingenious tech-niques such as END-seq with NGS provide biologists with an exceptional tool to resolve DNA structures and pro-cesses with unprecedentred resolution and sensitivity. END-seq, therefore, rep-resents a major advance for the DNA damage field with collateral implications across a broad spectrum of biological disciplines.

REFERENCES

Canela, A., Sridharan, S., Sciascia, N., Tubbs, A., Meltzer, P., Sleckman, B.P., and Nussenzweig, A. (2016). Mol. Cell *63*, this issue, 898–911.

Chiarle, R., Zhang, Y., Frock, R.L., Lewis, S.M., Molinie, B., Ho, Y.J., Myers, D.R., Choi, V.W.,

Compagno, M., Malkin, D.J., et al. (2011). Cell *147*, 107–119.

Crosetto, N., Mitra, A., Silva, M.J., Bienko, M., Dojer, N., Wang, Q., Karaca, E., Chiarle, R., Skrzypczak, M., Ginalski, K., et al. (2013). Nat. Methods *10*, 361–365.

Iacovoni, J.S., Caron, P., Lassadi, I., Nicolas, E., Massip, L., Trouche, D., and Legube, G. (2010). EMBO J. *29*, 1446–1457.

Leduc, F., Faucher, D., Bikond Nkoma, G., Gre´goire, M.C., Arguin, M., Wellinger, R.J., and Boissonneault, G. (2011). PLoS ONE *6*, e17353.

Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S., et al. (2015). Nature *520*, 186–191.

Rodriguez, H., Akman, S.A., Holmquist, G.P., Wil- son, G.L., Driggers, W.J., and LeDoux, S.P. (2000). Methods *22*, 148–156.

Schmid, M., Durussel, T., and Laemmli, U.K. (2004). Mol. Cell *16*, 147–157.

Sutherland, J.C., Monteleone, D.C., Mugavero, J.H., and Trunk, J. (1987). Anal. Biochem. *162*, 511–520.

Szilard, R.K., Jacques, P.E., Larame´e, L., Cheng, B., Galicia, S., Bataille, A.R., Yeung, M., Mendez, M., Bergeron, M., Robert, F., and Durocher, D. (2010). Nat. Struct. Mol. Biol. *17*, 299–305.

Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., et al. (2015). Nat. Biotechnol. *33*, 187–197.

1.7.5.6 Breaks labeling in situ and sequencing (BLISS)

This technique is the most recent technique to have emerged for identifying DSBs and has made significant improvements to address all the drawbacks of previous approaches. The DNA linkers used to label DSBs contain the T7 promoter which allows in vitro transcription of labeled DSBs instead of enrichment on streptavidin. The incorporation of a unique molecular index of short random sequences in the linkers allows the quantification of the absolute number of cells accumulating a DSB at a specific position. With these improvements BLISS has drastically reduced the number of cells needed for the assay from tens of millions of cells to a few thousand cells and from tissue sections. Through this technique, it has become possible to map breaks generated by etoposide and by CRISPR/Cas9 in cells and DSBs in liver sections in untreated mice (Yan et al., 2017).

1.7.5.7 Digested genome Sequencing (Digenome-seq)

This technique was designed to map cleavage sites after digesting the genome with specific endonucleases. The cleavage sites are identified as those regions having an accumulation of reads with identical 5' ends. This technique was used to identify off-target cleavage sites of Cas9 (Kim et al., 2015) (Fig. 15).

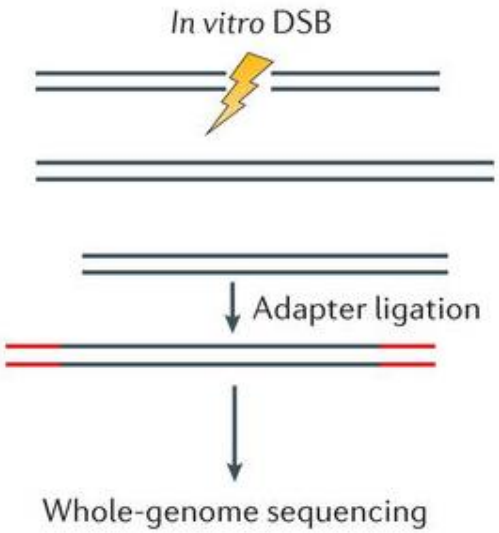


Figure 15. ***Digested genome sequencing (Digenome-seq).*** *Genomic DNA is isolated from cells and treated with Cas9 nuclease in vitro. Sequencing adapters are ligated and high-throughput sequencing is performed at standard whole-genome sequencing coverage. (Adapted from Tsai and Joung, 2016)*

1.7.6 Mapping repaired DSBs

The second category of techniques to map DSBs genome wide does not actually map the break site but follows the repair. The simplest method involves a whole genome sequencing approach at high coverage based on the assumption that some of the repaired DSBs would result in mutations. Off-target cleavage sites generated by Cas9 in stem cell clones have been detected by this method (Veres et al., 2014). The prohibitive cost of sequencing at such a high coverage coupled with the fact that the damage needs to be consistently present at the same site in a high proportion of the cells being analyzed makes this method only suitable for detecting high-frequency events in a cell population.

1.7.6.1 Genome-wide unbiased identification of DSBs enabled by sequencing (GUIDE-seq)

This technique tracks the rejoining of two broken DNA ends. It involves the transfection of blunt double-stranded DNA oligodeoxynucleotides (dsODNs), which can integrate at DSBs through cNHEJ (Tsai et al., 2014). The primers that are used to generate sequencing libraries specifically pair with the dsODNs and amplify the junction regions (Fig. 16). Since this technique requires an initial step of transfection, the length and sequence of the dsODNs are critical to ensure integration. Also, the cells need to have a proficient repair mechanism which limits the use of this technique to only some cell lines. This technique has been used to identify off-target sites of Cas9 (Tsai et al., 2015; Friedland et al., 2015).

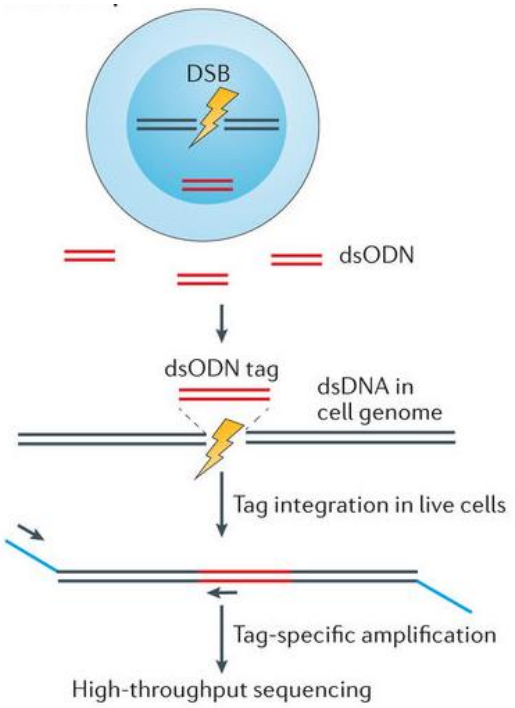


Figure 16. ***Genome-wide unbiased identification of DSBs enabled by sequencing (GUIDE-seq)****. An end-protected, short, double-stranded oligodeoxynucleotide (dsODN) is integrated into sites of nuclease-induced DSBs in living cells. This short sequence is used for tag-specific amplification followed by high-throughput sequencing to identify off-target cleavage sites. (Adapted from Tsai and Joung, 2016).*

1.7.6.2 Integrase-defective lentiviral vector capture (IDLV)

This technique is similar to GUIDE-Seq and relies on IDLVs which can enter target cells in a highly efficient manner but since they are integrase-deficient, they remain in the nuclei of target cells as episomal DNAs. This episomal DNA can be integrated into DSBs sites and are recovered by linear amplification mediated (LAM)-PCR (Fig. 17). This material is then sequenced and has been used to detect off-target sites generated by ZFNs, TALENs and Cas9 (Gabriel et al., 2011; Wang et al., 2015). This technique uses lentiviral DNA which can be used in cell lines which are difficult to transfect such as primary human cell lines. Since a large number of IDLV integrations can happen at non-relevant genomic loci, this assay requires appropriate controls for meaningful analysis.

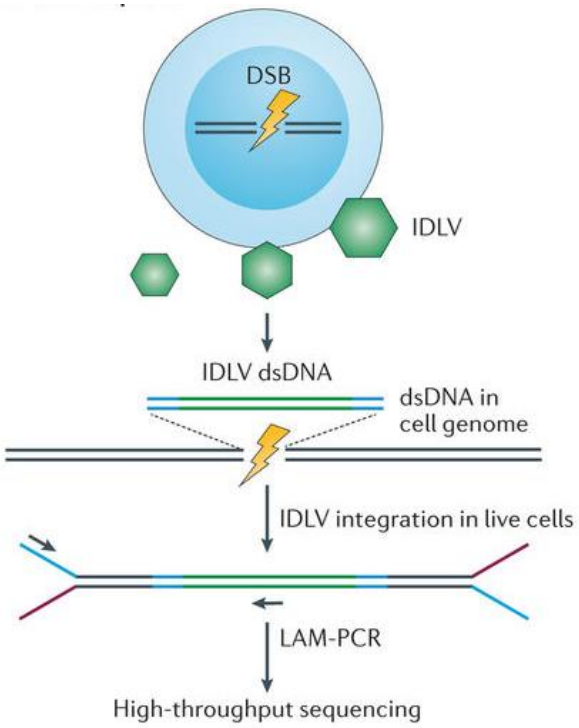


Figure 17. ***Integrase-defective lentiviral vector (IDLV) capture****. IDLVs (green) are integrated with a selectable marker into sites of nuclease-induced DSBs in living cells. Integration sites are recovered by linear amplification-mediated PCR (LAM-PCR), followed by high-throughput sequencing. (Adapted from Tsai and Joung, 2016).*

1.7.6.3 High-throughput genome-wide translocation sequencing (HTGTS) and linear amplified-mediated HTGTS (LAM-HTGTS)

These methods of tracking DSBs relied on the occurrence of translocation events after the completion of error-prone repair. A bait DSB which is produced endogenously or as a result of drug treatment serves as a translocation site for other target DSBs which may also be produced endogenously or due to drug treatment. Genomic DNA is then purified, sheared and sequencing libraries are prepared after PCR amplification (Chiarle et al., 2011; Hu et al., 2016). It has been used to study programmed DSBs in developing and mature lymphoid cells (Chiarle et al., 2011; Meng et al., 2014; Dong et al., 2015), recurrent DSB cluster generation in neural stem/progenitor cells and transcription associated breaks (Schwer et al., 2016) and off-target sites of Cas9 (Frock et al., 2014). Since translocations occur when DSBs are in close spatial proximity, HTGTS is suited for studying intrachromosomal translocations (Zhang et al., 2012). The major drawback of this approach is the rarity of a translocation event which increases the genome coverage needed during sequencing to detect such events.

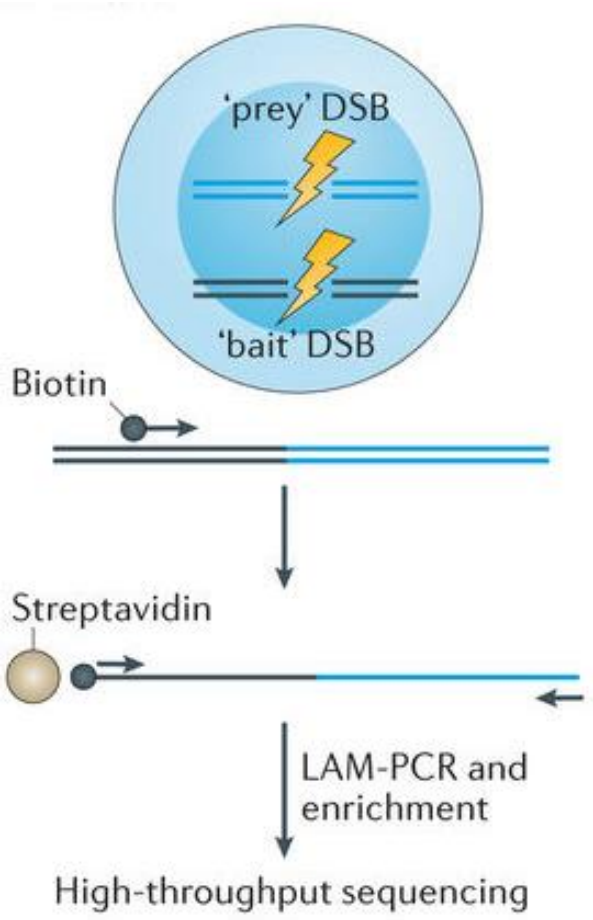


Figure 18. ***High-throughput genome-wide translocation sequencing (HTGTS)****. Two nucleases are expressed in a cell to generate a 'prey' and 'bait' DSB. Using a biotinylated primer designed against the bait DSB junction, translocations between prey and bait are recovered by LAM-PCR and streptavidin-based enrichment for high-throughput sequencing. Off-target cleavage sites are identified by analysis of these translocation junctions (Adapted from Tsai and Joung, 2016)*

1.7.6.4 Translocation Capture sequencing (TC-Seq)

This method was independently developed from HTGTS and was used to study AID-dependent chromosomal rearrangement in B cells (Klein et al., 2011; Oliveira et al., 2012). This technique is not suitable for single-nucleotide resolution mapping of translocation sites.

Techniques to study DSB occurrences and map regions of genomic instability have rapidly improved and will continue to do so. However, they are beset by some common drawbacks the most common of which is the high-coverage needed to obtain consistet results. Also since most of these techniques need a large number of cells, only extremely recurrent regions are mapped. Randomly generated breaks will result in a background level of signal indistinguishable from noise and low frequency recurrent DSBs will be masked. Given the non-random nature of chromatin

organization in the nucleus, it may be possible to have a higher incidence of DSBs and/or repair in specific nuclear compartments.

1.8 Chromosomal fragile sites and 53BP1

Increase in the rate of mutation accumulation is associated with an alteration of genetic factors such as genes involved in the sensing and repair of DNA damage. This increase in mutation rate can be due to a variety of reasons including exposure to carcinogens and genetic predisposition to tumorigenesis caused by an inheritance of altered genes. The ultimate outcome of this increase in mutation rate is an increase in the rate of tumorigenesis. Different loci in the human genome have varying degrees of susceptibility or response to mutagenesis. Some sites, referred to as fragile sites, are significantly more susceptible to certain types of DNA damage. These sites are particularly sensitive to replication stress and exhibit gaps, constrictions, or breaks on metaphase chromosomes when exposed to certain culture conditions. Over 120 fragile sites have been identified in the human genome (Shwartz et al., 2006). These fragile sites may be classified as common and rare fragile sites depending on their population frequency, inheritance pattern and method of induction (Durkin and Glover, 2007).

1.8.1 Rare fragile sites

These sites are observed in only a small proportion of the population (<5%) and segregate in a Mendelian fashion (Lukusa and Fryns, 2008; Sutherland and Richards, 1995). These sites can be induced by folate/thymidylate stress, distamycin A and bromodeoxyuridine (BrdU), which block DNA synthesis (Lukusa and Fryns, 2008). Analysis of replication timing of folate-sensitive rare fragile sites has shown that these regions replicate very late in the S-phase of untreated cells which is further delayed to G2 under thymidylate stress (Hansen et al., 1997; Subramanian et al., 1996). The rare fragile site FRAXA is associated with fragile X syndrome which causes severe mental retardation (Verkerk et al., 1991). FRAXE, another rare fragile site, is responsible for non-specific mental retardation (Gu et al., 1996).

1.8.2 Common fragile sites (CFS)

Unlike rare fragile sites, common fragile sites (CFS) are loci with recurrent breaks which are visible under conditions of replicative stress. Unlike rare fragile sites, CFSs are present in all individuals and constitute a component of normal chromosome

structure (Durkin and Glover, 2007; Freudenreich, 2007). Some CFSs can be specifically induced to breakage by treatment with aphidicolin (APH) a DNA polymerase inhibitor. CFSs are characterized by late replication during the cell cycle. They are maintained in different species suggesting a biological role. Investigation into the occurrence of CFSs has shown that they play a key role in chromosome stability and genome dynamics. They are associated with sister chromatid exchange hotspots, act as sites of deletion, amplification and translocations in various cancers, and as sites for viral integration (Glover and Stein, 1987; Arlt et al. 2006; Durkin et al. 2008; Burrow et al. 2011; Bester et al. 2006; Dall et al. 2008). They have also been shown to be sites of structural variation in stem cells (Hussein et al. 2011).

In humans, CFSs have been mapped in lymphocytes, fibroblasts, epithelial colon cells, breast cancer cells and erythroid cells. Different CFS maps show that these sites are tissue specific and not all of thses sites having the same frequency of breaks with some exhibiting a higher breaking frequency than others. CFSs are significantly associated with large genes (> 300 kb). This is true for the sites FRA3B and FRA16D, which contain the large genes FHIT (1.5 Mb) and WWOX (1.1 Mb) respectively.

1.8.2.1 Conservation of CFS during evolution.

CFSs have been conserved during evolution and human orthologs of CFS have been reported in other primates (Durkin and Glover, 2007), as well as in other species such as cats, dogs, pigs, horses, cattle, rats and mice (Elder and Robinson, 1989). Human orthologs of CFSs have been identified in mouse lymphocytes: Fra14A2, (FRA3B) (Glover et al., 1998), Fra8E1, (FRA16D) (Krummel et al., 2002), Fra6C1, (FRA4F) (Rozier et al., 2004), Fra12C1, (FRA7K), Fra2D, (FRA2G), Fra6A3, (FRA7G), Fra6B1, (FRA7H), Fra4C2, (FRA9E) (Helmrich et al., 2006). Fragility of these regions may be due to the conservation of the large genes within these sites. This has indeed been reported in the case of murine Fhit and Wwox gene orthologs (Krummel et al., 2002). Moreover, in *S. cerevisae*, it was also shown that DSBs occur recurrently at specific regions that have been called replication slow zones which can be considered as analogous to CFSs (Cha and Kleckner, 2002). A conservation of these sites suggests a biological role and function common to these sites. It was reported that these areas were retained for promoting homologous recombination (Glover and Stein, 1987). It was also reported that the replication of CFSs is a signal that marks the end of the S-phase and passage into mitosis (Debatisse et al., 2006).

1.8.2.2 Late refplication of CFS

Studies on the temporal program of replication have linked fragility with late replication. It has been reported that the CFS FRA3B in human lymphocytes is replicated late during the S-phase and its replication is again delayed by replication stress induced by aphidicolin with 16.5% of FRA3B sites not being able to complete replication (Le Beau et al., 1998). Later it was also shown that both FHIT alleles in human lymphocytes do not have the same timing of replication and the latest allele is the most fragile allele (Wang et al., 1999). Similarly, it was also shown that the replication timing of FRA7H is specific for each allele and that aphidicololin treatment delays the completion of the replication of this region relative to reference regions (Hellman et al., 2000). Further studies on the replication timing of WWOX on cells sorted according to different phases of the cell cycle showed that FRA16D also replicates later in human lymphocytes (Palakodeti et al., 2004). A study also showed that the early condensation of DNA in cells in phase G2, induced by calyculin A treatment, induces breaks in CFSs indicating the late replication of these sites (El Achkar et al., 2005). More recently, data obtained by Repli-Seq highlighted replication timing of CFSs in a comprehensive manner (Hansen et al., 2010). Chromosomal breakage has therefore been attributed to an entry into mitosis while the replication of these sites is not completed.

A new category of fragile sites called Early Replicating Fragile sites (ERFS) have also been described, which unlike CFSs, replicate early (Barlow et al., 2013). These sites were identified by Chromatin immunoprecipitation of regions enriched in RPA protein in murine B lymphocytes blocked at the beginning of the S-phase by high doses of hydroxyurea, which stalls replication forks which ultimately collapse leading to the formation of a DSB. RPA protein covers ssDNA at the replication forks. ERFS are preferentially located in regions rich in highly transcribed genes. The instability of the ERFS has been attributed to conflicts between the machinery of transcription and replication forks at these sites. These ERFS have been implicated in diffuse large B-cell lymphoma with more than 50% of amplifications and deletions occurring at ERFS (Lenz et al., 2008).

1.8.2.3 CFS are tissue specific

It was initially believed that CFS characterized in human lymphocytes were universal, i.e., they were present in all human cell types. It was later reported that FRA3B and

FRA16D, which were fragile in lymphocytes are not fragile in fibroblasts suggesting that CFSs are tissue specific and their fragility also depended on epigenetic traits (Letessier et al., 2011). Further exhaustive analyses of human fibroblast line (MRC5) (Le Tallec et al., 2011), three colon cancer epithelial cell lines (HCT116, LoVo, LS174T), two normal epithelial cell lines (MCF10A), tumour cell line (CAL51) and an erythroid cell line (K562) (Le Tallec et al., 2013) have shown that two different cell types share less than 20% of their CFS, and even if a CFS is shared between two cell types, frequency of breaks is not the same between the two. These results indicate that there is a set of chromosomal regions that can become fragile in a given cell type according to the epigenetic context. This raises the question of the existence of one or more common characteristics common to these loci that predispose them to become fragile. It also raises the question of the existence and localization of CFSs in stem cells and their fate during differentiation.

1.8.2.4 CFS contain large genes

It has been observed that CFSs are generally associated with large genes covering several hundred kb (Helmrich et al., 2006; Helmrich et al., 2011). It has been confirmed that 80-100% of human CFS and 100% of murine CFS harbour a large gene (> 300 kb) (Le Tallec et al., 2013). Since the average size of human genes is 20kb, these genes are 15 times larger. It may be possible that part of the CFS not associated with long genes does not become fragile.

1.8.2.5 Involvement of CFS in cancer and other diseases

Genetic instability is a well-established hallmark of tumour cells. This instability can be classified into several categories. The most common form of instability in human cancers is chromosomal instability manifested by large deletions or amplifications (from a few hundred kb to several Mb) initiated by DSBs and is very predominant in sporadic cancers. Other forms of instability include the accumulation of point mutations or microdeletions and microduplications which are common in hereditary cancers.

Many studies show that CFSs are preferred regions for chromosome rearrangements in a variety of cancers including esophageal cancer, gastric cancer (Lee et al., 2001), lung cancer (Sozzi et al., 1996), lymphomas (Kameoka et al., 2004), breast cancer (Negrini et al., 1996), pancreatic cancer (Shridhar et al., 1996) and renal cancer (Ohta

et al., 1996). These rearrangements mainly correspond to deletions and often lead to the inactivation of the gene which is within the CFS. CFSs have been reported as being sites for preferential breaks during chromosome amplification (Coquelle et al., 1997). The human CMYC oncogene is located between two fragile sites (FRA8C and FRA8D) and is amplified in many cancers (Ferber et al., 2003, Ferber et al., 2004). CFSs are also preferential sites of viral DNA integration which is involved in the development of some cancers. This has been reported in the case of HPV16 papillomavirus where the viral DNA is integrated in CFSs in more than 50% of the cases of cervical cancer (Thorland et al., 2003).

An additional relationship between CFSs and cancer is the presence of oncosupressors in CFSs (Drusco et al., 2011). The FHIT and WWOX genes, located in FRA3B and FRA16D respectively, have been reported to be frequently altered in several types of cancers (Bignell et al., 2010). However, it is unknown if this is a consequence of instability due to the CFS or if their inactivation has actively participated in tumour progression. Mutations at the level of onco-suppressive genes or oncogenes are initiating mutations that offer a selective advantage to the cell. On the other hand, deletions at CFSs and large genes contained within them would be rather frequent but passive alterations, expressing an instability of these regions in the tumoural cells and which would be preserved because of the clonal nature of the tumors (Dereli-Oz et al., 2011). The model of oncogenes-induced replicative stress provides an explanation for the accumulation of these deletions (Negrini et al., 2010). In this model, activation of an oncogene induces uncontrolled proliferation and replicative stress. Since CFSs are particularly sensitive to this type of stress, they would be more frequently affected than other areas of the genome. Replicative stress would therefore induce genetic instability, a property common to all cancer cells.

1.8.2.6 Oncogene activation and replication stress

Two mechanisms have been proposed for the onset of replicative stress by oncogene activation. Activation of an oncogene leads to phosphorylation of the retinoblastoma (Rb) protein in the absence of external mitogenic stimulus and activation of E2F transcription, causing forced passage of S-phase cells. This would occur without the cell having had time to accumulate enough precursors nucleotides and create endogenous replicative stress. Replication stress could also result from a disturbed loading and/or triggering of the origins of replication. The loading of origins onto pre-

replication complex (pre-RC) normally takes place only in the G1 phase when Cyclin Dependent Kinases (CDKs) are not active. Activation of origins during S-phase requires active CDKs. Oncogenes and oncosuppressive genes control the activity of CDKs and could thus disrupt the loading and triggering of origins (Macheret and Halazonetis, 2015).

1.8.2.7 Models to explain fragility

The mechanisms responsible for fragility of CFS has been the subject of intense debate. Three models have been proposed for the presence of CFSs. The first model attributes the fragility of CFSs to the presence of AT rich sequences which tend to form secondary structures and block replication fork progression (Durkin and Glover, 2007). The second model proposes that CFSs are regions which replicate late and have low replication origin sites (Letessier et al., 2011). The third model associates the fragility of CFSs with the collision of replication forks and transcription complexes in large genes (Helmrich et al., 2011).

A. Fragility due to the sequence

According to this model, the fragility of CFS is thought to be due to the intrinsic characteristics of their sequences that would promote blocking of replication forks under replicative stress conditions (Durkin and Glover, 2007). This is based on the following observations:

i. The sequences of some CFSs are rich in AT. This is true for FRA3B (Boldog et al., 1997), FRA7H (Mishmar et al., 1998), FRA16D (Ried et al., 2000), FRAXB (Arlt et al., 2002). These sequences would have the ability to form secondary structures that would block the progression of the replication forks.

ii. Torsion angle measurements of DNA sequences at some CFSs show that their flexibility is high compared to non-fragile control sequences (Shah et al., 2010; Zhang and Freudenreich, 2007). It is the case for FRA2G (Limongi et al., 2003), FRA3B (Mimori et al., 1999), FRAXB (Arlt et al., 2002), FRA7H (Mishmar et al., 1998), FRA8C (Ferber et al., 2004), and FRA16D (Ried et al., 2000). These regions may act as sinks for the superhelical density generated ahead of the replication fork, hindering efficient topoisomerase activity and decreasing the processivity of the polymerase complex (Zlotorynski et al., 2003).

iii. The deletion in two tumor lineages of a 500kb region including AT-rich sequences removes the fragility of the FRAXB site (Arlt et al., 2002). In support of this model, it has been reported that the ectopic integration of part of the FRA3B sequence is accompanied by an increase in the fragility of the insertion region (Ragland et al., 2008). Other studies have also shown that the number of forks blocked at AT-rich sequences of FRA16C (Ozeri-Galai et al., 2011) or FRA16D (Zhang and Freudenreich, 2007) increases under replicative stress.

However, arguments have been advanced against this model. Other studies failed to demonstrate an ability of the AT-rich sequences to form secondary structures in comparison to non-fragile regions having the same base composition (Helmrich et al., 2006). In addition, it has been shown that deletion of AT-rich sequences in FRA3B (Corbin et al., 2002) or FRA16D (Finnis et al., 2005) does not diminish their fragility. The study of replication dynamics at FRA3B (Letessier et al., 2011) and FRA6E (Palumbo et al., 2010) could not highlight fork stalling at a significantly higher level at these sites. Finally, the tissue specificity of fragile sites points against the fact that fragility is dictated by the sequence of DNA.

B. Late completion of replication

This second model was proposed based on the analysis of replication timing of the FRA3B site by DNA-combing coupled with FISH in human lymphoblastoid and fibroblast cells (Letessier et al., 2011). FRA3B is a major CFS in lymphocytes but not in fibroblasts. Different parameters of replication dynamics including speed of replication, mapping of initiation and termination of replication and fork stalling were studied in these two cell types under normal culture conditions and under conditions of replication stress. The results obtained have shown that the replication rate of FHIT is the same as the rest of the genome and there is no increase in forks blocked at FHIT compared to the rest of the genome under replicative stress. These results do not agree with the first model presented above which predicts slowdowns and/or fork stalling at the fragile region.

Replication origin mapping shows that, in lymphocytes, the FHIT gene has a core region of 700 kb which is poor in replication origins and which is replicated by forks progressing from the two flanking regions. This is not the case in fibroblasts

where the core region does not exist and a normal density of replication origins has been demonstrated throughout the gene. Under replicative stress conditions, termination events are excluded from the core region in lymphocytes whereas they are present throughout the gene in fibroblasts. As already mentioned, the completion of FRA3B replication occurs late in S-phase, or even in G2. The model proposes that fragility is due to the presence of a region poor in replication initiation events that is replicated by forks progressing from adjacent regions. When these forks are slowed down by replication stress, they cannot finish replication of this region, which causes fragility. In support of this, Repli-seq analysis of two major CFSs (FRA1L and FRA3L) in fibroblasts also showed the same replication profile as FRA3B in lymphocytes, with a core region replicated in late S-phase by replication forks progressing from regions bordering the site. In lymphocytes, where these regions are not fragile, such a core region does not exist (Le Tallec et al., 2011). So, fragility can be attributed to a lack of replication origins and late completion of replication. Since replication timing and replication origins are demarcated during cell differentiation by epigenetic mechanisms, this model also explains why fragility of different loci varies according to the cell type.

C. Replication versus transcription

DNA serves as a template for replication and transcription. In order to avoid collisions between the transcription machinery and the replication forks, these two processes usually take place separately in time and space (Aguilera and Garcia-Muse, 2013). Unlike prokaryotes, DNA polymerases and RNA polymerases advance at comparable rates in eukaryotes: 17-33 nt/s for the replication fork and 17-72 nt/s for the transcription machinery, which limits the probability of intersections between the two. One of the mechanisms further reducing this probability is the existence of break points of the replication fork in the genome. These pause points result from the existence of barriers preventing the progression of replication forks through highly transcribed sequences such as genes that code for ribosomal RNAs (Lopez-Estrano et al., 1998). There is also a temporal regulation of replication with transcribed genes being replicated early during the S phase (Ryba et al., 2010). Another form of regulation concerns multi-copy genes. The different copies do not have the same regulatory sequences and transcription

of some copies can continue while other copies are replicated. This is true for genes encoding histone H4 (Holmes et al., 2005). Nevertheless, despite these regulation systems, multiple studies show that conflicts between transcription and replication exist in vivo and that the absence of their resolution leads to DNA damage and DSBs (Kim and Jinks-Robertson, 2012, Tuduri et al., 2009).

The third model proposed to explain the mechanism of CFS instability is the collisions between replication forks slowed by replicative stress and the transcription machinery (Helmrich et al., 2011). As has already been mentioned, a large majority of CFSs harbour large genes. If the transcription of these genes plays a role in instability of the CFSs, this would explain the tissue specificity of the latter. This report showed a correlation between the level of mRNA of large genes (CNTNAP2, DMD, FHIT, WWOX, IMMP2L) in two cell types (lymphocytes and myoblasts) and the level of fragility of the corresponding CFS. It was further reported that transcription of these large genes, because of their size, lasts more than one cell cycle and so must lead to collisions between the replication forks and the transcription machinery at these sites. These collisions would lead to the formation of DNA-RNA hybrid structures called R-loops. It has been reported that such structures may be responsible for genetic instability (Lin and Pasero, 2012). The elimination of these structures by RNase H reduces the fragility of CFSs induced by aphidicolin. It was theorized that the slowing down of the replication fork by replication stress would lead to the formation and stabilization of R-loops which would disrupt the progression of the replication forks and hence induce the CFS to breakage (Helmrich et al., 2011).

However, it was reported that there is no correlation between the level of mRNA of large genes and the fragility of CFSs (Le Tallec et al., 2013). Notably, there are large genes transcribed at a high level that are stable and do not correspond to CFSs. The collision model does not seem to explain the mechanism of fragility by itself. In addition, if R-loops are frequently formed at CFS under replicative stress, this should result in a very significant increase in the number of blocked forks. However, the study of the FRA3B replication dynamics in lymphocytes and the FHIT gene expression, have shown that there is no increase in forks blocked at this site compared to the rest of the genome (Letessier et al., 2011).

1.8.2.8 Mapping fragile sites and recurrent double strand breaks

Advances in DSB mapping techniques (covered in section 1.7.4) have identified the recurrent DSB landscape in different cell types. HTGTS has been used to study the occurrence of DSBs in developing and mature lymphoid cells (Chiarle et al., 2011; Dong et al., 2015; Meng et al., 2014) and the formation of recurrent DSB clusters in neural progenitor cells and transcription-associated breaks (Schwer et al., 2016). DSB clusters identified in neural progenitor cells specifically have characteristics similar to CFSs, i.e., they fall in long genes which have a late replicating time (Wei et al., 2016). BLESS has been used to study replication-stress induced DSB formation after aphidicolin treatment, and has shown a clear correlation between gene length and sensitivity to aphidicolin (Crosetto et al., 2013).

1.8.2.9 Repair at CFS

It has been shown that under replication stress, CFSs colocalize with γH2AX, RAD51 and DNA-PK foci (Schwartz et al., 2006) suggesting that DSBs are formed at CFSs. The activation of the DNA damage response has multiple consequences: the stabilization of the replication forks, restarting stalled forks, firing of additional origins of replication, activation of the various repair routes as well as cell-cycle arrest. A significant number of factors in this response have been implicated in the protection of CFSs. It has been shown that the depletion of these factors increases the instability of CFS. This is true for ATR (Casper et al., 2002), CHK1 (Durkin et al., 2006), BRCA1 (Arlt et al., 2004), Rad51, Claspin (Focarelli et al. 2009), FANCD2 (Howlett et al. 2005), BLM (Naim et al. 2013) and others. However, in order to correctly interpret these data, it is necessary to bear in mind that the lack or inhibition of some of these factors by themselves induce replication stress.

Specifically, it has been shown that BLM helps maintain the pyrimidine precursor pool in cells and that its depletion reduces the rate of replication (Chabosseau et al., 2011). It has also been shown that depletion of CHK1 (Koundrioukoff et al., 2013; Petermann et al., 2006), RAD51 (Daboussi et al., 2008) or Claspin (Petermann et al., 2008) decreases the speed of replication and therefore their effects on CFS instability can be explained, at least in part, by the replication stress induced by their inhibition. On the other hand, depletion of ATR affects the stability of the CFSs in addition to its effect via speed reduction (Koundrioukoff et al., 2013).

Given the danger that unreplicated regions and DSBs represent for the cell, the question arises that under mild replication stress induced by a low dose of aphidicolin that causes CFSs to break, the damage response is not activated and the cell enters mitosis with unreplicated sites? On the one hand, it is possible that as in yeast, the DNA damage response does not occur when a fraction of the genome is unreplicated (Torres-Rosell et al., 2007). Additionally, several results strongly suggest that lesions resulting from under-replication can be supported by the cell until the next cell-cycle.

Breaks are created mechanically at the intermediate structures resulting from resolution of under-replicated regions by specific nucleases. It has been shown that the endonucleases MUS81/EME1 and ERCC1 which participate in the resolution of inappropriate DNA structures such as stalled forks or Holliday junctions, are recruited at the CFSs at the end of G2-phase and at the beginning of mitosis. Depletion of these nucleases leads to a decrease in breaks at the CFSs suggesting that they are actively involved in creating breaks (Naim et al., 2013, Ying et al., 2013). It has been shown that FANCD2 and FANCI are recruited at the CFSs and that if the aberrant structures resulting from under-replication are not resolved by nucleases, interchromosomal bridges are formed during anaphase. These bridges are protected by the helicase BLM in combination with topoisomerase IIIα, RMI1 and RMI2 factors (RecQ mediated genome instability 1 and 2) and PICH (PLK1 interacting checkpoint helicase) while the extremities of the bridge are associated with FANCD2 and FANCI factors (Chan et al., 2009, Naim and Rosselli, 2009) which allow chromosomal segregation to take place. Incomplete resolution of these anomalous structures, downregulation of BLM or PICH enhances the formation of 53BP1 nuclear bodies in G1 daughter cells. This suggests that breaks in CFSs can be repaired in the next G1-phase (Fig. 19).

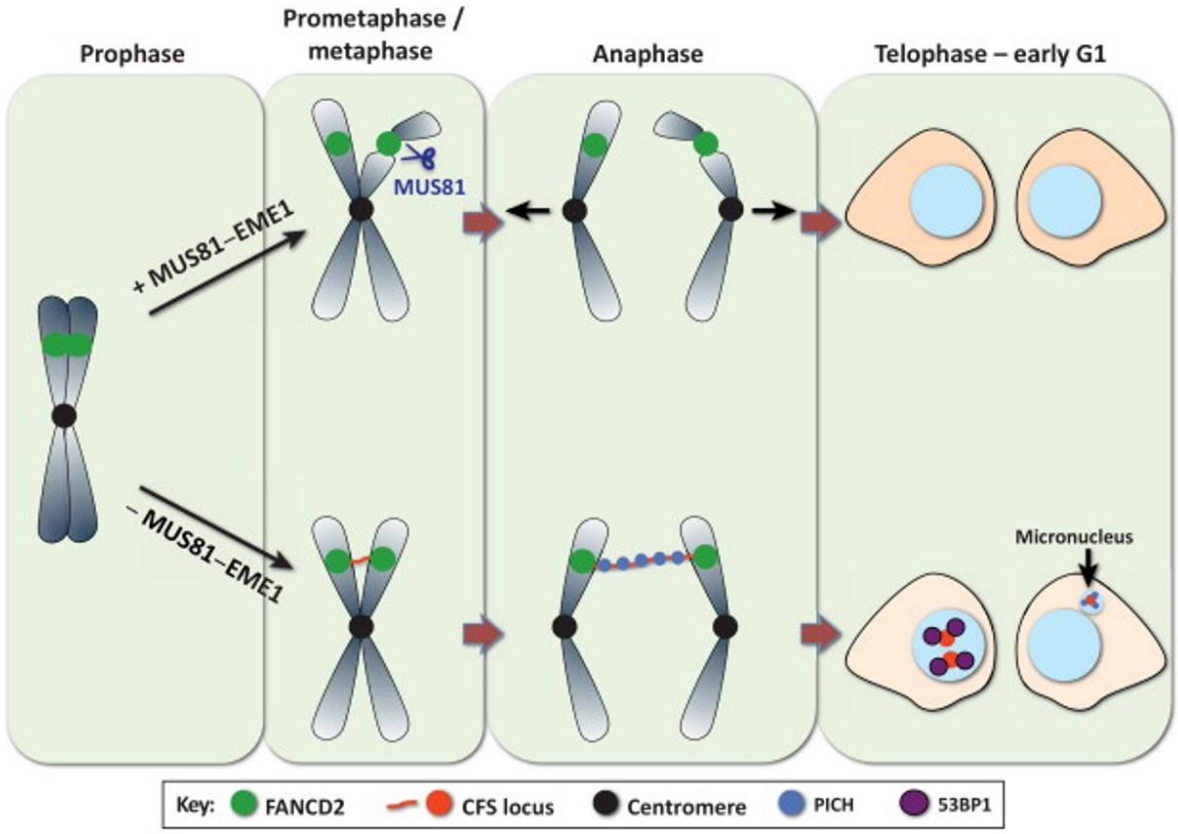


Figure 19. ***Model representing the molecular events leading up to common fragile site expression.*** *MUS81-dependent cleavage is required for the generation of breaks or gaps at common fragile site (CFS) loci, which are marked by the recruitment of a FANCD2 focus on each segregating sister chromatid in mitosis. FANCD2 foci associate with CFS loci throughout G2 and mitosis. The upper panel (+MUS81–EME1) shows normal cells, and the lower panel (−MUS81–EME1) shows the situation in cells lacking MUS81–EME1. Depletion of MUS81 results in a reduction in the incidence of breaks or gaps at CFS loci, promoting the occurrence of CFS-associated sister chromatid nondisjunction. CFS loci that are not cleaved by MUS81 in early mitosis persist as bulky anaphase bridges or ultra-fine anaphase DNA bridges (UFBs) marked by FANCD2 foci at their termini, which are then processed by the helicase BLM and the PICH (Plk1-interacting checkpoint helicase) translocase in anaphase. This permits cell division to take place, albeit with structural abnormalities that manifest as an increased frequency of CFS-associated, PICH-positive micronuclei in G1-daughter cells. In addition, the DNA repair factor 53BP1 forms nuclear bodies in G1 cells, potentially shielding CFS regions. (Minocherhomji and Hickson, 2014).*

1.9 DNA damage effector: 53BP1

The NHEJ protein 53BP1 is a large 1972 aa long protein which was first characterized as a binding partner of the tumor suppressor protein p53 (Iwabuchi et al., 1994). This interaction is dependent on the tandem BRCT domain at the carboxy-terminus of 53BP1 with the DNA binding domain of p53 and promotes the transcriptional activity of p53 (Iwabuchi et al., 1998). 53BP1 also contains an oligomerisation domain and a tandem Tudor domain, upstream of the BRCT domains, which are required for its relocalization at DSBs (Botuyan et al., 2006; Huyen et al., 2004; Zgheib et al., 2009). The Tudor domain spanning amino acids 1486-1540 (human 53BP1) has a high degree of conservation throughout evolution and orthologues Crb2 and Rad9 are present in

S.pombe and S.cerevisiae respectively (Huyen et al., 2004). Multiple SQ/TQ phosphorylation sites are present on the amino-terminus of 53BP1 (Fig. 19).

Upon damage induction, 53BP1 is hyper-phosphorylated by the PIKK proteins ATM, ATR and possibly DNA-PK (Jowsey et al., 2007; Ward et al., 2003a). It has been shown that this hyper-phosphorylation is not required for 53BP1 re-localization to DNA damage sites (Ward et al., 2003a). Under normal culture conditions, a few but particularly bright foci called "53BP1 nuclear bodies" have been observed (Lukas et al., 2011). Upon damage induction, 53BP1 is redistributed at damage sites and colocalizes with several other repair factors such as the MRN complex, γH2AX, MDC1 and BRCA1 (Bekker-Jensen et al., 2005; Schultz et al., 2000; Stewart et al., 2003; Wang et al., 2002).

53BP1 recruitment involves two distinct components: MDC1 signaling cascade involving RNF8 and RNF168 (Bohgaki et al., 2011; Stewart et al., 2003); and the association of the Tudor domain with the histone modification H4K20me2 and/or H3K79me2 (Botuyan et al., 2006; Huyen et al., 2004). The identity of the histone interaction partners of 53BP1 is controversial. Structural studies have shown that the Tudor domains interact specifically with H3K79me2 (Huyen et al., 2004). This is supported by studies where mutations to the Tudor domain abolished 53BP1 foci formation. However, subsequent studies involving the deletion of Dot1, which is responsible for the dimethylation of H3K79, did not impair foci formation (Botuyan et al., 2006). An alternate interaction partner, H4K20me2 was identified with H4K20me1 having a lower binding affinity (Botuyan et al., 2006). Dimethylation of these residues is catalyzed by the histone methyltransferases SUV4-20H1 and SUV4-20H2 but a double knockout of these enzymes did not affect 53BP1 foci formation except at very early timepoints (Schotta et al., 2008). It was also reported that recruitment of 53BP1 at the foci was dependent on MMSET mediated methylation of H4K20 and downregulation of this histone methyltransferase reduces H4K20 methylation and subsequent recruitment of 53BP1 (Pei et al., 2011). This interaction could control the repair pathway choice in different cell cycle phases. Possible reasons for these observations could include a complementary role of H3K79me2 and H4K20me2 in binding the Tudor domain to chromatin or the presence of a hitherto unknown histone modification which acts as the binding partner for the Tudor domain. The minimal region required for focus formation spans aa 1220-1711 and includes the oligomerization domain, the tandem Tudor domains and the nuclear localization

sequence (Iwabuchi et al., 2003; Ward et al., 2003). Overexpression of the minimal region reduces HR (Xie et al., 2007).

53BP1 interacts with RNF168 ubiquitin chains and is highly selective for H2AK15ub through its ubiquitylation-dependent recruitment motif which is adjacent to the Tudor domain (Fradet-Turcotte et al., 2013). Mutations in this domain impair the formation of 53BP1 foci and prevent its interaction with H2AK15ub but not H4K20me2. This indicates that 53BP1 binding to H4K20me2 is independent of RNF168. This observation proves that the revelation of H4K20me2 for interaction with 53BP1 occurs in an RNF8-UBCH8 dependent manner (Fradet-Turcotte et al., 2013; Mallette et al., 2012; Acs et al., 2011). Studies through quantitative chemical proteomics have shown that BRCT repeats interact with γH2AX and the interactions of 53BP1 with H4K20me2 and H2AK15ub were weak interactions (Kleiner et al., 2015). From these reports it can be theorized that the localization of 53BP1 to damaged sites depends on multiple interactions with no single target or mark providing sufficient interaction for stable interaction. It has been reported that 53BP1 facilitates ATM phosphorylation of various substrates in DDR including CHK2 (Wang et al., 2002; Ward et al., 2003b), SMC1, RPA2 and BRCA1 (Wang et al., 2002), reflecting its role as a mediator of the ATM signaling response.

Deletion of 53BP1 in mice increases IR-sensitivity of the mice and makes them tumour prone (Morales et al., 2003; Ward et al., 2003). They also exhibit reduced isotype switching in mature B cells, revealing a defect in class switch recombination and defects in V(D)J recombination (Manis et al., 2004; Morales et al., 2003; Ward et al., 2003). Studies of 53BP1 at deprotected telomeres have shown that it may be involved in promoting telomere fusions (Dimitrova et al., 2008). 53BP1 has also been identified as having a role in DNA end resection during HR (Bunting et al., 2010; Dimitrova et al., 2008; Zimmermann et al., 2013). During DDR, BRCA1 and 53BP1 are competing to either promote or inhibit end resection. Deletion of 53BP1 in BRCA1 deficient cells rescues the HR defect (Bunting et al., 2010).

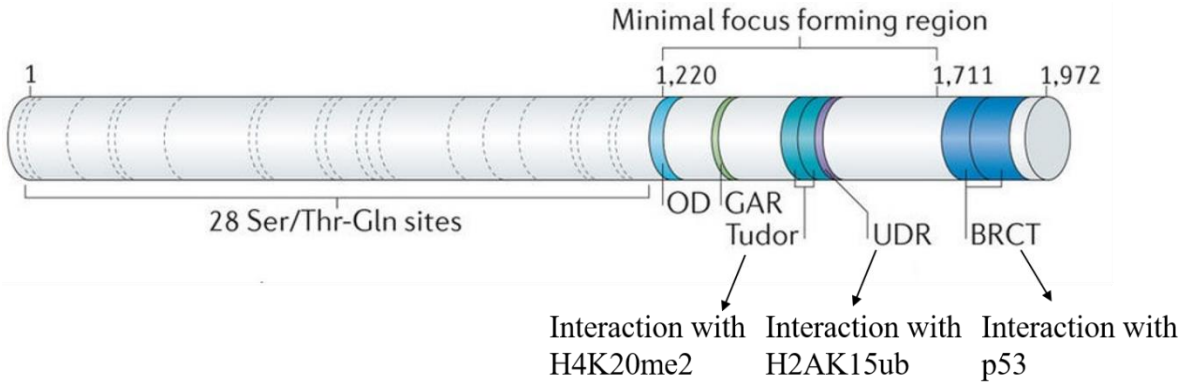


Figure 20. Domains of p53-binding protein 1 (53BP1). 53BP1 *contains 28 Ser/Thr-Gln (S/T-Q) sites in its amino terminus that match the target sequence of the ataxia-telangiectasia mutated (ATM) kinase. ATM-mediated phosphorylation of the 53BP1 N terminus promotes the binding of the effector RIF1. The 53BP1 carboxyl terminus contains tandem BRCT (BRCA1 carboxy-terminal) domains that binds to p53. The minimal focus-forming region in 53BP1 contains an oligomerization domain (OD), a Glycine and Arginine rich (GAR) motif, a tandem Tudor motif that binds to H4K20me2 and a ubiquitylation-dependent recruitment (UDR) motif that interacts with H2AK15ub. Amino acid positions are indicated. (Adapted from Panier and Boulton, 2013).(Panier and Boulton, 2013)*

1.9.1 Association of 53BP1 with chromosomal fragile sites

The accumulation of endogenously occurring DSBs in untreated cells has been observed in laboratory cell cultures for many years. These breaks have been well characterized by microscopy and are known to accumulate DSB response factors such as MDC1, 53BP1 and γH2AX. It has been proposed that unresolved replication intermediates produced during normal replication and under replication stress are converted to DSBs during mitotic progression. This conversion can be due to aberrant resolution of ultra-fine chromatin bridges or due to breaking during chromosome hypercondensation (Gelot et al., 2015; Liu et al., 2014; Ying et al., 2013; Naim et al., 2013). These lesions have been shown to be shielded by 53BP1-Nuclear bodies (53BP1-NBs) (Harrigan et al., 2011, Lukas et al., 2011) (Fig. 20). 53BP1-NBs are ATM and γH2AX dependent and are mainly present in cells in G1 and disappear during S-phase. These foci have been shown to be associated with CFSs that are difficult to replicate and their incidence increases after a mild replication stress. 53BP1-NBs were proposed to protect fragile sites and shield unrepaired DNA lesions left over from the previous cell cycle from erosion and sequester them until the next cell cycle for repair (Lukas et al., 2011). It may be possible that these 53BP1-NBs arise before S/G2 phase and progressively diminish as the majority of S phase damage is repaired. The remaining foci may be present at sites containing late replicating or unreplicated DNA which escape detection by cell cycle checkpoints. The escape from cell cycle checkpoints is not surprising since unreplicated DNA in itself is not an

aberrant DNA structure (Branzei and Foiani, 2010). Recently, it was reported that there exists an inverse relation between replication status and 53BP1 which is bound to H4K20me2 in a replication dependent manner and this interaction decreases as the cell progresses through the S-phase (Pellegrino et al., 2017). This interaction could control the repair pathway choice in different cell cycle phases. It is unclear if 53BP1-NBs remain at the fragile sites at mitosis or come off and are reloaded onto the CFSs at the next cell cycle since no foci are observed during mitosis. If cells are continuously exposed to sub-lethal levels of replication stress, it is possible for cells to be transformed and become cancerous. This has been observed in human cells with early stages of cancer arising as a direct consequence of improper DNA replication.

An important question which arises from these observations is that even though fragile sites have been associated with increasing the chances for mutations and in the development of cancer, they have still been maintained in the genome instead of being eliminated through evolution. In other words, what incentive does the cell have for maintaining a fragile site, given its predisposition to replication stress induced DNA damage.

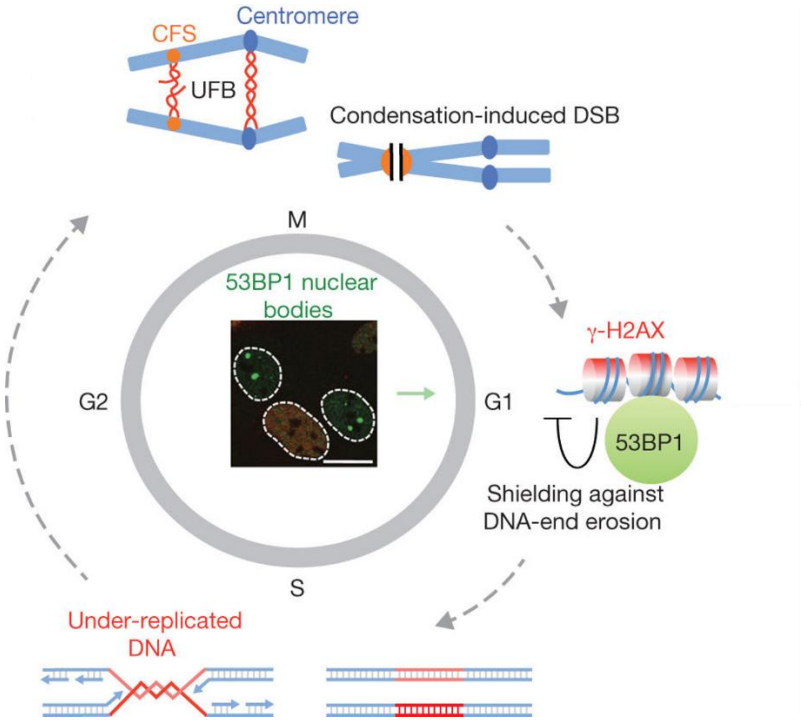


Figure 21. ***53BP1 nuclear bodies****. Due to a scarcity of replication origins at loci called common fragile sites (CFS), some regions of the genome might remain underreplicated. The DNA within such regions is converted to gaps or breaks through chromosome condensation or dissolution of ultrafine bridges (UFB), when the chromosomes enter mitosis.. A fraction of such lesions can be transmitted to daughter cells, where they are sequestered in large chromatin domains enriched in 53BP1 and other markers associated with the DNA-damage-modified chromatin. These chromatin domains have been proposed to shield the DNA lesions against adverse erosion by cellular nucleases and thus protect such loci until repair mechanisms become available. Inset Image shows 53BP1 nuclear bodies (green) in a pair of daughter cells shortly after cell division and one cell in G2 (Cyclin A in red) devoid of 53BP1 nuclear bodies . Scale bar, 10 μm. (Adapted from Lukas et al., 2011a; Lukas et al., 2011b).*

THESIS OBJECTIVES

3-D Genome Organization of DNA damage repair

Inaccurate DNA repair causes mutations, chromosomal rearrangements and alteration of gene expression that predispose the cell to cancer and aging. This emphasizes the importance of the study of DDR to identify novel potential therapeutic targets for cancer therapy. As mentioned previously, the eukaryotic nucleus is organized into distinct subnuclear compartments which segregate distinct nuclear functions (Misteli and Soutoglou, 2009). The non-random organization of the nucleus raises questions of whether the occurrence of DNA lesions and their subsequent repair maybe influenced by the compartmentalization of the nucleus. Observations in yeast have suggested that DNA repair centres exist at preferential sites of repair (Lisby et al., 2003). It has been shown that DSBs in yeast are mobile which facilitates homology search during HR (Mine-Hattab and Rothstein, 2012; Dion et al., 2012). Although mammalian cell DSBs possess limited mobility (Misteli and Soutoglou, 2009), DSBs have been shown to cluster into larger repair units, both in euchromatin and heterochromatin (Aymard et al., 2017). These clusters could indicate halted HR until cells enter the S-phase.

The objective of my first project was to investigate the influence of 3D genome organization on DNA damage repair and the genome-wide relationship of epigenetic features with factors involved in DNA damage repair. I used two genome-wide strategies to follow the induction of breaks (BLESS) and follow the repair kinetics (γH2AX ChIP-Seq). This was done by the following approaches:

i. Genome-wide investigation of double strand break distribution under physiological conditions and in response to radiomimetic drug treatment in mouse embryonic stem cells.
ii. Correlation of induced DNA lesions and genomic features in a genome-wide manner.
iii. Analysis of DNA damage distribution and repair kinetics in response to global DNA damage.

The questions that I have tried to answer include:

i. Are certain regions more susceptible to damage i.e. does damage occur at specific loci in the genome?
ii. Does damage persist in specific nuclear compartments? If yes, then in which compartments?
iii. What is the kinetics of repair?

The data from ChIP-Seq and BLESS were analysed and correlated with publicly available data to answer these questions and will be further validated by other experiments.

Elucidation of the genomic binding targets of 53BP1 by BioChIP

53BP1 is a key DNA damage repair protein recruited at DSB sites and has been observed as large nuclear bodies in a sub-population of cells without any exogenous stress. These 53BP1 nuclear bodies (53BP1-NBs) are cell cycle dependent and have only been observed in G0/G1 cells. 53BP1-NBs have a key role as shielding factors against endogenous DNA damage. Due to their unique cell-cycle dependent recruitment and appearance, they are believed to represent a link between stress encountered by a mother cell and resultant signalling in daughter cells. Common Fragile sites (CFS) are chromosomal regions which are prone to breakage when exposed to a mild replication stress. Chromatin immuno-precipitation (ChIP) experiments, after a mild replication stress induced by a low dose of aphidicolin, have shown an enrichment of 53BP1 at CFS sites such as FRA3B and FRA16D (Lukas et al., 2011). A similar study on G0-arrested cells (BJ fibroblasts) also showed an enrichment of γH2AX on many CFS (Harrigan et al., 2011). 53BP1-NBs are believed to facilitate repair of those fragile sites that failed to complete DNA replication in the previous cell cycle. Despite the importance of 53BP1-NBs at fragile sites, only a few of the 53BP1-NB shielded loci have been described by ChIP-qPCR and specific sites bound by 53BP1 have not been identified.

The aim of my second project was to identify and characterize novel genomic loci shielded by 53BP1-NBs. Previous attempts to study 53BP1-NBs were hampered by the lack of a highly specific antibody which could be well suited for these purposes. This drawback was circumvented by using the Bio-ID system which allows efficient isolation and identification through affinity purification (Roux et al., 2012). To study the genome-wide footprint of 53BP1, a plasmid containing the Tudor domain of 53BP1 (minimal functional region of 53BP1 composed of amino acids 1220-1771) fused with BioTag (a small 23 a.a. peptide serving as substrate for the *E.coli* biotin ligase enzyme BirA) was generated. The plasmid was transfected into U2OS cells to raise stable cell lines.

This model system was used to identify 53BP1-NB shielded loci in asynchronous cells and in cells arrested at the G1 stage of the cell cycle. The specific questions that I have tried to answer in this project include:

i. Where are these loci located and whether they span any annotated CFS?
ii. Do these loci have any similarities with CFS, such as late-replication or spanning long genes?
iii. What is the transcription of these loci?
iv. Do these loci cluster together?

RESULTS

3-D Genome Organization of DNA damage repair

DNA double strand breaks are cytotoxic forms of DNA damage which need to be efficiently repaired to avoid mutations and the onset of cancer. Given the increasing amount of evidence regarding the role of chromatin organization in regulating DNA damage repair, it is important to establish the functional relationship between genome organization and DNA repair pathways.

Mouse embryonic stem (mES) cells were chosen for this project because of their unique chromatin organization. In embryonic cells, the chromatin is globally decondensed and enriched in active histone marks (Meshorer and Misteli, 2006). During differentiation, the cell responds to internal and external cues which are responsible for adding repressive marks to genes responsible for maintaining pluripotency (Mohn and Schuebler, 2009). ES cells can be coaxed to differentiate into the cell type of choice by the application of proper differentiating factors such as retinoic acid (Bibel et al., 2008), thus making them an ideal model system to investigate the DDR mechanism and its relationship with epigenomic features. The same genomic locus can be studied in ES cells (with an open chromatin structure) or in differentiated cells (with repressive marks added to some genes). Given the non-random organization of the genome, I explored the role of chromatin organization in regulating the DNA damage repair kinetics in mES cells after global induction of DNA damage.

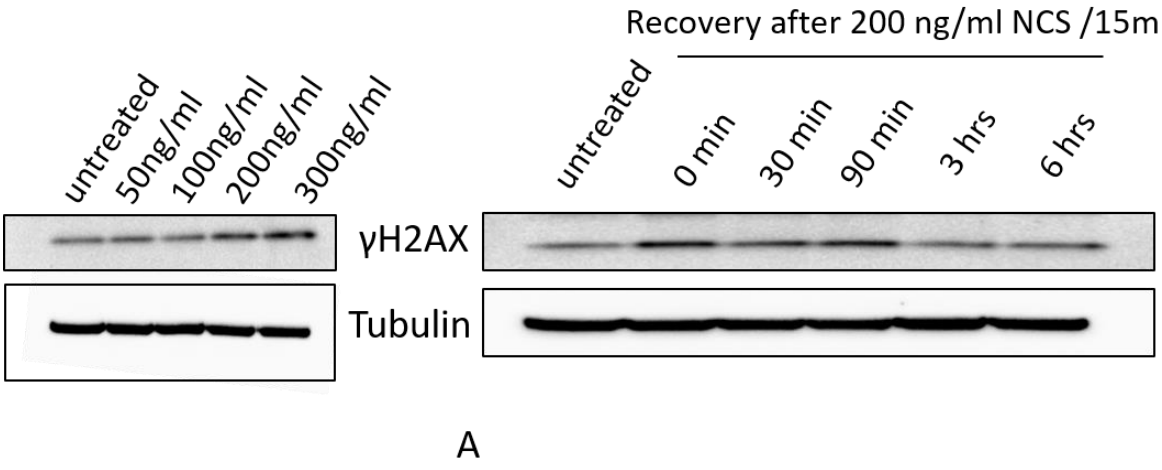
With improvements in Next-generation sequencing technologies, it is possible to sequence whole genomes in a single experiment. This approach allows us to study the interaction of chromatin with several factors in parallel including DDR factors. Using two parallel sequencing strategies, I have investigated the DNA damage response (ChIP-Seq) and the localization of damage (BLESS) in mouse embryonic stem cells exposed to global DNA damage.

3.1 Experimental setup and validation

To induce a global DNA damage response, we did a titration of the concentration of the radiomimetic drug neocarzinostatin (NCS) which would give the maximum induction from which the cells would recover. We used the 200ng/ml dose for further experiments (Fig. 22A left panel). 200ng/ml of NCS was used to induce a global DNA damage response at early (0m, 30m of recovery), middle (90m, 3h of recovery) and late (6h of recovery) time-points after drug treatment. To quantify the DDR, Western Blot analysis was performed on mouse embryonic

(mES) cell extracts against γH2AX. The γH2AX was significantly increased after drug treatment and damage persisted until the 3h timepoint before returning to basal levels. γH2AX could also be detected in the untreated sample which can be attributed to endogenous unrepaired DSBs or damage due to replication stress (Fig. 22A right panel).

To validate that the drug-treatment did not induce cell death during the experiment which would affect further analysis, apoptosis analysis was performed. No significant variation in the number of viable cells was observed (Fig. 22B, quantification at right). However, it was observed that mES cells enter a G2 phase cell cycle arrest after drug treatment (Fig. 22C, quantification at right). The cells recovered from the cell cycle arrest and continued to proliferate after the repair was completed.



A

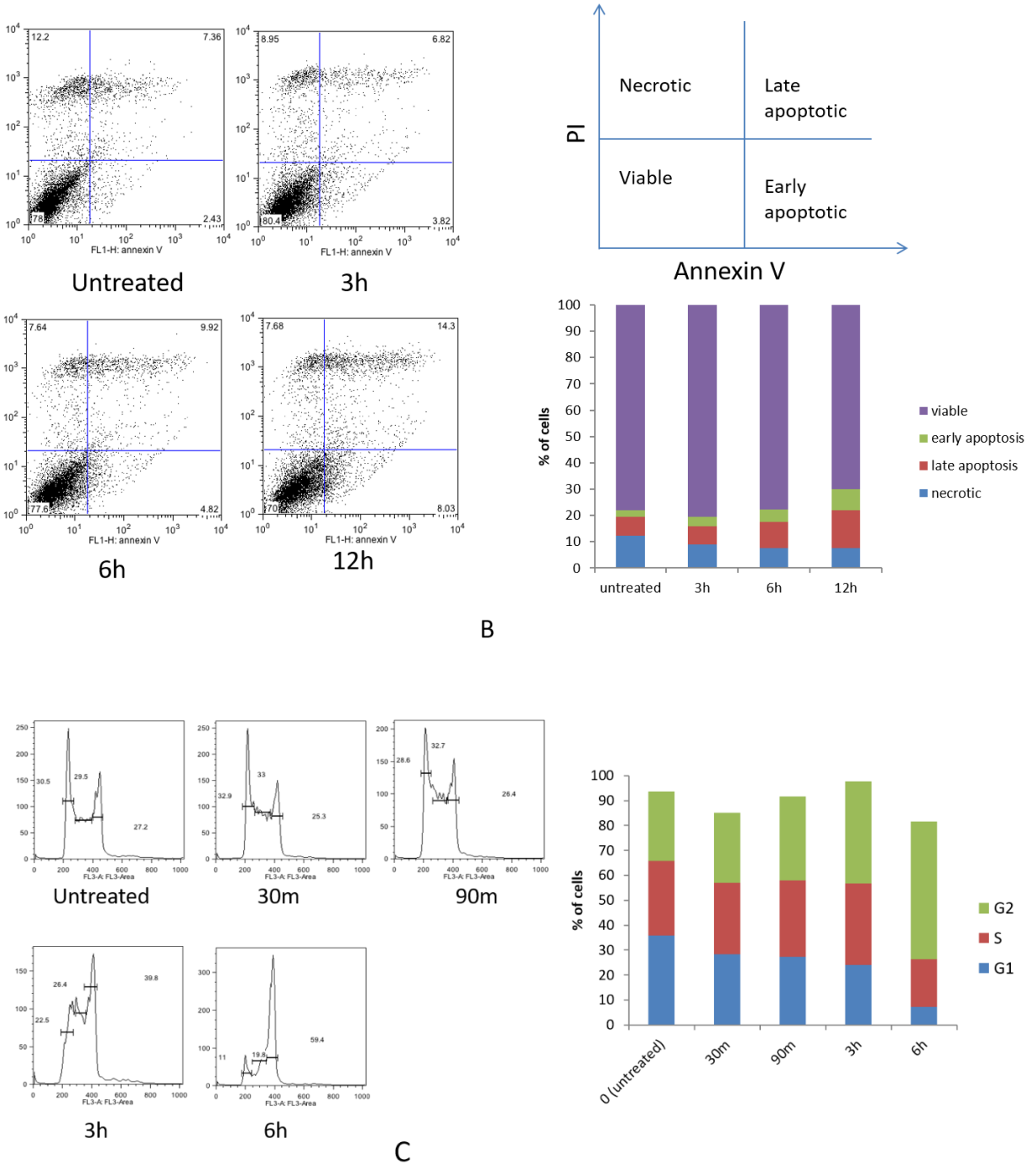


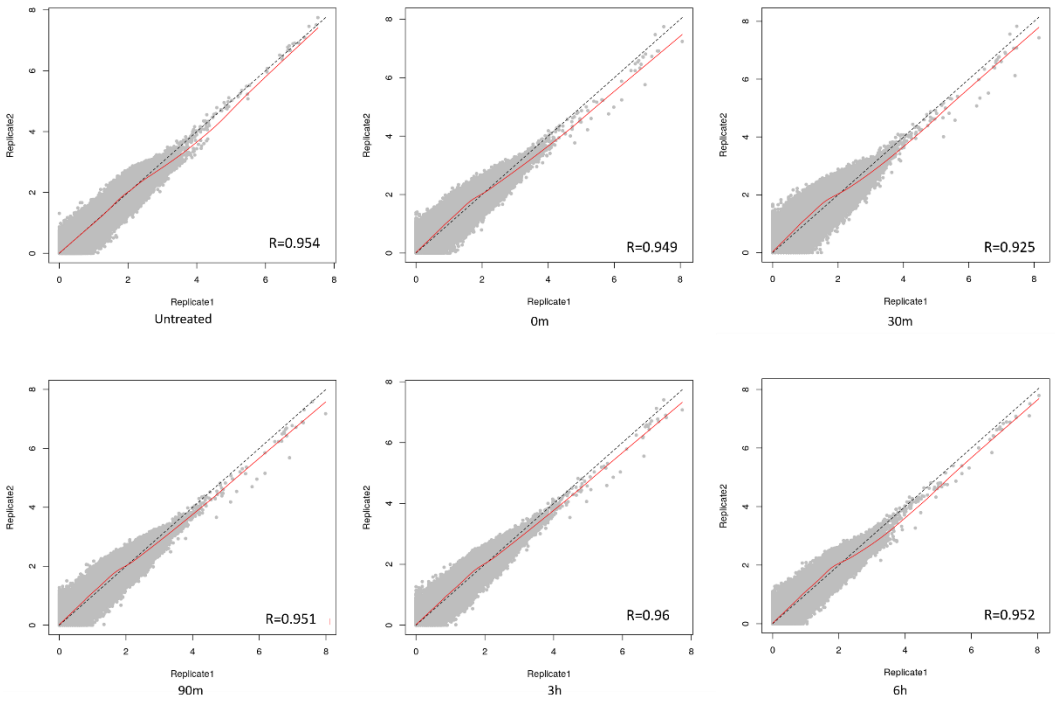
Figure 22 ***Experimental setup and validation****. A. Western Blot analysis of DNA damage response, (left) dosage titration of Neocarzinostatin (NCS), (right) analysis of γH2AX levels after treatment with 200ng/ml of NCS at indicated timepoints; B. Apoptosis analysis of treated cells 3h, 6h and 12h after treatment with 200ng/ml of NCS; quantification at right. C. Cell cycle analysis of treated cells at indicated timepoints; quantification at right.*

It has been reported that γH2AX has a broad enrichment pattern around a break site (Nakamura et al., 2010; Iacovoni et al., 2010; Aymard et al., 2014). Peak calling packages such as MACS (Zhang et al., 2008) cannot correctly identify a broad region of enrichment since they are designed specifically for identifying transcription factor localization and instead will give a series of peaks falling within a broad enrichment region which need to be manually checked for actual enrichment. To identify broad patterns genome-wide, a custom script was written which identified regions of enrichment. The genome was divided into 10kbp bins resulting in 265502 bins for the mouse genome. To be comparable across samples, RPKM (reads mapped per kb per million base pairs) was calculated for each of these 10kbp bins (Mortazavi et al., 2008). Enrichment index was calculated as [(ChIP RPKM $_{interval}$/ Input RPKM $_{interval}$) - (ChIP RPKM $_{Average}$/ Input RPKM $_{Average}$)]. Positive values represent enrichment while negative values indicate depletion at the bin. This enrichment was plotted out to identify overall enrichment and depletion. Plots were smoothed by LOESS (Locally weighted scatter-plot smoother) (span=.025) (Natale et al., 2017).

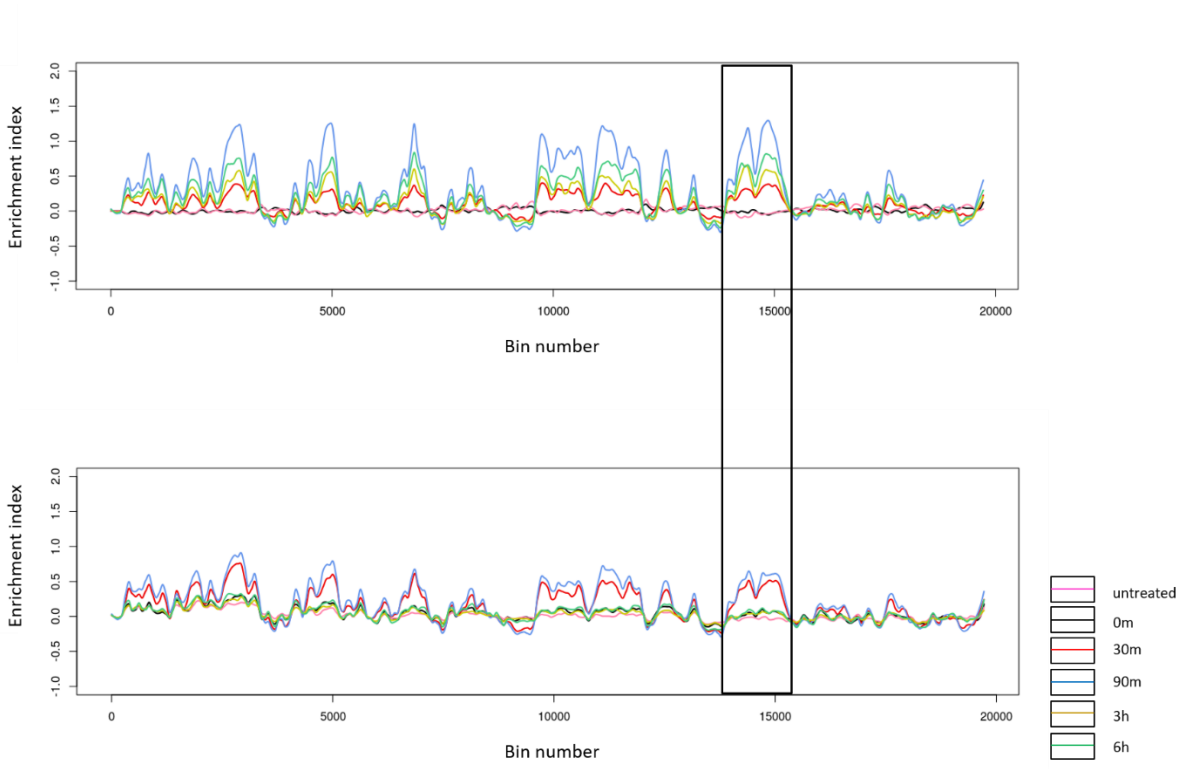
3.2 γH2AX enrichment during DDR occurs at defined loci in the mouse genome

Prior immunofluorescence experiments have shown that γH2AX occurs as distinct foci which localize to the lesion after damage induction. We wanted to identify the genomic location of such foci and whether the distinct foci that are observed in microscopy experiments can also be observed through ChIP-Seq. We noticed a correlation between the enrichment indices of the two replicates at similar time points (Fig. 23A). We investigated whether this correlation did translate to the formation of domains on the genome. A distinctive γH2AX enrichment pattern was observed at specific loci in the mouse genome, indicating that the damage and corresponding repair occurs non-randomly (Fig. 23B). The enrichment domains show a prominent level of similarity across replicates and can be considered as damage repair centers in the genome. It was also observed that the enrichment was detectable after 30 minutes of recovery. This could be due to a lag between damage sensing and actual repair factor recruitment to the break site. Generally, a higher global enrichment score was observed at the early-and middle-timepoints with maximum enrichment occurring at 90 minutes post-treatment in the first replicate and 30 minutes in the second replicate (Fig. 23C). This variation in the enrichment and the timepoint at which

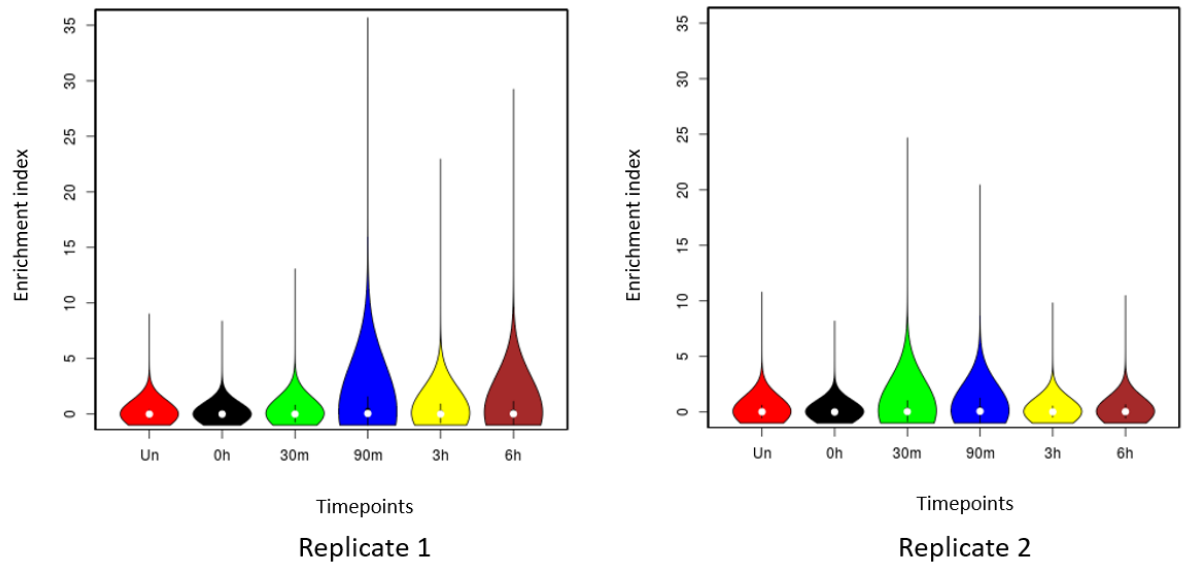
maximal enrichment occurs may be due to batch variations in the NCS used. To study the kinetics of the repair, differences in the enrichment index between every timepoint and the preceding timepoint were calculated (for example, index at 30m-index at 0m) (Fig. 23D). The maximum amount of damage repair occurred at the 90m to 3h interval, although the repair domains remain discernible at the 3h timepoint. Interestingly, our experiments have shown that at the initial timepoint when domains were detected, the domain borders had already been demarcated and although enrichment did increase and then decrease during the experiment, no domain expansion was detected (Fig. 23B). Since domain expansion is a very early and quick process, it is believed that the domain expansion around the lesion occurred between the 0m and 30m timepoints and could not be captured.



A


Enrichment index
Bin number
untreated
0m
30m
90m
3h
6h

B


Enrichment index
Un
0h
30m
90m
3h
6h
Timepoints
Replicate 1
Replicate 2

C

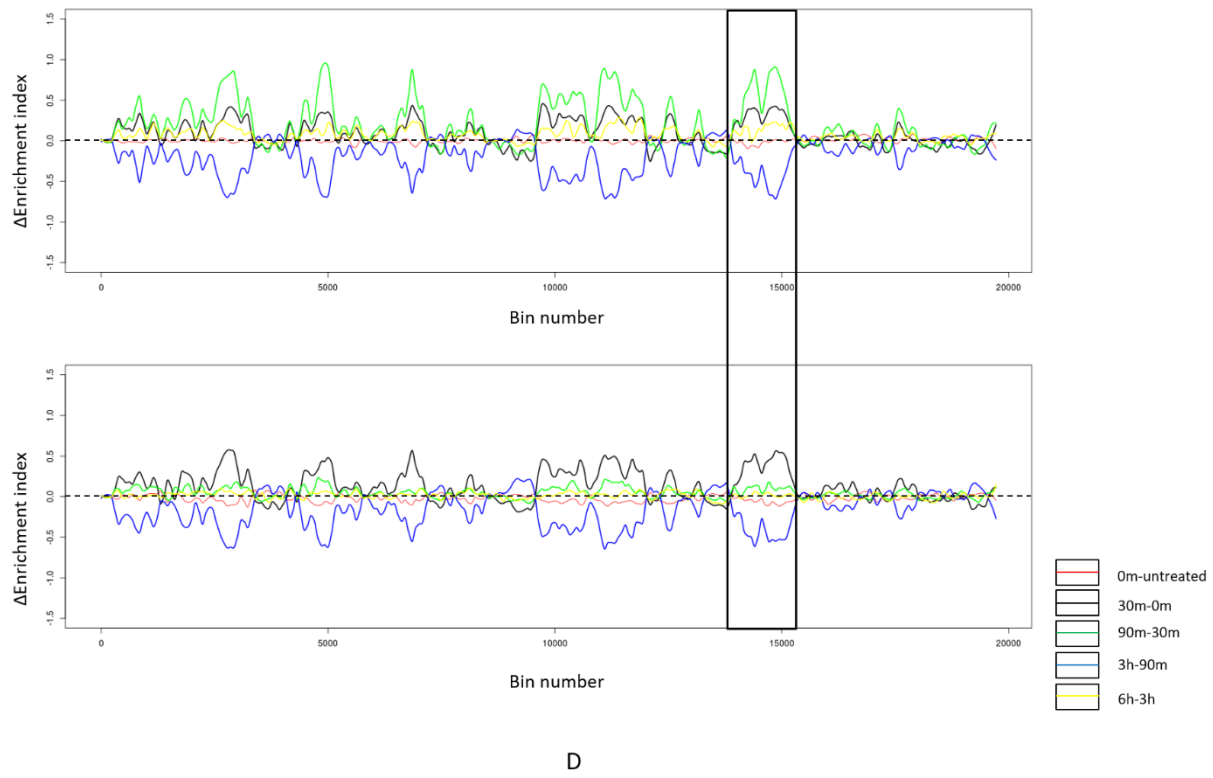


Figure 23. ***γH2AX enrichment occurs at defined genomic loci****. A. Correlation plots between replicates; X-axis-replicate 1, Y-axis-replicate 2; Pearson correlation coefficient values inset; broken black line is diagonal; red line indicates best fit curve. B. Chromosome-wide plot of chromosome 1, replicate 1 (top) and replicate 2 (bottom), box indicates the presence of a γH2AX domain. C. Quantification of γH2AX enrichment at indicated timepoints, (left) replicate 1, (right) replicate 2. D. Difference plot for chromosome 1. Difference in enrichment index calculated between every timepoint and preceding timepoint as indicated in plot key; negative values indicate repair has occurred; positive value indicates further recruitment of γH2AX; top panel- replicate 1, bottom panel- replicate 2; box indicates one repair domain.*

3.3 γH2AX enrichment occurs preferentially at regions devoid of H3K36me3

Given the distinctive pattern of γH2AX enrichment, we investigated if these domains of γH2AX correspond to specific epigenomic features. Since the genome can be broadly divided into two compartments, euchromatin and heterochromatin, which play important roles in regulating DDR, we compared our ChIP-Seq data with publicly available data to correlate repair with nuclear compartments. The histone mark H3K36me3 is associated with transcriptional activation and open chromatin, and is considered as a marker for euchromatin while H3K9me3 and H3K27me3 are associated with constitutive and facultative heterochromatin respectively. Histone modification data for mES cells raised in serum was retrieved from publicly available datasets for identifying these regions (Marks et al., 2012). RPKM enrichments of these datasets were calculated and compared with γH2AX ChIP-Seq data. It was observed that γH2AX domains

were recruited specifically at regions devoid of H3K36me3 (Fig. 24; Annex). In mES cells, a substantial proportion of the genome can be considered as open transcriptionally active euchromatin with repressive marks being added as the cell differentiates (Zhu et al., 2013). This was observed during the analysis as fewer regions with heterochromatin peaks were seen. The heterochromatin peaks which were observed did show an accumulation of γH2AX.

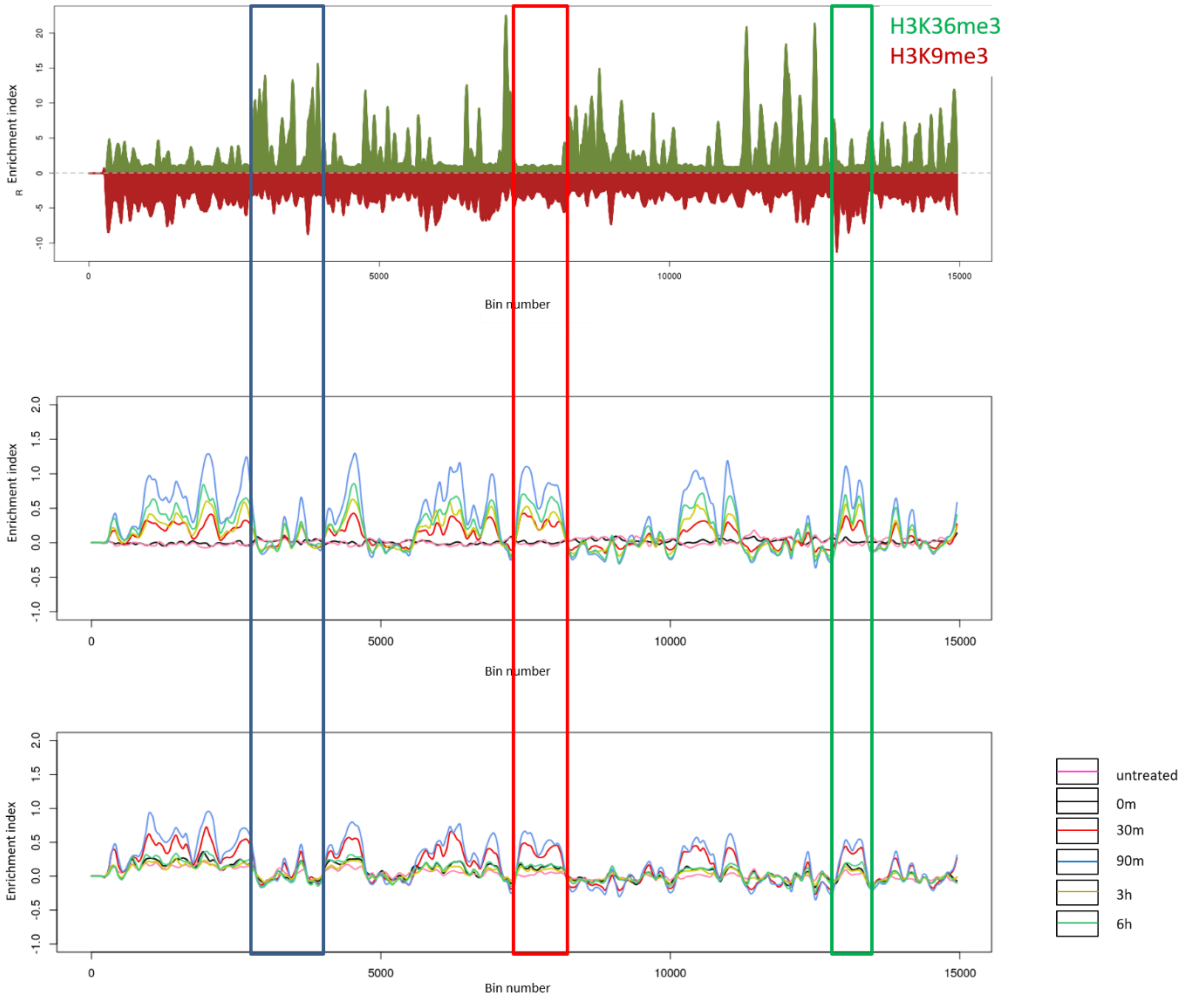


Figure 24. ***Correlation with histone marks****. Chromosome-wide plot of chromosome 6. Top panel: H3K36me3 (green) and H3K9me3 (red) enrichment values; Middle panel: replicate 1; Bottom panel: replicate 2. Blue box shows lack of enrichment of γH2AX at an H3K36me3 rich locus (euchromatin); red box shows enrichment of γH2AX at a region devoid of H3K36me3; green box shows enrichment of γH2AX at an H3K9me3 rich locus (heterochromatin).*

3.4 Endogenous DNA damage in mES cells

Since it was difficult to identify peak regions which would demarcate heterochromatin boundaries, an alternate strategy was adopted to identify such regions in mES. A cutoff value

which signified a high confidence of enrichment over background was calculated for H3K36me3 (euchromatin) and H3K9me3 (constitutive heterochromatin) 10kb bins. Through this approach, 26861 euchromatin and 23467 heterochromatin non-overlapping bins were identified resulting in ~20% genome coverage (Fig. 26A). The γH2AX enrichment index was then plotted as a heat-map for all bins and the specific bins for euchromatin and heterochromatin were used for analysis. From this approach, it was seen that in the untreated cells, the endogenous damage was majorly located in regions which could not be called either euchromatin or heterochromatin by our approach (Fig. 25A). Quantification of endogenous damage in the bins which we had labeled did show an enrichment in the heterochromatin (Fig. 26B). A similar approach was also applied for H3K27me3 but since it was not possible to have a high genome coverage with bins identified for this mark, only the global effect could be studied. This analysis also showed a similar tendency with endogenous damage being more persistent in heterochromatin (Fig. 25B).

3.5 DNA damage sensing and repair is faster in euchromatin than heterochromatin

Based on the approach from the previous analysis, we wanted to address how the kinetics of repair processes is affected by chromatin organization. It was observed that after an initial lag between damage sensing and repair factor recruitment, most of the bins had an enrichment of γH2AX at the timepoint where maximal enrichment occurred regardless of whether they belonged to the euchromatin compartment or heterochromatin compartment. The reduction in γH2AX enrichment occurred faster in the euchromatin compartment than in heterochromatin (Fig. 26C, D).

3.6 DNA damage persists in compacted transcriptionally inactive chromatin

Enrichment of γH2AX in bins falling in regions with a low enrichment of euchromatin mark, but high enrichment of heterochromatin marks indicates that γH2AX persistence occurs in transcriptionally inactive chromatin (Fig. 26B). This adds to earlier reports and strengthens our hypothesis that damage persists in the heterochromatin compartment (Kim et al., 2007; Goodarzi et al., 2010).

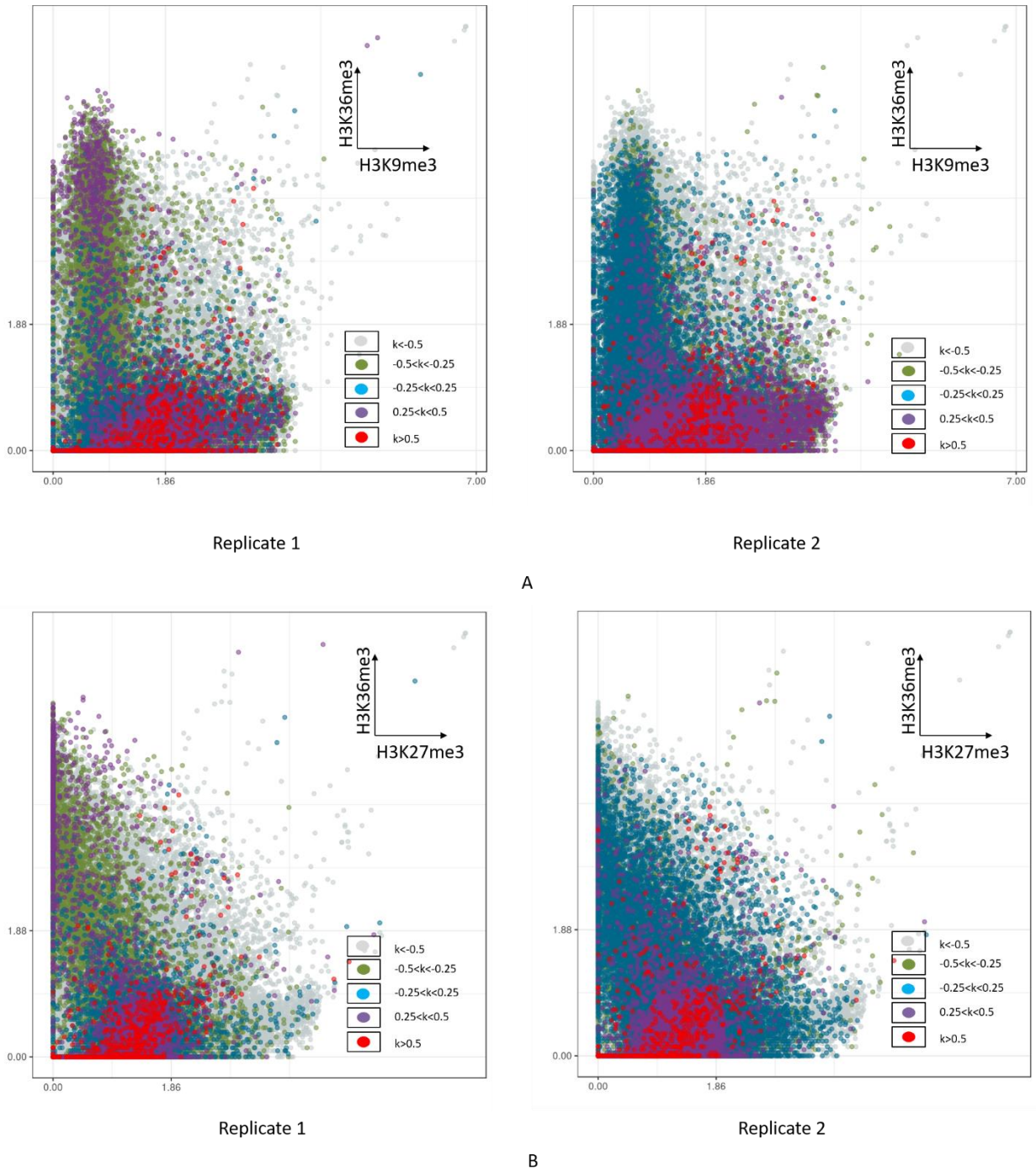
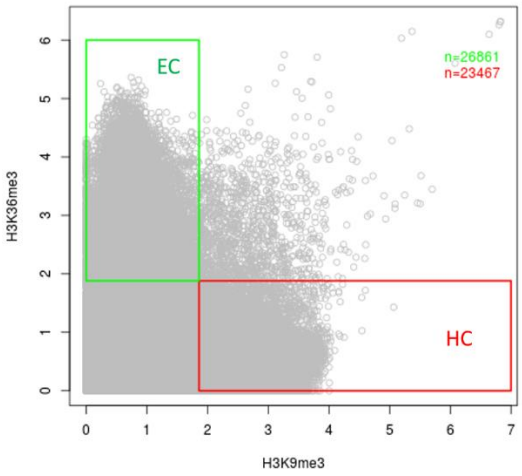
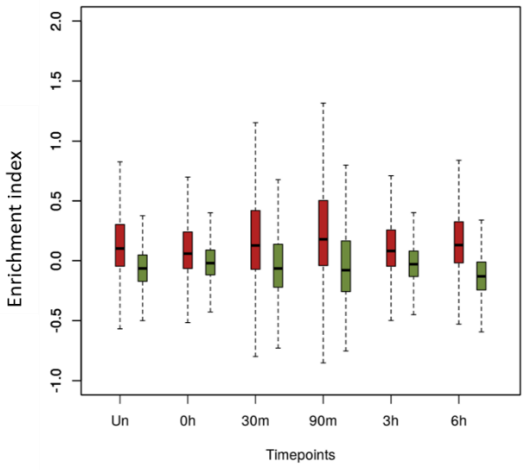


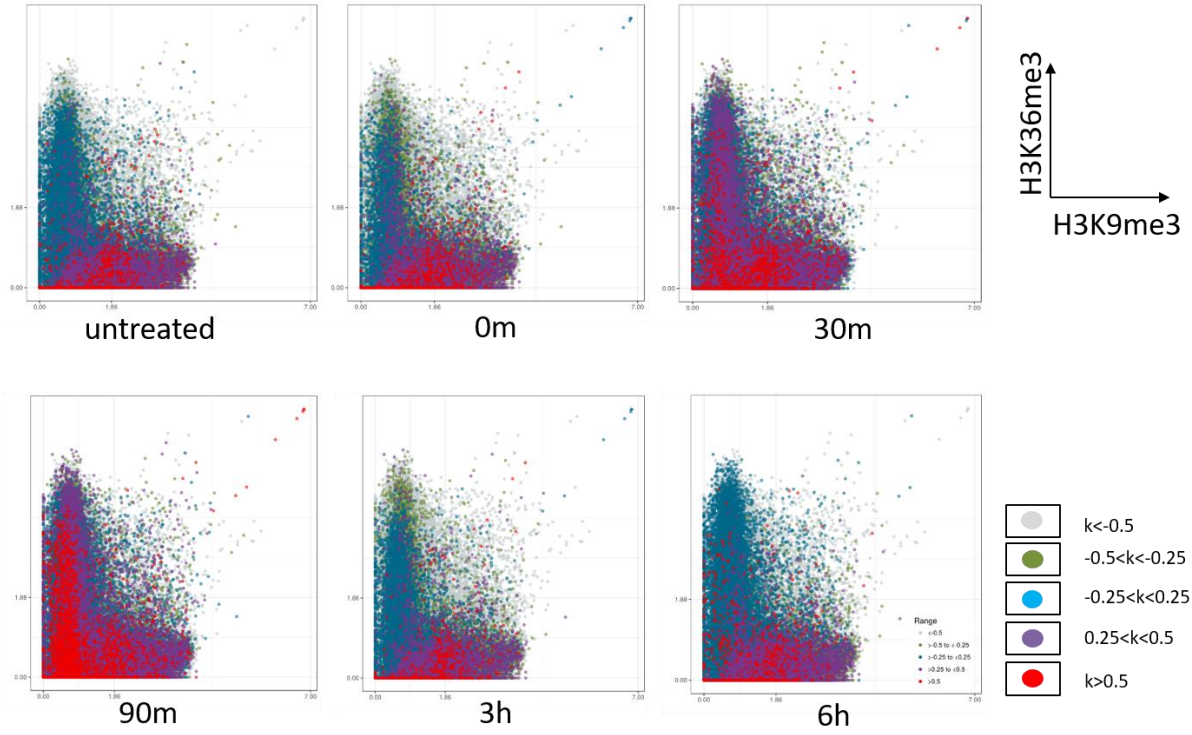
Figure 25 ***Endogenous DNA damage in mouse embryonic stem cells****. 10kb bins plotted according to enrichment of H3K36me3 and H3K9me3 (top row) and H3K36me3 and H3K27me3 (bottom row)(refer text). γH2AX enrichment values plotted as a heatmap (grey-low enrichment, red- high enrichment) for replicate 1 (left column) and replicate 2 (right column). A. γH2AX enrichment based on H3K36me3 and H3K9me3 levels. B. γH2AX enrichment based on H3K36me3 and H3K27me3 levels.*


EC
HC
n=26861
n=23467
H3K36me3
H3K9me3

A


Enrichment index
Un
0h
30m
90m
3h
6h
Timepoints

B


H3K36me3
H3K9me3
untreated
0m
30m
90m
3h
6h
k<-0.5
-0.5<k<-0.25
-0.25<k<0.25
0.25<k<0.5
k>0.5

C

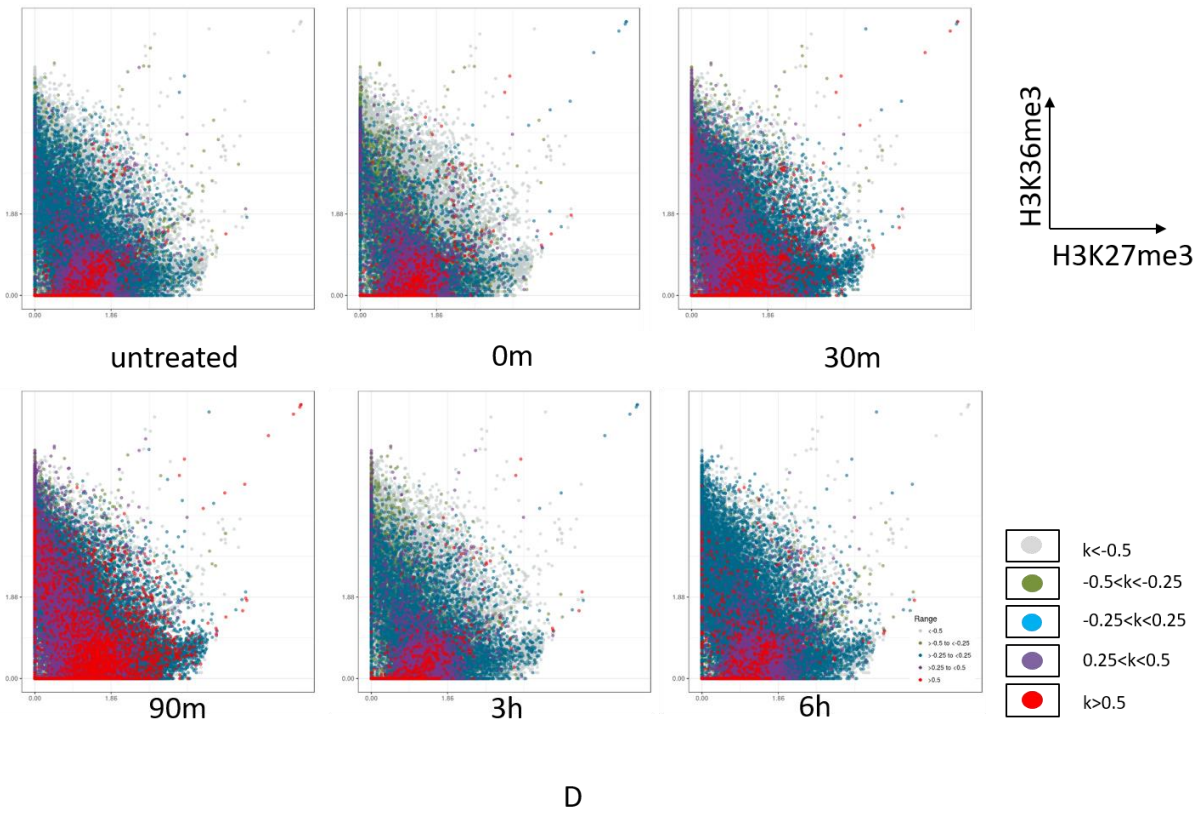


Figure 26. ***DNA damage repair in euchromatin and heterochromatin****. A. Gating euchromatin and heterochromatin bins based on H3K36me3 and H3K9me3 enrichment scores; bins with high H3K36me3 (>1.88) and low H3K9me3(<1.86) were assigned euchromatin; bins with high H3K9me3 (>1.86) and low H3K36me3(<1.88) were assigned heterochromatin. B. Quantification of γH2AX enrichment in euchromatin (EC) and heterochromatin (HC) for replicate 2. C. Heatmap of γH2AX enrichment on H3K36me3 vs. H3K9me3; and D. Heatmap of γH2AX enrichment on H3K36me3 vs. H3K27me3.*

3.7 DNA damage persists in Lamina-Associated Domains

Lamina-associated domains (LADs) have several properties like heterochromatin. Genes in LADs are generally transcriptionally silent (Guelen et al., 2008; Peric-Hupkes et al., 2010). Furthermore, LADs are also enriched for heterochromatin marks such as H3K9me3 and H3K27me3 (Guelen et al., 2008; Wen et al., 2009; Harr et al., 2015). Because of these similarities with heterochromatin, we also investigated the occupancy of γH2AX in defined LAD compartments compared to occupancy in non-LAD compartments in mES cells (Handoko et al., 2011). Like the previous results from heterochromatin analysis, a persistence of γH2AX was also observed in the LAD compartment (Fig. 27). Interestingly, a lower amount of endogenous damage was observed in the LAD compartment in untreated cells.

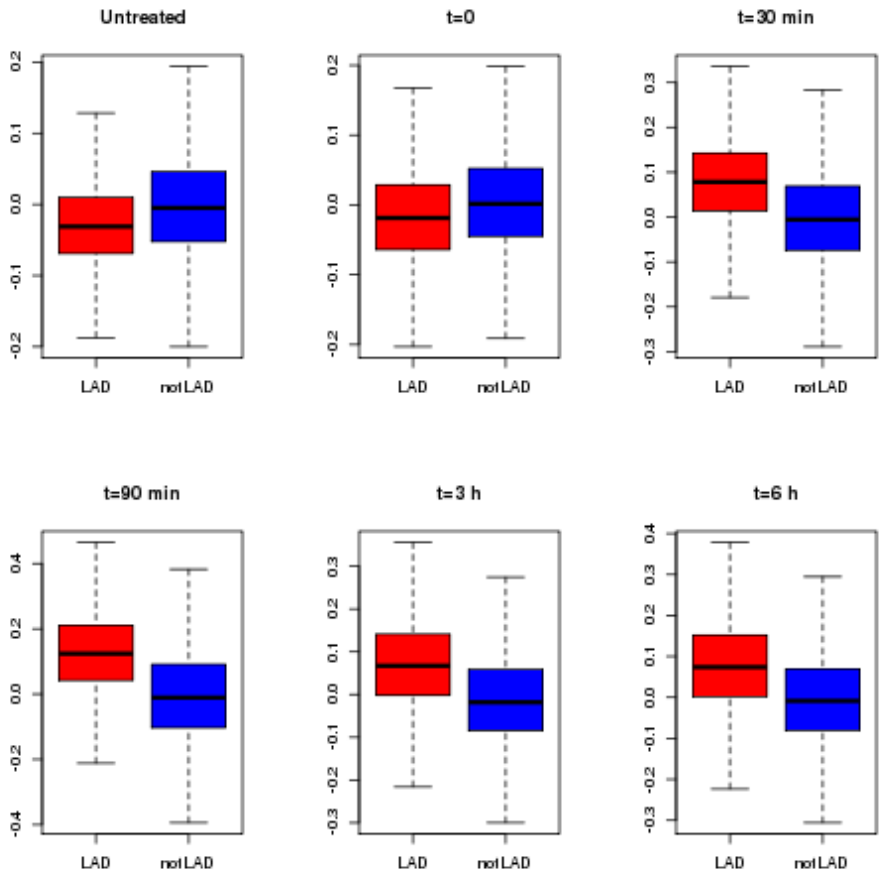
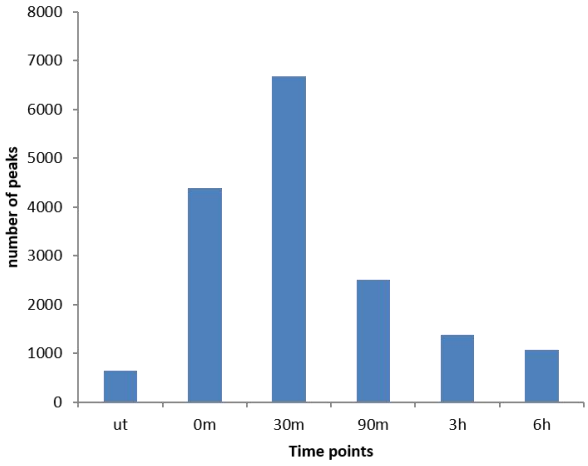


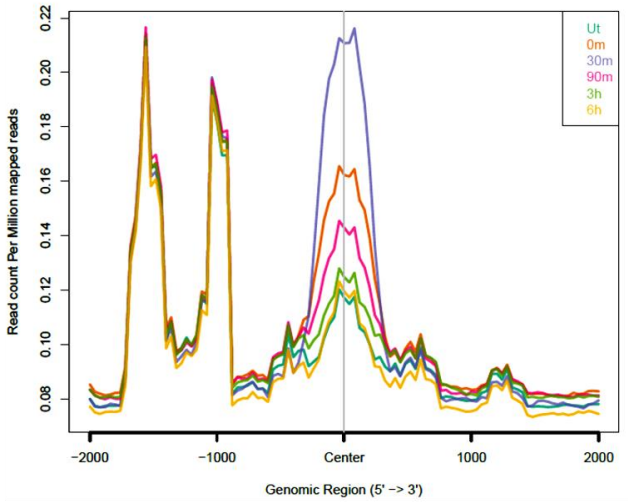
Figure 27. ***γH2AX enrichment in Lamina-associated domains****. Enrichment scores of γH2AX in regions classified as LADs and not LADs. (data from replicate 2).*

3.8 Distribution in genomic features

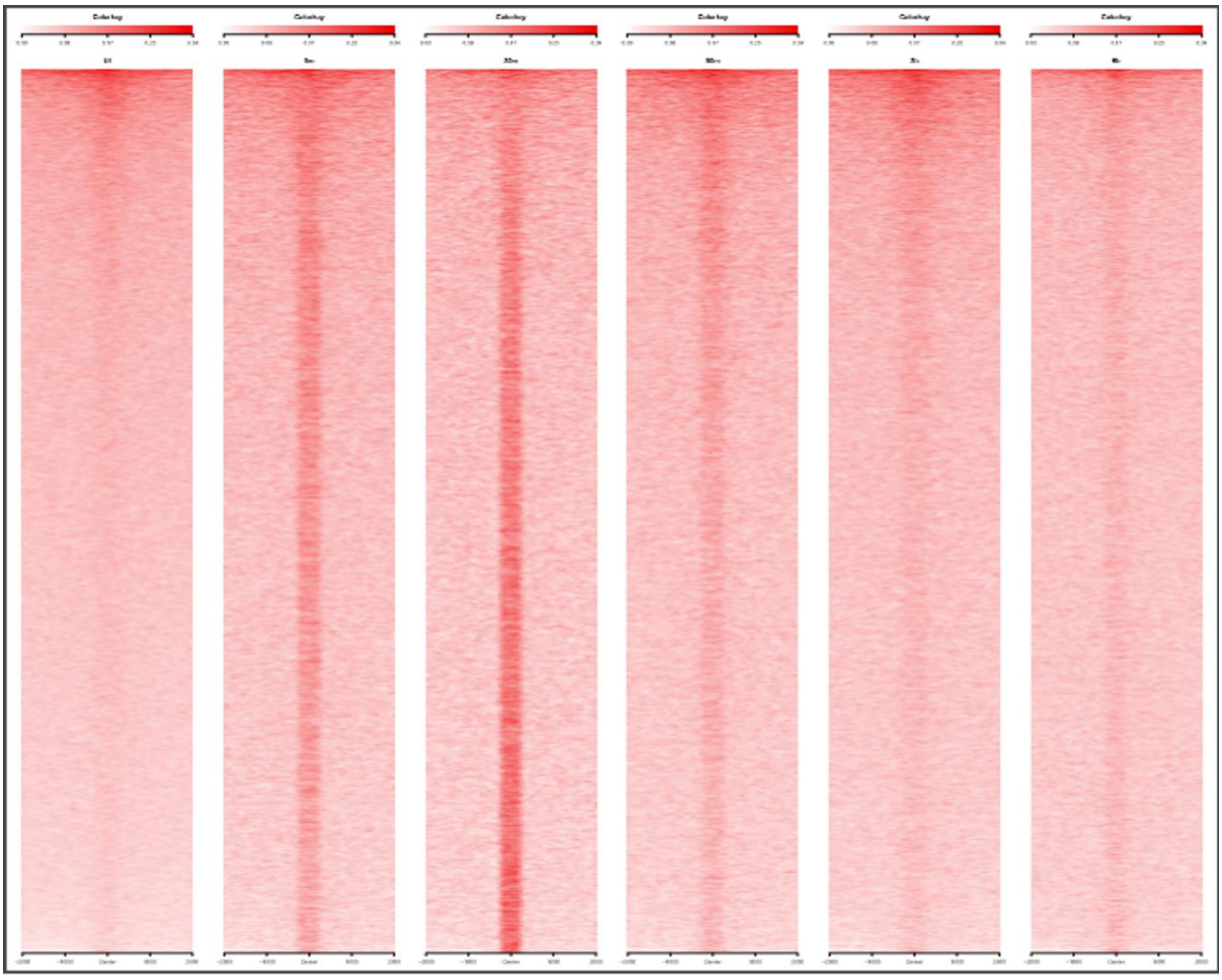
We also evaluated if certain genomic features have a preferential enrichment of γH2AX. MACS was used to identify peak regions for doing this analysis since it gives a discrete number of peaks with defined start and end coordinates. The enrichment levels of these peaks were similar to what was observed in the earlier analyses (Fig. 28A, B, C). This analysis revealed that in the untreated cells, nearly a quarter of the identified peaks were located near the transcription start site (TSS) (Fig. 28D). This would indicate a link between γH2AX and transcriptional activation.


8000
7000
6000
5000
4000
3000
2000
1000
0
number of peaks
ut
0m
30m
90m
3h
6h
Time points

A


Ut
0m
30m
90m
3h
6h
Read count Per Million mapped reads
0.22
0.20
0.18
0.16
0.14
0.12
0.10
0.08
−2000
−1000
Center
1000
2000
Genomic Region (5' −> 3')

B


UT
0m
30m
90m
3h
6h

C

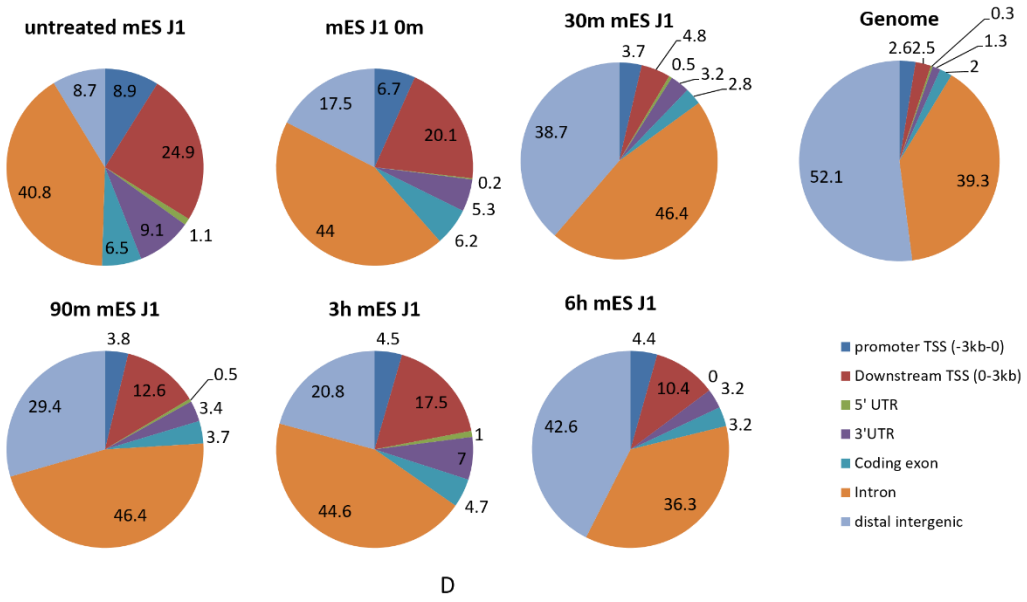


*Figure 28. **γH2AX enrichment in genomic features**. A. Peak counts from MACS. B. Average intensity of γH2AX enrichment around peak regions. C. Total enrichment of γH2AX around all peaks (sorted in descending order from top to bottom based on untreated sample (UT) intensity); intensity expressed as heatmap from white (low) to red (high). D. Distribution of γH2AX in genomic factors. (UTR- Untranslated region) (Data from replicate 2).*

3.9 Distribution around Transcription start site

Given the localization of γH2AX around transcription start sites, we further explored the status of γH2AX with respect to transcription start sites. We looked at γH2AX occupancy 0-5kb, 5-50kb, 50-500kb and more than 500kb up-and downstream from transcription start sites. In the untreated sample, we noted a higher enrichment downstream of TSS in the 0-5kb and 5-50kb windows (Fig. 29). After induction of damage, we noticed a shift in the γH2AX enrichment to the 50-500kb window (both in terms of absolute number of peaks and percentage of peaks) at the maximal enrichment timepoint. During repair, the enrichment of γH2AX increased further downstream of the TSS before increasing closer to the TSS. This agrees with previous observations that damage accumulates and persists in regions away from active chromatin.

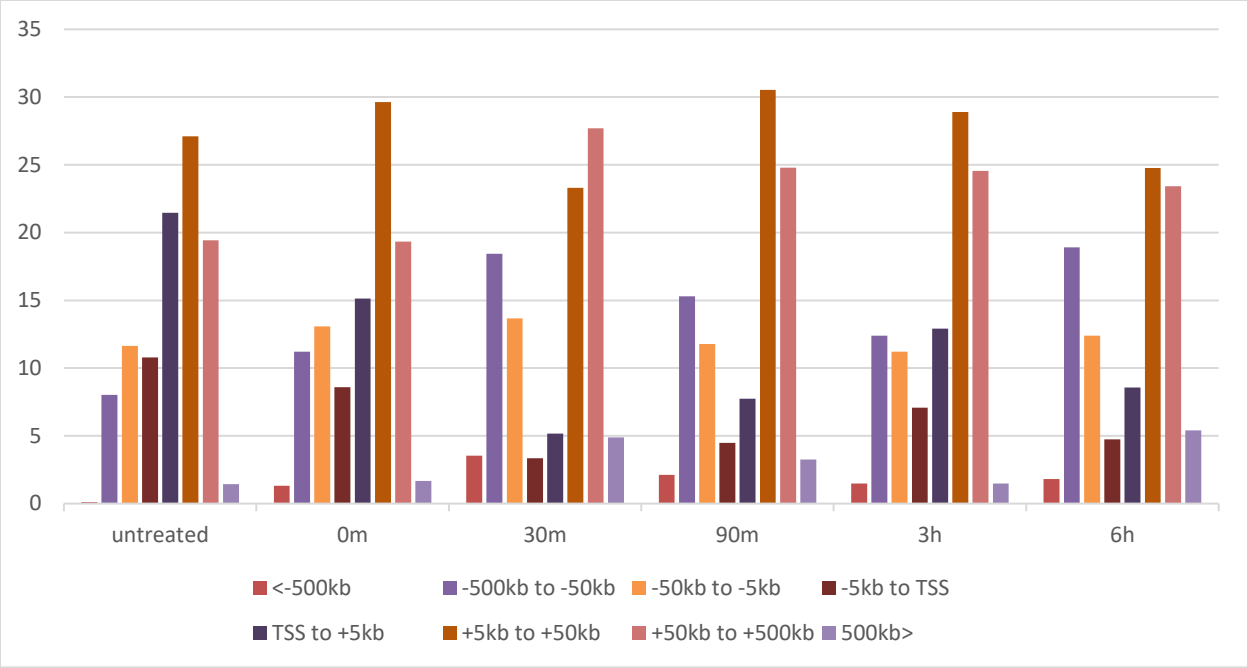
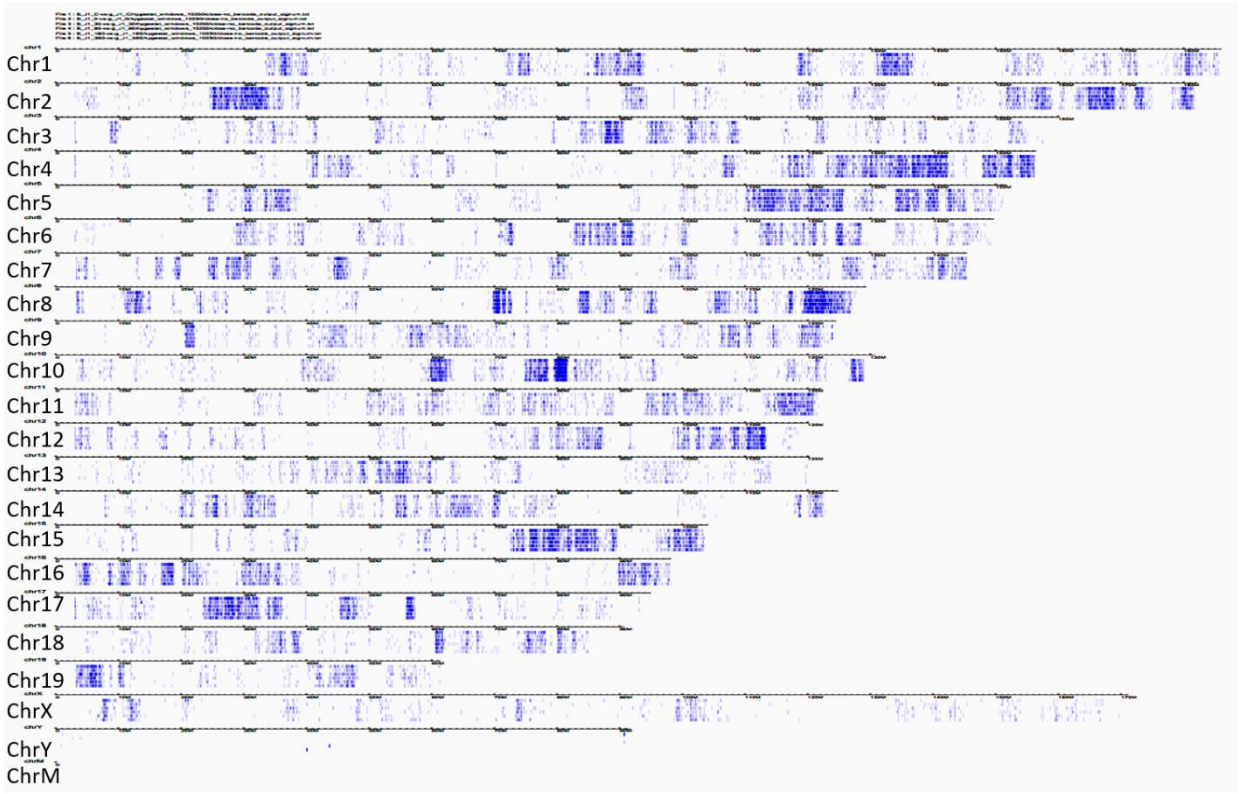


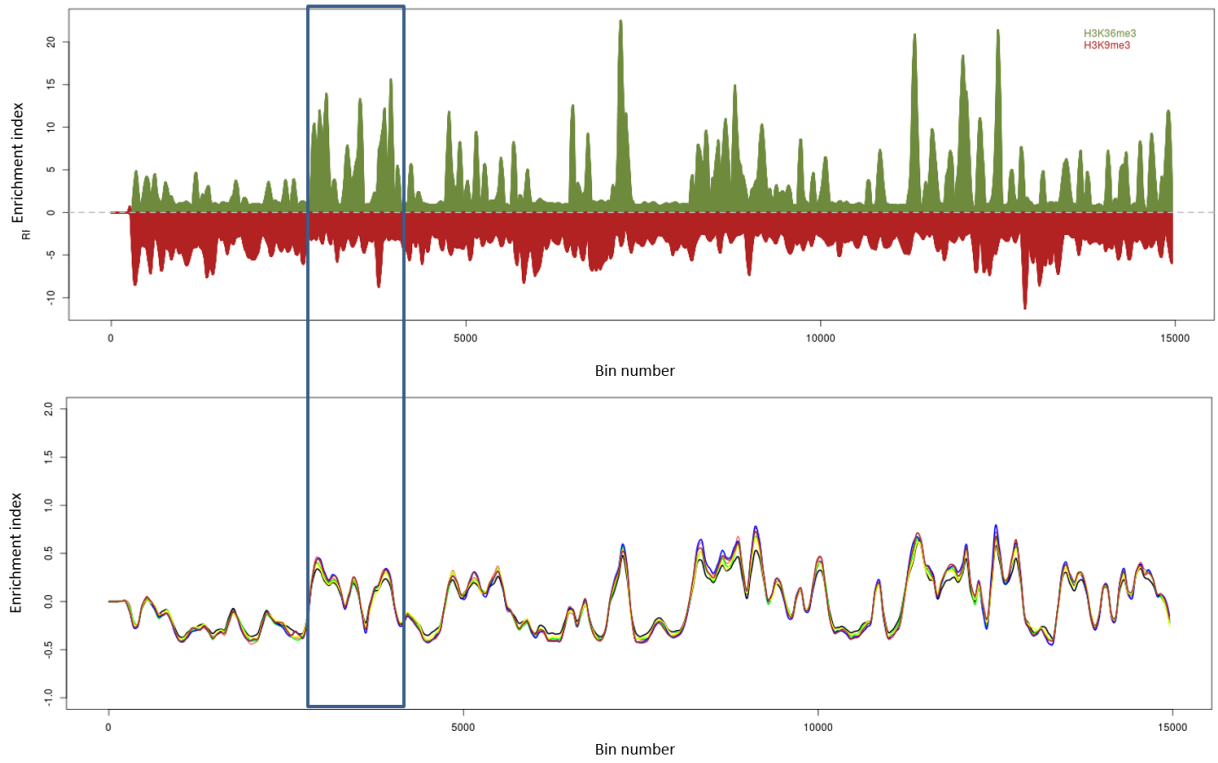
Figure 29. ***Distribution around Transcription start site.*** *Enrichment of γH2AX up to 500kb up-and downstream of transcription start site at indicated timepoints (Number of breaks expressed as percentage of total breaks observed).*

3.10 Correlation with BLESS data

γH2AX enrichment is a measure of the cellular response to a DSB and does not actually measure the "physical" damage to the DNA, i.e., it does not give information about the actual break-site. In order to map the break-site, we performed BLESS (in collaboration with Krzysztof Ginalski and Magdalena Skrypczak, University of Warsaw and Maga Rowicka, UTMB) on mES cells with the same drug treatment and timepoints. A similar analysis of the BLESS data for enrichment of DSBs has been inconclusive. It does however appear that the technique has only managed to identify endogenous breaks in the cells and not the induced breaks due to drug treatment (Fig. 30A, B). The endogenous breaks were present in the active compartment (euchromatin) (Fig. 30B).


Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
ChrX
ChrY
ChrM

A


H3K36me3
H3K9me3
Enrichment index
Bin number
Enrichment index
Bin number

B

Figure 30. ***Correlation with BLESS****. A. Genome-wide view of DSB locations in mouse embryonic stem cells. 1pixel width corresponds to 100kb. Blue lines indicate occurrence of at least one break in the 100kb bin. B. Correlation of histone marks with BLESS data. Top panel: H3K36me3 (green) and H3K9me3 (red) enrichment values; Bottom panel: enrichment values for DSBs. Box shows enrichment of DSBs in euchromatin region. No variation in DSB enrichment levels is observed during the experiment.*

3.11 Establishment of Damage inducible via AsiSI (DIvA) system in mES cells

In order to study damage induction and repair at specific loci, we tried to establish the Damage inducible via AsiSI (DIvA) system in mES cells. This system contains the restriction enzyme AsiSI, an 8-base pair cutter with ~1100 sites in the mouse genome, fused to a modified estrogen receptor (ER) hormone-binding domain, which binds to 4-hydroxy tamoxifen (4OHT). An auxin-inducible degron (AID) domain is also fused to the AsiSI-ER domain which allows for the degradation of the restriction enzyme by the addition of the plant hormone auxin (Iacovoni et al., 2010; Aymard et al., 2014) (Fig. 31A). With this system we wanted to follow the damage induction and repair kinetics at specific loci and try to differentiate repair occurring by different pathways in different compartments.

The establishment of this cell line was only partially successful. Although in transient transfections, an induction of damage after 4h of treatment with 4OHT could be observed which was reduced within 2h by the addition of auxin (Fig. 31B), a stable clonal mES cell line where AID-AsiSI-ER could consistently be induced with 4OHT and degraded with auxin could not be generated. This may be due to silencing of the plasmid in mES cells. This problem precluded the use of this system in further analysis.

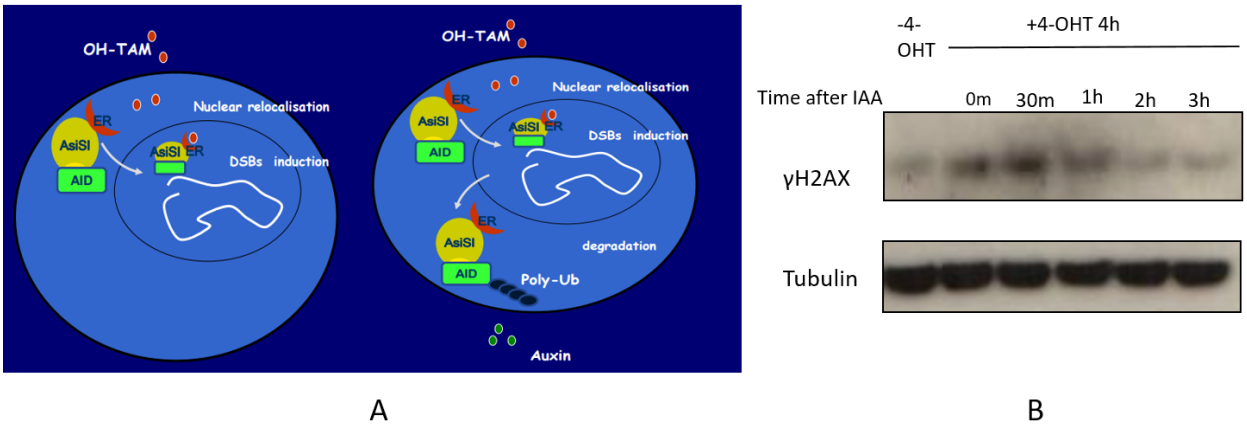


Figure 31. ***DIvA system in mouse embryonic stem cells****. A. Mechanism for induction of breaks using the DIvA system. B. Induction of damage by treatment with tamoxifen and subsequent repair by addition of auxin in transiently transfected embryonic stem cells.*

4. Discussion

My doctoral dissertation project has explored DNA damage response (DDR) in the context of chromatin organization. Mouse embryonic stem cells (mES) were chosen because of their pluripotent differentiation capacity that is characterized by a unique chromatin structure. ES cells have an open chromatin configuration with a large number of bivalent genes that are marked by both active and repressive marks which resolve upon lineage commitment. I performed ChIP Sequencing against the DNA damage sensor protein γH2AX as a proxy of the DNA damage in the genome. The results indicate that the enrichment of γH2AX occurs at defined genomic loci in the mouse genome and this enrichment persists in the non-euchromatin regions (regions devoid of H3K36me3). The persistence of damage is localized to heterochromatin and other compacted regions of the genome such as Lamina-associated domains (LADs). Analysis of the kinetics of the enrichment of γH2AX have shown that damage sensing and repair occurs faster in the euchromatin as compared to heterochromatin. This project was based completely on next generation sequencing (NGS) approaches to study the kinetics of DDR. To validate the results from this project, further experiments using NGS and microscopy will be required.

4.1 Enrichment of γH2AX at defined loci

During cell cycle analysis, we observed that mES cells entered a G2 cell cycle arrest. This indicates a slower component of repair had been used to repair damage. It has been suggested that entering cell cycle arrest may be a strategy to utilize the high fidelity homologous recombination pathway to avoid mutations caused by using error-prone but faster repair pathways such as NHEJ. Since the slow component of repair has been proposed to be involved in the repair of DSBs in heterochromatin (Goodarzi et al., 2010; Tsouroula et al., 2016; Biehs et al., 2017), the cell cycle arrest could be ascribed to persistent damage in this compartment. The enrichment of γH2AX occurred in a distinctive reproducible manner across the genome. This enrichment indicates that the damage sensing and repair is organized at specific loci. It was also observed that the borders of these γH2AX domains were already present at the timepoint when they were first detected, and DDR kinetics happened normally within these domains, with an initial increase of enrichment followed by a gradual decrease. A possible explanation for the formation of these domains could be due to cohesin binding that prevents γH2AX spreading and helps to isolate active genes (Caron et al., 2012). This also suggests that the regulation of γH2AX spreading is dependent on higher order chromatin organization. The open conformation of euchromatin facilitates the spread of the γH2AX signal to the neighborhood which in turn permits faster repair while for condensed chromatin found in heterochromatin and LADs, this spread would be hampered which would explain why damage persistence and delayed repair occurs in these regions. This is in concurrence with other observations which suggest that damage is persistent in LADs and heterochromatin compartments and there is a euchromatin-to-

heterochromatin tendency for repair kinetics (Lemaitre et al., 2014; Natale et al., 2017). We observed that the euchromatin compartment (H3K36me3-rich) is not associated with these damage domains and such regions appear to be shielded from damage. An evolutionary strategy developed by the cells to protect this compartment which contains actively transcribed genes from damage may also be a reason for this observation. Interestingly, similar results were observed in mouse renal cells exposed to high salt stress with damage being concentrated in gene deserts (Dmitrieva et al., 2011). These results together may indicate that higher order chromatin structure evolved in a way to protect the genome from exogenous damage.

Another reason why the active compartment appears to be relatively damage free may be the presence of nucleosome free regions around promoters of active and poised genes. It has been reported that in ES cells, there is a higher dissociation of histones in promoters and transcription end sites which permit transcription initiation and termination (Ha et al., 2014).

4.2 Repair pathway choice after damage

The choice of repair pathway depends on the transcriptional status of the break site. Homologous recombination (HR) is used to repair breaks occurring in transcriptionally active regions (Aymard et al., 2014). H3K36me3 has been implicated in this repair pathway choice and promotes resection by recruiting CtIP through LEDGF, a chromatin binding protein that binds H3K36me3. Breaks at inactive regions would be unable to recruit resection factors and hence would be repaired by non-homologous end joining (NHEJ). Our results have shown that DNA damage persists in regions devoid of H3K36me3. Since H3K36me3 is involved in regulating HR (Aymard et al., 2014) and given that mouse embryonic stem cells have elevated levels of the HR proteins RPA and Rad51 coupled with a short G1 phase (Ahuja et al., 2016), it can be argued that this persistence is occurring in heterochromatin which triggers a cell cycle arrest. The euchromatin compartment is either protected from the damage due to the chromatin organization or is rapidly repaired so that the experimental approaches used in this study could not capture these events. To study whether the active compartment is affected by the drug treatment and elicits a damage response, an intermediate timepoint is needed between the 0m and 30m timepoints. We also observed an increased level of γH2AX enrichment around transcription start sites in untreated cells. An explanation for this observation could be the formation of DSBs during transcriptional activation and elongation in active genes (Bunch et al., 2015; Singh et al., 2016).

4.3 Chromatin mobility and clustering after damage induction

Recently it was shown that breaks in transcriptionally active regions of the genome exhibited increased clustering during G1 phase so that they could be repaired by HR in G2. Clustering has been proposed as a

mechanism to sequester and/or prepare breaks for faithful repair (Aymard et al., 2017). On the other hand, some breaks are held together by the NHEJ factors Ku70-Ku80 or XRCC4-XLF and are hence immobile (Soutoglou et al., 2007; Brouwer et al., 2016). The damage domains observed in our data occur specifically in regions devoid of H3K36me3 and so may not be prone to clustering. To answer whether persistent damage domains cluster together, and to check the overall mobility of mES cell chromatin after damage, Hi-C or 4-C approaches should be taken which will additionally reveal translocation prone regions which arise due to improper repair.

4.4 Studying DNA damage repair at specific loci

To study DNA damage repair at specific loci, we had tried to establish the Damage Inducible via AsiSI (DIvA) system in mES cells. Through this approach, we wanted to compare repair kinetics at different loci and infer the role of chromatin organization at the break-site in regulating repair. A stable cell line expressing the AID-AsiSI-ER system could not be established in mES cells although the system was shown to be functional during transient transfections. A likely reason for this is the silencing of plasmids which do not contain stem cell specific promoters such as EF1α (the AID-AsiSI-ER plasmid contained SV40 and CMV promoters) during cell passaging.

4.5 Mapping of double strand breaks

BLESS was performed to map the position of both endogenous and induced DSBs. It appears that this technique could only map endogenous breaks and not induced breaks. It was seen that the data from this technique was extremely noisy and inconclusive. It may be possible that the breaks induced in cells were at different locations. These loci would produce a higher level of background which would make identifying low-level recurrent breaks difficult (Canela et al., 2016). Although it remains unclear why this technique did not work, improvements have been made in mapping DSBs. BLISS (Yan et al., 2017) was developed specifically to address the shortcomings of BLESS and prevents the formation of additional breaks during the experimental procedure and needs a lesser amount of starting material. It is recommended to use BLISS to map DSBs in continuation of this project.

5. Perspectives

We have observed that the enrichment of γH2AX happens at specific loci on the mouse genome. It would be essential to validate the observations of ChIP sequencing by immunofluorescence experiments. The status of γH2AX at euchromatin (H3K9me3) and heterochromatin (H3K36me3, H3K27me3) will be studied using super-resolution microscopy. An integration of high-resolution microscopy data with sequencing data will help in getting a clear understanding of the compartmentalization of the DNA damage response in chromatin compartments.

Since γH2AX spreads around the break site, further investigation on which factors regulate the demarcation of these domain boundaries will need to be carried out. One way to approach this would be to study the relative distribution of the histone variant H2AX in euchromatin and heterochromatin. Studies have shown a lower abundance of H2AX in heterochromatin in human cancer cell lines (Cowell et al., 2007).

Bioinformatically, comparison of the data with datasets related to chromosome conformation, gene expression levels, replication timing, origin of replication and other histone marks is being performed. Repeat elements such as satellite repeats which are commonly found in heterochromatin could not be studied since the analysis pipeline only considered uniquely mapped reads. A new analysis pipeline will be designed to study these repeat elements.

We have observed that the persistence of damage occurs at regions devoid of H3K36me3 while H3K36me3 rich regions are protected from damage or have very fast repair kinetics. It would be important to study the localization of specific HR/NHEJ factors at regions of fast and slow repair. This analysis can be done by both microscopy based approaches and sequencing approaches (ChIP-Seq) using Rad51 (for HR) and XRCC4 (for NHEJ) as probes to identify repair pathway choice at specific loci.

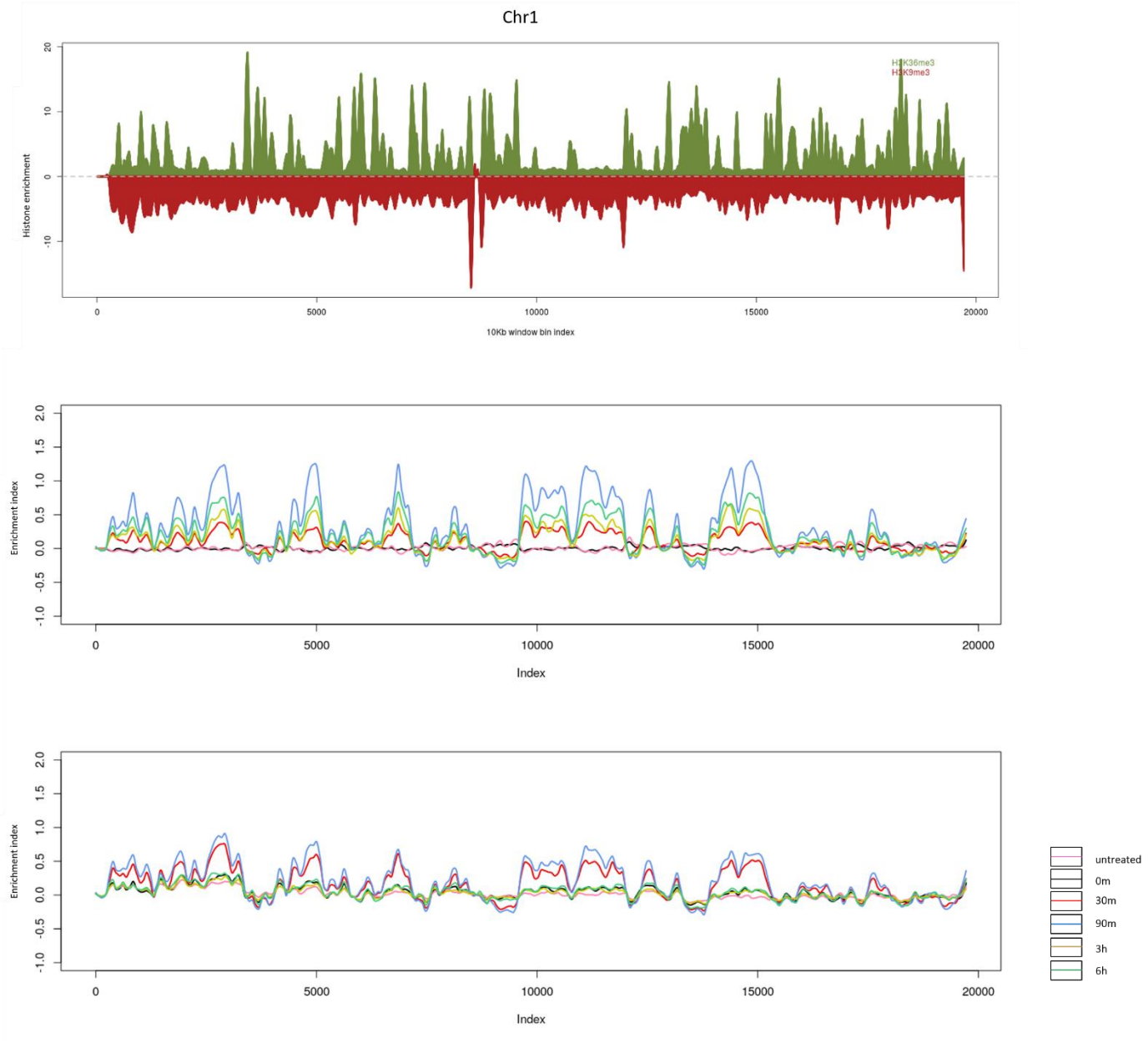
The methyltransferases Suv39h1 and Suv39h2 are responsible for H3K9 methylation. It would be interesting to study where damage will persist in Suv39h double negative mES cells which have decreased methylation of pericentric heterochromatin.

mES cells were chosen for this project since the chromatin is in an open conformation. Upon differentiation, repressive marks are added to most regions of the genome when the cell specializes to only carry out specific activities. To study how this change in the chromatin organization locally and globally would affect DDR, a direction for further investigation would be to differentiate mES cells into a definite cell type and perform a similar analysis. Since current differentiation protocols used (Bibel et al., 2007) generate a mixed population of cells which would make data analysis confounding, a standard cell line such as mouse embryonic fibroblasts or 3T3 cells should be used for such studies. Damage persistence and kinetics data in such differentiated cell lines, when compared with mES cells, will establish the role of chromatin organization at the same genomic locus in regulating DDR.

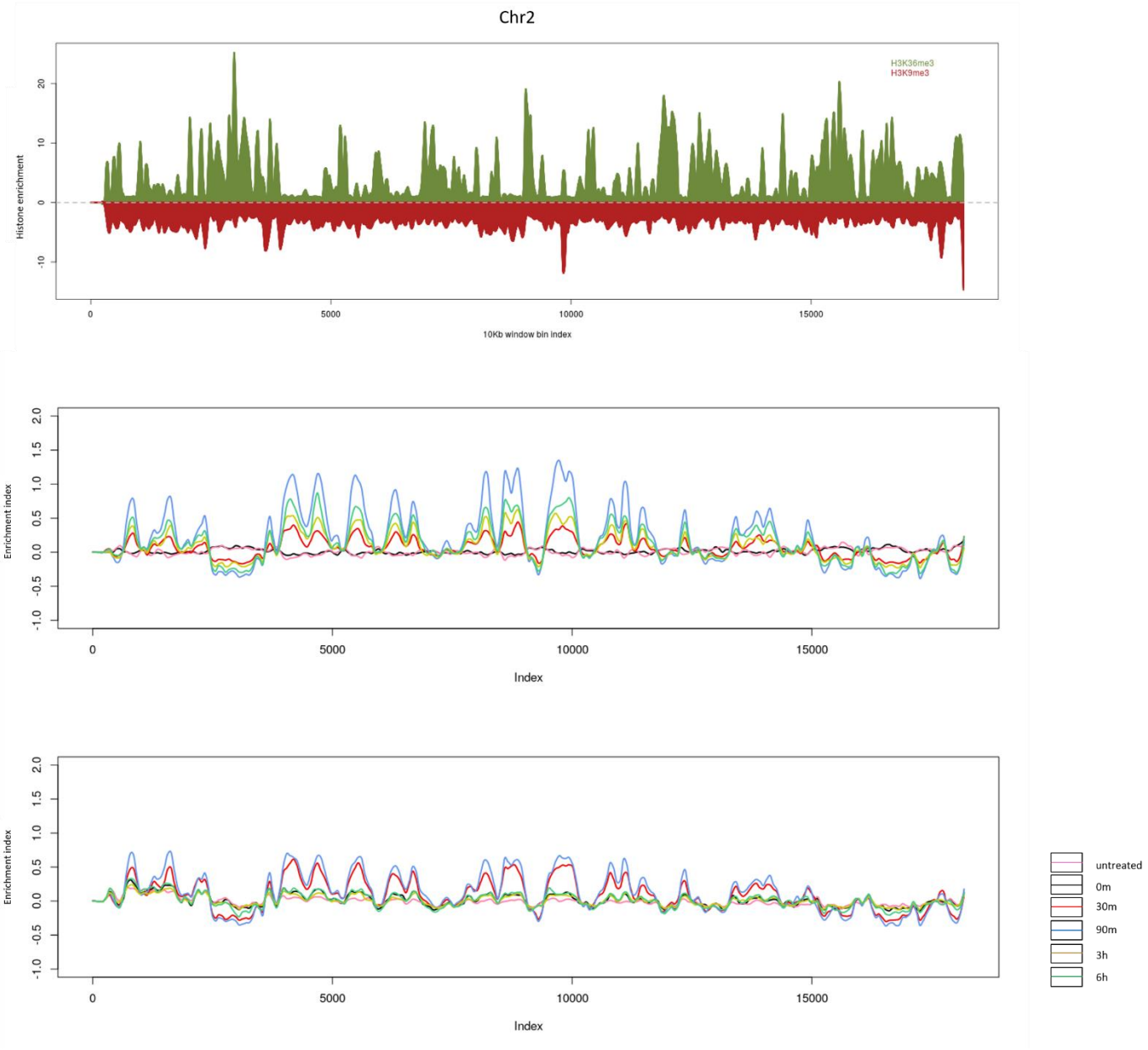
Attempts to establish a stable mES cell line containing the AID-AsiSI-ER plasmid were not completely successful although the plasmid was shown to be functional in transient transfections. This may be due to the silencing of non-ES specific promoters present on this plasmid. To establish the DIvA system in mES cells, the promoters will be changed to the mES specific EF1α promoter. Another solution to this problem would be to co-transfect the existing plasmid with another plasmid containing mCherry or EGFP, sorting mCherry-or EGFP-positive cells and performing experiments on these transiently transfected cells.

Alternately, to study break repair kinetics at different loci, a candidate based approach using CRISPR-Cas9, can be established. Specific loci can be targeted for break induction, depending on chromatin organization (euchromatin, heterochromatin, LADs, transcription status, etc.) and repair kinetics can be followed. Applying this strategy on targets identified from the previous analysis would serve as a proof of principle for the role of chromatin organization in regulating the DNA damage response.

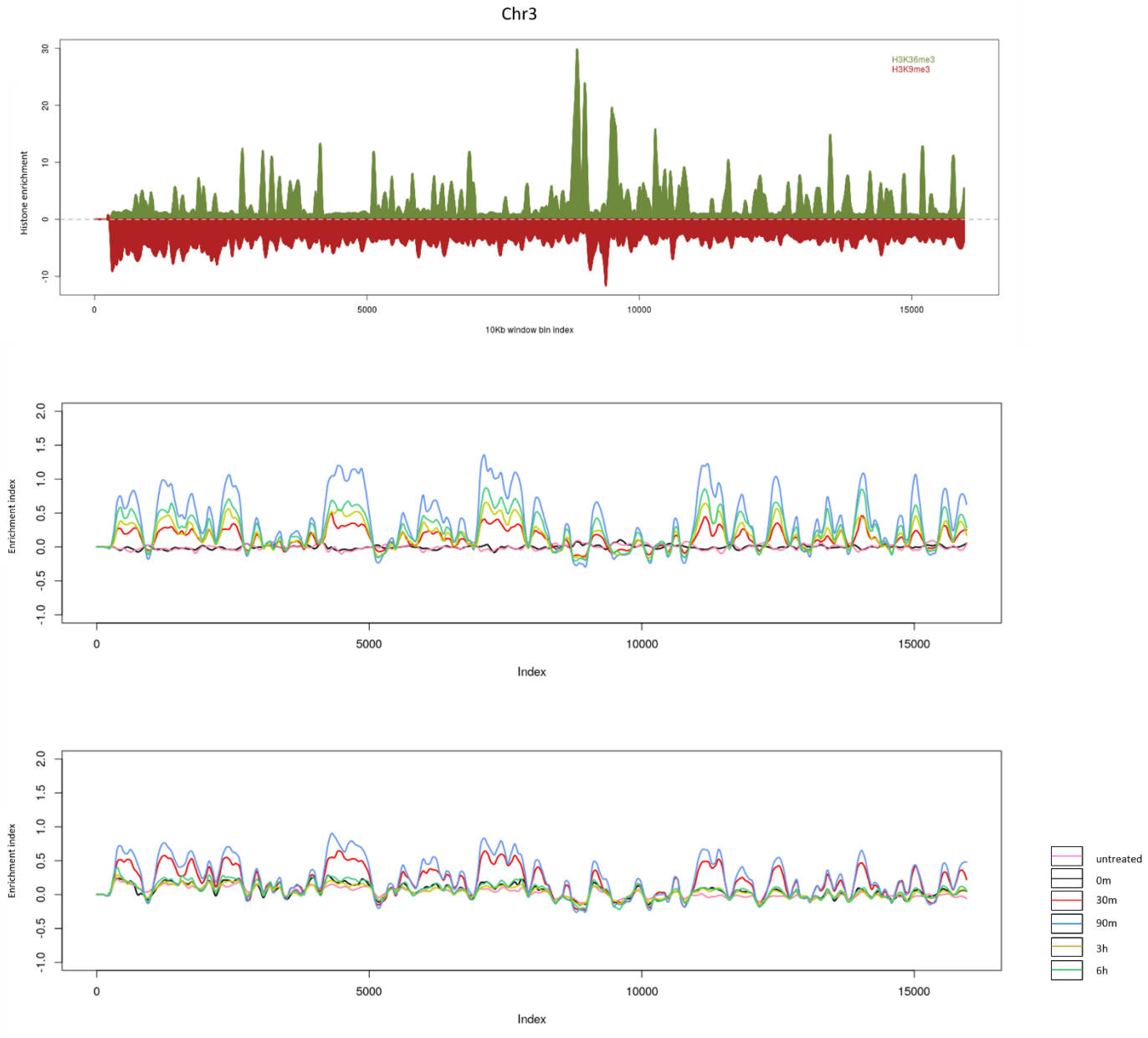
ANNEX



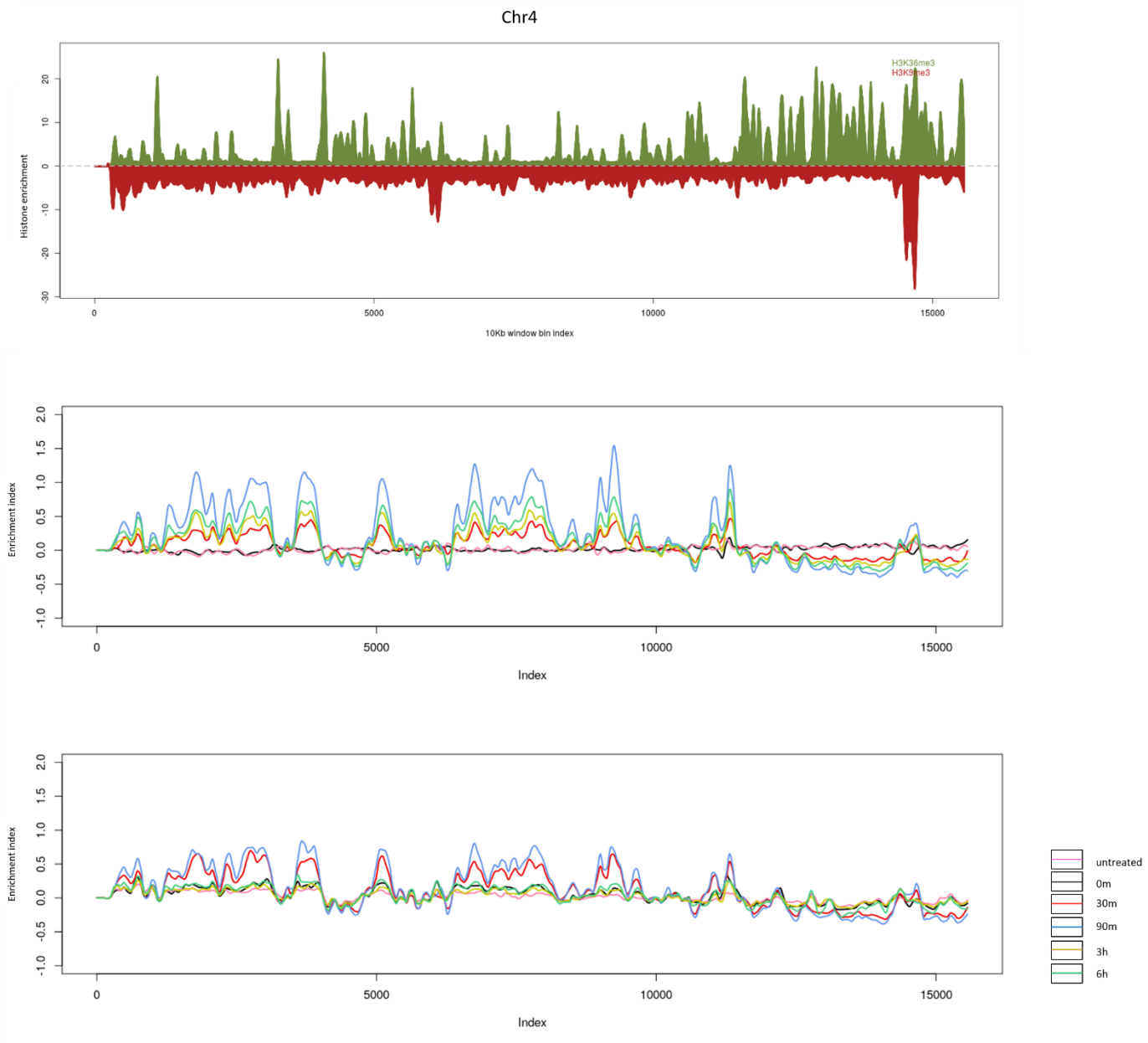
Correlation with histone marks. Chromosome-wide plot of chromosome 1. Top panel: H3K36me3 (green) and H3K9me3 (red) enrichment values; Middle panel: γH2AX enrichment in replicate 1; Bottom panel: γH2AX enrichment in replicate 2.



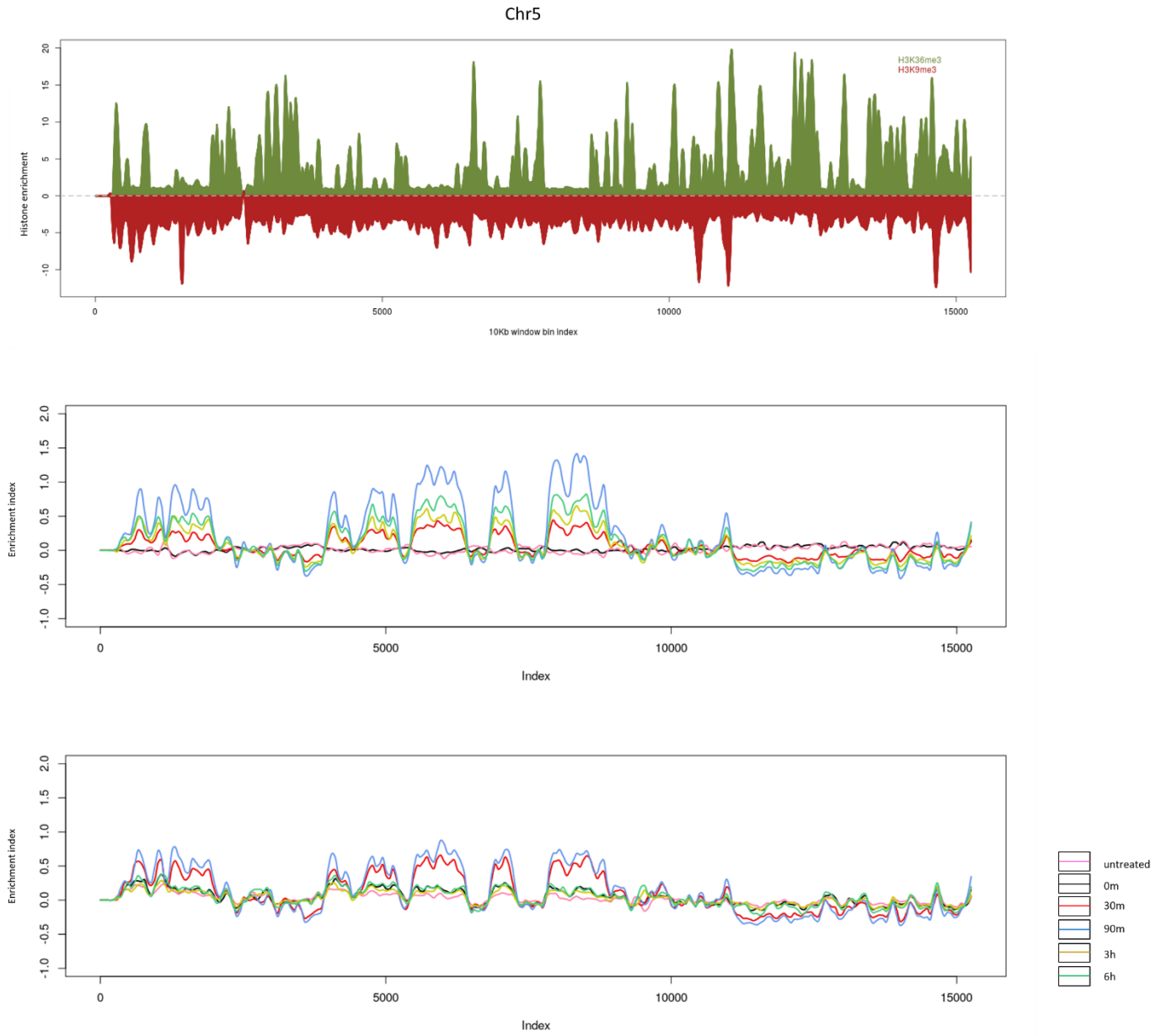
Correlation with histone marks. Chromosome-wide plot of chromosome 2. Top panel: H3K36me3 (green) and H3K9me3 (red) enrichment values; Middle panel: γH2AX enrichment in replicate 1; Bottom panel: γH2AX enrichment in replicate 2.



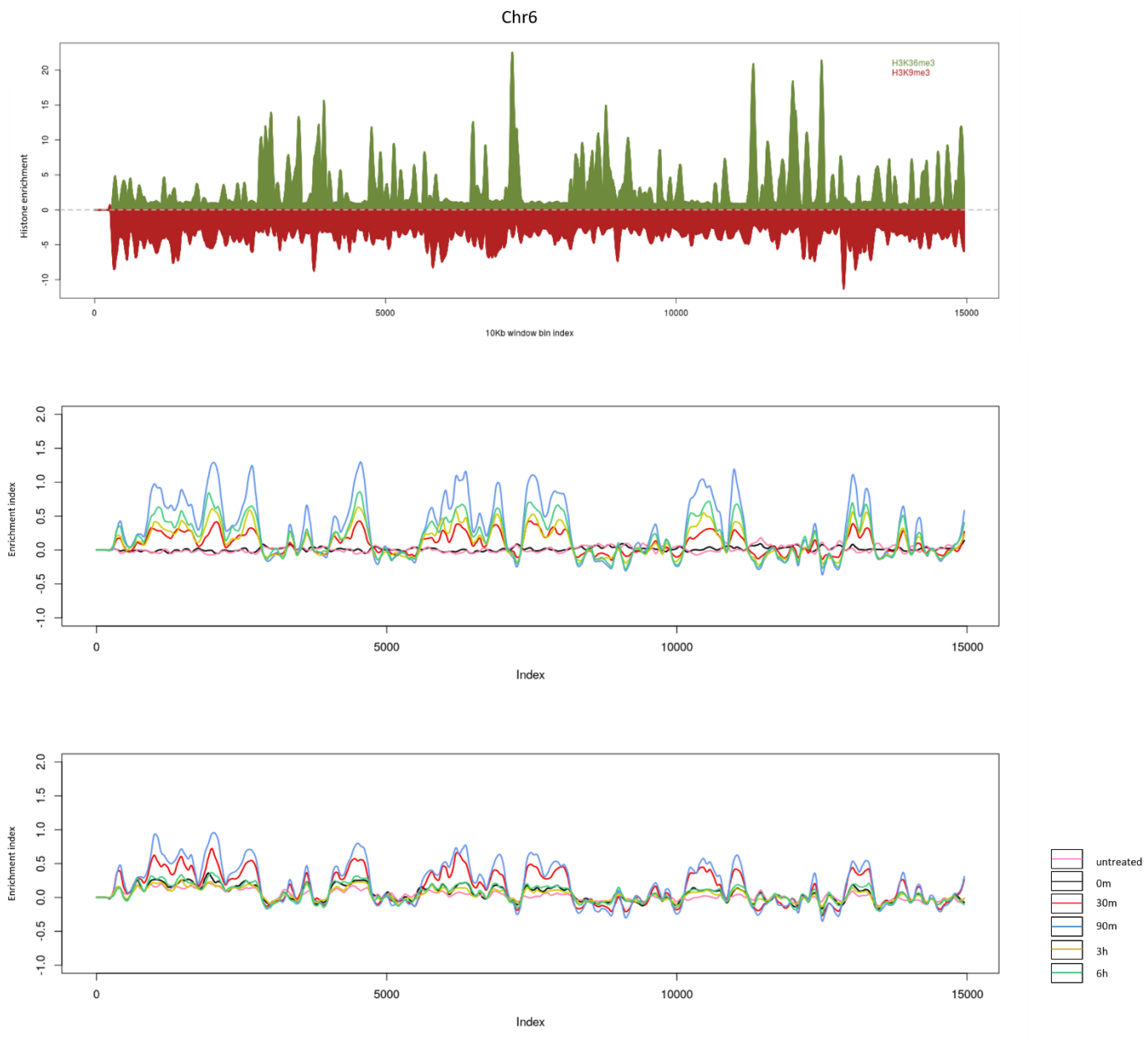
Correlation with histone marks. Chromosome-wide plot of chromosome 3. Top panel: H3K36me3 (green) and H3K9me3 (red) enrichment values; Middle panel: γH2AX enrichment in replicate 1; Bottom panel: γH2AX enrichment in replicate 2.



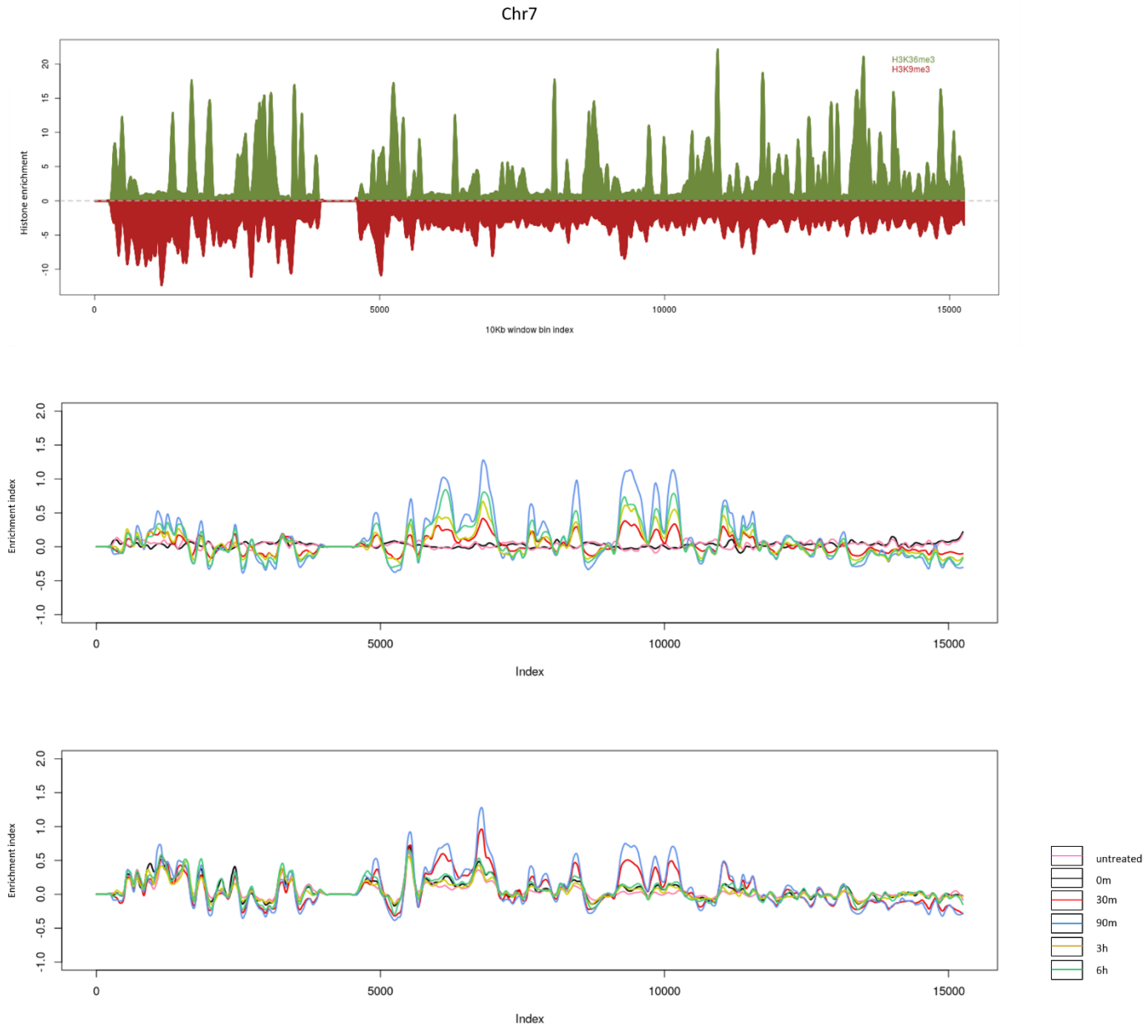
Correlation with histone marks. Chromosome-wide plot of chromosome 4. Top panel: H3K36me3 (green) and H3K9me3 (red) enrichment values; Middle panel: γH2AX enrichment in replicate 1; Bottom panel: γH2AX enrichment in replicate 2.



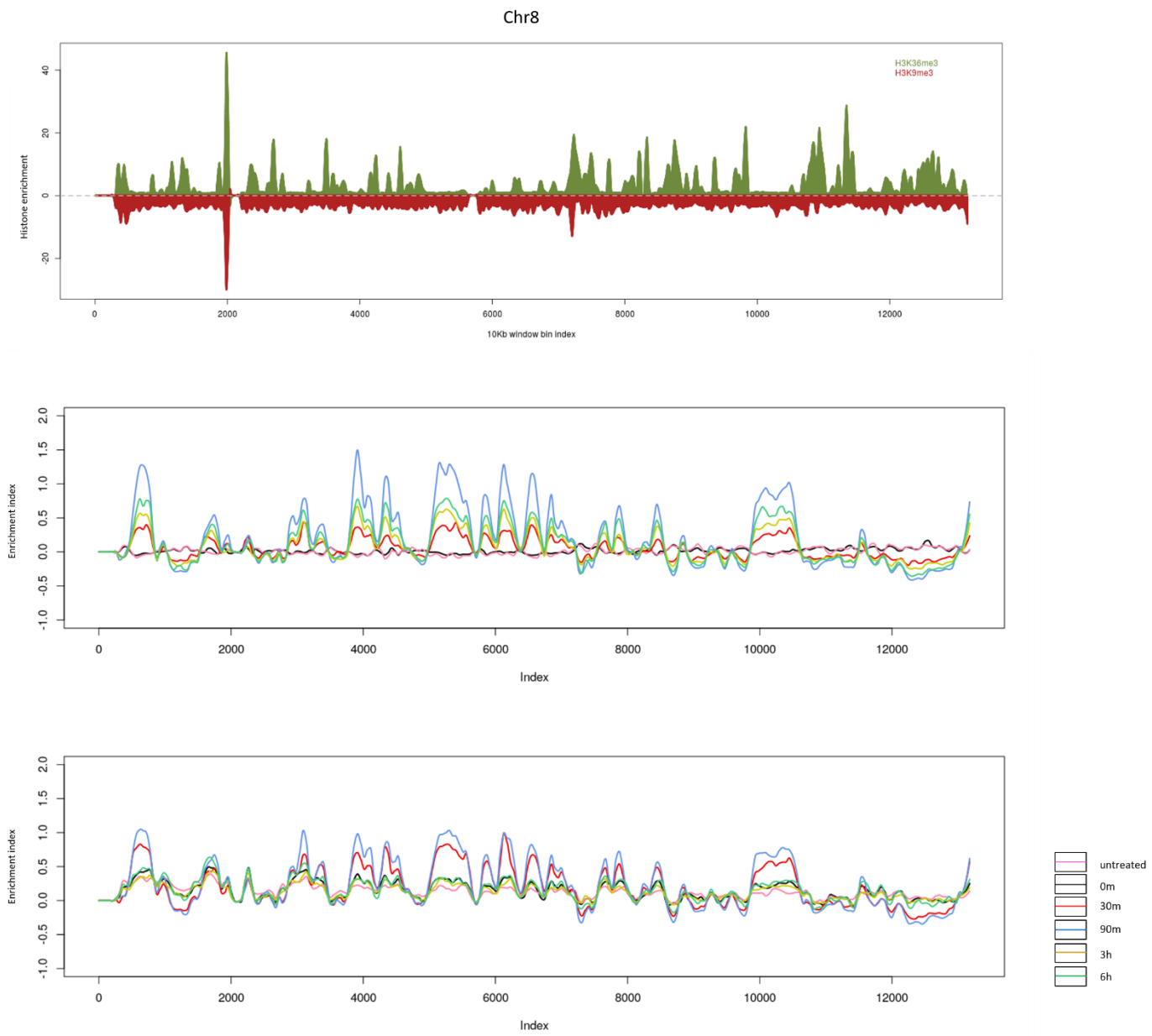
Correlation with histone marks. Chromosome-wide plot of chromosome 5. Top panel: H3K36me3 (green) and H3K9me3 (red) enrichment values; Middle panel: γH2AX enrichment in replicate 1; Bottom panel: γH2AX enrichment in replicate 2.



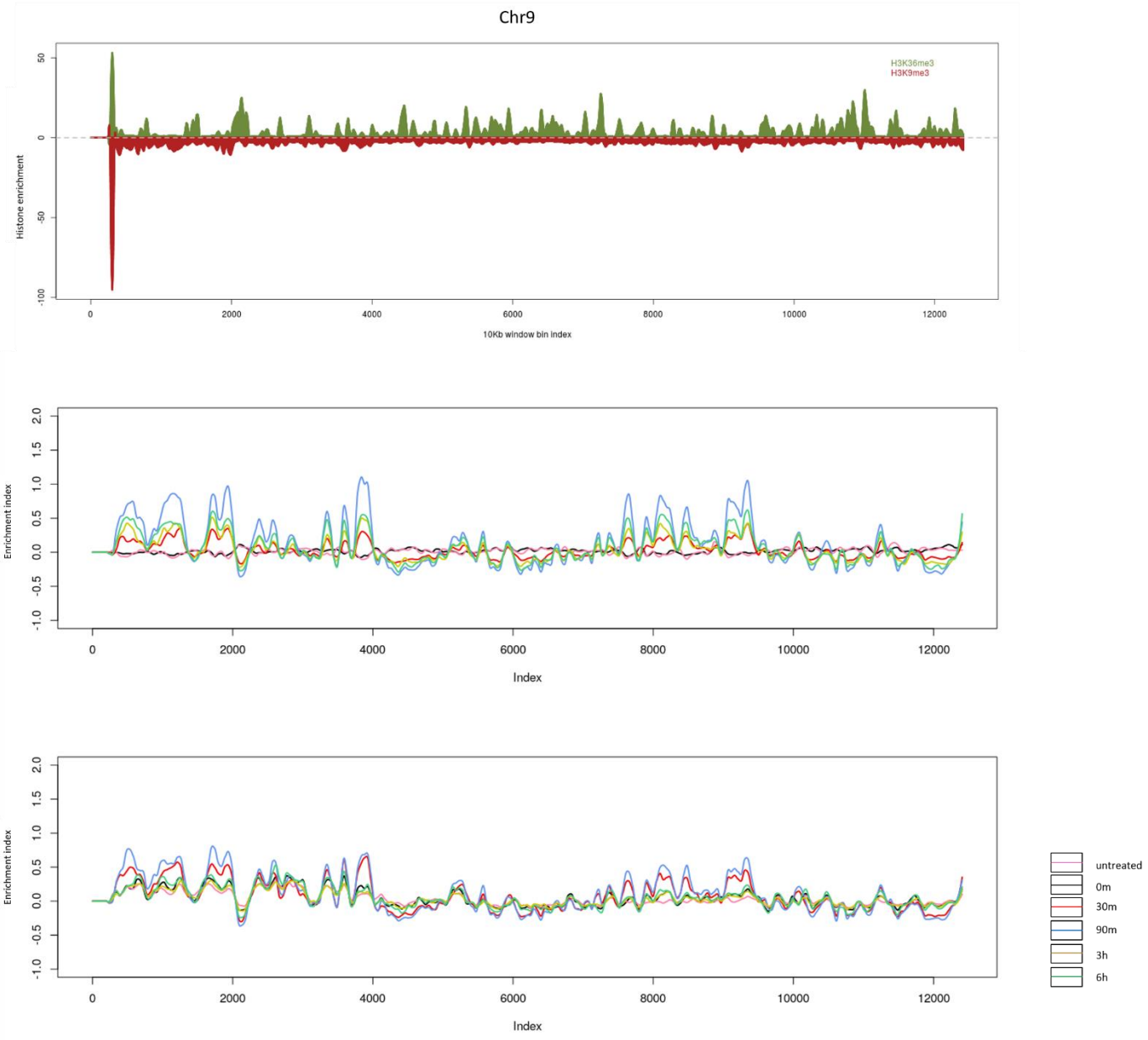
Correlation with histone marks. Chromosome-wide plot of chromosome 6. Top panel: H3K36me3 (green) and H3K9me3 (red) enrichment values; Middle panel: γH2AX enrichment in replicate 1; Bottom panel: γH2AX enrichment in replicate 2.



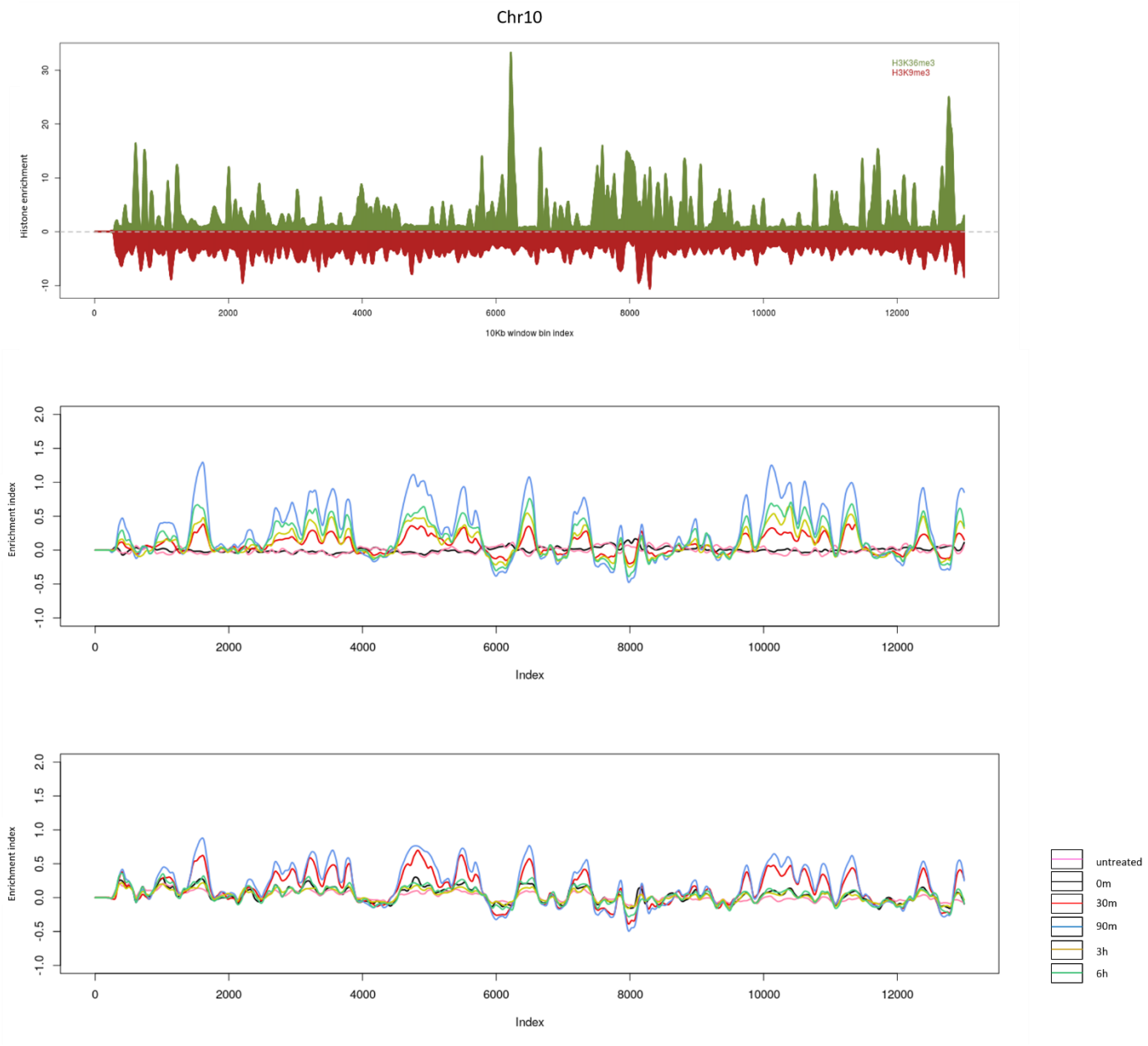
Correlation with histone marks. Chromosome-wide plot of chromosome 7. Top panel: H3K36me3 (green) and H3K9me3 (red) enrichment values; Middle panel: γH2AX enrichment in replicate 1; Bottom panel: γH2AX enrichment in replicate 2.



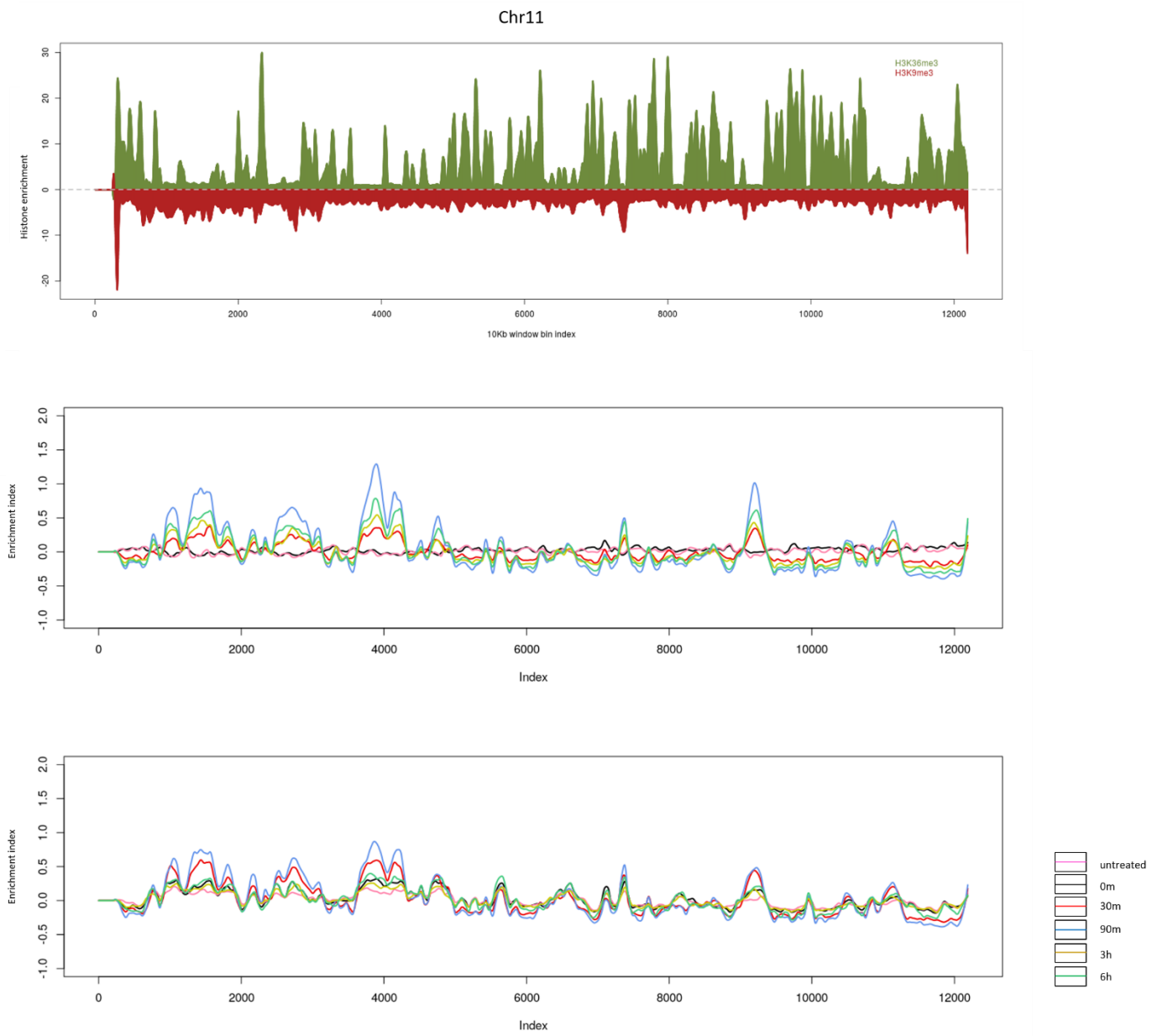
Correlation with histone marks. Chromosome-wide plot of chromosome 8. Top panel: H3K36me3 (green) and H3K9me3 (red) enrichment values; Middle panel: γH2AX enrichment in replicate 1; Bottom panel: γH2AX enrichment in replicate 2.



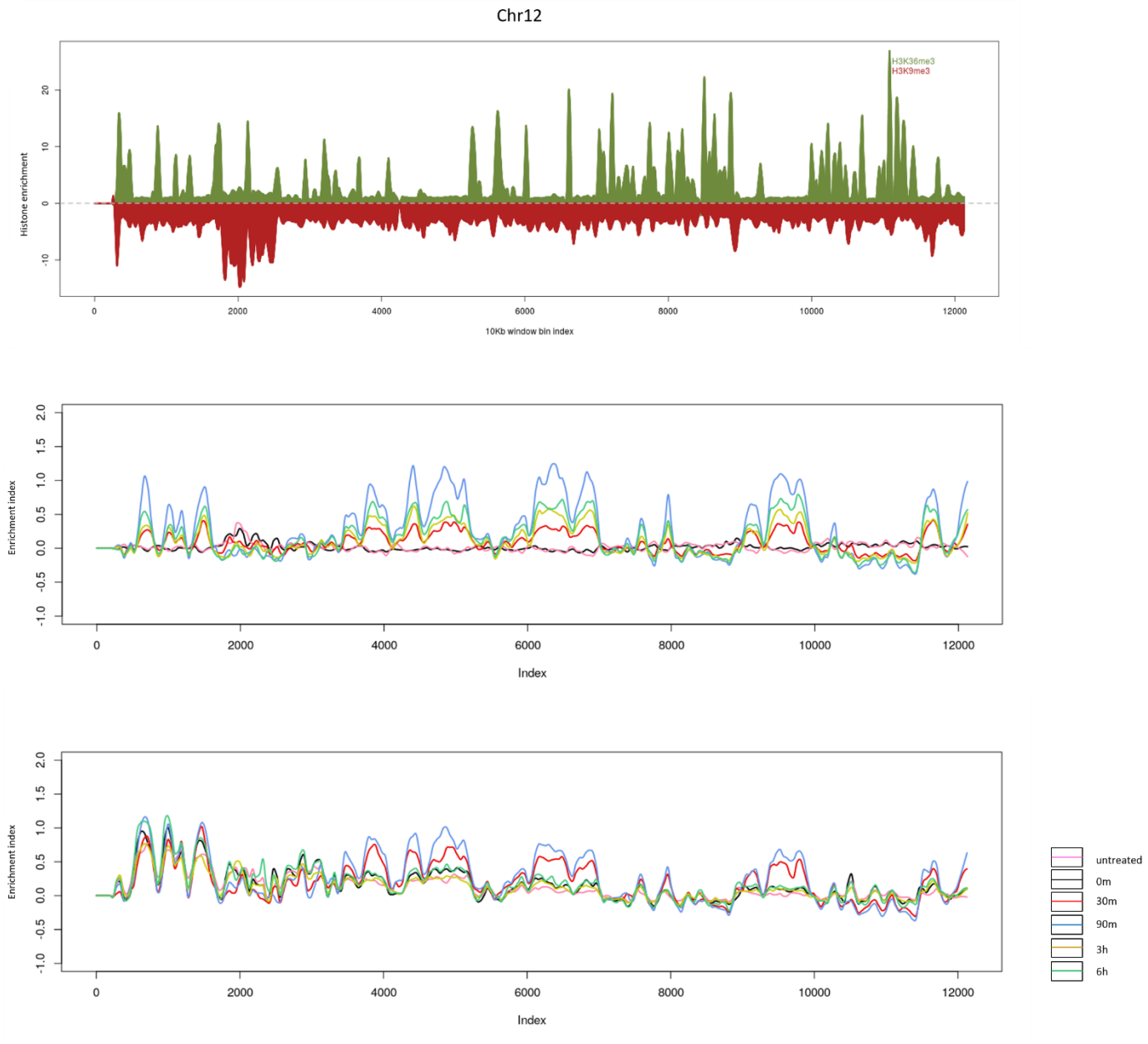
Correlation with histone marks. Chromosome-wide plot of chromosome 9. Top panel: H3K36me3 (green) and H3K9me3 (red) enrichment values; Middle panel: γH2AX enrichment in replicate 1; Bottom panel: γH2AX enrichment in replicate 2.



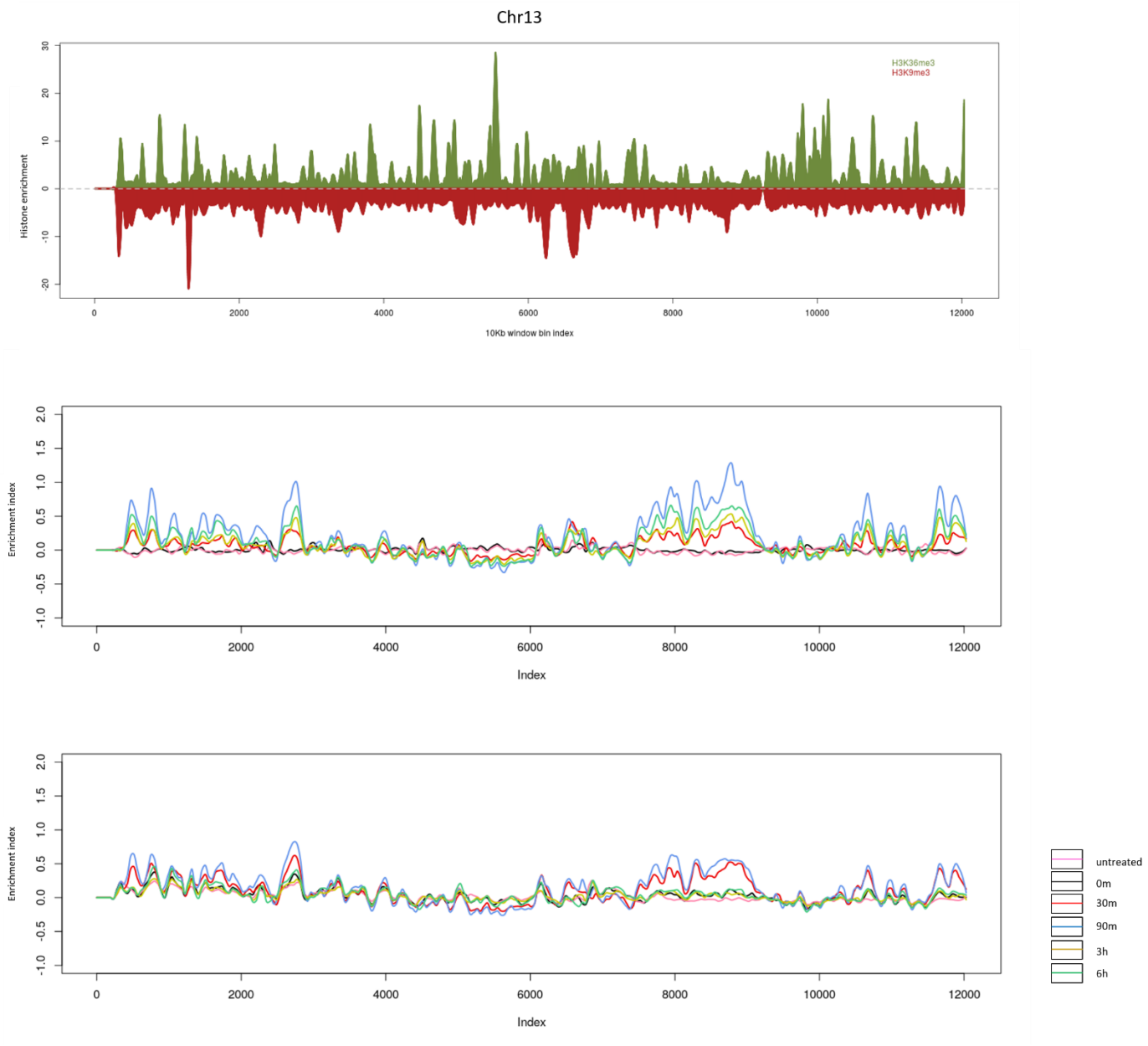
Correlation with histone marks. Chromosome-wide plot of chromosome 10. Top panel: H3K36me3 (green) and H3K9me3 (red) enrichment values; Middle panel: γH2AX enrichment in replicate 1; Bottom panel: γH2AX enrichment in replicate 2.



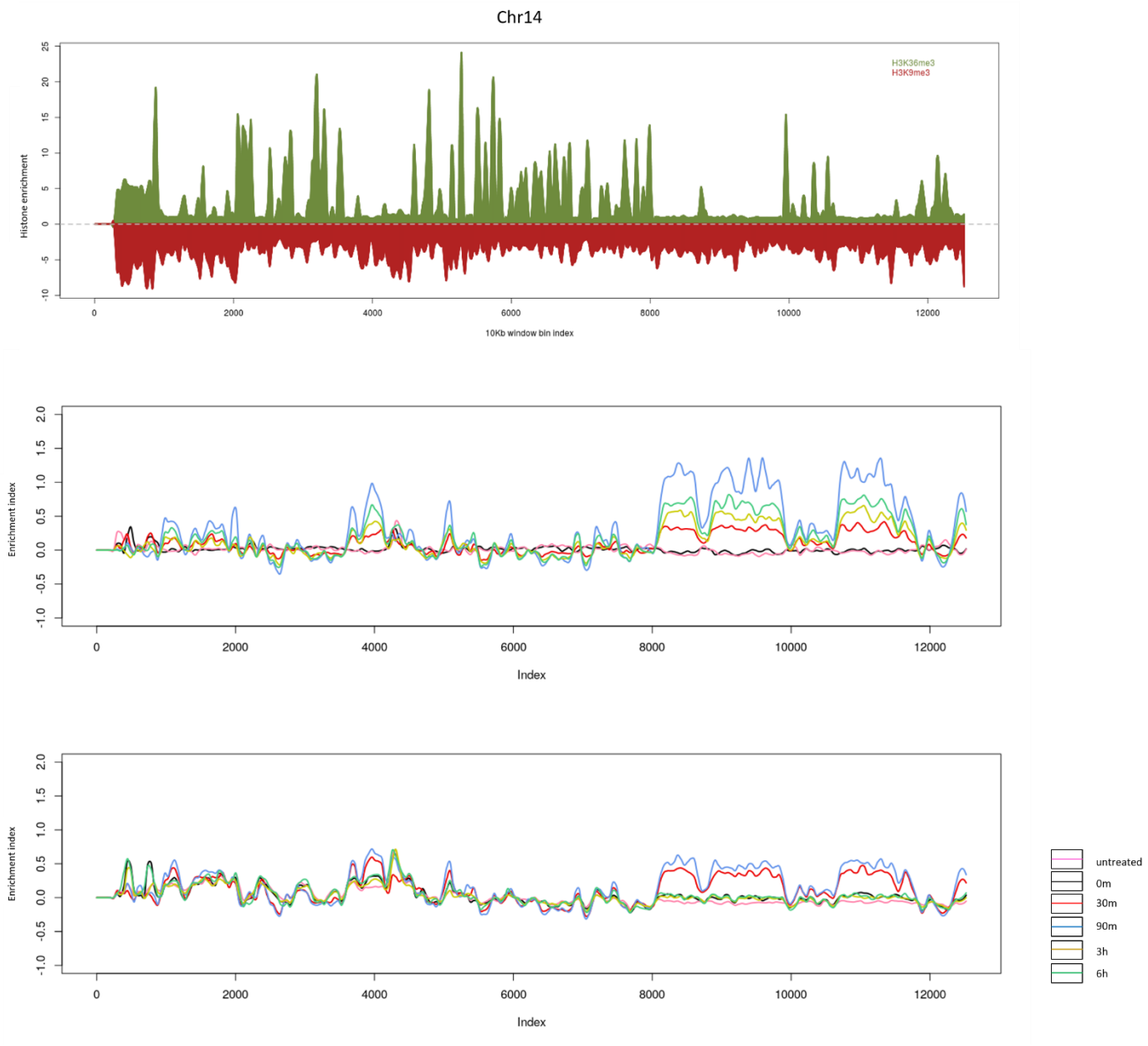
Correlation with histone marks. Chromosome-wide plot of chromosome 11. Top panel: H3K36me3 (green) and H3K9me3 (red) enrichment values; Middle panel: γH2AX enrichment in replicate 1; Bottom panel: γH2AX enrichment in replicate 2.



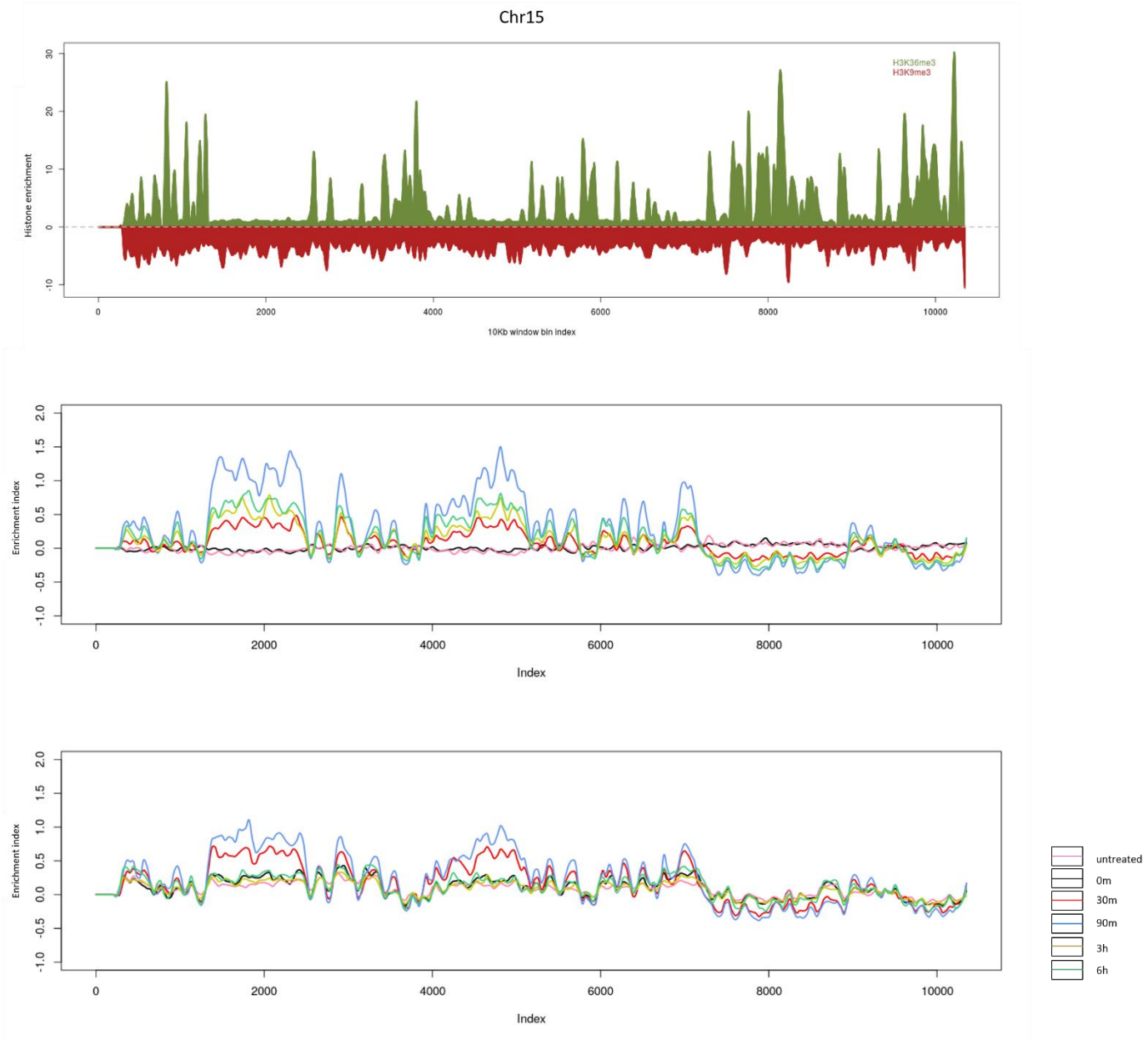
Correlation with histone marks. Chromosome-wide plot of chromosome 12. Top panel: H3K36me3 (green) and H3K9me3 (red) enrichment values; Middle panel: γH2AX enrichment in replicate 1; Bottom panel: γH2AX enrichment in replicate 2.



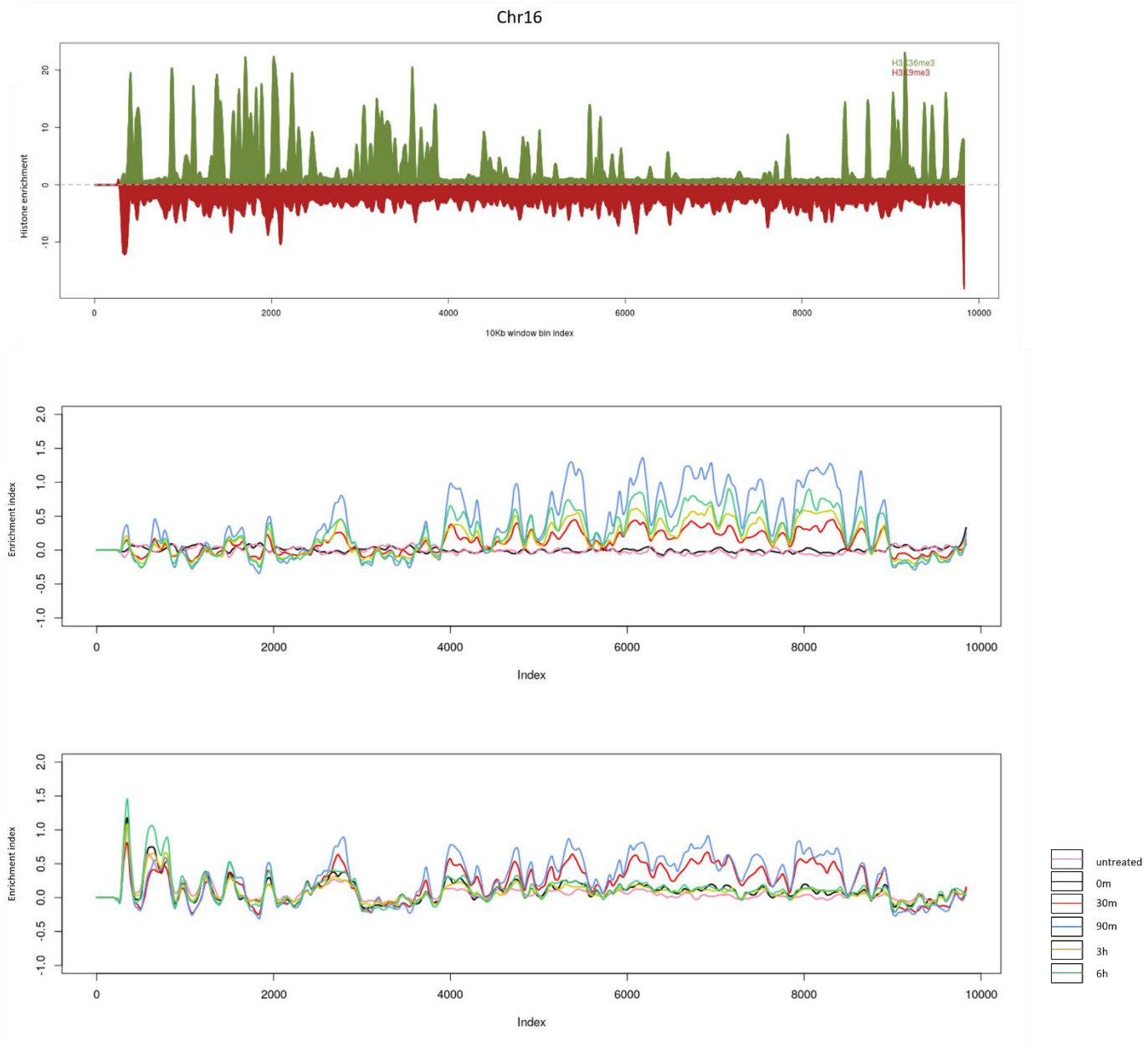
Correlation with histone marks. Chromosome-wide plot of chromosome 13. Top panel: H3K36me3 (green) and H3K9me3 (red) enrichment values; Middle panel: γH2AX enrichment in replicate 1; Bottom panel: γH2AX enrichment in replicate 2.



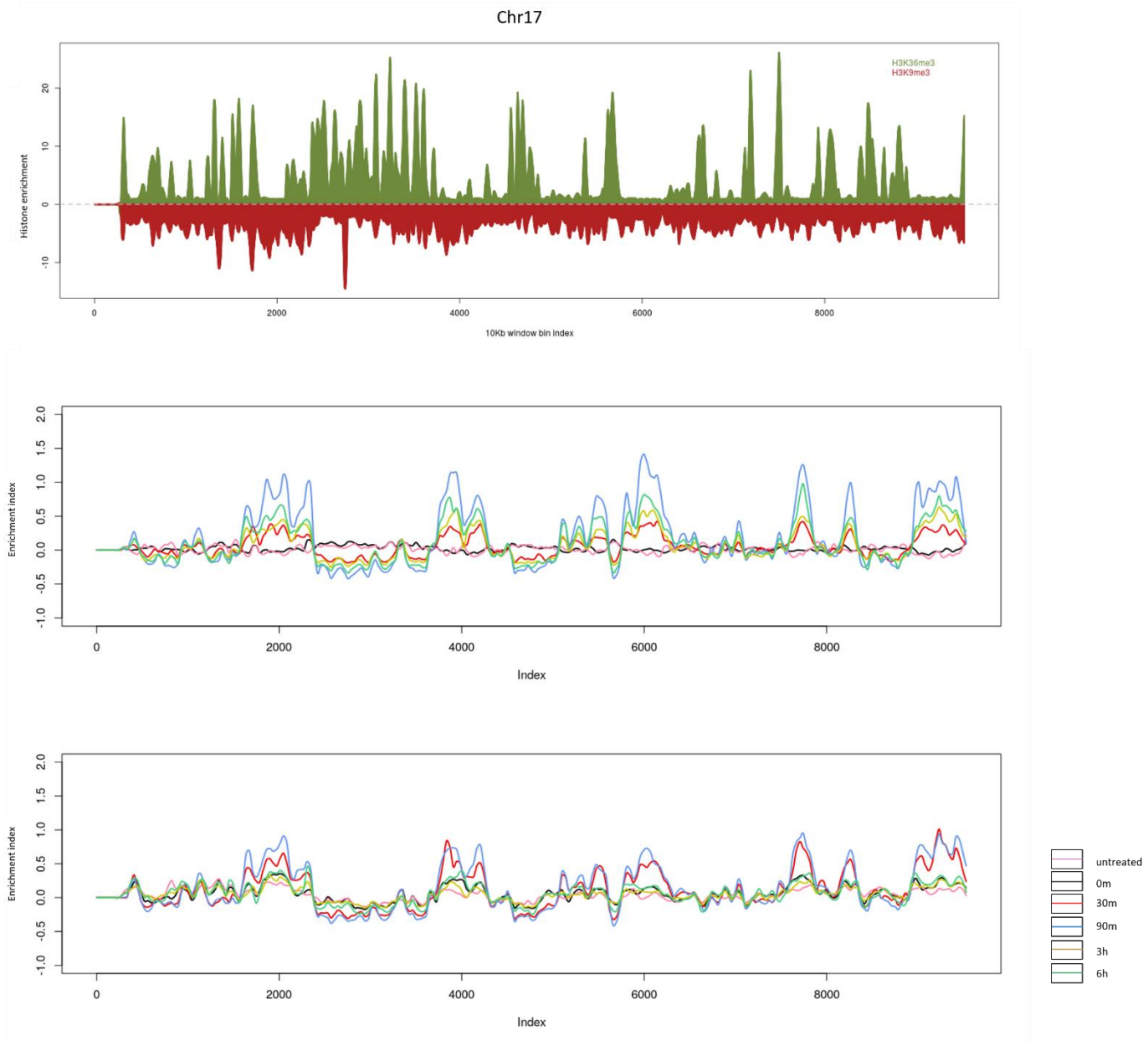
Correlation with histone marks. Chromosome-wide plot of chromosome 14. Top panel: H3K36me3 (green) and H3K9me3 (red) enrichment values; Middle panel: γH2AX enrichment in replicate 1; Bottom panel: γH2AX enrichment in replicate 2.



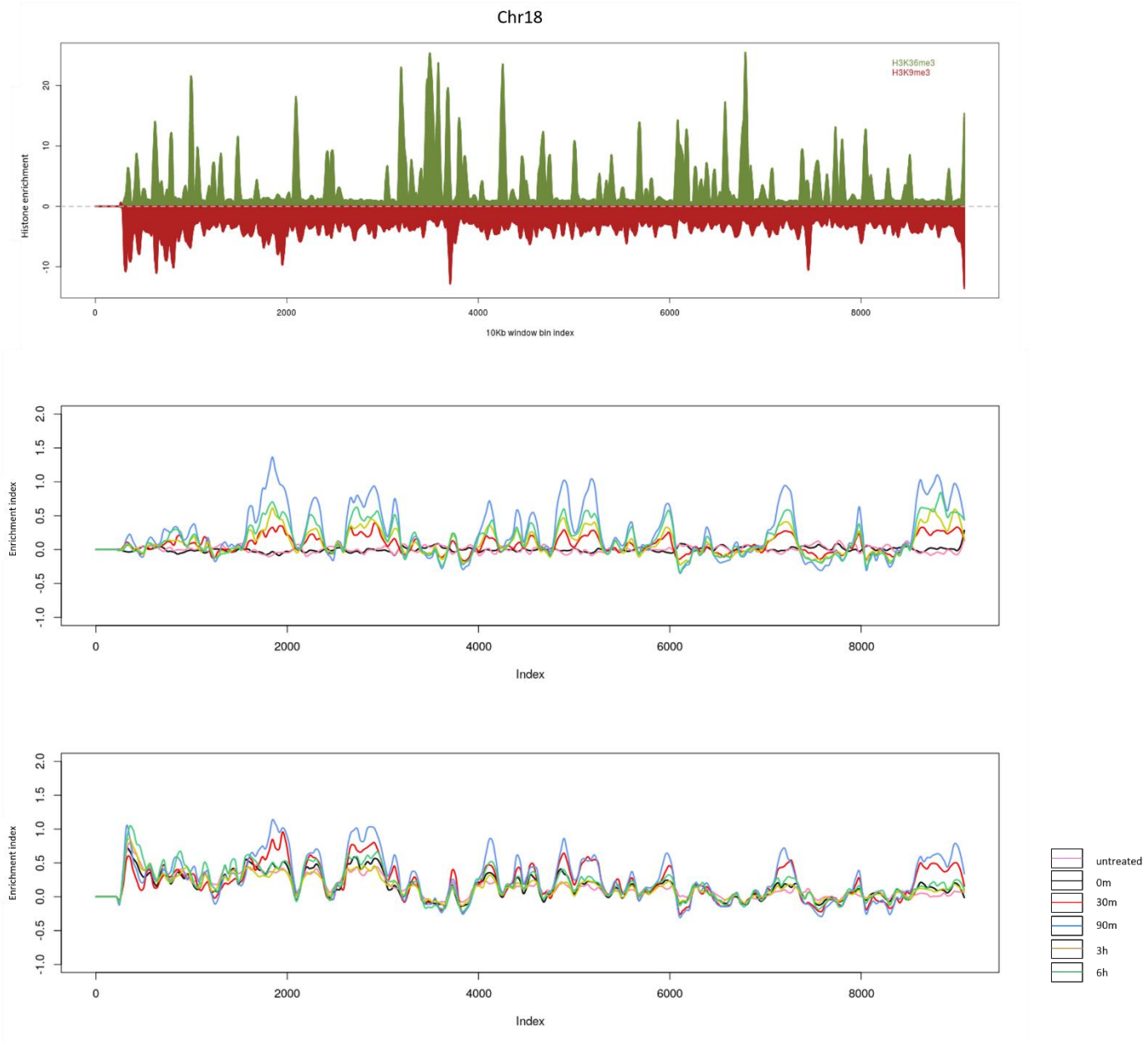
Correlation with histone marks. Chromosome-wide plot of chromosome 15. Top panel: H3K36me3 (green) and H3K9me3 (red) enrichment values; Middle panel: γH2AX enrichment in replicate 1; Bottom panel: γH2AX enrichment in replicate 2.



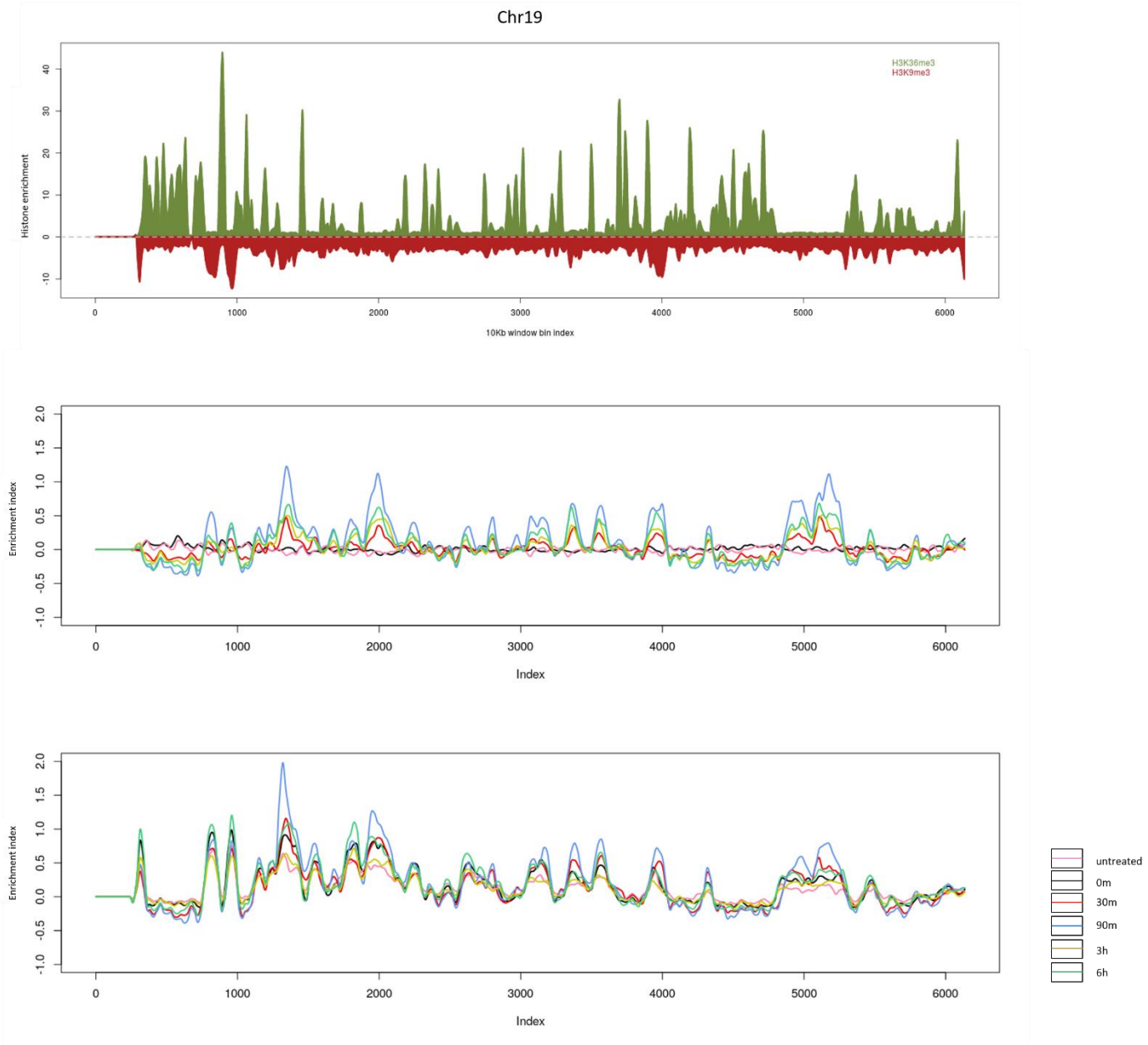
Correlation with histone marks. Chromosome-wide plot of chromosome 16. Top panel: H3K36me3 (green) and H3K9me3 (red) enrichment values; Middle panel: γH2AX enrichment in replicate 1; Bottom panel: γH2AX enrichment in replicate 2.



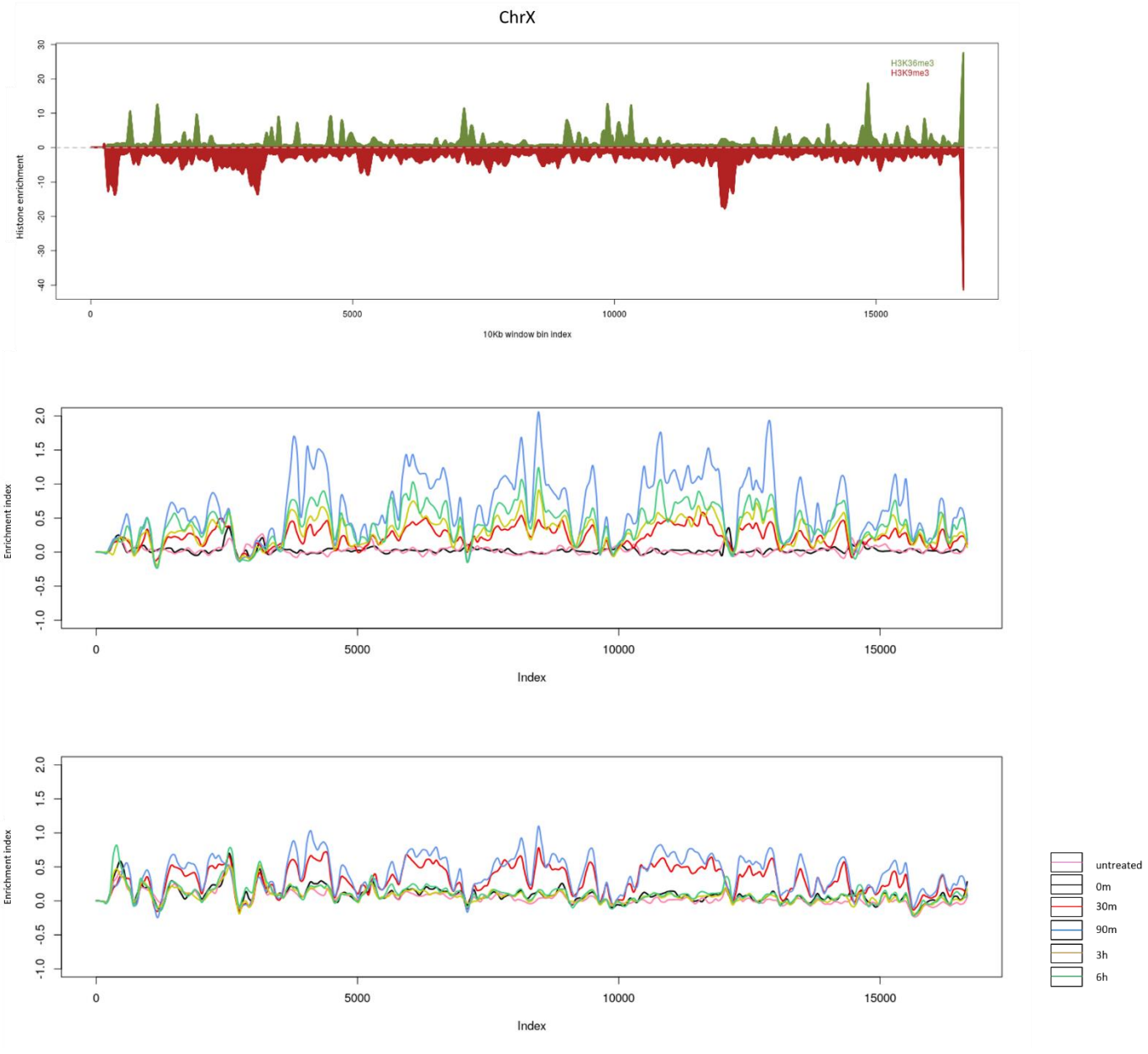
Correlation with histone marks. Chromosome-wide plot of chromosome 17. Top panel: H3K36me3 (green) and H3K9me3 (red) enrichment values; Middle panel: γH2AX enrichment in replicate 1; Bottom panel: γH2AX enrichment in replicate 2.



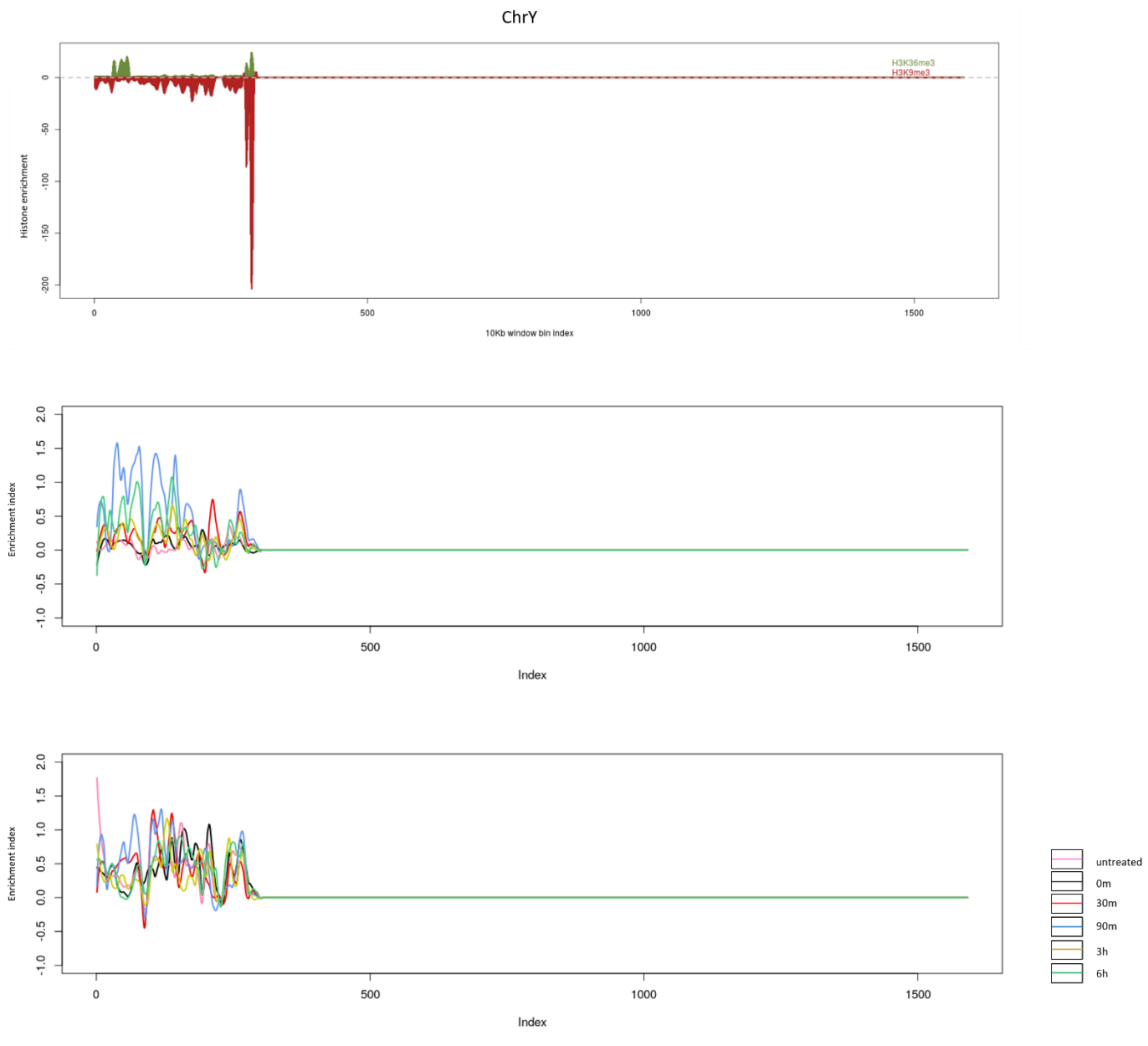
Correlation with histone marks. Chromosome-wide plot of chromosome 18. Top panel: H3K36me3 (green) and H3K9me3 (red) enrichment values; Middle panel: γH2AX enrichment in replicate 1; Bottom panel: γH2AX enrichment in replicate 2.



Correlation with histone marks. Chromosome-wide plot of chromosome 19. Top panel: H3K36me3 (green) and H3K9me3 (red) enrichment values; Middle panel: γH2AX enrichment in replicate 1; Bottom panel: γH2AX enrichment in replicate 2.



Correlation with histone marks. Chromosome-wide plot of chromosome X. Top panel: H3K36me3 (green) and H3K9me3 (red) enrichment values; Middle panel: γH2AX enrichment in replicate 1; Bottom panel: γH2AX enrichment in replicate 2.



Correlation with histone marks. Chromosome-wide plot of chromosome Y. Top panel: H3K36me3 (green) and H3K9me3 (red) enrichment values; Middle panel: γH2AX enrichment in replicate 1; Bottom panel: γH2AX enrichment in replicate 2.

Elucidation of the binding targets of 53BP1 by BioChIP

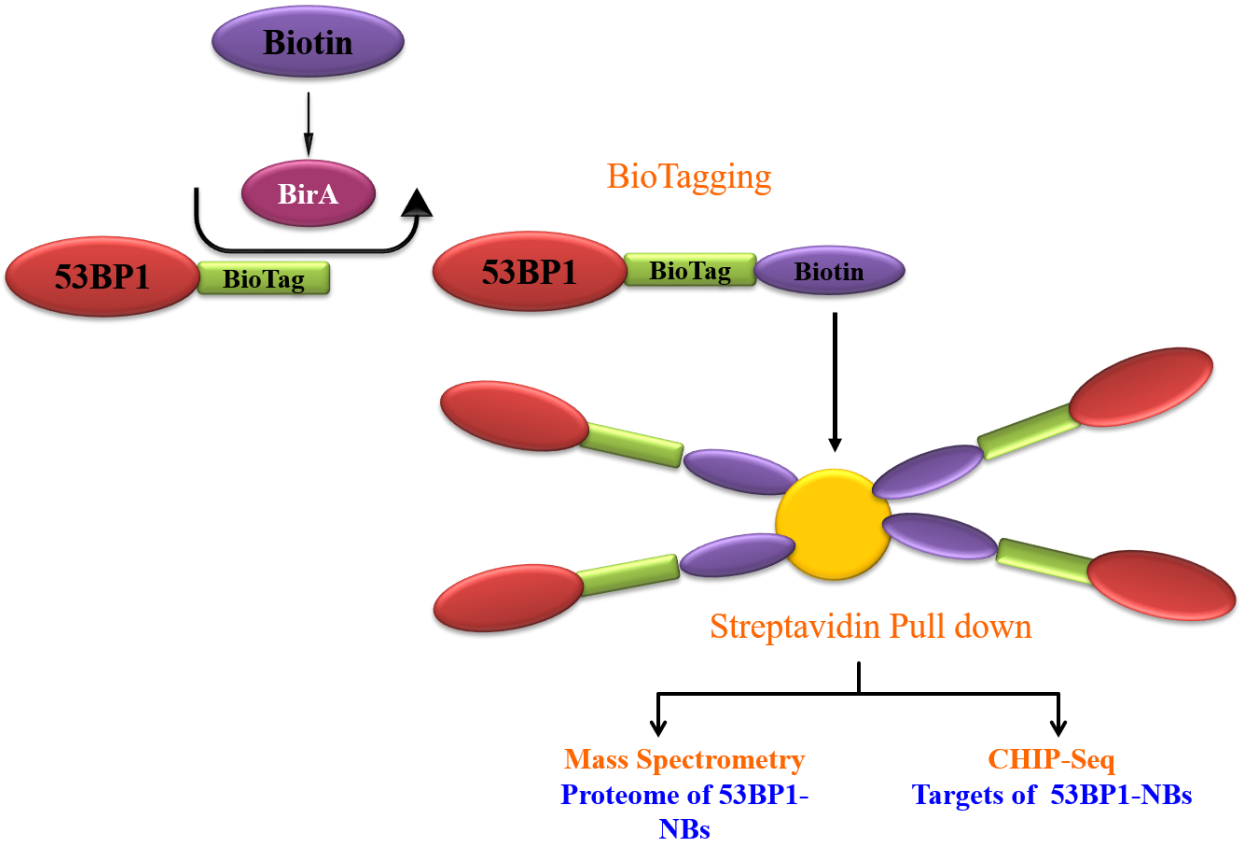
Accurate DNA replication is critical to prevent the transmission of mutations to the daughter cells. Replication is a programmed evet with some regions of the genome replicating earlier followed by other regions. Some regions, termed as chromosomal fragile sites (CFSs), are in regions which contain late-firing replication origins or are located far away from replication origin sites causing these fragile regions to remain unreplicated. Since unreplicated DNA does not have an aberrant structure which would trigger DNA damage checkpoints, the cell can enter mitosis. The sequestering of common fragile sites by 53BP1-nuclear bodies (53BP1-NBs) has been well documented (Harrigan et al., 2011; Lukas et al., 2011). These 53BP1-NBs were proposed to protect fragile sites and shield unrepaired DNA lesions left over from the previous cell cycle from erosion and sequester them until the next cell cycle for repair (Lukas et al., 2011).

Previous studies have proposed that 53BP1-NBs arise before S/G2-phase and progressively diminish as the majority of S phase damage is repaired. The remaining foci may be present at sites containing late replicating or unreplicated DNA which escape detection by cell cycle checkpoints. Chromosomal fragility may result from several genomic properties characterizing a locus (Durkin and Glover, 2007). In my second project, I identified novel genomic loci shielded by 53BP1-NBs in exponentially growing cells and cells arrested in G1 through Next-generation sequencing approaches. I will be elaborating on the preliminary results below.

6.1 Experimental setup and validation

Previous attempts to study the genomic binding pattern of 53BP1 relied on the use of the replication inhibitors such as aphidicolin to induce the formation of 53BP1-NBs. This strategy enabled the identification of 53BP1-NBs binding to fragile sites at G1 but is an artificial way of studying 53BP1 biology. Since antibody variability is a very real concern for immunoprecipitation based methods such as ChIP sequencing and mass spectrometry, we designed a novel system for following the kinetics of 53BP1-NBs in the cell. To this end, we have used the Bio-ID system (Roux et al., 2012) as a proximity-based tool to identify potential interactions of 53BP1 with neighboring proteins and chromatin in vivo. We have fused the Tudor domain (1220-1771 aa) of 53BP1 with the wild type *E.coli* biotin ligase (BirA) and an mCherry tag. In the presence of endogenous biotin in the medium, BirA catalyses biotinylation of the BioTag. This allows for the chromatin or other proteins bound to the protein of interest to be immunoprecipitated onto streptavidin beads for genomic (BioChIP) (He et al., 2010) or proteomic

analyses (Fig. 32A, B). The mCherry tag fused to the Tudor domain permits visualization of the tagging system (Fig. 32C). We are interested in studying 53BP1-NBs in asynchronous cells as well as in different cell-cycle stages. We used lovastatin (40µM) to synchronize cells in G0 followed by release with mevalonate (4mM) to obtain cells in G1 (Fig. 32D).



A

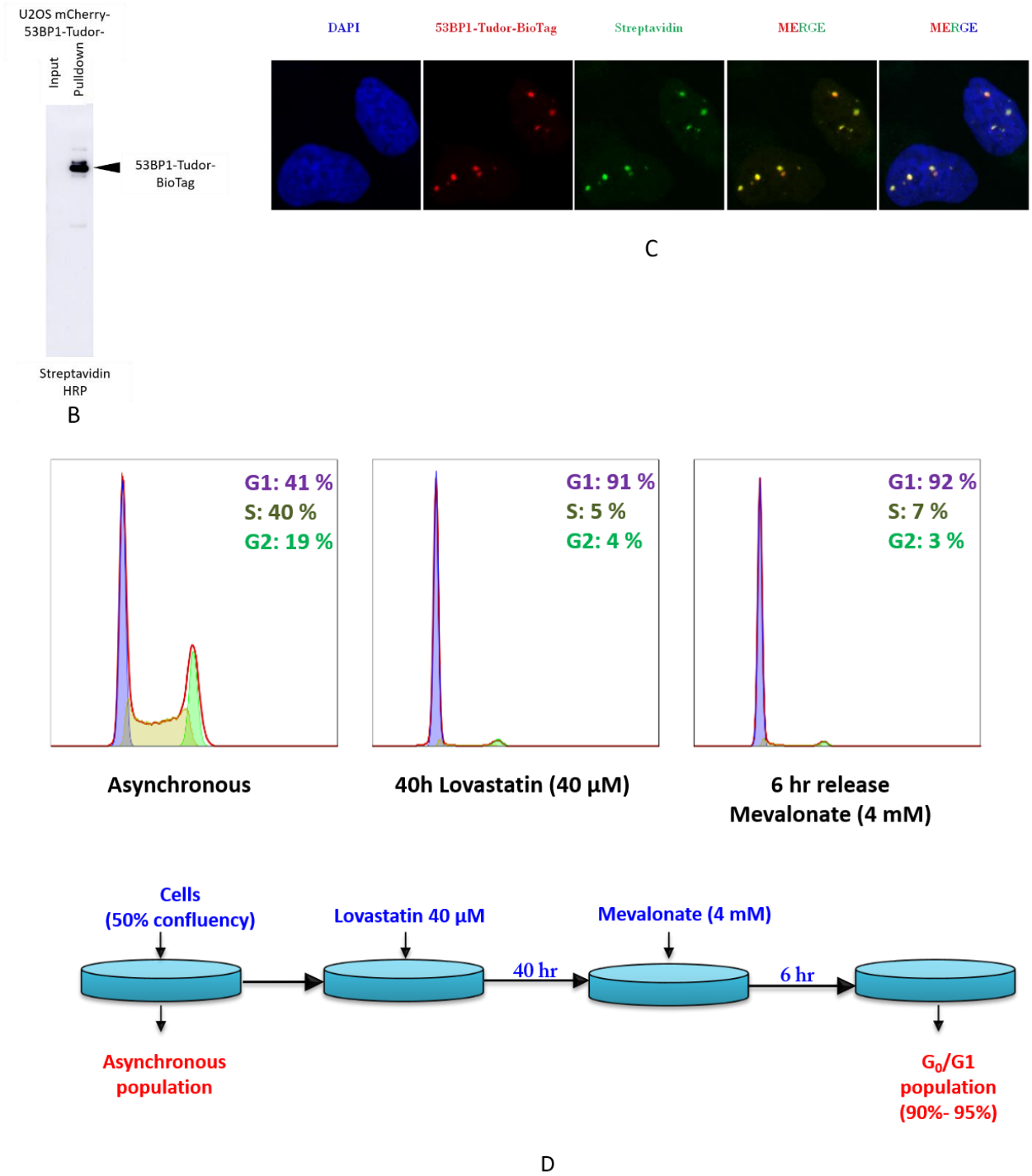
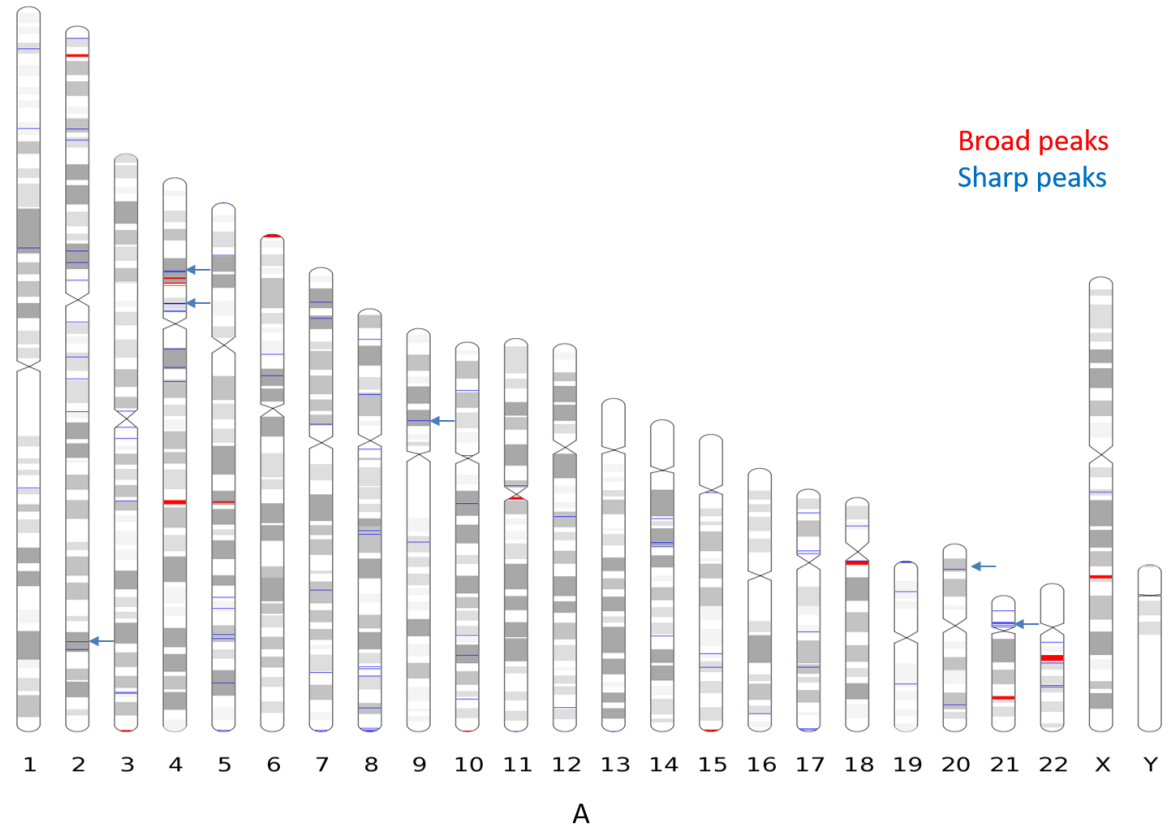


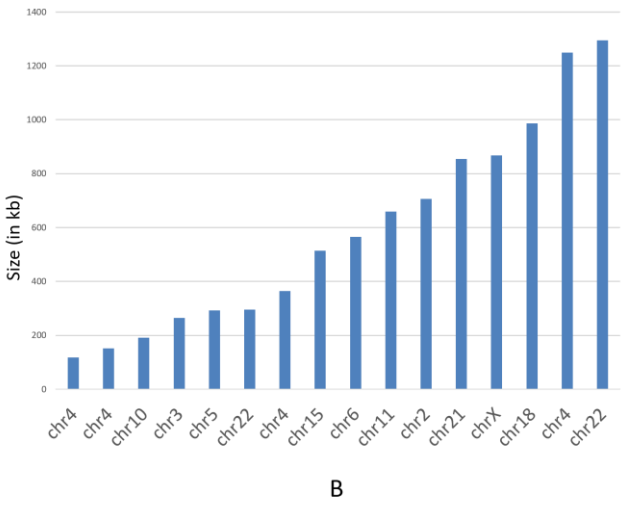
Figure 32. ***Experimental setup and validation****. A. 53BP1 BioTag system. BirA (violet oval) biotinylates BioTag (green block) fused with 53BP1-Tudor domain (red oval). Biotinylated complex is pulled down on streptavidin beads (gold circle). B. Western blot indicating 53BP1-Tudor domain immunoprecipitation from stably transfected U2OS cells. C. Visualization of BioTag system. Immunofluorescence image of cells expressing 53BP1 BioTag. D. Cell cycle analysis of asynchronous cells and after arrest with 40µM Lovastatin for 40h (G0) and release with mevalonate for 6h (G1). Samples collected at indicated timepoints. (Experiments performed by Indrajeet Ghodke; reported here for completion).*

6.2 Identification of binding sites of 53BP1

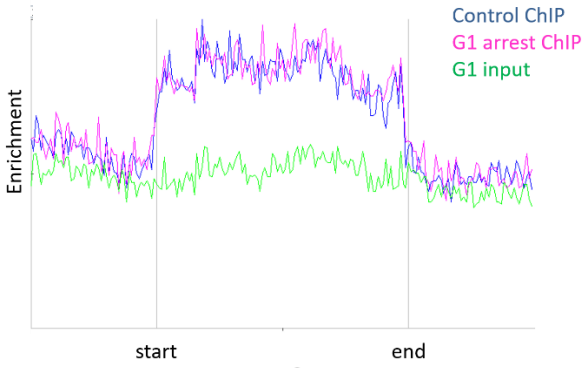
53BP1-NBs have been observed previously in cells during the G1 stage of the cell cycle. To understand the molecular mechanisms regulating their formation we performed a BioChIP on asynchronous and G1-arrested cells. After performing BioChIP, we have attempted to identify the binding targets of 53BP1 genome-wide by sequencing the BioChIP'ed product. Peak calling by MACS showed the presence of numerous very prominent peaks in very close proximity. These peaks were merged together to form a broad genomic binding region. We identified 16 broad genomic binding regions of 53BP1 in U2OS cells stably expressing the Tudor-BioTag-BirA cell line (Fig. 33A). These targets are present in asynchronous cells (2 replicates) as well as in G1 arrested cells. We noticed a high degree of similarity between the peaks from the two replicates of asynchronous cells and the G1 arrested cells. This may be due to a proportion of asynchronous cells being in G1 stage of the cell cycle (Fig. 32D). The targets identified cover long stretches of the genome (average size 0.5Mb) with the largest being 1.3Mb wide (Fig. 33B). These regions have domain-like characteristics with a definite border where the enrichment abruptly ends (Fig. 33C, D, E). These regions will be referred to as “broad peaks” hereafter. Some of these broad peaks span annotated CFS such as FRA1A. In addition to these broad peaks, we were also able to identify 16 previously annotated CFS such as FRA7B in our data which had not been identified by MACS but did have local enrichment. Smaller binding sites (~1kb wide) were also identified which will be referred to as “sharp peaks” (Fig. 33A).


Broad peaks
Sharp peaks
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y

A


Size (in kb)
1400
1200
1000
800
600
400
200
0
chr4 chr4 chr10 chr3 chr5 chr22 chr4 chr15 chr6 chr11 chr2 chr21 chrX chr18 chr4 chr22

B


Control ChIP
G1 arrest ChIP
G1 input
Enrichment
start
end

C

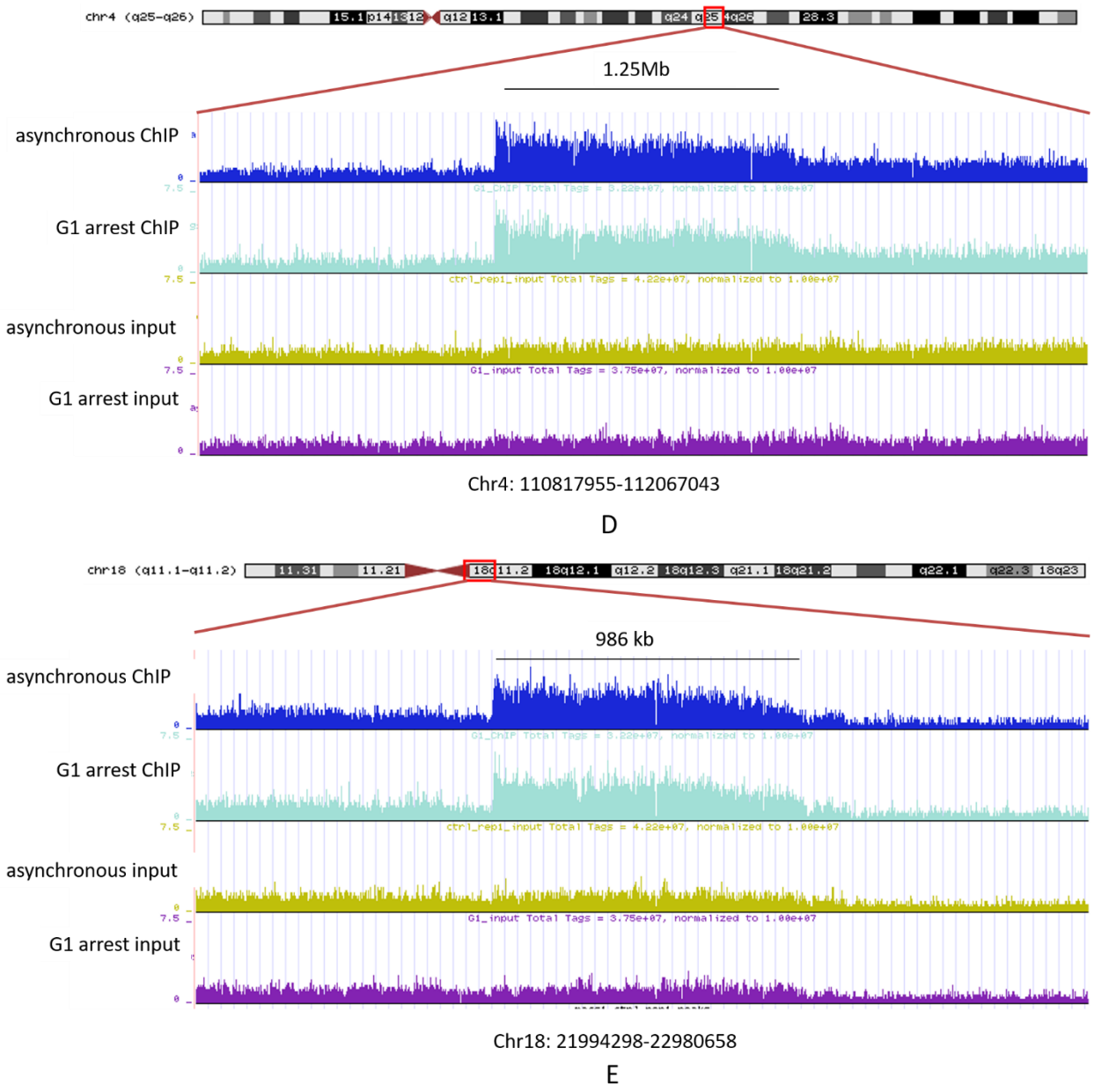
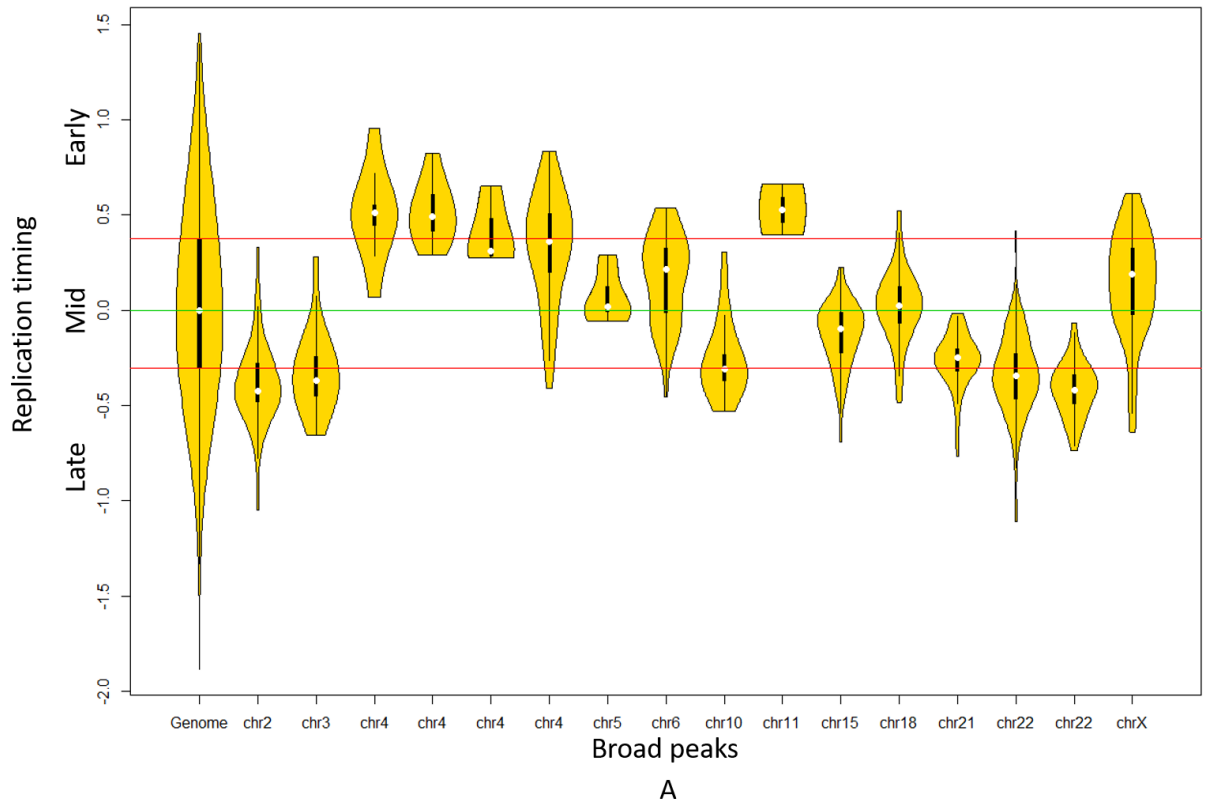


Figure 33. ***53BP1 binding sites****. A. Ideogram showing location of broad (red) and sharp (blue) peaks. Arrows indicate presence of representative sharp peaks. B. Size distribution of broad peaks (in kb). C. Average enrichment in all broad peaks observed. D, E. Representative images of two broad peaks. Data tracks arranged from top to bottom as asynchronous ChIP (dark blue), G1 arrest ChIP (light blue), asynchronous input (gold), G1 arrest input (purple). Coordinates below data track indicate start and end coordinates of broad peaks. Data tracks scaled to same limits. Bar indicates size of the peak observed.*

6.3 Correlation with replication timing

Chromosomal fragility has been associated with regions undergoing delayed replication and under-replication. We decided to verify if the peaks that we identified fell within late-replicating regions. To this

end, we retrieved replication timing data (Hadjadj et al., 2016) for U2OS cells and divided it into three categories signifying early-, mid- and late-replicating regions. We then classified our broad and sharp peaks into these three categories. Since the sharp peaks are smaller than the resolution of the replication timing data (13kb), the score of the complete bin was used for classification. The broad peaks spanned a large interval, so scores for the total peak have been reported. Only a small number of broad peaks were completely present in the early-replicating region of the genome while the rest of the peaks occupied mid- and late-replicating regions (Fig. 34A). This indicates that some of the peaks identified in the analysis may be bonafide targets for 53BP1 binding. Since CFSs are near late-replicating origins, we tested whether the sharp peaks observed were specific late firing origins by comparing the frequency of their occurrence in the late replicating regions. The short peaks however, were more randomly distributed on the replication timing scale and will need to be further validated (Fig. 34B).



A

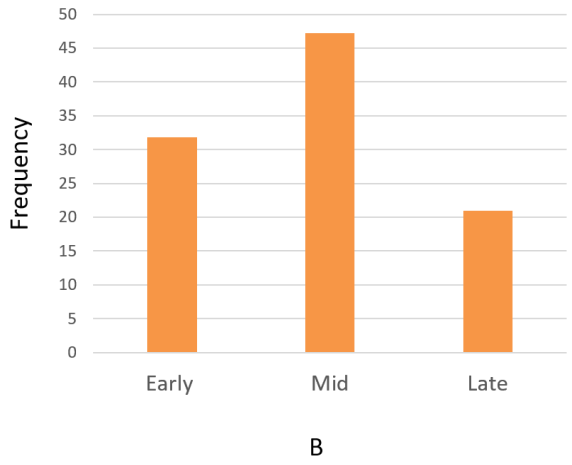
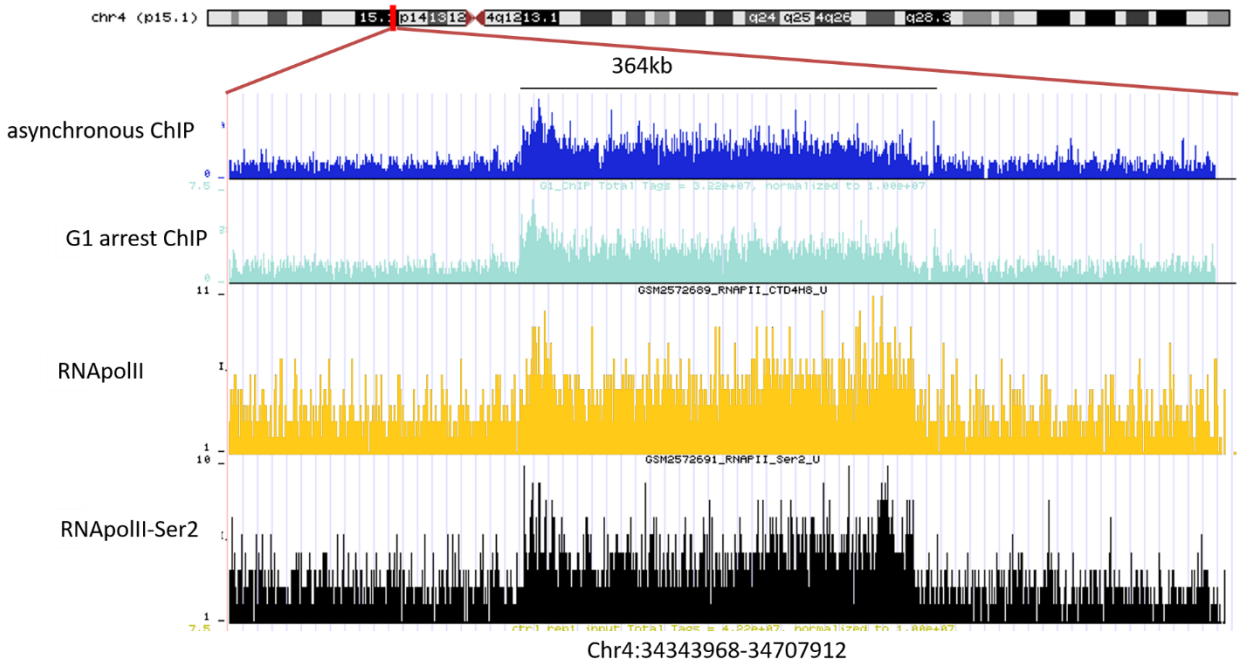


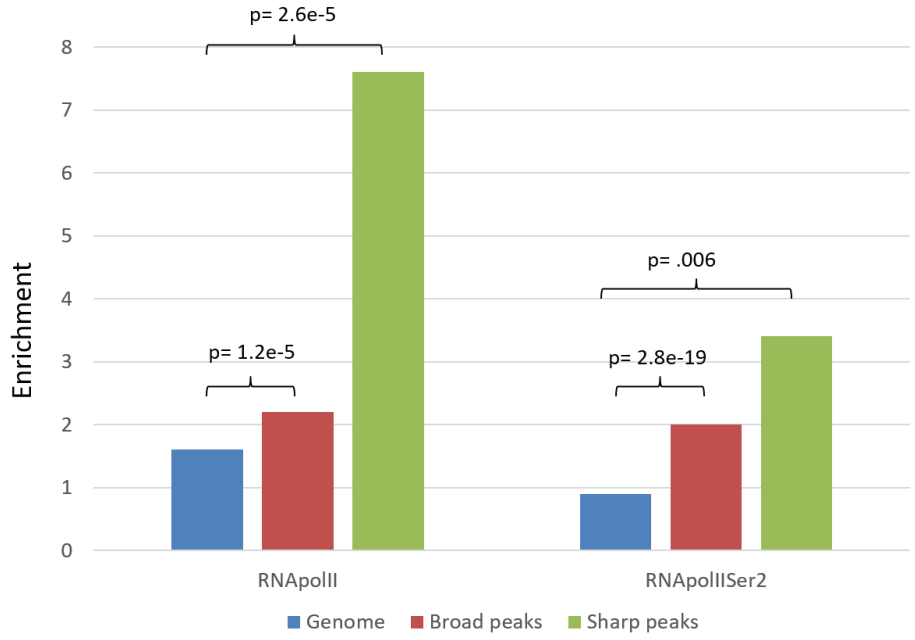
Figure 34. ***Correlation with replication timing****. Replication timing divided into three compartments early (timing index> 0.37639), mid (-0.30598<timing index<0.37639) and late (timing index<-0.30598). A. Replication time distribution of whole genome (first column) and all broad peaks. B. Frequency distribution of sharp peaks in early-, mid- and late-replicating regions.*

6.4 Correlation with transcription status

CFS become more apparent after mild replication stress induced by aphidicolin due to the collision of the transcription machinery of RNA polymerase II complex with replication forks. Longer genes have longer replication and transcription times which may result in collision of these two mechanisms. To explore the role of 53BP1 in transcription, we investigated the recruitment of RNA polymerase II (RNApolII) and elongating RNA polymerase II (RNApolII-Ser2) at 53BP1 binding sites. We retrieved RNApolII and RNApolII-Ser2 ChIP-Seq data (Iannelli et al., 2017) and compared them with our data. We noticed an overlap between enrichment of 53BP1 at some of the broad peaks and enrichment of RNApolII and RNApolII-Ser2 (Fig. 35A). This enrichment was also observed for sharp peaks (Fig. 35B). To check whether the broad peaks we identified also harbor long genes which may be more prone to transcription-replication collisions, we annotated all the genes which were present within the broad peaks. We observed that nearly 35% of the genes which fell within broad peaks had sizes greater than 10kb (Fig. 35C). The longest gene within the broad peaks was the RUNX1 gene with a size of 1.2Mb.


chr4 (p15.1)
364kb
asynchronous ChIP
G1 arrest ChIP
RNApolII
RNApolII-Ser2
GSM2572689_RNAPII_CTD4H8_U
GSM2572691_RNAPII_Ser2_U
Chr4:34343968-34707912

A


p= 2.6e-5
p= 1.2e-5
p= .006
p= 2.8e-19
Enrichment
RNApolII
RNApolIISer2
Genome
Broad peaks
Sharp peaks

B

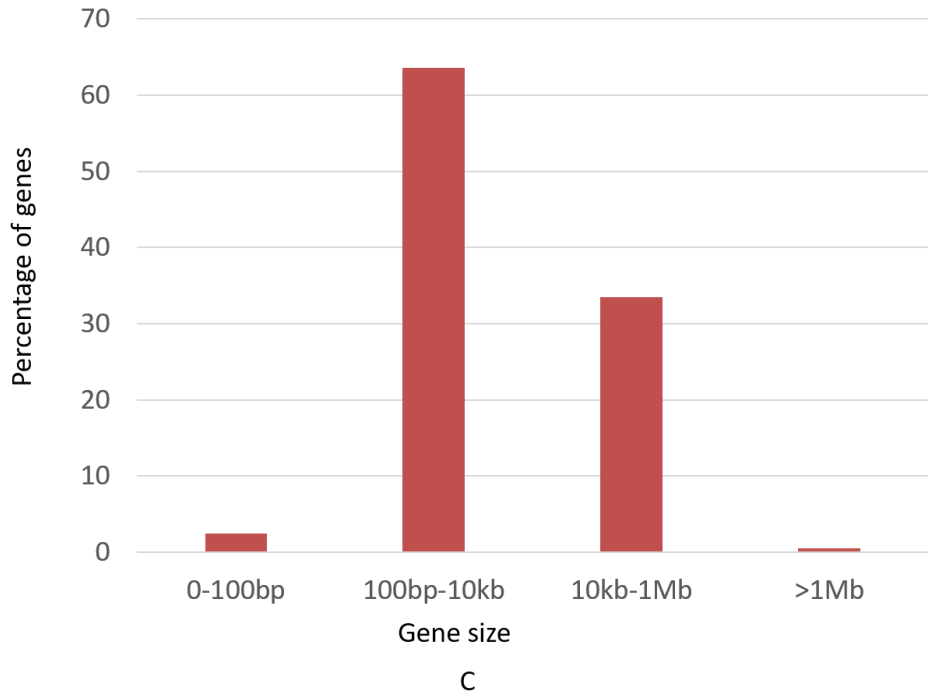


Figure 35. ***Correlation with transcription status.*** *A. Broad peaks are enriched in RNApolII and RNApolII-Ser2. Data tracks arranged from top to bottom as asynchronous ChIP (dark blue), G1 arrest ChIP (light blue), RNApolII (yellow) and RNApolII-Ser2 (black). B. Enrichment of RNApolII and RNApolII-Ser2 in broad peaks (red) and sharp peaks (green) over genome (blue). C. Gene size distribution of genes within broad peaks (expressed as percentage of all genes identified within broad peaks)..*

7. Discussion

Multiple lines of evidence have been presented for replication stress as a major contributor of genome instability. Under-replication of DNA does not trigger a cell cycle checkpoint and the cell proceeds into mitosis. Chromosomal fragile sites (CFSs) are considered to be the most replication-stress sensitive sites in the genome (Durkin and Glover, 2007). Although no single mechanism has been pinpointed to account for this instability, a number of factors have been hypothesized to contribute to their fragility. These factors include late replication, co-occurrence with large genes and a low density of replication origins. 53BP1 nuclear bodies (53BP1-NBs) play a key role in shielding such regions for maintenance of genome integrity (Lukas et al., 2011; Harrigan et al., 2011). I have used a genome-wide approach to uncover the functional significance of (53BP1-NBs) in maintaining genome integrity during replication and exploring the role of 53BP1 during the progression of the cell cycle.

7.1 Identification of novel 53BP1-binding sites

We have established the Bio-ID system in a U2OS cell line where the Tudor domain of 53BP1 is fused with the wild type *E.coli* biotin ligase (BirA) and BioTag. This system circumvents problems related to antibody variability and permits us to track the kinetics of 53BP1 in vivo. We used this system to identify binding sites of 53BP1 in asynchronous cells and G1 arrested cells after performing a BioChIP-sequencing. We have identified 16 broad binding domains of 53BP1 in addition to other short binding domains. In addition to these domains, we have also identified 16 annotated common fragile sites which are also bound by 53BP1. One of these sites is the fragile site FRA1A (1p36) whose deletion is associated with growth and developmental defects. The fragile sites we captured are clinically relevant and further characterization of the broad peaks identified by us is being done.

7.2 Correlation with replication timing

Delayed replication timing has been considered as a primary reason for the presence of CFS in the genome. CFS occur in regions which are difficult to replicate due to a paucity of replication origins, and therefore fragile sites may not complete replication during a given S phase. Late replicating regions contain elements such as heterochromatin which contain repetitive sequences. CFSs are revealed by inducing a mild replication stress through aphidicolin. If such a replicative stress occurs chronically at sub-lethal levels, it can contribute to cellular transformation. We used replication timing data to determine whether the broad peaks we had identified would be bound with 53BP1. We also wanted to determine if the sharp peaks we had identified would be indicative of late origin firing. Our analysis has shown that the majority of the broad peaks we have identified fall in the mid-to late-replicating regions. The sharp peaks had a more random distribution with respect to replication timing. It may be possible that the sharp

and broad peaks with an early replication timing signify early replicating fragile sites. This observation along with our previous observations indicates that the broad peaks (with late replication timing) we have identified may indeed be novel fragile sites which are shielded by 53BP1.

7.3 Correlation with transcription status and gene size

CFSs are also produced due to the collision of the replication and transcription machinery. Our observations have shown that nearly 35% of the genes which fall within the broad peaks have a size greater than 10kb. These genes together account for nearly 80% of the total regions bound by 53BP1, thereby establishing a correlation between 53BP1 binding and longer genes where replication fork stalling might occur. It was reported that more time is taken to transcribe human genes larger than 800kb than the length of one cell cycle while the transcription rates are equal to smaller genes (Helmrich et al., 2011). Since long genes replicate later, this causes a collision between transcription and replication machineries where RNA-DNA hybrid structures called R-loops form. These R-loops provoke CFS instability which is suppressed by RNase H1 (Helmrich et al., 2011).

We also observed a recruitment of RNA polymerase II and elongating RNA polymerase II at some of the 53BP1 binding sites. 53BP1 is an interactor of the tumor suppressor protein p53 through the BRCT domain. It was recently reported that the roles of 53BP1 in DNA damage response and in the regulation of p53-dependent transcription responses are independent and distinct (Cuella-Martin et al., 2016). It would be of interest to study if this accumulation of RNA polymerase II at some of our targets is due to a collision of the transcription-replication machinery or if it is due to p53-mediated transcriptional regulation at these sites.

8. Perspectives

We have identified novel 53BP1 genomic targets in asynchronous and G1-arrested cells. These targets cover long stretches of the genome and some of these regions have also shown an overlap with previously annotated chromosomal fragile sites. The number of 53BP1-NBs observed during immunofluorescence is less than the number of broad peaks that we have identified. This might be due to broad regions of the genome being held together by 53BP1. To test this hypothesis, we are currently validating the broad peaks through fluorescence in situ hybridization experiments. The results from these experiments will help in verifying if the targets identified by the BioChIP-sequencing approach are bound with 53BP1 and will also answer whether these broad peaks are clustered at the focus. If the results do indicate that they are clustered together, the next line of inquiry would be on the occurrence of translocations at these fragile sites.

Recruitment of 53BP1 is an ATM dependent process. Since ATM inhibition decreases 53BP1 accumulation in nuclear bodies, it would be of interest to study whether the decrease in 53BP1 enrichment occurs equally and linearly across all the binding sites or if 53BP1 is still accumulated around specific regions. This which would indicate the inherent fragility of this region and help in identifying potential break points in the chromosome.

It has been reported that 53BP1 is bound to H4K20me2 in a replication dependent manner (Pellegrino et al., 2017). This establishes an inverse relationship between replication status and 53BP1 levels. A decrease in H4K20me2 during cell cycle progression has been established as the cause of diminishing 53BP1 foci during the S phase. Currently, opinion is divided on the identity of the histone modifications which act as a binding target of 53BP1. There is evidence of a complementary role of H3K79me2 and H4K20me2 in binding the Tudor domain to chromatin. Performing BioChIP on our cell line after depletion of Dot1 (responsible for dimethylation of H3K79) and Suv4-20H1, Suv4-20h2 and MMSET (responsible for dimethylation of H4K20) would help in identifying which of these histone modifications serve as targets for recruitment of 53BP1.

R-loops have been reported at fragile sites which increase chromosomal instability (Helmrich et al., 2011). A DNA-RNA hybrid immunoprecipitation and sequencing approach (DRIP-Seq) would help in answering if R-loops are present at 53BP1 binding sites. Another approach would be over-expression or depletion of RNase H1 followed by BioChIP to see a corresponding decrease or increase in 53BP1 enrichment.

We have observed that some of the 53BP1 binding sites appear at chromosome ends. It has been shown that 53BP1 limits mobility of dysfunctional telomeres (Dimitrova et al., 2008). It would be interesting to investigate why there is only a subset of telomeres that are bound by 53BP1. A study of the sheltrin

proteins TRF1 and TRF2 by immunofluorescence would help in identifying if these telomeres are indeed deprotected.

To have a better understanding of the status of 53BP1 throughout the cell cycle, we are also performing another BioChIP on cells which are arrested in the S/G2 phase of the cell cycle. This approach would help in understanding if 53BP1-NBs remain at the fragile sites at mitosis or come off and are reloaded onto the CFSs at the next cell cycle.

The present study has been carried out in an osteosarcoma cell line. To validate our targets, we are also currently generating MCF7 (breast cancer cell line) and IMR90 (human primary lung fibroblast) cell lines with the Bio-ID system. We will perform a similar analysis on these cell lines to investigate if there is a difference in 53BP1 biology between primary cells and cancerous cells. Another advantage of using IMR90 and MCF7 cell lines is that Hi-C maps are available for these cell lines, which will enable us to answer if fragile sites are held in proximity by 53BP1.

9. Concluding remarks

In my doctoral thesis projects, I have used Next-generation sequencing approaches to examine the binding patterns of DNA damage repair factors in response to exogenous damage and at separate phases of the cell cycle.

My first project has shown that DNA damage response is organized at distinct loci in the mouse genome. I have also shown that the damage persists in the heterochromatin and Lamina-associated domains. Both regions have a higher level of chromatin compaction which may be a cause for this persistence of damage. Our future experiments using confocal and super-resolution microscopy will shed light on the DNA damage response specifically in these compartments. We will also map the incidence of double strand breaks in these compartments by means of improved break mapping techniques such as END-Seq and BLISS. We will also perform similar experiments on differentiated mouse cells such as mouse embryonic fibroblasts and 3T3 cells to answer if the results that we have observed in stem cells is a consequence of the chromatin organization in these cells. Our future experiments will reveal the exact role of chromatin organization in regulating the DNA damage response.

My second project has identified novel binding sites of 53BP1 nuclear bodies. These binding sites cover long stretches of genomic DNA and occur in mid-to late-replicating regions. Our future experiments will use the Bio-ID system to study the status of 53BP1 in cancer and primary human cell lines at different cell cycle stages. We are also validating the targets revealed from our BioChIP-Seq experiments by immuno-FISH. Furthermore, we are currently in the process of establishing the BioID system in primary cell lines as well in order to study differences in 53BP1 biology between cancerous and primary cells. Combining the results of BioChIP-Seq with proteomic studies will provide a better understanding of the regulatory aspect of 53BP1 nuclear bodies in the physiological context.

MATERIALS AND METHODS

Mouse embryonic stem cell culture and transfection

J1 mouse embryonic stem (mES) cells were cultured on CD1 feeder cells (MEFs inactivated with 10µg/ml mitomycin C (Sigma)) for two passages before performing experiments. mES cells were separated from feeder cells by trypsinization (0.05% Trypsin, 0.53mM EDTA; Invitrogen) and centrifugation and grown on gelatinized (Gelatin 0.1% in PBS; PAN BIOTECH) tissue culture dishes to avoid signals from contaminating feeder cells. Feeder cells were raised in feeder medium containing Dulbecco's Modified Eagle Medium (DMEM) (4.5g/l glucose) (Sigma), 15% Foetal Calf Serum (ES tested, heat inactivated), 1x Non-Essential Amino Acids (Sigma), 1x Gentamycin and 0.1mM β-Mercaptoethanol. The medium for mES cell culture was the same as the feeder medium with the addition of 2x Leukemia Inhibiting factor (LIF) (Millipore).

Transient transfections were carried out on feeder free mES cells using Lipofectamine 2000 (Life Technologies) according to manufacturer's protocol and cultured on appropriate feeder cells with antibiotic resistance.

Cell were treated with Neocarzinostatin (NCS) for 15 mins at 37°C by addition of NCS to the medium. At the end of the treatment, NCS was removed and plates were washed 2x with PBS and fresh medium was added to the cells. Samples were collected 0min, 30min, 90min, 3h and 6h after the end of treatment with NCS. Untreated cells were used as a control.

U2OS cell culture and transfection

Human osteosarcoma U2OS cells were cultured in DMEM (1g/l glucose) supplemented with 10% foetal Calf Serum and 40µg/ml gentamycin. Transfections were done using FuGene 6 (Promega) according to manufacturer's protocol. Transfected cells were grown in medium containing 800µg/ml of G418 before being sorted by FACS for mCherry positive cells to generate a clonal cell line.

Cells were synchronized to G0 by the addition of 40µM Lovastatin for 40h. Medium was then removed and replaced with 4mM mevalonate for 6h to release cells from G0 and enter into G1 phase. Asynchronous cells were used as a control.

Western Blot

Cells were harvested in RIPA buffer containing HEPES pH 7.6 (50mM), EDTA (1mM), Sodium deoxycholate (0.5%v/v), NP40 (1%), LiCL (0.5M) and supplemented with protease inhibitor cocktail (PIC) and phosSTOP (Roche). Lysis was done on ice for 30min before centrifugation (30min, 14000 rpm, 4°C). Quantification of protein extracts was done by Bradford assay (Bio-Rad) according to manufacturer's protocol. 4x loading buffer was added to protein extracts before boiling for 10 min at 95°C. Samples were loaded on pre-cast 4-12% gradient gels (Invitrogen) and migrated. Wet transfer onto nitrocellulose membrane was done at 400mA for 90 min at 4°C. Membrane was blocked in 5% non-fat dry milk in PBS for 1 h at room temperature. Primary antibodies were diluted in 1% non-fat dry milk in PBS-0.1% Tween 20 (γH2AX, Abcam ab2893, rabbit, 1:1000; αtubulin, Sigma T9026, mouse, 1:10000) and membranes were incubated at 4°C overnight. Membranes were washed 3 times for 10 mins each in PBS-0.1% Tween 20. Secondary antibody was also diluted in 1% non-fat dry milk in PBS-0.1% Tween 20 and membranes were incubated at room temperature for 1 h. Membranes were again washed 3 times for 10 mins each in PBS-0.1% Tween 20 before incubating for 3 mins in chemiluminescent reagent (ECL, Fisher Scientific). Membranes were exposed to photographic film (GE Healthcare) which was then developed.

Cell cycle and apoptosis analysis

Cells were fixed in 70% EtOH overnight at -20°C, rehydrated with PBS, treated with RNase to remove RNA and stained with 50 mg/mL propidium iodide (PI). The acquisition was performed on a FACSCalibur. Apoptosis analysis was performed using Dead Cell Apoptosis Kit with Annexin V FITC and PI (Thermo-Fisher Scientific, Catalogue number 13242). Results were analysed using FlowJo software.

Chromatin Immunoprecipitation (ChIP)

Cells were treated as indicated above and medium was removed. For every timepoint, 35 million cells were used. After the treatment, cells were washed 1x with room temperature PBS to remove any remaining medium. Cells were fixed with fixation buffer (1% paraformaldehyde (PFA) in PBS) at 37°C for 15 mins with shaking every 5 mins to redistribute the fixation buffer. Cross-linking was stopped by adding 0.125M glycine and incubating for a further 10 mins. Cells were washed 2x with cold PBS to remove residual fixation buffer and plates were

transferred to ice for scraping. Scraping was done using 5ml of scraping buffer (PBS with PIC and PhosSTOP). Scraped cells for each sample were pooled together and centrifuged at 4°C for 5 mins at 1200 rpm. Supernatant was removed, and Lysis Buffer I was added and cells were incubated in this buffer for 10 mins on ice with agitation every 5 mins. Cells were centrifuged at 4°C for 2 mins at 6000 rpm. Supernatant was removed and Lysis Buffer II was added to the cell pellet and cells were incubated in this buffer for 10 mins on ice with agitation every 5 mins. Cells were centrifuged at 4°C for 2 mins at 6000 rpm. Supernatant was removed, and sonication buffer was added to the cell pellet. The samples were then transferred into Covaris sonication tubes and sonicated for 15 mins to have fragments between 300 and 500 bp. A 50µl aliquot was reserved for quantification of the material and was reverse crosslinked and used as the Input sample for sequencing.

Immunoprecipitation

50µl of sheep anti-rabbit magnetic IgG Dynabeads (ThermoFisher Scientific) was used for each pull-down reaction. Beads were washed 3 times with cold PBS containing 5mg/ml of BSA with each wash of 5 mins on a rotating wheel. Beads were then blocked with PBS-BSA (5mg/ml) for 30 mins at 4°C before being coated with 2.5µg of γH2AX antibody (ab2893; Abcam) for 4h at 4°C. Immunoprecipitation was carried out on 30µg of sonicated material. Volume sufficient for this quantity of material was diluted in ChIP dilution buffer and immunoprecipitated overnight at 4°C. Samples were then washed sequentially: 2x in SDS wash buffer, 1x in high salt buffer, 1x in LiCl buffer and 1x in TE buffer. All washes were done on a Dynal Magnetic rack (Invitrogen). The immunoprecipitated chromatin was incubated with elution buffer overnight at 65°C with vigorous shaking to elute DNA fragments. The eluted fragments were further purified on Quiagen PCR Cleanup columns according to manufacturer's protocol. The eluate was quantified using Qubit dsDNA High Sensitivity Assay Kit and used for preparing sequencing libraries at the IGBMC Sequencing platform.

BioChIP

A protocol similar to ChIP was performed with the following modifications. 50 µl of magnetic Protein A beads (Invitrogen, 10002D) were used per reaction for preclearing the sonicated chromatin. The beads were washed 3x in PBS-BSA (1%) and then added to diluted chromatin and incubated at 4°C for 1 h. 50 µl Streptavidin M-280 magnetic Dynabeads (11205D) were used per reaction and were blocked for 1 h with PBS-BSA (1%) before being added to

precleared chromatin. Immunoprecipitation was carried out on 30µg of sonicated material overnight at 4°C.

Buffers used in ChIP and BioChIP:

Lysis Buffer I

Tris-HCl pH 8.0	10mM
Triton X-100	0.25% v/v
EDTA	100mM

Lysis Buffer II

Tris-HCl pH 8.0	10mM
NaCl	200mM
EDTA	10mM

Sonication Buffer

Tris-HCl pH 8.0	10mM
NaCl	100mM
EDTA	1mM
SDS	1% v/v

SDS Wash Buffer

SDS	2% v/v (in water)

High Salt Buffer

HEPES pH 7.5	50mM
NaCl	500mM
EDTA	1mM
Sodium deoxycholate	0.1% v/v
Triton X-100	1% v/v

LiCl Buffer

Tris-HCl pH 8	10mM
LiCl	250mM
EDTA	1mM
NP40	0.5% v/v
Sodium deoxycholate	0.5% v/v

TE Buffer

Tris-HCl pH 7.5	10mM
EDTA	1mM

Elution Buffer

Tris-HCl pH8	50mM
EDTA	10mM
SDS	1% v/v

Cell preparation for BLESS

Cells were treated as previously indicated, trypsinized, washed 2x with PBS and fixed with 2% PFA for 30 mins. The samples were then sent to the Laboratory of Bioinformatics and Systems Biology, Centre of New Technologies, University of Warsaw, Warsaw, Poland, for the subsequent steps of the protocol and sequencing. Preliminary analysis for distribution of double strand breaks was done in the lab of Dr. Maga Rowicka, UTMB, Galveston, Texas, USA. Further analysis steps are described below.

Next-generation sequencing and data analyses

The libraries were sequenced on the Illumina Hiseq 2500 as single-end 50 base reads following Illumina's instructions. Reads were mapped to the mouse genome (mm9) assembly (for γH2AX project) and the human genome (hg38) (for BioChIP project) using Bowtie. Only uniquely mapped reads were considered for further analysis. Peak calling was done using MACS at the default settings. Further analysis was done on the local instance of the data analysis platform

Galaxy (Galaxeast). Peak annotation was done using HOMER. Distribution of peaks in genomic features was calculated using the CEAS (Cis-regulatory element annotation system) tool. Peaks were manually checked on UCSC Genome Browser to identify true peaks and exclude false positives. Reads per kilobase per million reads (RPKM) were calculated for non-overlapping 10 kb genomic intervals for all sequence tracks using BamCoverage from the DeepTools suite.

Analysis of γH2AX data

Enrichment index was calculated as [(ChIP RPKM $_{interval}$/ Input RPKM $_{interval}$) - (ChIP RPKM $_{Average}$/ Input RPKM $_{Average}$)] (Natale et al., 2017). Calculations and plotting were done in R. Positive values represent enrichment while negative values indicate depletion at the bin. Data tracks were smoothed by LOESS (Locally weighted scatter-plot smoother) (span=.025). For deciding threshold values for defining euchromatin and heterochromatin, a Poisson distribution of the reads was done and threshold was calculated to separate the background from the signal.

Specific analysis of BioChIP data

After manually validating all peaks called by MACS, peaks located in close proximity (<5kb) were merged together. Peaks not in close proximity were treated as sharp peaks.

Replication timing data was split into three categories based on the values of the first and third quartiles of the replication timing. Regions with replication timing values less than the first quartile value (< -0.30598) were categorized as late replicating regions while values greater than the third quartile value (>0.37639) were categorized as early replicating regions. All intermediate values of replication timing were categorized as mid-replicating regions.

Chromosome ideograms were generated with Ideographica (Kin and Ono, 2007).

BIBLIOGRAPHY

Ababou, M., Dutertre, S., Lecluse, Y., Onclercq, R., Chatton, B., and Amor-Gueret, M. (2000). ATM-dependent phosphorylation and accumulation of endogenous BLM protein in response to ionizing radiation. Oncogene *19*, 5955-5963.

Acharya, S., Foster, P.L., Brooks, P., and Fishel, R. (2003). The coordinated functions of the E. coli MutS and MutL proteins in mismatch repair. Molecular cell *12*, 233-246.

Acs, K., Luijsterburg, M.S., Ackermann, L., Salomons, F.A., Hoppe, T., and Dantuma, N.P. (2011). The AAA-ATPase VCP/p97 promotes 53BP1 recruitment by removing L3MBTL1 from DNA double-strand breaks. Nature structural & molecular biology *18*, 1345-1350.

Adam, S., Polo, S.E., and Almouzni, G. (2013). Transcription recovery after DNA damage requires chromatin priming by the H3.3 histone chaperone HIRA. Cell *155*, 94-106.

Adams, B.R., Hawkins, A.J., Povirk, L.F., and Valerie, K. (2010). ATM-independent, high-fidelity nonhomologous end joining predominates in human embryonic stem cells. Aging (Albany NY) *2*, 582-596.

Adar, S., Hu, J., Lieb, J.D., and Sancar, A. (2016). Genome-wide kinetics of DNA excision repair in relation to chromatin state and mutagenesis. Proceedings of the National Academy of Sciences of the United States of America *113*, E2124-2133.

Aguilera, A., and Garcia-Muse, T. (2013). Causes of genome instability. Annual review of genetics *47*, 1-32.

Ahuja, A.K., Jodkowska, K., Teloni, F., Bizard, A.H., Zellweger, R., Herrador, R., Ortega, S., Hickson, I.D., Altmeyer, M., Mendez, J.*, et al.* (2016). A short G1 phase imposes constitutive replication stress and fork remodelling in mouse embryonic stem cells. Nature communications *7*, 10660.

Aladjem, M.I., Spike, B.T., Rodewald, L.W., Hope, T.J., Klemm, M., Jaenisch, R., and Wahl, G.M. (1998). ES cells do not activate p53-dependent stress responses and undergo p53-independent apoptosis in response to DNA damage. Curr Biol *8*, 145-155.

Andres, S.N., Modesti, M., Tsai, C.J., Chu, G., and Junop, M.S. (2007). Crystal structure of human XLF: a twist in nonhomologous DNA end-joining. Molecular cell *28*, 1093-1101.

Arlt, M.F., Durkin, S.G., Ragland, R.L., and Glover, T.W. (2006). Common fragile sites as targets for chromosome rearrangements. DNA Repair (Amst) *5*, 1126-1135.

Arlt, M.F., Miller, D.E., Beer, D.G., and Glover, T.W. (2002). Molecular characterization of FRAXB and comparative common fragile site instability in cancer cells. Genes, chromosomes & cancer *33*, 82-92.

Arlt, M.F., Xu, B., Durkin, S.G., Casper, A.M., Kastan, M.B., and Glover, T.W. (2004). BRCA1 is required for common-fragile-site stability via its G2/M checkpoint function. Molecular and cellular biology *24*, 6701-6709.

Armstrong, L. (2012). Epigenetic control of embryonic stem cell differentiation. Stem

Cell Rev *8*, 67-77.

Assenmacher, N., and Hopfner, K.P. (2004). MRE11/RAD50/NBS1: complex activities. Chromosoma *113*, 157-166.

Aymard, F., Aguirrebengoa, M., Guillou, E., Javierre, B.M., Bugler, B., Arnould, C., Rocher, V., Iacovoni, J.S., Biernacka, A., Skrzypczak, M.*, et al.* (2017). Genome-wide mapping of long-range contacts unveils clustering of DNA double-strand breaks at damaged active genes. Nature structural & molecular biology *24*, 353-361.

Aymard, F., Bugler, B., Schmidt, C.K., Guillou, E., Caron, P., Briois, S., Iacovoni, J.S., Daburon, V., Miller, K.M., Jackson, S.P.*, et al.* (2014). Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. Nature structural & molecular biology *21*, 366-374.

Azqueta, A., Meier, S., Priestley, C., Gutzkow, K.B., Brunborg, G., Sallette, J., Soussaline, F., and Collins, A. (2011). The influence of scoring method on variability in results obtained with the comet assay. Mutagenesis *26*, 393-399.

Bachrati, C.Z., and Hickson, I.D. (2009). Dissolution of double Holliday junctions by the concerted action of BLM and topoisomerase IIIalpha. Methods Mol Biol *582*, 91-102.

Bak, S.T., Sakellariou, D., and Pena-Diaz, J. (2014). The dual nature of mismatch repair as antimutator and mutator: for better or for worse. Front Genet *5*, 287.

Bakkenist, C.J., and Kastan, M.B. (2003). DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. Nature *421*, 499-506.

Ball, H.L., Ehrhardt, M.R., Mordes, D.A., Glick, G.G., Chazin, W.J., and Cortez, D. (2007). Function of a conserved checkpoint recruitment domain in ATRIP proteins. Molecular and cellular biology *27*, 3367-3377.

Banath, J.P., Banuelos, C.A., Klokov, D., MacPhail, S.M., Lansdorp, P.M., and Olive, P.L. (2009). Explanation for excessive DNA single-strand breaks and endogenous repair foci in pluripotent mouse embryonic stem cells. Exp Cell Res *315*, 1505-1520.

Banuelos, C.A., Banath, J.P., MacPhail, S.H., Zhao, J., Eaves, C.A., O'Connor, M.D., Lansdorp, P.M., and Olive, P.L. (2008). Mouse but not human embryonic stem cells are deficient in rejoining of ionizing radiation-induced DNA double-strand breaks. DNA Repair (Amst) *7*, 1471-1483.

Baranello, L., Kouzine, F., Wojtowicz, D., Cui, K., Przytycka, T.M., Zhao, K., and Levens, D. (2014). DNA break mapping reveals topoisomerase II activity genome-wide. Int J Mol Sci *15*, 13111-13122.

Barlow, J.H., Faryabi, R.B., Callen, E., Wong, N., Malhowski, A., Chen, H.T., Gutierrez-Cruz, G., Sun, H.W., McKinnon, P., Wright, G.*, et al.* (2013). Identification of early replicating fragile sites that contribute to genome instability. Cell *152*, 620-632.

Bartek, J., and Lukas, J. (2003). Chk1 and Chk2 kinases in checkpoint control and cancer. Cancer Cell *3*, 421-429.

Bassing, C.H., and Alt, F.W. (2004). H2AX may function as an anchor to hold broken chromosomal DNA ends in close proximity. Cell cycle (Georgetown, Tex *3*, 149-153.

Beck, C., Robert, I., Reina-San-Martin, B., Schreiber, V., and Dantzer, F. (2014). Poly(ADP-ribose) polymerases in double-strand break repair: focus on PARP1, PARP2 and PARP3. Exp Cell Res *329*, 18-25.

Bekker-Jensen, S., Lukas, C., Kitagawa, R., Melander, F., Kastan, M.B., Bartek, J., and Lukas, J. (2006). Spatial organization of the mammalian genome surveillance machinery in response to DNA strand breaks. The Journal of cell biology *173*, 195-206.

Bekker-Jensen, S., Lukas, C., Melander, F., Bartek, J., and Lukas, J. (2005). Dynamic assembly and sustained retention of 53BP1 at the sites of DNA damage are controlled by Mdc1/NFBD1. The Journal of cell biology *170*, 201-211.

Bekker-Jensen, S., Rendtlew Danielsen, J., Fugger, K., Gromova, I., Nerstedt, A., Lukas, C., Bartek, J., Lukas, J., and Mailand, N. (2010). HERC2 coordinates ubiquitin-dependent assembly of DNA repair factors on damaged chromosomes. Nature cell biology *12*, 80-86; sup pp 81-12.

Benhamou, S., and Sarasin, A. (2000). Variability in nucleotide excision repair and cancer risk: a review. Mutation research *462*, 149-158.

Berezney, R., and Wei, X. (1998). The new paradigm: integrating genomic function and nuclear architecture. J Cell Biochem Suppl *30-31*, 238-242.

Bermudez, V.P., Lindsey-Boltz, L.A., Cesare, A.J., Maniwa, Y., Griffith, J.D., Hurwitz, J., and Sancar, A. (2003). Loading of the human 9-1-1 checkpoint complex onto DNA by the checkpoint clamp loader hRad17-replication factor C complex in vitro. Proceedings of the National Academy of Sciences of the United States of America *100*, 1633-1638.

Bester, A.C., Schwartz, M., Schmidt, M., Garrigue, A., Hacein-Bey-Abina, S., Cavazzana-Calvo, M., Ben-Porat, N., Von Kalle, C., Fischer, A., and Kerem, B. (2006). Fragile sites are preferential targets for integrations of MLV vectors in gene therapy. Gene Ther *13*, 1057-1059.

Beucher, A., Birraux, J., Tchouandong, L., Barton, O., Shibata, A., Conrad, S., Goodarzi, A.A., Krempler, A., Jeggo, P.A., and Lobrich, M. (2009). ATM and Artemis promote homologous recombination of radiation-induced DNA double-strand breaks in G2. EMBO J *28*, 3413-3427.

Bewersdorf, J., Bennett, B.T., and Knight, K.L. (2006). H2AX chromatin structures and their response to DNA damage revealed by 4Pi microscopy. Proceedings of the National Academy of Sciences of the United States of America *103*, 18137-18142.

Bibel, M., Richter, J., Lacroix, E., and Barde, Y.A. (2007). Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. Nat Protoc *2*, 1034-1043.

Biehs, R., Steinlage, M., Barton, O., Juhasz, S., Kunzel, J., Spies, J., Shibata, A., Jeggo, P.A., and Lobrich, M. (2017). DNA Double-Strand Break Resection Occurs during Non-homologous End Joining in G1 but Is Distinct from Resection during

Homologous Recombination. Molecular cell *65*, 671-684 e675.

Bignell, G., Micklem, G., Stratton, M.R., Ashworth, A., and Wooster, R. (1997). The BRC repeats are conserved in mammalian BRCA2 proteins. Human molecular genetics *6*, 53-58.

Bignell, G.R., Greenman, C.D., Davies, H., Butler, A.P., Edkins, S., Andrews, J.M., Buck, G., Chen, L., Beare, D., Latimer, C*., et al.* (2010). Signatures of mutation and selection in the cancer genome. Nature *463*, 893-898.

Blanco, M.G., Matos, J., Rass, U., Ip, S.C., and West, S.C. (2010). Functional overlap between the structure-specific nucleases Yen1 and Mus81-Mms4 for DNA-damage repair in S. cerevisiae. DNA Repair (Amst) *9*, 394-402.

Blier, P.R., Griffith, A.J., Craft, J., and Hardin, J.A. (1993). Binding of Ku protein to DNA. Measurement of affinity for ends and demonstration of binding to nicks. The Journal of biological chemistry *268*, 7594-7601.

Block, W.D., Yu, Y., and Lees-Miller, S.P. (2004). Phosphatidyl inositol 3-kinase-like serine/threonine protein kinases (PIKKs) are required for DNA damage-induced phosphorylation of the 32 kDa subunit of replication protein A at threonine 21. Nucleic Acids Res *32*, 997-1005.

Bohgaki, T., Bohgaki, M., Cardoso, R., Panier, S., Zeegers, D., Li, L., Stewart, G.S., Sanchez, O., Hande, M.P., Durocher, D*., et al.* (2011). Genomic instability, defective spermatogenesis, immunodeficiency, and cancer in a mouse model of the RIDDLE syndrome. PLoS genetics *7*, e1001381.

Boldog, F., Gemmill, R.M., West, J., Robinson, M., Robinson, L., Li, E., Roche, J., Todd, S., Waggoner, B., Lundstrom, R*., et al.* (1997). Chromosome 3p14 homozygous deletions and sequence analysis of FRA3B. Human molecular genetics *6*, 193-203.

Bonner, W.M., Redon, C.E., Dickey, J.S., Nakamura, A.J., Sedelnikova, O.A., Solier, S., and Pommier, Y. (2008). GammaH2AX and cancer. Nat Rev Cancer *8*, 957-967.

Bork, P., Blomberg, N., and Nilges, M. (1996). Internal repeats in the BRCA2 protein sequence. Nature genetics *13*, 22-23.

Botuyan, M.V., Lee, J., Ward, I.M., Kim, J.E., Thompson, J.R., Chen, J., and Mer, G. (2006). Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. Cell *127*, 1361-1373.

Boye, E., Skjolberg, H.C., and Grallert, B. (2009). Checkpoint regulation of DNA replication. Methods Mol Biol *521*, 55-70.

Boyle, S., Gilchrist, S., Bridger, J.M., Mahy, N.L., Ellis, J.A., and Bickmore, W.A. (2001). The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. Human molecular genetics *10*, 211-219.

Branzei, D., and Foiani, M. (2010). Maintaining genome stability at the replication fork. Nature reviews *11*, 208-219.

Branzei, D., Vanoli, F., and Foiani, M. (2008). SUMOylation regulates Rad18-mediated template switch. Nature *456*, 915-920.

Broderick, R., Nieminuszczy, J., Baddock, H.T., Deshpande, R., Gileadi, O., Paull, T.T., McHugh, P.J., and Niedzwiedz, W. (2016). EXD2 promotes homologous recombination by facilitating DNA end resection. Nature cell biology *18*, 271-280.

Brouwer, I., Sitters, G., Candelli, A., Heerema, S.J., Heller, I., de Melo, A.J., Zhang, H., Normanno, D., Modesti, M., Peterman, E.J.*, et al.* (2016). Sliding sleeves of XRCC4-XLF bridge DNA and connect fragments of broken DNA. Nature *535*, 566-569.

Brown, E.J., and Baltimore, D. (2003). Essential and dispensable roles of ATR in cell cycle arrest and genome maintenance. Genes Dev *17*, 615-628.

Bryant, H.E., Petermann, E., Schultz, N., Jemth, A.S., Loseva, O., Issaeva, N., Johansson, F., Fernandez, S., McGlynn, P., and Helleday, T. (2009). PARP is activated at stalled forks to mediate Mre11-dependent replication restart and recombination. EMBO J *28*, 2601-2615.

Buermeyer, A.B., Deschenes, S.M., Baker, S.M., and Liskay, R.M. (1999). Mammalian DNA mismatch repair. Annual review of genetics *33*, 533-564.

Bunch, H., Lawney, B.P., Lin, Y.F., Asaithamby, A., Murshid, A., Wang, Y.E., Chen, B.P., and Calderwood, S.K. (2015). Transcriptional elongation requires DNA break-induced signalling. Nature communications *6*, 10191.

Bunting, S.F., Callen, E., Wong, N., Chen, H.T., Polato, F., Gunn, A., Bothmer, A., Feldhahn, N., Fernandez-Capetillo, O., Cao, L.*, et al.* (2010). 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. Cell *141*, 243-254.

Burrow, A.A., Williams, L.E., Pierce, L.C., and Wang, Y.H. (2009). Over half of breakpoints in gene pairs involved in cancer-specific recurrent translocations are mapped to human chromosomal fragile sites. BMC Genomics *10*, 59.

Canela, A., Sridharan, S., Sciascia, N., Tubbs, A., Meltzer, P., Sleckman, B.P., and Nussenzweig, A. (2016). DNA Breaks and End Resection Measured Genome-wide by End Sequencing. Molecular cell *63*, 898-911.

Caron, P., Aymard, F., Iacovoni, J.S., Briois, S., Canitrot, Y., Bugler, B., Massip, L., Losada, A., and Legube, G. (2012). Cohesin protects genes against gammaH2AX Induced by DNA double-strand breaks. PLoS genetics *8*, e1002460.

Caron, P., Choudjaye, J., Clouaire, T., Bugler, B., Daburon, V., Aguirrebengoa, M., Mangeat, T., Iacovoni, J.S., Alvarez-Quilon, A., Cortes-Ledesma, F.*, et al.* (2015). Non-redundant Functions of ATM and DNA-PKcs in Response to DNA Double-Strand Breaks. Cell reports *13*, 1598-1609.

Carreira, A., and Kowalczykowski, S.C. (2011). Two classes of BRC repeats in BRCA2 promote RAD51 nucleoprotein filament function by distinct mechanisms. Proceedings of the National Academy of Sciences of the United States of America *108*, 10448-10453.

Carson, C.T., Schwartz, R.A., Stracker, T.H., Lilley, C.E., Lee, D.V., and Weitzman, M.D. (2003). The Mre11 complex is required for ATM activation and the G2/M checkpoint. EMBO J *22*, 6610-6620.

Casper, A.M., Nghiem, P., Arlt, M.F., and Glover, T.W. (2002). ATR regulates fragile site stability. Cell *111*, 779-789.

Cejka, P., Cannavo, E., Polaczek, P., Masuda-Sasa, T., Pokharel, S., Campbell, J.L., and Kowalczykowski, S.C. (2010). DNA end resection by Dna2-Sgs1-RPA and its stimulation by Top3-Rmi1 and Mre11-Rad50-Xrs2. Nature *467*, 112-116.

Celeste, A., Difilippantonio, S., Difilippantonio, M.J., Fernandez-Capetillo, O., Pilch, D.R., Sedelnikova, O.A., Eckhaus, M., Ried, T., Bonner, W.M., and Nussenzweig, A. (2003). H2AX haploinsufficiency modifies genomic stability and tumor susceptibility. Cell *114*, 371-383.

Celeste, A., Petersen, S., Romanienko, P.J., Fernandez-Capetillo, O., Chen, H.T., Sedelnikova, O.A., Reina-San-Martin, B., Coppola, V., Meffre, E., Difilippantonio, M.J.*, et al.* (2002). Genomic instability in mice lacking histone H2AX. Science (New York, NY *296*, 922-927.

Cha, R.S., and Kleckner, N. (2002). ATR homolog Mec1 promotes fork progression, thus averting breaks in replication slow zones. Science (New York, NY *297*, 602-606.

Chabosseau, P., Buhagiar-Labarchede, G., Onclercq-Delic, R., Lambert, S., Debatisse, M., Brison, O., and Amor-Gueret, M. (2011). Pyrimidine pool imbalance induced by BLM helicase deficiency contributes to genetic instability in Bloom syndrome. Nature communications *2*, 368.

Chan, D.W., Chen, B.P., Prithivirajsingh, S., Kurimasa, A., Story, M.D., Qin, J., and Chen, D.J. (2002). Autophosphorylation of the DNA-dependent protein kinase catalytic subunit is required for rejoining of DNA double-strand breaks. Genes Dev *16*, 2333-2338.

Chan, K.L., Palmai-Pallag, T., Ying, S., and Hickson, I.D. (2009). Replication stress induces sister-chromatid bridging at fragile site loci in mitosis. Nature cell biology *11*, 753-760.

Chan, Y.W., and West, S. (2015). GEN1 promotes Holliday junction resolution by a coordinated nick and counter-nick mechanism. Nucleic Acids Res *43*, 10882-10892.

Chapman, J.R., Barral, P., Vannier, J.B., Borel, V., Steger, M., Tomas-Loba, A., Sartori, A.A., Adams, I.R., Batista, F.D., and Boulton, S.J. (2013). RIF1 is essential for 53BP1-dependent nonhomologous end joining and suppression of DNA double-strand break resection. Molecular cell *49*, 858-871.

Chaturvedi, P., Eng, W.K., Zhu, Y., Mattern, M.R., Mishra, R., Hurle, M.R., Zhang, X., Annan, R.S., Lu, Q., Faucette, L.F.*, et al.* (1999). Mammalian Chk2 is a downstream effector of the ATM-dependent DNA damage checkpoint pathway. Oncogene *18*, 4047-4054.

Chen, C.C., Carson, J.J., Feser, J., Tamburini, B., Zabaronick, S., Linger, J., and Tyler,

J.K. (2008). Acetylated lysine 56 on histone H3 drives chromatin assembly after repair and signals for the completion of repair. Cell *134*, 231-243.

Chen, H., Lisby, M., and Symington, L.S. (2013). RPA coordinates DNA end resection and prevents formation of DNA hairpins. Molecular cell *50*, 589-600.

Chen, X., Cui, D., Papusha, A., Zhang, X., Chu, C.D., Tang, J., Chen, K., Pan, X., and Ira, G. (2012). The Fun30 nucleosome remodeller promotes resection of DNA double-strand break ends. Nature *489*, 576-580.

Chiarle, R., Zhang, Y., Frock, R.L., Lewis, S.M., Molinie, B., Ho, Y.J., Myers, D.R., Choi, V.W., Compagno, M., Malkin, D.J., *et al.* (2011). Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. Cell *147*, 107-119.

Chiolo, I., Minoda, A., Colmenares, S.U., Polyzos, A., Costes, S.V., and Karpen, G.H. (2011). Double-strand breaks in heterochromatin move outside of a dynamic HP1a domain to complete recombinational repair. Cell *144*, 732-744.

Chowdhury, D., Keogh, M.C., Ishii, H., Peterson, C.L., Buratowski, S., and Lieberman, J. (2005). gamma-H2AX dephosphorylation by protein phosphatase 2A facilitates DNA double-strand break repair. Molecular cell *20*, 801-809.

Chowdhury, D., Xu, X., Zhong, X., Ahmed, F., Zhong, J., Liao, J., Dykxhoorn, D.M., Weinstock, D.M., Pfeifer, G.P., and Lieberman, J. (2008). A PP4-phosphatase complex dephosphorylates gamma-H2AX generated during DNA replication. Molecular cell *31*, 33-46.

Christmann, M., Tomicic, M.T., Roos, W.P., and Kaina, B. (2003). Mechanisms of human DNA repair: an update. Toxicology *193*, 3-34.

Ciabrelli, F., and Cavalli, G. (2015). Chromatin-driven behavior of topologically associating domains. Journal of molecular biology *427*, 608-625.

Ciccia, A., and Elledge, S.J. (2010). The DNA damage response: making it safe to play with knives. Molecular cell *40*, 179-204.

Ciechanover, A., Elias, S., Heller, H., and Hershko, A. (1982). "Covalent affinity" purification of ubiquitin-activating enzyme. The Journal of biological chemistry *257*, 2537-2542.

Cimprich, K.A., and Cortez, D. (2008). ATR: an essential regulator of genome integrity. Nature reviews *9*, 616-627.

Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. Annu Rev Biochem *78*, 273-304.

Cliby, W.A., Roberts, C.J., Cimprich, K.A., Stringer, C.M., Lamb, J.R., Schreiber, S.L., and Friend, S.H. (1998). Overexpression of a kinase-inactive ATR protein causes sensitivity to DNA-damaging agents and defects in cell cycle checkpoints. EMBO J *17*, 159-169.

Collins, A.R. (2004). The comet assay for DNA damage and repair: principles, applications, and limitations. Mol Biotechnol *26*, 249-261.

Cook, P.J., Ju, B.G., Telese, F., Wang, X., Glass, C.K., and Rosenfeld, M.G. (2009). Tyrosine dephosphorylation of H2AX modulates apoptosis and survival decisions. Nature *458*, 591-596.

Corbin, S., Neilly, M.E., Espinosa, R., 3rd, Davis, E.M., McKeithan, T.W., and Le Beau, M.M. (2002). Identification of unstable sequences within the common fragile site at 3p14.2: implications for the mechanism of deletions within fragile histidine triad gene/common fragile site at 3p14.2 in tumors. Cancer research *62*, 3477-3484.

Cortez, D., Guntuku, S., Qin, J., and Elledge, S.J. (2001). ATR and ATRIP: partners in checkpoint signaling. Science (New York, NY *294*, 1713-1716.

Costelloe, T., Louge, R., Tomimatsu, N., Mukherjee, B., Martini, E., Khadaroo, B., Dubois, K., Wiegant, W.W., Thierry, A., Burma, S.*, et al.* (2012). The yeast Fun30 and human SMARCAD1 chromatin remodellers promote DNA end resection. Nature *489*, 581-584.

Cowell, I.G., Sunter, N.J., Singh, P.B., Austin, C.A., Durkacz, B.W., and Tilby, M.J. (2007). gammaH2AX foci form preferentially in euchromatin after ionising-radiation. PLoS One *2*, e1057.

Craig, J.M. (2005). Heterochromatin--many flavours, common themes. Bioessays *27*, 17-28.

Cremer, M., von Hase, J., Volm, T., Brero, A., Kreth, G., Walter, J., Fischer, C., Solovei, I., Cremer, C., and Cremer, T. (2001). Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. Chromosome Res *9*, 541-567.

Cremer, T., Kreth, G., Koester, H., Fink, R.H., Heintzmann, R., Cremer, M., Solovei, I., Zink, D., and Cremer, C. (2000). Chromosome territories, interchromatin domain compartment, and nuclear matrix: an integrated view of the functional nuclear architecture. Crit Rev Eukaryot Gene Expr *10*, 179-212.

Croft, J.A., Bridger, J.M., Boyle, S., Perry, P., Teague, P., and Bickmore, W.A. (1999). Differences in the localization and morphology of chromosomes in the human nucleus. The Journal of cell biology *145*, 1119-1131.

Crosetto, N., Mitra, A., Silva, M.J., Bienko, M., Dojer, N., Wang, Q., Karaca, E., Chiarle, R., Skrzypczak, M., Ginalski, K.*, et al.* (2013). Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. Nat Methods *10*, 361-365.

Cruz-Garcia, A., Lopez-Saavedra, A., and Huertas, P. (2014). BRCA1 accelerates CtIP-mediated DNA-end resection. Cell reports *9*, 451-459.

Cuella-Martin, R., Oliveira, C., Lockstone, H.E., Snellenberg, S., Grolmusova, N., and Chapman, J.R. (2016). 53BP1 Integrates DNA Repair and p53-Dependent Cell Fate Decisions via Distinct Mechanisms. Molecular cell *64*, 51-64.

Culligan, K.M., Robertson, C.E., Foreman, J., Doerner, P., and Britt, A.B. (2006). ATR

and ATM play both distinct and additive roles in response to ionizing radiation. Plant J *48*, 947-961.

Daboussi, F., Courbet, S., Benhamou, S., Kannouche, P., Zdzienicka, M.Z., Debatisse, M., and Lopez, B.S. (2008). A homologous recombination defect affects replication-fork progression in mammalian cells. Journal of cell science *121*, 162-166.

d'Adda di Fagagna, F., Reaper, P.M., Clay-Farrace, L., Fiegler, H., Carr, P., Von Zglinicki, T., Saretzki, G., Carter, N.P., and Jackson, S.P. (2003). A DNA damage checkpoint response in telomere-initiated senescence. Nature *426*, 194-198.

Dall, K.L., Scarpini, C.G., Roberts, I., Winder, D.M., Stanley, M.A., Muralidhar, B., Herdman, M.T., Pett, M.R., and Coleman, N. (2008). Characterization of naturally occurring HPV16 integration sites isolated from cervical keratinocytes under noncompetitive conditions. Cancer research *68*, 8249-8259.

D'Amours, D., Desnoyers, S., D'Silva, I., and Poirier, G.G. (1999). Poly(ADP-ribosyl)ation reactions in the regulation of nuclear functions. Biochem J *342 (Pt 2)*, 249-268.

D'Amours, D., and Jackson, S.P. (2002). The Mre11 complex: at the crossroads of dna repair and checkpoint signalling. Nature reviews *3*, 317-327.

Daugaard, M., Baude, A., Fugger, K., Povlsen, L.K., Beck, H., Sorensen, C.S., Petersen, N.H., Sorensen, P.H., Lukas, C., Bartek, J.*, et al.* (2012). LEDGF (p75) promotes DNA-end resection and homologous recombination. Nature structural & molecular biology *19*, 803-810.

Davis, A.J., and Chen, D.J. (2013). DNA double strand break repair via non-homologous end-joining. Transl Cancer Res *2*, 130-143.

de Boer, J., and Hoeijmakers, J.H. (2000). Nucleotide excision repair and human syndromes. Carcinogenesis *21*, 453-460.

De Bont, R., and van Larebeke, N. (2004). Endogenous DNA damage in humans: a review of quantitative data. Mutagenesis *19*, 169-185.

De Cecco, M., Jeyapalan, J., Zhao, X., Tamamori-Adachi, M., and Sedivy, J.M. (2011). Nuclear protein accumulation in cellular senescence and organismal aging revealed with a novel single-cell resolution fluorescence microscopy assay. Aging (Albany NY) *3*, 955-967.

de Jager, M., van Noort, J., van Gent, D.C., Dekker, C., Kanaar, R., and Wyman, C. (2001). Human Rad50/Mre11 is a flexible complex that can tether DNA ends. Molecular cell *8*, 1129-1135.

de Waard, H., de Wit, J., Gorgels, T.G., van den Aardweg, G., Andressoo, J.O., Vermeij, M., van Steeg, H., Hoeijmakers, J.H., and van der Horst, G.T. (2003). Cell type-specific hypersensitivity to oxidative damage in CSB and XPA mice. DNA Repair (Amst) *2*, 13-25.

Debatisse, M., El Achkar, E., and Dutrillaux, B. (2006). Common fragile sites nested at

the interfaces of early and late-replicating chromosome bands: cis acting components of the G2/M checkpoint? Cell cycle (Georgetown, Tex *5*, 578-581.

Deckbar, D., Birraux, J., Krempler, A., Tchouandong, L., Beucher, A., Walker, S., Stiff, T., Jeggo, P., and Lobrich, M. (2007). Chromosome breakage after G2 checkpoint release. The Journal of cell biology *176*, 749-755.

Deckbar, D., Jeggo, P.A., and Lobrich, M. (2011). Understanding the limitations of radiation-induced cell cycle checkpoints. Crit Rev Biochem Mol Biol *46*, 271-283.

Deng, C., Zhang, P., Harper, J.W., Elledge, S.J., and Leder, P. (1995). Mice lacking p21CIP1/WAF1 undergo normal development, but are defective in G1 checkpoint control. Cell *82*, 675-684.

Deng, X., Prakash, A., Dhar, K., Baia, G.S., Kolar, C., Oakley, G.G., and Borgstahl, G.E. (2009). Human replication protein A-Rad52-single-stranded DNA complex: stoichiometry and evidence for strand transfer regulation by phosphorylation. Biochemistry *48*, 6633-6643.

Dereli-Oz, A., Versini, G., and Halazonetis, T.D. (2011). Studies of genomic copy number changes in human cancers reveal signatures of DNA replication stress. Molecular oncology *5*, 308-314.

Di Micco, R., Fumagalli, M., Cicalese, A., Piccinin, S., Gasparini, P., Luise, C., Schurra, C., Garre, M., Nuciforo, P.G., Bensimon, A.*, et al.* (2006). Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. Nature *444*, 638-642.

Di Virgilio, M., Callen, E., Yamane, A., Zhang, W., Jankovic, M., Gitlin, A.D., Feldhahn, N., Resch, W., Oliveira, T.Y., Chait, B.T.*, et al.* (2013). Rif1 prevents resection of DNA breaks and promotes immunoglobulin class switching. Science (New York, NY *339*, 711-715.

Dimitrova, N., Chen, Y.C., Spector, D.L., and de Lange, T. (2008). 53BP1 promotes non-homologous end joining of telomeres by increasing chromatin mobility. Nature *456*, 524-528.

Dion, V., Kalck, V., Horigome, C., Towbin, B.D., and Gasser, S.M. (2012). Increased mobility of double-strand breaks requires Mec1, Rad9 and the homologous recombination machinery. Nature cell biology *14*, 502-509.

DiTullio, R.A., Jr., Mochan, T.A., Venere, M., Bartkova, J., Sehested, M., Bartek, J., and Halazonetis, T.D. (2002). 53BP1 functions in an ATM-dependent checkpoint pathway that is constitutively activated in human cancer. Nature cell biology *4*, 998-1002.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376-380.

Dmitrieva, N.I., Cui, K., Kitchaev, D.A., Zhao, K., and Burg, M.B. (2011). DNA double-strand breaks induced by high NaCl occur predominantly in gene deserts. Proceedings of the National Academy of Sciences of the United States of America *108*,

20796-20801.

Doil, C., Mailand, N., Bekker-Jensen, S., Menard, P., Larsen, D.H., Pepperkok, R., Ellenberg, J., Panier, S., Durocher, D., Bartek, J.*, et al.* (2009). RNF168 binds and amplifies ubiquitin conjugates on damaged chromosomes to allow accumulation of repair proteins. Cell *136*, 435-446.

Dominguez-Bendala, J., Priddle, H., Clarke, A., and McWhir, J. (2003). Elevated expression of exogenous Rad51 leads to identical increases in gene-targeting frequency in murine embryonic stem (ES) cells with both functional and dysfunctional p53 genes. Exp Cell Res *286*, 298-307.

Dong, J., Panchakshari, R.A., Zhang, T., Zhang, Y., Hu, J., Volpi, S.A., Meyers, R.M., Ho, Y.J., Du, Z., Robbiani, D.F.*, et al.* (2015). Orientation-specific joining of AID-initiated DNA breaks promotes antibody class switching. Nature *525*, 134-139.

Dornreiter, I., Erdile, L.F., Gilbert, I.U., von Winkler, D., Kelly, T.J., and Fanning, E. (1992). Interaction of DNA polymerase alpha-primase with cellular replication protein A and SV40 T antigen. EMBO J *11*, 769-776.

Downs, J.A., and Jackson, S.P. (2004). A means to a DNA end: the many roles of Ku. Nature reviews *5*, 367-378.

Drusco, A., Pekarsky, Y., Costinean, S., Antenucci, A., Conti, L., Volinia, S., Aqeilan, R.I., Huebner, K., and Zanesi, N. (2011). Common fragile site tumor suppressor genes and corresponding mouse models of cancer. Journal of biomedicine & biotechnology *2011*, 984505.

Dundr, M., and Misteli, T. (2001). Functional architecture in the cell nucleus. Biochem J *356*, 297-310.

Durkin, S.G., Arlt, M.F., Howlett, N.G., and Glover, T.W. (2006). Depletion of CHK1, but not CHK2, induces chromosomal instability and breaks at common fragile sites. Oncogene *25*, 4381-4388.

Durkin, S.G., and Glover, T.W. (2007). Chromosome fragile sites. Annual review of genetics *41*, 169-192.

Durkin, S.G., Ragland, R.L., Arlt, M.F., Mulle, J.G., Warren, S.T., and Glover, T.W. (2008). Replication stress induces tumor-like microdeletions in FHIT/FRA3B. Proceedings of the National Academy of Sciences of the United States of America *105*, 246-251.

Eapen, V.V., Sugawara, N., Tsabar, M., Wu, W.H., and Haber, J.E. (2012). The Saccharomyces cerevisiae chromatin remodeler Fun30 regulates DNA end resection and checkpoint deactivation. Molecular and cellular biology *32*, 4727-4740.

Eggler, A.L., Inman, R.B., and Cox, M.M. (2002). The Rad51-dependent pairing of long DNA substrates is stabilized by replication protein A. The Journal of biological chemistry *277*, 39280-39288.

El Achkar, E., Gerbault-Seureau, M., Muleris, M., Dutrillaux, B., and Debatisse, M.

(2005). Premature condensation induces breaks at the interface of early and late replicating chromosome bands bearing common fragile sites. Proceedings of the National Academy of Sciences of the United States of America *102*, 18069-18074.

Elder, F.F., and Robinson, T.J. (1989). Rodent common fragile sites: are they conserved? Evidence from mouse and rat. Chromosoma *97*, 459-464.

El-Khamisy, S.F., Masutani, M., Suzuki, H., and Caldecott, K.W. (2003). A requirement for PARP-1 for the assembly or stability of XRCC1 nuclear foci at sites of oxidative DNA damage. Nucleic Acids Res *31*, 5526-5533.

Ellison, V., and Stillman, B. (2003). Biochemical characterization of DNA damage checkpoint complexes: clamp loader and clamp complexes with specificity for 5' recessed DNA. PLoS Biol *1*, E33.

Enoch, T., and Norbury, C. (1995). Cellular responses to DNA damage: cell-cycle checkpoints, apoptosis and the roles of p53 and ATM. Trends Biochem Sci *20*, 426-430.

Escribano-Diaz, C., Orthwein, A., Fradet-Turcotte, A., Xing, M., Young, J.T., Tkac, J., Cook, M.A., Rosebrock, A.P., Munro, M., Canny, M.D.*, et al.* (2013). A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. Molecular cell *49*, 872-883.

Falbo, K.B., and Shen, X. (2006). Chromatin remodeling in DNA replication. J Cell Biochem *97*, 684-689.

Falck, J., Coates, J., and Jackson, S.P. (2005). Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. Nature *434*, 605-611.

Fan, J., Robert, C., Jang, Y.Y., Liu, H., Sharkis, S., Baylin, S.B., and Rassool, F.V. (2011). Human induced pluripotent cells resemble embryonic stem cells demonstrating enhanced levels of DNA repair and efficacy of nonhomologous end-joining. Mutation research *713*, 8-17.

Fanning, E., Klimovich, V., and Nager, A.R. (2006). A dynamic model for replication protein A (RPA) function in DNA processing pathways. Nucleic Acids Res *34*, 4126-4137.

Farrell, A.W., Halliday, G.M., and Lyons, J.G. (2011). Chromatin structure following UV-induced DNA damage-repair or death? Int J Mol Sci *12*, 8063-8085.

Fasulo, B., Koyama, C., Yu, K.R., Homola, E.M., Hsieh, T.S., Campbell, S.D., and Sullivan, W. (2012). Chk1 and Wee1 kinases coordinate DNA replication, chromosome condensation, and anaphase entry. Molecular biology of the cell *23*, 1047-1057.

Fell, V.L., and Schild-Poulter, C. (2015). The Ku heterodimer: function in DNA repair and beyond. Mutat Res Rev Mutat Res *763*, 15-29.

Felsenfeld, G., and Groudine, M. (2003). Controlling the double helix. Nature *421*, 448-453.

Feng, L., and Chen, J. (2012). The E3 ligase RNF8 regulates KU80 removal and NHEJ

repair. Nature structural & molecular biology *19*, 201-206.

Feng, W., Di Rienzi, S.C., Raghuraman, M.K., and Brewer, B.J. (2011). Replication stress-induced chromosome breakage is correlated with replication fork progression and is preceded by single-stranded DNA formation. G3 (Bethesda) *1*, 327-335.

Ferber, M.J., Eilers, P., Schuuring, E., Fenton, J.A., Fleuren, G.J., Kenter, G., Szuhai, K., Smith, D.I., Raap, A.K., and Brink, A.A. (2004). Positioning of cervical carcinoma and Burkitt lymphoma translocation breakpoints with respect to the human papillomavirus integration cluster in FRA8C at 8q24.13. Cancer genetics and cytogenetics *154*, 1-9.

Ferber, M.J., Thorland, E.C., Brink, A.A., Rapp, A.K., Phillips, L.A., McGovern, R., Gostout, B.S., Cheung, T.H., Chung, T.K., Fu, W.Y.*, et al.* (2003). Preferential integration of human papillomavirus type 18 near the c-myc locus in cervical carcinoma. Oncogene *22*, 7233-7242.

Fernandez-Capetillo, O., Celeste, A., and Nussenzweig, A. (2003). Focusing on foci: H2AX and the recruitment of DNA-damage response factors. Cell cycle (Georgetown, Tex *2*, 426-427.

Finnis, M., Dayan, S., Hobson, L., Chenevix-Trench, G., Friend, K., Ried, K., Venter, D., Woollatt, E., Baker, E., and Richards, R.I. (2005). Common chromosomal fragile site FRA16D mutation in cancer cells. Human molecular genetics *14*, 1341-1349.

Fishel, R., and Wilson, T. (1997). MutS homologs in mammalian cells. Curr Opin Genet Dev *7*, 105-113.

FitzGerald, J., Moureau, S., Drogaris, P., O'Connell, E., Abshiru, N., Verreault, A., Thibault, P., Grenon, M., and Lowndes, N.F. (2011). Regulation of the DNA damage response and gene expression by the Dot1L histone methyltransferase and the 53Bp1 tumour suppressor. PLoS One *6*, e14714.

Fluckiger, A.C., Marcy, G., Marchand, M., Negre, D., Cosset, F.L., Mitalipov, S., Wolf, D., Savatier, P., and Dehay, C. (2006). Cell cycle features of primate embryonic stem cells. Stem Cells *24*, 547-556.

Focarelli, M.L., Soza, S., Mannini, L., Paulis, M., Montecucco, A., and Musio, A. (2009). Claspin inhibition leads to fragile site expression. Genes, chromosomes & cancer *48*, 1083-1090.

Fousteri, M., and Mullenders, L.H. (2008). Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. Cell Res *18*, 73-84.

Fradet-Turcotte, A., Canny, M.D., Escribano-Diaz, C., Orthwein, A., Leung, C.C., Huang, H., Landry, M.C., Kitevski-LeBlanc, J., Noordermeer, S.M., Sicheri, F.*, et al.* (2013). 53BP1 is a reader of the DNA-damage-induced H2A Lys 15 ubiquitin mark. Nature *499*, 50-54.

Freeman, S.E., Hacham, H., Gange, R.W., Maytum, D.J., Sutherland, J.C., and Sutherland, B.M. (1989). Wavelength dependence of pyrimidine dimer formation in DNA

of human skin irradiated in situ with ultraviolet light. Proceedings of the National Academy of Sciences of the United States of America *86*, 5605-5609.

Freudenreich, C.H. (2007). Chromosome fragility: molecular mechanisms and cellular consequences. Front Biosci *12*, 4911-4924.

Friedland, A.E., Baral, R., Singhal, P., Loveluck, K., Shen, S., Sanchez, M., Marco, E., Gotta, G.M., Maeder, M.L., Kennedy, E.M.*, et al.* (2015). Characterization of Staphylococcus aureus Cas9: a smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications. Genome Biol *16*, 257.

Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.J., Kii, E., and Alt, F.W. (2015). Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. Nat Biotechnol *33*, 179-186.

Gabriel, R., Lombardo, A., Arens, A., Miller, J.C., Genovese, P., Kaeppel, C., Nowrouzi, A., Bartholomae, C.C., Wang, J., Friedman, G.*, et al.* (2011). An unbiased genome-wide analysis of zinc-finger nuclease specificity. Nat Biotechnol *29*, 816-823.

Galanty, Y., Belotserkovskaya, R., Coates, J., Polo, S., Miller, K.M., and Jackson, S.P. (2009). Mammalian SUMO E3-ligases PIAS1 and PIAS4 promote responses to DNA double-strand breaks. Nature *462*, 935-939.

Galkin, V.E., Esashi, F., Yu, X., Yang, S., West, S.C., and Egelman, E.H. (2005). BRCA2 BRC motifs bind RAD51-DNA filaments. Proceedings of the National Academy of Sciences of the United States of America *102*, 8537-8542.

Gamper, A.M., Rofougaran, R., Watkins, S.C., Greenberger, J.S., Beumer, J.H., and Bakkenist, C.J. (2013). ATR kinase activation in G1 phase facilitates the repair of ionizing radiation-induced DNA damage. Nucleic Acids Res *41*, 10334-10344.

Garcia, V., Phelps, S.E., Gray, S., and Neale, M.J. (2011). Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. Nature *479*, 241-244.

Gatti, M., Pinato, S., Maiolica, A., Rocchio, F., Prato, M.G., Aebersold, R., and Penengo, L. (2015). RNF168 promotes noncanonical K27 ubiquitination to signal DNA damage. Cell reports *10*, 226-238.

Gelot, C., Magdalou, I., and Lopez, B.S. (2015). Replication stress in Mammalian cells and its consequences for mitosis. Genes (Basel) *6*, 267-298.

Germann, M.W., Johnson, C.N., and Spring, A.M. (2012). Recognition of damaged DNA: structure and dynamic markers. Med Res Rev *32*, 659-683.

Giachino, C., Orlando, L., and Turinetto, V. (2013). Maintenance of genomic stability in mouse embryonic stem cells: relevance in aging and disease. Int J Mol Sci *14*, 2617-2636.

Gillet, L.C., and Scharer, O.D. (2006). Molecular mechanisms of mammalian global genome nucleotide excision repair. Chem Rev *106*, 253-276.

Glover, T.W., Hoge, A.W., Miller, D.E., Ascara-Wilke, J.E., Adam, A.N., Dagenais, S.L.,

Wilke, C.M., Dierick, H.A., and Beer, D.G. (1998). The murine Fhit gene is highly similar to its human orthologue and maps to a common fragile site region. Cancer research *58*, 3409-3414.

Glover, T.W., and Stein, C.K. (1987). Induction of sister chromatid exchanges at common fragile sites. Am J Hum Genet *41*, 882-890.

Glover, T.W., and Stein, C.K. (1988). Chromosome breakage and recombination at fragile sites. Am J Hum Genet *43*, 265-273.

Goldberg, I.H. (1991). Mechanism of neocarzinostatin action: role of DNA microstructure in determination of chemistry of bistranded oxidative damage. Accounts of Chemical Research *24*, 191-198.

Gong, F., Kwon, Y., and Smerdon, M.J. (2005). Nucleotide excision repair in chromatin and the right of entry. DNA Repair (Amst) *4*, 884-896.

Goodarzi, A.A., Jeggo, P., and Lobrich, M. (2010). The influence of heterochromatin on DNA double strand break repair: Getting the strong, silent type to relax. DNA Repair (Amst) *9*, 1273-1282.

Goodarzi, A.A., and Jeggo, P.A. (2013). The repair and signaling responses to DNA double-strand breaks. Adv Genet *82*, 1-45.

Goodarzi, A.A., Kurka, T., and Jeggo, P.A. (2011). KAP-1 phosphorylation regulates CHD3 nucleosome remodeling during the DNA double-strand break response. Nature structural & molecular biology *18*, 831-839.

Goodarzi, A.A., Noon, A.T., Deckbar, D., Ziv, Y., Shiloh, Y., Lobrich, M., and Jeggo, P.A. (2008). ATM signaling facilitates repair of DNA double-strand breaks associated with heterochromatin. Molecular cell *31*, 167-177.

Goodarzi, A.A., Yu, Y., Riballo, E., Douglas, P., Walker, S.A., Ye, R., Harer, C., Marchetti, C., Morrice, N., Jeggo, P.A.*, et al.* (2006). DNA-PK autophosphorylation facilitates Artemis endonuclease activity. EMBO J *25*, 3880-3889.

Goodhead, D.T. (1999). Mechanisms for the biological effectiveness of high-LET radiations. J Radiat Res *40 Suppl*, 1-13.

Govan, H.L., 3rd, Valles-Ayoub, Y., and Braun, J. (1990). Fine-mapping of DNA damage and repair in specific genomic segments. Nucleic Acids Res *18*, 3823-3830.

Gravel, S., Chapman, J.R., Magill, C., and Jackson, S.P. (2008). DNA helicases Sgs1 and BLM promote DNA double-strand break resection. Genes Dev *22*, 2767-2772.

Grawunder, U., Wilm, M., Wu, X., Kulesza, P., Wilson, T.E., Mann, M., and Lieber, M.R. (1997). Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. Nature *388*, 492-495.

Gregoire, M.C., Massonneau, J., Leduc, F., Arguin, M., Brazeau, M.A., and Boissonneault, G. (2016). Quantification and genome-wide mapping of DNA double-strand breaks. DNA Repair (Amst) *48*, 63-68.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W.*, et al.* (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature *453*, 948-951.

Ha, M., Kraushaar, D.C., and Zhao, K. (2014). Genome-wide analysis of H3.3 dissociation reveals high nucleosome turnover at distal regulatory regions of embryonic stem cells. Epigenetics Chromatin *7*, 38.

Haince, J.F., Kozlov, S., Dawson, V.L., Dawson, T.M., Hendzel, M.J., Lavin, M.F., and Poirier, G.G. (2007). Ataxia telangiectasia mutated (ATM) signaling network is modulated by a novel poly(ADP-ribose)-dependent pathway in the early response to DNA-damaging agents. The Journal of biological chemistry *282*, 16441-16453.

Haince, J.F., McDonald, D., Rodrigue, A., Dery, U., Masson, J.Y., Hendzel, M.J., and Poirier, G.G. (2008). PARP1-dependent kinetics of recruitment of MRE11 and NBS1 proteins to multiple DNA damage sites. The Journal of biological chemistry *283*, 1197-1208.

Handoko, L., Xu, H., Li, G., Ngan, C.Y., Chew, E., Schnapp, M., Lee, C.W., Ye, C., Ping, J.L., Mulawadi, F.*, et al.* (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. Nature genetics *43*, 630-638.

Hansen, R.S., Canfield, T.K., Fjeld, A.D., Mumm, S., Laird, C.D., and Gartler, S.M. (1997). A variable domain of delayed replication in FRAXA fragile X chromosomes: X inactivation-like spread of late replication. Proceedings of the National Academy of Sciences of the United States of America *94*, 4587-4592.

Hansen, R.S., Thomas, S., Sandstrom, R., Canfield, T.K., Thurman, R.E., Weaver, M., Dorschner, M.O., Gartler, S.M., and Stamatoyannopoulos, J.A. (2010). Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proceedings of the National Academy of Sciences of the United States of America *107*, 139-144.

Hardman, R.A., Afshari, C.A., and Barrett, J.C. (2001). Involvement of mammalian MLH1 in the apoptotic response to peroxide-induced oxidative stress. Cancer research *61*, 1392-1397.

Harley, C.B., Futcher, A.B., and Greider, C.W. (1990). Telomeres shorten during ageing of human fibroblasts. Nature *345*, 458-460.

Harper, J.W., Adami, G.R., Wei, N., Keyomarsi, K., and Elledge, S.J. (1993). The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. Cell *75*, 805-816.

Harper, J.W., Elledge, S.J., Keyomarsi, K., Dynlacht, B., Tsai, L.H., Zhang, P., Dobrowolski, S., Bai, C., Connell-Crowley, L., Swindell, E.*, et al.* (1995). Inhibition of cyclin-dependent kinases by p21. Molecular biology of the cell *6*, 387-400.

Harr, J.C., Luperchio, T.R., Wong, X., Cohen, E., Wheelan, S.J., and Reddy, K.L. (2015). Directed targeting of chromatin to the nuclear lamina is mediated by chromatin

state and A-type lamins. The Journal of cell biology *208*, 33-52.

Harrigan, J.A., Belotserkovskaya, R., Coates, J., Dimitrova, D.S., Polo, S.E., Bradshaw, C.R., Fraser, P., and Jackson, S.P. (2011). Replication stress induces 53BP1-containing OPT domains in G1 cells. The Journal of cell biology *193*, 97-108.

Hartlerode, A.J., Guan, Y., Rajendran, A., Ura, K., Schotta, G., Xie, A., Shah, J.V., and Scully, R. (2012). Impact of histone H4 lysine 20 methylation on 53BP1 responses to chromosomal double strand breaks. PLoS One *7*, e49211.

Hegan, D.C., Narayanan, L., Jirik, F.R., Edelmann, W., Liskay, R.M., and Glazer, P.M. (2006). Differing patterns of genetic instability in mice deficient in the mismatch repair genes Pms2, Mlh1, Msh2, Msh3 and Msh6. Carcinogenesis *27*, 2402-2408.

Hellman, A., Rahat, A., Scherer, S.W., Darvasi, A., Tsui, L.C., and Kerem, B. (2000). Replication delay along FRA7H, a common fragile site on human chromosome 7, leads to chromosomal instability. Molecular and cellular biology *20*, 4420-4427.

Helmrich, A., Ballarino, M., and Tora, L. (2011). Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. Molecular cell *44*, 966-977.

Helmrich, A., Stout-Weider, K., Hermann, K., Schrock, E., and Heiden, T. (2006). Common fragile sites are conserved features of human and mouse chromosomes and relate to large active genes. Genome research *16*, 1222-1230.

Helt, C.E., Cliby, W.A., Keng, P.C., Bambara, R.A., and O'Reilly, M.A. (2005). Ataxia telangiectasia mutated (ATM) and ATM and Rad3-related protein exhibit selective target specificities in response to different forms of DNA damage. The Journal of biological chemistry *280*, 1186-1192.

Hemann, M.T., Strong, M.A., Hao, L.Y., and Greider, C.W. (2001). The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability. Cell *107*, 67-77.

Herbig, U., Jobling, W.A., Chen, B.P., Chen, D.J., and Sedivy, J.M. (2004). Telomere shortening triggers senescence of human cells through a pathway involving ATM, p53, and p21(CIP1), but not p16(INK4a). Molecular cell *14*, 501-513.

Hershko, A., Heller, H., Elias, S., and Ciechanover, A. (1983). Components of ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown. The Journal of biological chemistry *258*, 8206-8214.

Hitomi, K., Iwai, S., and Tainer, J.A. (2007). The intricate structural chemistry of base excision repair machinery: implications for DNA damage recognition, removal, and repair. DNA Repair (Amst) *6*, 410-428.

Ho, C.K., Mazon, G., Lam, A.F., and Symington, L.S. (2010). Mus81 and Yen1 promote reciprocal exchange during mitotic recombination to maintain genome integrity in budding yeast. Molecular cell *40*, 988-1000.

Hoffman, E.A., McCulley, A., Haarer, B., Arnak, R., and Feng, W. (2015). Break-seq

reveals hydroxyurea-induced chromosome fragility as a result of unscheduled conflict between DNA replication and transcription. Genome research *25*, 402-412.

Holmes, W.F., Braastad, C.D., Mitra, P., Hampe, C., Doenecke, D., Albig, W., Stein, J.L., van Wijnen, A.J., and Stein, G.S. (2005). Coordinate control and selective expression of the full complement of replication-dependent histone H4 genes in normal and cancer cells. The Journal of biological chemistry *280*, 37400-37407.

Hong, Y., Cervantes, R.B., Tichy, E., Tischfield, J.A., and Stambrook, P.J. (2007). Protecting genomic integrity in somatic cells and embryonic stem cells. Mutation research *614*, 48-55.

Hong, Y., and Stambrook, P.J. (2004). Restoration of an absent G1 arrest and protection from apoptosis in embryonic stem cells after ionizing radiation. Proceedings of the National Academy of Sciences of the United States of America *101*, 14443-14448.

Horn, P.J., and Peterson, C.L. (2002). Molecular biology. Chromatin higher order folding--wrapping up transcription. Science (New York, NY *297*, 1824-1827.

Howlett, N.G., Taniguchi, T., Durkin, S.G., D'Andrea, A.D., and Glover, T.W. (2005). The Fanconi anemia pathway is required for the DNA replication stress response and for the regulation of common fragile site stability. Human molecular genetics *14*, 693-701.

Hu, J., Adebali, O., Adar, S., and Sancar, A. (2017). Dynamic maps of UV damage formation and repair for the human genome. Proceedings of the National Academy of Sciences of the United States of America *114*, 6758-6763.

Hu, J., Meyers, R.M., Dong, J., Panchakshari, R.A., Alt, F.W., and Frock, R.L. (2016). Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. Nat Protoc *11*, 853-871.

Huang, T.T., and D'Andrea, A.D. (2006). Regulation of DNA repair by ubiquitylation. Nature reviews *7*, 323-334.

Huen, M.S., Grant, R., Manke, I., Minn, K., Yu, X., Yaffe, M.B., and Chen, J. (2007). RNF8 transduces the DNA-damage signal via histone ubiquitylation and checkpoint protein assembly. Cell *131*, 901-914.

Huertas, P., and Jackson, S.P. (2009). Human CtIP mediates cell cycle control of DNA end resection and double strand break repair. The Journal of biological chemistry *284*, 9558-9565.

Hussein, S.M., Batada, N.N., Vuoristo, S., Ching, R.W., Autio, R., Narva, E., Ng, S., Sourour, M., Hamalainen, R., Olsson, C.*, et al.* (2011). Copy number variation and selection during reprogramming to pluripotency. Nature *471*, 58-62.

Huyen, Y., Zgheib, O., Ditullio, R.A., Jr., Gorgoulis, V.G., Zacharatos, P., Petty, T.J., Sheston, E.A., Mellert, H.S., Stavridi, E.S., and Halazonetis, T.D. (2004). Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. Nature *432*,

406-411.

Iacovoni, J.S., Caron, P., Lassadi, I., Nicolas, E., Massip, L., Trouche, D., and Legube, G. (2010). High-resolution profiling of gammaH2AX around DNA double strand breaks in the mammalian genome. EMBO J *29*, 1446-1457.

Iannelli, F., Galbiati, A., Capozzo, I., Nguyen, Q., Magnuson, B., Michelini, F., D'Alessandro, G., Cabrini, M., Roncador, M., Francia, S.*, et al.* (2017). A damaged genome's transcriptional landscape through multilayered expression profiling around in situ-mapped DNA double-strand breaks. Nature communications *8*, 15656.

Iliakis, G., Murmann, T., and Soni, A. (2015). Alternative end-joining repair pathways are the ultimate backup for abrogated classical non-homologous end-joining and homologous recombination repair: Implications for the formation of chromosome translocations. Mutat Res Genet Toxicol Environ Mutagen *793*, 166-175.

Iwabuchi, K., Bartel, P.L., Li, B., Marraccino, R., and Fields, S. (1994). Two cellular proteins that bind to wild-type but not mutant p53. Proceedings of the National Academy of Sciences of the United States of America *91*, 6098-6102.

Iwabuchi, K., Basu, B.P., Kysela, B., Kurihara, T., Shibata, M., Guan, D., Cao, Y., Hamada, T., Imamura, K., Jeggo, P.A.*, et al.* (2003). Potential role for 53BP1 in DNA end-joining repair through direct interaction with DNA. The Journal of biological chemistry *278*, 36487-36495.

Iwabuchi, K., Li, B., Massa, H.F., Trask, B.J., Date, T., and Fields, S. (1998). Stimulation of p53-mediated transcriptional activation by the p53-binding proteins, 53BP1 and 53BP2. The Journal of biological chemistry *273*, 26061-26068.

Jackson, D., Dhar, K., Wahl, J.K., Wold, M.S., and Borgstahl, G.E. (2002). Analysis of the human replication protein A:Rad52 complex: evidence for crosstalk between RPA32, RPA70, Rad52 and DNA. Journal of molecular biology *321*, 133-148.

Jackson, D.A., and Pombo, A. (1998). Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. The Journal of cell biology *140*, 1285-1295.

Jackson, S.P. (2009). The DNA-damage response: new molecular insights and new approaches to cancer therapy. Biochem Soc Trans *37*, 483-494.

Jacquet, K., Fradet-Turcotte, A., Avvakumov, N., Lambert, J.P., Roques, C., Pandita, R.K., Paquet, E., Herst, P., Gingras, A.C., Pandita, T.K.*, et al.* (2016). The TIP60 Complex Regulates Bivalent Chromatin Recognition by 53BP1 through Direct H4K20me Binding and H2AK15 Acetylation. Molecular cell *62*, 409-421.

Jascur, T., and Boland, C.R. (2006). Structure and function of the components of the human DNA mismatch repair system. Int J Cancer *119*, 2030-2035.

Jeggo, P.A., and Downs, J.A. (2014). Roles of chromatin remodellers in DNA double strand break repair. Exp Cell Res *329*, 69-77.

Jeggo, P.A., and Lobrich, M. (2007). DNA double-strand breaks: their cellular and

clinical impact? Oncogene *26*, 7717-7719.

Jensen, R.B., Carreira, A., and Kowalczykowski, S.C. (2010). Purified human BRCA2 stimulates RAD51-mediated recombination. Nature *467*, 678-683.

Jiricny, J. (2006). The multifaceted mismatch-repair system. Nature reviews *7*, 335-346.

Joo, H.Y., Zhai, L., Yang, C., Nie, S., Erdjument-Bromage, H., Tempst, P., Chang, C., and Wang, H. (2007). Regulation of cell cycle progression and gene expression by H2A deubiquitination. Nature *449*, 1068-1072.

Jost, K.L., Bertulat, B., and Cardoso, M.C. (2012). Heterochromatin and gene positioning: inside, outside, any side? Chromosoma *121*, 555-563.

Jowsey, P., Morrice, N.A., Hastie, C.J., McLauchlan, H., Toth, R., and Rouse, J. (2007). Characterisation of the sites of DNA damage-induced 53BP1 phosphorylation catalysed by ATM and ATR. DNA Repair (Amst) *6*, 1536-1544.

Jun, S.H., Kim, T.G., and Ban, C. (2006). DNA mismatch repair system. Classical and fresh roles. FEBS J *273*, 1609-1619.

Kalousi, A., Hoffbeck, A.S., Selemenakis, P.N., Pinder, J., Savage, K.I., Khanna, K.K., Brino, L., Dellaire, G., Gorgoulis, V.G., and Soutoglou, E. (2015). The nuclear oncogene SET controls DNA repair by KAP1 and HP1 retention to chromatin. Cell reports *11*, 149-163.

Kalousi, A., and Soutoglou, E. (2016). Nuclear compartmentalization of DNA repair. Curr Opin Genet Dev *37*, 148-157.

Kameoka, Y., Tagawa, H., Tsuzuki, S., Karnan, S., Ota, A., Suguro, M., Suzuki, R., Yamaguchi, M., Morishima, Y., Nakamura, S.*, et al.* (2004). Contig array CGH at 3p14.2 points to the FRA3B/FHIT common fragile region as the target gene in diffuse large B-cell lymphoma. Oncogene *23*, 9148-9154.

Karakoula, A., Evans, M.D., Podmore, I.D., Hutchinson, P.E., Lunec, J., and Cooke, M.S. (2003). Quantification of UVR-induced DNA damage: global- versus gene-specific levels of thymine dimers. J Immunol Methods *277*, 27-37.

Kastan, M.B., and Bartek, J. (2004). Cell-cycle checkpoints and cancer. Nature *432*, 316-323.

Khanna, K.K., and Jackson, S.P. (2001). DNA double-strand breaks: signaling, repair and the cancer connection. Nature genetics *27*, 247-254.

Kiianitsa, K., Solinger, J.A., and Heyer, W.D. (2006). Terminal association of Rad54 protein with the Rad51-dsDNA filament. Proceedings of the National Academy of Sciences of the United States of America *103*, 9767-9772.

Kim, C., Paulus, B.F., and Wold, M.S. (1994). Interactions of human replication protein A with oligonucleotides. Biochemistry *33*, 14197-14206.

Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H.R., Hwang, J., Kim, J.I., and Kim, J.S. (2015). Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in

human cells. Nat Methods *12*, 237-243, 231 p following 243.

Kim, N., and Jinks-Robertson, S. (2012). Transcription as a source of genome instability. Nat Rev Genet *13*, 204-214.

Kin, T., and Ono, Y. (2007). Idiographica: a general-purpose web application to build idiograms on-demand for human, mouse and rat. Bioinformatics (Oxford, England) *23*, 2945-2946.

Kind, J., and van Steensel, B. (2010). Genome-nuclear lamina interactions and gene regulation. Curr Opin Cell Biol *22*, 320-325.

Kirsch, I.R. (1994). V(D)J recombination and ataxia-telangiectasia: a review. Int J Radiat Biol *66*, S97-108.

Klein, I.A., Resch, W., Jankovic, M., Oliveira, T., Yamane, A., Nakahashi, H., Di Virgilio, M., Bothmer, A., Nussenzweig, A., Robbiani, D.F.*, et al.* (2011). Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. Cell *147*, 95-106.

Kleiner, R.E., Verma, P., Molloy, K.R., Chait, B.T., and Kapoor, T.M. (2015). Chemical proteomics reveals a gammaH2AX-53BP1 interaction in the DNA damage response. Nat Chem Biol *11*, 807-814.

Kolas, N.K., Chapman, J.R., Nakada, S., Ylanko, J., Chahwan, R., Sweeney, F.D., Panier, S., Mendez, M., Wildenhain, J., Thomson, T.M.*, et al.* (2007). Orchestration of the DNA-damage response by the RNF8 ubiquitin ligase. Science (New York, NY *318*, 1637-1640.

Kolpashchikov, D.M., Khodyreva, S.N., Khlimankov, D.Y., Wold, M.S., Favre, A., and Lavrik, O.I. (2001). Polarity of human replication protein A binding to DNA. Nucleic Acids Res *29*, 373-379.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science (New York, NY *184*, 868-871.

Koundrioukoff, S., Carignon, S., Techer, H., Letessier, A., Brison, O., and Debatisse, M. (2013). Stepwise activation of the ATR signaling pathway upon increasing replication stress impacts fragile site integrity. PLoS genetics *9*, e1003643.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Krummel, K.A., Denison, S.R., Calhoun, E., Phillips, L.A., and Smith, D.I. (2002). The common fragile site FRA16D and its associated gene WWOX are highly conserved in the mouse at Fra8E1. Genes, chromosomes & cancer *34*, 154-167.

Kumagai, A., Lee, J., Yoo, H.Y., and Dunphy, W.G. (2006). TopBP1 activates the ATR-ATRIP complex. Cell *124*, 943-955.

Kumari, A., Lim, Y.X., Newell, A.H., Olson, S.B., and McCullough, A.K. (2012). Formaldehyde-induced genome instability is suppressed by an XPF-dependent pathway. DNA Repair (Amst) *11*, 236-246.

Kuroda, M., Tanabe, H., Yoshida, K., Oikawa, K., Saito, A., Kiyuna, T., Mizusawa, H., and Mukai, K. (2004). Alteration of chromosome positioning during adipocyte differentiation. Journal of cell science *117*, 5897-5903.

Lan, M.L., Acharya, M.M., Tran, K.K., Bahari-Kashani, J., Patel, N.H., Strnadel, J., Giedzinski, E., and Limoli, C.L. (2012). Characterizing the radioresponse of pluripotent and multipotent human stem cells. PLoS One *7*, e50048.

Lavin, M.F. (2004). The Mre11 complex and ATM: a two-way functional interaction in recognising and signaling DNA double strand breaks. DNA Repair (Amst) *3*, 1515-1520.

Le Beau, M.M., Rassool, F.V., Neilly, M.E., Espinosa, R., 3rd, Glover, T.W., Smith, D.I., and McKeithan, T.W. (1998). Replication of a common fragile site, FRA3B, occurs late in S phase and is delayed further upon induction: implications for the mechanism of fragile site induction. Human molecular genetics *7*, 755-761.

Le Tallec, B., Dutrillaux, B., Lachages, A.M., Millot, G.A., Brison, O., and Debatisse, M. (2011). Molecular profiling of common fragile sites in human fibroblasts. Nature structural & molecular biology *18*, 1421-1423.

Le Tallec, B., Millot, G.A., Blin, M.E., Brison, O., Dutrillaux, B., and Debatisse, M. (2013). Common fragile site profiling in epithelial and erythroid cells reveals that most recurrent cancer deletions lie in fragile sites hosting large genes. Cell reports *4*, 420-428.

Leduc, F., Faucher, D., Bikond Nkoma, G., Gregoire, M.C., Arguin, M., Wellinger, R.J., and Boissonneault, G. (2011). Genome-wide mapping of DNA strand breaks. PLoS One *6*, e17353.

Lee, J.H., and Paull, T.T. (2005). ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. Science (New York, NY *308*, 551-554.

Lee, J.H., and Paull, T.T. (2007). Activation and regulation of ATM kinase activity in response to DNA double-strand breaks. Oncogene *26*, 7741-7748.

Lee, S.H., Kim, W.H., Kim, H.K., Woo, K.M., Nam, H.S., Kim, H.S., Kim, J.G., and Cho, M.H. (2001). Altered expression of the fragile histidine triad gene in primary gastric adenocarcinomas. Biochemical and biophysical research communications *284*, 850-855.

Lemaitre, C., Grabarz, A., Tsouroula, K., Andronov, L., Furst, A., Pankotai, T., Heyer, V., Rogier, M., Attwood, K.M., Kessler, P.*, et al.* (2014). Nuclear position dictates DNA repair pathway choice. Genes Dev *28*, 2450-2463.

Lensing, S.V., Marsico, G., Hansel-Hertsch, R., Lam, E.Y., Tannahill, D., and Balasubramanian, S. (2016). DSBCapture: in situ capture and sequencing of DNA breaks. Nat Methods *13*, 855-857.

Lenz, G., Wright, G.W., Emre, N.C., Kohlhammer, H., Dave, S.S., Davis, R.E., Carty, S., Lam, L.T., Shaffer, A.L., Xiao, W.*, et al.* (2008). Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. Proceedings of the National Academy of Sciences of the United States of America *105*, 13520-13525.

Letessier, A., Millot, G.A., Koundrioukoff, S., Lachages, A.M., Vogt, N., Hansen, R.S., Malfoy, B., Brison, O., and Debatisse, M. (2011). Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. Nature *470*, 120-123.

Lewis, K.A., Mullany, S., Thomas, B., Chien, J., Loewen, R., Shridhar, V., and Cliby, W.A. (2005). Heterozygous ATR mutations in mismatch repair-deficient cancer cells have functional significance. Cancer research *65*, 7091-7095.

Li, G.M. (2008). Mechanisms and functions of DNA mismatch repair. Cell Res *18*, 85-98.

Li, S., Chang, H.H., Niewolik, D., Hedrick, M.P., Pinkerton, A.B., Hassig, C.A., Schwarz, K., and Lieber, M.R. (2014). Evidence that the DNA endonuclease ARTEMIS also has intrinsic 5'-exonuclease activity. The Journal of biological chemistry *289*, 7825-7834.

Li, V.C., Ballabeni, A., and Kirschner, M.W. (2012). Gap 1 phase length and mouse embryonic stem cell self-renewal. Proceedings of the National Academy of Sciences of the United States of America *109*, 12550-12555.

Li, W., Hu, J., Adebali, O., Adar, S., Yang, Y., Chiou, Y.Y., and Sancar, A. (2017). Human genome-wide repair map of DNA damage caused by the cigarette smoke carcinogen benzo[a]pyrene. Proceedings of the National Academy of Sciences of the United States of America *114*, 6752-6757.

Li, X., Corsa, C.A., Pan, P.W., Wu, L., Ferguson, D., Yu, X., Min, J., and Dou, Y. (2010). MOF and H4 K16 acetylation play important roles in DNA damage repair by modulating recruitment of DNA damage repair protein Mdc1. Molecular and cellular biology *30*, 5335-5347.

Li, X., and Heyer, W.D. (2009). RAD54 controls access to the invading 3'-OH end after RAD51-mediated DNA strand invasion in homologous recombination in Saccharomyces cerevisiae. Nucleic Acids Res *37*, 638-646.

Li, X., Stith, C.M., Burgers, P.M., and Heyer, W.D. (2009). PCNA is required for initiation of recombination-associated DNA synthesis by DNA polymerase delta. Molecular cell *36*, 704-713.

Lichtman, M.A. (2008). Battling the hematological malignancies: the 200 years' war. Oncologist *13*, 126-138.

Lieber, M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu Rev Biochem *79*, 181-211.

Lieber, M.R., and Wilson, T.E. (2010). SnapShot: Nonhomologous DNA end joining (NHEJ). Cell *142*, 496-496 e491.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O*., et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science (New York, NY *326*, 289-293.

Limongi, M.Z., Pelliccia, F., and Rocchi, A. (2003). Characterization of the human

common fragile site FRA2G. Genomics *81*, 93-97.

Lin, T., Chao, C., Saito, S., Mazur, S.J., Murphy, M.E., Appella, E., and Xu, Y. (2005). p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression. Nature cell biology *7*, 165-171.

Lin, Y.F., Shih, H.Y., Shang, Z., Matsunaga, S., and Chen, B.P. (2014). DNA-PKcs is required to maintain stability of Chk1 and Claspin for optimal replication stress response. Nucleic Acids Res *42*, 4463-4473.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. Nature *362*, 709-715.

Lindahl, T., and Nyberg, B. (1972). Rate of depurination of native deoxyribonucleic acid. Biochemistry *11*, 3610-3618.

Lindahl, T., and Wood, R.D. (1999). Quality control by DNA repair. Science (New York, NY *286*, 1897-1905.

Linke, S.P., Clarkin, K.C., and Wahl, G.M. (1997). p53 mediates permanent arrest over multiple cell cycles in response to gamma-irradiation. Cancer research *57*, 1171-1179.

Lisby, M., Mortensen, U.H., and Rothstein, R. (2003). Colocalization of multiple DNA double-strand breaks at a single Rad52 repair centre. Nature cell biology *5*, 572-577.

Lisby, M., and Rothstein, R. (2005). Localization of checkpoint and repair proteins in eukaryotes. Biochimie *87*, 579-589.

Liu, J., Doty, T., Gibson, B., and Heyer, W.D. (2010). Human BRCA2 protein promotes RAD51 filament formation on RPA-covered single-stranded DNA. Nature structural & molecular biology *17*, 1260-1262.

Liu, Q., Guntuku, S., Cui, X.S., Matsuoka, S., Cortez, D., Tamai, K., Luo, G., Carattini-Rivera, S., DeMayo, F., Bradley, A.*, et al.* (2000). Chk1 is an essential kinase that is regulated by Atr and required for the G(2)/M DNA damage checkpoint. Genes Dev *14*, 1448-1459.

Liu, S., Opiyo, S.O., Manthey, K., Glanzer, J.G., Ashley, A.K., Amerin, C., Troksa, K., Shrivastav, M., Nickoloff, J.A., and Oakley, G.G. (2012). Distinct roles for DNA-PK, ATM and ATR in RPA phosphorylation and checkpoint activation in response to replication stress. Nucleic Acids Res *40*, 10780-10794.

Liu, W.F., Yu, S.S., Chen, G.J., and Li, Y.Z. (2006). DNA damage checkpoint, damage repair, and genome stability. Yi Chuan Xue Bao *33*, 381-390.

Liu, Y., Nielsen, C.F., Yao, Q., and Hickson, I.D. (2014). The origins and processing of ultra fine anaphase DNA bridges. Curr Opin Genet Dev *26*, 1-5.

Lobrich, M., Shibata, A., Beucher, A., Fisher, A., Ensminger, M., Goodarzi, A.A., Barton, O., and Jeggo, P.A. (2010). gammaH2AX foci analysis for monitoring DNA double-strand break repair: strengths, limitations and optimization. Cell cycle (Georgetown, Tex *9*, 662-669.

Lomax, M.E., Folkes, L.K., and O'Neill, P. (2013). Biological consequences of radiation-induced DNA damage: relevance to radiotherapy. Clin Oncol (R Coll Radiol) *25*, 578-585.

Lopez-estrano, C., Schvartzman, J.B., Krimer, D.B., and Hernandez, P. (1998). Co-localization of polar replication fork barriers and rRNA transcription terminators in mouse rDNA. Journal of molecular biology *277*, 249-256.

Luger, K., Dechassa, M.L., and Tremethick, D.J. (2012). New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? Nature reviews *13*, 436-447.

Luijsterburg, M.S., de Krijger, I., Wiegant, W.W., Shah, R.G., Smeenk, G., de Groot, A.J.L., Pines, A., Vertegaal, A.C.O., Jacobs, J.J.L., Shah, G.M.*, et al.* (2016). PARP1 Links CHD2-Mediated Chromatin Expansion and H3.3 Deposition to DNA Repair by Non-homologous End-Joining. Molecular cell *61*, 547-562.

Lukas, C., Melander, F., Stucki, M., Falck, J., Bekker-Jensen, S., Goldberg, M., Lerenthal, Y., Jackson, S.P., Bartek, J., and Lukas, J. (2004). Mdc1 couples DNA double-strand break recognition by Nbs1 with its H2AX-dependent chromatin retention. EMBO J *23*, 2674-2683.

Lukas, C., Savic, V., Bekker-Jensen, S., Doil, C., Neumann, B., Pedersen, R.S., Grofte, M., Chan, K.L., Hickson, I.D., Bartek, J.*, et al.* (2011a). 53BP1 nuclear bodies form around DNA lesions generated by mitotic transmission of chromosomes under replication stress. Nature cell biology *13*, 243-253.

Lukas, J., Lukas, C., and Bartek, J. (2011b). More than just a focus: The chromatin response to DNA damage and its role in genome integrity maintenance. Nature cell biology *13*, 1161-1169.

Lukusa, T., and Fryns, J.P. (2008). Human chromosome fragility. Biochim Biophys Acta *1779*, 3-16.

Lydeard, J.R., Jain, S., Yamaguchi, M., and Haber, J.E. (2007). Break-induced replication and telomerase-independent telomere maintenance require Pol32. Nature *448*, 820-823.

Ma, H., Samarabandu, J., Devdhar, R.S., Acharya, R., Cheng, P.C., Meng, C., and Berezney, R. (1998). Spatial and temporal dynamics of DNA replication sites in mammalian cells. The Journal of cell biology *143*, 1415-1425.

Macheret, M., and Halazonetis, T.D. (2015). DNA replication stress as a hallmark of cancer. Annual review of pathology *10*, 425-448.

Macurek, L., Lindqvist, A., Voets, O., Kool, J., Vos, H.R., and Medema, R.H. (2010). Wip1 phosphatase is associated with chromatin and dephosphorylates gammaH2AX to promote checkpoint inhibition. Oncogene *29*, 2281-2291.

Madabhushi, R., Pan, L., and Tsai, L.H. (2014). DNA damage and its links to neurodegeneration. Neuron *83*, 266-282.

Magdalou, I., Lopez, B.S., Pasero, P., and Lambert, S.A. (2014). The causes of replication stress and their consequences on genome stability and cell fate. Semin Cell Dev Biol *30*, 154-164.

Mailand, N., Bekker-Jensen, S., Faustrup, H., Melander, F., Bartek, J., Lukas, C., and Lukas, J. (2007). RNF8 ubiquitylates histones at DNA double-strand breaks and promotes assembly of repair proteins. Cell *131*, 887-900.

Mallette, F.A., Mattiroli, F., Cui, G., Young, L.C., Hendzel, M.J., Mer, G., Sixma, T.K., and Richard, S. (2012). RNF8- and RNF168-dependent degradation of KDM4A/JMJD2A triggers 53BP1 recruitment to DNA damage sites. EMBO J *31*, 1865-1878.

Maloisel, L., Fabre, F., and Gangloff, S. (2008). DNA polymerase delta is preferentially recruited during homologous recombination to promote heteroduplex DNA extension. Molecular and cellular biology *28*, 1373-1382.

Manis, J.P., Morales, J.C., Xia, Z., Kutok, J.L., Alt, F.W., and Carpenter, P.B. (2004). 53BP1 links DNA damage-response pathways to immunoglobulin heavy chain class-switch recombination. Nat Immunol *5*, 481-487.

Mansour, W.Y., Borgmann, K., Petersen, C., Dikomey, E., and Dahm-Daphi, J. (2013). The absence of Ku but not defects in classical non-homologous end-joining is required to trigger PARP1-dependent end-joining. DNA Repair (Amst) *12*, 1134-1142.

Mansour, W.Y., Rhein, T., and Dahm-Daphi, J. (2010). The alternative end-joining pathway for repair of DNA double-strand breaks requires PARP1 but is not dependent upon microhomologies. Nucleic Acids Res *38*, 6065-6077.

Marechal, A., and Zou, L. (2013). DNA damage sensing by the ATM and ATR kinases. Cold Spring Harb Perspect Biol *5*.

Margueron, R., and Reinberg, D. (2010). Chromatin structure and the inheritance of epigenetic information. Nat Rev Genet *11*, 285-296.

Matsuoka, S., Huang, M., and Elledge, S.J. (1998). Linkage of ATM to cell cycle regulation by the Chk2 protein kinase. Science (New York, NY *282*, 1893-1897.

Matsuoka, S., Rotman, G., Ogawa, A., Shiloh, Y., Tamai, K., and Elledge, S.J. (2000). Ataxia telangiectasia-mutated phosphorylates Chk2 in vivo and in vitro. Proceedings of the National Academy of Sciences of the United States of America *97*, 10389-10394.

Mattiroli, F., Vissers, J.H., van Dijk, W.J., Ikpa, P., Citterio, E., Vermeulen, W., Marteijn, J.A., and Sixma, T.K. (2012). RNF168 ubiquitinates K13-15 on H2A/H2AX to drive DNA damage signaling. Cell *150*, 1182-1195.

Maynard, S., Schurman, S.H., Harboe, C., de Souza-Pinto, N.C., and Bohr, V.A. (2009). Base excision repair of oxidative DNA damage and association with cancer and aging. Carcinogenesis *30*, 2-10.

Mazin, A.V., Mazina, O.M., Bugreev, D.V., and Rossi, M.J. (2010). Rad54, the motor of homologous recombination. DNA Repair (Amst) *9*, 286-302.

Mazon, G., Mimitou, E.P., and Symington, L.S. (2010). SnapShot: Homologous recombination in DNA double-strand break repair. Cell *142*, 646, 646 e641.

Mekeel, K.L., Tang, W., Kachnic, L.A., Luo, C.M., DeFrank, J.S., and Powell, S.N. (1997). Inactivation of p53 results in high rates of homologous recombination. Oncogene *14*, 1847-1857.

Meng, F.L., Du, Z., Federation, A., Hu, J., Wang, Q., Kieffer-Kwon, K.R., Meyers, R.M., Amor, C., Wasserman, C.R., Neuberg, D.*, et al.* (2014). Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. Cell *159*, 1538-1548.

Michishita, E., McCord, R.A., Boxer, L.D., Barber, M.F., Hong, T., Gozani, O., and Chua, K.F. (2009). Cell cycle-dependent deacetylation of telomeric histone H3 lysine K56 by human SIRT6. Cell cycle (Georgetown, Tex *8*, 2664-2666.

Mimori, K., Druck, T., Inoue, H., Alder, H., Berk, L., Mori, M., Huebner, K., and Croce, C.M. (1999). Cancer-specific chromosome alterations in the constitutive fragile region FRA3B. Proceedings of the National Academy of Sciences of the United States of America *96*, 7456-7461.

Mimori, T., and Hardin, J.A. (1986). Mechanism of interaction between Ku protein and DNA. The Journal of biological chemistry *261*, 10375-10379.

Mine-Hattab, J., and Rothstein, R. (2012). Increased chromosome mobility facilitates homology search during recombination. Nature cell biology *14*, 510-517.

Minocherhomji, S., and Hickson, I.D. Structure-specific endonucleases: guardians of fragile site stability. Trends in cell biology *24*, 321-327.

Mishmar, D., Rahat, A., Scherer, S.W., Nyakatura, G., Hinzmann, B., Kohwi, Y., Mandel-Gutfroind, Y., Lee, J.R., Drescher, B., Sas, D.E.*, et al.* (1998). Molecular characterization of a common fragile site (FRA7H) on human chromosome 7 by the cloning of a simian virus 40 integration site. Proceedings of the National Academy of Sciences of the United States of America *95*, 8141-8146.

Misteli, T., and Soutoglou, E. (2009). The emerging role of nuclear architecture in DNA repair and genome maintenance. Nature reviews *10*, 243-254.

Mitra, A., Skrzypczak, M., Ginalski, K., and Rowicka, M. (2015). Strategies for achieving high sequencing accuracy for low diversity samples and avoiding sample bleeding using illumina platform. PLoS One *10*, e0120520.

Mohn, F., and Schubeler, D. (2009). Genetics and epigenetics: stability and plasticity during cellular differentiation. Trends Genet *25*, 129-136.

Mol, C.D., Hosfield, D.J., and Tainer, J.A. (2000). Abasic site recognition by two apurinic/apyrimidinic endonuclease families in DNA base excision repair: the 3' ends justify the means. Mutation research *460*, 211-229.

Momcilovic, O., Knobloch, L., Fornsaglio, J., Varum, S., Easley, C., and Schatten, G. (2010). DNA damage responses in human induced pluripotent stem cells and

embryonic stem cells. PLoS One *5*, e13410.

Morales, J.C., Xia, Z., Lu, T., Aldrich, M.B., Wang, B., Rosales, C., Kellems, R.E., Hittelman, W.N., Elledge, S.J., and Carpenter, P.B. (2003). Role for the BRCA1 C-terminal repeats (BRCT) protein 53BP1 in maintaining genomic stability. The Journal of biological chemistry *278*, 14971-14977.

Mordes, D.A., Glick, G.G., Zhao, R., and Cortez, D. (2008). TopBP1 activates ATR through ATRIP and a PIKK regulatory domain. Genes Dev *22*, 1478-1489.

Morris, J.R., Boutell, C., Keppler, M., Densham, R., Weekes, D., Alamshah, A., Butler, L., Galanty, Y., Pangon, L., Kiuchi, T.*, et al.* (2009). The SUMO modification pathway is involved in the BRCA1 response to genotoxic stress. Nature *462*, 886-890.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods *5*, 621-628.

Mortusewicz, O., Rothbauer, U., Cardoso, M.C., and Leonhardt, H. (2006). Differential recruitment of DNA Ligase I and III to DNA repair sites. Nucleic Acids Res *34*, 3523-3532.

Munoz-Galvan, S., Jimeno, S., Rothstein, R., and Aguilera, A. (2013). Histone H3K56 acetylation, Rad52, and non-DNA repair factors control double-strand break repair choice with the sister chromatid. PLoS genetics *9*, e1003237.

Murphy, A.K., Fitzgerald, M., Ro, T., Kim, J.H., Rabinowitsch, A.I., Chowdhury, D., Schildkraut, C.L., and Borowiec, J.A. (2014). Phosphorylated RPA recruits PALB2 to stalled DNA replication forks to facilitate fork recovery. The Journal of cell biology *206*, 493-507.

Nagai, S., Dubrana, K., Tsai-Pflugfelder, M., Davidson, M.B., Roberts, T.M., Brown, G.W., Varela, E., Hediger, F., Gasser, S.M., and Krogan, N.J. (2008). Functional targeting of DNA damage to a nuclear pore-associated SUMO-dependent ubiquitin ligase. Science (New York, NY *322*, 597-602.

Naim, V., and Rosselli, F. (2009). The FANC pathway and BLM collaborate during mitosis to prevent micro-nucleation and chromosome abnormalities. Nature cell biology *11*, 761-768.

Naim, V., Wilhelm, T., Debatisse, M., and Rosselli, F. (2013). ERCC1 and MUS81-EME1 promote sister chromatid separation by processing late replication intermediates at common fragile sites during mitosis. Nature cell biology *15*, 1008-1015.

Nakamura, A.J., Rao, V.A., Pommier, Y., and Bonner, W.M. (2010). The complexity of phosphorylated H2AX foci formation and DNA repair assembly at DNA double-strand breaks. Cell cycle (Georgetown, Tex *9*, 389-397.

Nakamura, H., Morita, T., and Sato, C. (1986). Structural organizations of replicon domains during DNA synthetic phase in the mammalian nucleus. Exp Cell Res *165*, 291-297.

Nakayasu, H., and Berezney, R. (1989). Mapping replicational sites in the eucaryotic

cell nucleus. The Journal of cell biology *108*, 1-11.

Narlikar, G.J., Fan, H.Y., and Kingston, R.E. (2002). Cooperation between complexes that regulate chromatin structure and transcription. Cell *108*, 475-487.

Natale, F., Rapp, A., Yu, W., Maiser, A., Harz, H., Scholl, A., Grulich, S., Anton, T., Horl, D., Chen, W.*, et al.* (2017). Identification of the elementary structural units of the DNA damage response. Nature communications *8*, 15760.

Nazarov, I.B., Smirnova, A.N., Krutilina, R.I., Svetlova, M.P., Solovjeva, L.V., Nikiforov, A.A., Oei, S.L., Zalenskaya, I.A., Yau, P.M., Bradbury, E.M.*, et al.* (2003). Dephosphorylation of histone gamma-H2AX during repair of DNA double-strand breaks in mammalian cells and its inhibition by calyculin A. Radiat Res *160*, 309-317.

Negrini, M., Monaco, C., Vorechovsky, I., Ohta, M., Druck, T., Baffa, R., Huebner, K., and Croce, C.M. (1996). The FHIT gene at 3p14.2 is abnormal in breast carcinomas. Cancer research *56*, 3173-3179.

Negrini, S., Gorgoulis, V.G., and Halazonetis, T.D. (2010). Genomic instability--an evolving hallmark of cancer. Nature reviews *11*, 220-228.

Nghiem, P., Park, P.K., Kim, Y., Vaziri, C., and Schreiber, S.L. (2001). ATR inhibition selectively sensitizes G1 checkpoint-deficient cells to lethal premature chromatin condensation. Proceedings of the National Academy of Sciences of the United States of America *98*, 9092-9097.

Nicassio, F., Corrado, N., Vissers, J.H., Areces, L.B., Bergink, S., Marteijn, J.A., Geverts, B., Houtsmuller, A.B., Vermeulen, W., Di Fiore, P.P.*, et al.* (2007). Human USP3 is a chromatin modifier required for S phase progression and genome stability. Curr Biol *17*, 1972-1977.

Nimonkar, A.V., Genschel, J., Kinoshita, E., Polaczek, P., Campbell, J.L., Wyman, C., Modrich, P., and Kowalczykowski, S.C. (2011). BLM-DNA2-RPA-MRN and EXO1-BLM-RPA-MRN constitute two DNA end resection machineries for human DNA break repair. Genes Dev *25*, 350-362.

Nimonkar, A.V., Sica, R.A., and Kowalczykowski, S.C. (2009). Rad52 promotes second-end DNA capture in double-stranded break repair to form complement-stabilized joint molecules. Proceedings of the National Academy of Sciences of the United States of America *106*, 3077-3082.

Niu, H., Erdjument-Bromage, H., Pan, Z.Q., Lee, S.H., Tempst, P., and Hurwitz, J. (1997). Mapping of amino acid residues in the p34 subunit of human single-stranded DNA-binding protein phosphorylated by DNA-dependent protein kinase and Cdc2 kinase in vitro. The Journal of biological chemistry *272*, 12634-12641.

Noon, A.T., Shibata, A., Rief, N., Lobrich, M., Stewart, G.S., Jeggo, P.A., and Goodarzi, A.A. (2010). 53BP1-dependent robust localized KAP-1 phosphorylation is essential for heterochromatic DNA double-strand break repair. Nature cell biology *12*, 177-184.

Norbury, C.J., and Hickson, I.D. (2001). Cellular responses to DNA damage. Annu Rev Pharmacol Toxicol *41*, 367-401.

Oberdoerffer, P., and Sinclair, D.A. (2007). The role of nuclear architecture in genomic instability and ageing. Nature reviews *8*, 692-702.

Ochi, T., Blackford, A.N., Coates, J., Jhujh, S., Mehmood, S., Tamura, N., Travers, J., Wu, Q., Draviam, V.M., Robinson, C.V.*, et al.* (2015). DNA repair. PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. Science (New York, NY *347*, 185-188.

Ohta, M., Inoue, H., Cotticelli, M.G., Kastury, K., Baffa, R., Palazzo, J., Siprashvili, Z., Mori, M., McCue, P., Druck, T.*, et al.* (1996). The FHIT gene, spanning the chromosome 3p14.2 fragile site and renal carcinoma-associated t(3;8) breakpoint, is abnormal in digestive tract cancers. Cell *84*, 587-597.

Olins, A.L., and Olins, D.E. (1974). Spheroid chromatin units (v bodies). Science (New York, NY *183*, 330-332.

Olive, P.L., and Banath, J.P. (2006). The comet assay: a method to measure DNA damage in individual cells. Nat Protoc *1*, 23-29.

Oliveira, T.Y., Resch, W., Jankovic, M., Casellas, R., Nussenzweig, M.C., and Klein, I.A. (2012). Translocation capture sequencing: a method for high throughput mapping of chromosomal rearrangements. J Immunol Methods *375*, 176-181.

Osley, M.A., Tsukuda, T., and Nickoloff, J.A. (2007). ATP-dependent chromatin remodeling factors and DNA damage repair. Mutation research *618*, 65-80.

Oza, P., Jaspersen, S.L., Miele, A., Dekker, J., and Peterson, C.L. (2009). Mechanisms that regulate localization of a DNA double-strand break to the nuclear periphery. Genes Dev *23*, 912-927.

Ozeri-Galai, E., Lebofsky, R., Rahat, A., Bester, A.C., Bensimon, A., and Kerem, B. (2011). Failure of origin activation in response to fork stalling leads to chromosomal instability at fragile sites. Molecular cell *43*, 122-131.

Palakodeti, A., Han, Y., Jiang, Y., and Le Beau, M.M. (2004). The role of late/slow replication of the FRA16D in common fragile site induction. Genes, chromosomes & cancer *39*, 71-76.

Palakodeti, A., Lucas, I., Jiang, Y., Young, D.J., Fernald, A.A., Karrison, T., and Le Beau, M.M. (2010). Impaired replication dynamics at the FRA3B common fragile site. Human molecular genetics *19*, 99-110.

Paliwal, S., Kanagaraj, R., Sturzenegger, A., Burdova, K., and Janscak, P. (2014). Human RECQ5 helicase promotes repair of DNA double-strand breaks by synthesis-dependent strand annealing. Nucleic Acids Res *42*, 2380-2390.

Palumbo, E., Matricardi, L., Tosoni, E., Bensimon, A., and Russo, A. (2010). Replication dynamics at common fragile site FRA6E. Chromosoma *119*, 575-587.

Panier, S., and Boulton, S.J. (2014). Double-strand break repair: 53BP1 comes into focus. Nature reviews *15*, 7-18.

Papamichos-Chronakis, M., and Peterson, C.L. (2013). Chromatin and the genome integrity network. Nat Rev Genet *14*, 62-75.

Parada, L.A., McQueen, P.G., and Misteli, T. (2004). Tissue-specific spatial organization of genomes. Genome Biol *5*, R44.

Paull, T.T., and Gellert, M. (1999). Nbs1 potentiates ATP-driven DNA unwinding and endonuclease cleavage by the Mre11/Rad50 complex. Genes Dev *13*, 1276-1288.

Paull, T.T., Rogakou, E.P., Yamazaki, V., Kirchgessner, C.U., Gellert, M., and Bonner, W.M. (2000). A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. Curr Biol *10*, 886-895.

Paulsen, R.D., and Cimprich, K.A. (2007). The ATR pathway: fine-tuning the fork. DNA Repair (Amst) *6*, 953-966.

Pei, H., Zhang, L., Luo, K., Qin, Y., Chesi, M., Fei, F., Bergsagel, P.L., Wang, L., You, Z., and Lou, Z. (2011). MMSET regulates histone H4K20 methylation and 53BP1 accumulation at DNA damage sites. Nature *470*, 124-128.

Pellegrino, S., Michelena, J., Teloni, F., Imhof, R., and Altmeyer, M. (2017). Replication-Coupled Dilution of H4K20me2 Guides 53BP1 to Pre-replicative Chromatin. Cell reports *19*, 1819-1831.

Pelliccia, F., Bosco, N., Curatolo, A., and Rocchi, A. (2008). Replication timing of two human common fragile sites: FRA1H and FRA2G. Cytogenet Genome Res *121*, 196-200.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W., Solovei, I., Brugman, W., Graf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M.*, et al.* (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Molecular cell *38*, 603-613.

Petermann, E., Helleday, T., and Caldecott, K.W. (2008). Claspin promotes normal replication fork rates in human cells. Molecular biology of the cell *19*, 2373-2378.

Petermann, E., Maya-Mendoza, A., Zachos, G., Gillespie, D.A., Jackson, D.A., and Caldecott, K.W. (2006). Chk1 requirement for high global rates of replication fork progression during normal vertebrate S phase. Molecular and cellular biology *26*, 3319-3326.

Peterson, C.L., and Almouzni, G. (2013). Nucleosome dynamics as modular systems that integrate DNA damage and repair. Cold Spring Harb Perspect Biol *5*.

Peterson, C.L., and Cote, J. (2004). Cellular machineries for chromosomal DNA repair. Genes Dev *18*, 602-616.

Peterson, S.E., Li, Y., Wu-Baer, F., Chait, B.T., Baer, R., Yan, H., Gottesman, M.E., and Gautier, J. (2013). Activation of DSB processing requires phosphorylation of CtIP by ATR. Molecular cell *49*, 657-667.

Pfister, S.X., Ahrabi, S., Zalmas, L.P., Sarkar, S., Aymard, F., Bachrati, C.Z., Helleday,

T., Legube, G., La Thangue, N.B., Porter, A.C.*, et al.* (2014). SETD2-dependent histone H3K36 trimethylation is required for homologous recombination repair and genome stability. Cell reports *7*, 2006-2018.

Pickersgill, H., Kalverda, B., de Wit, E., Talhout, W., Fornerod, M., and van Steensel, B. (2006). Characterization of the Drosophila melanogaster genome at the nuclear lamina. Nature genetics *38*, 1005-1014.

Piette, J., Merville-Louis, M.P., and Decuyper, J. (1986). Damages induced in nucleic acids by photosensitization. Photochem Photobiol *44*, 793-802.

Piret, B., Schoonbroodt, S., and Piette, J. (1999). The ATM protein is required for sustained activation of NF-kappaB following DNA damage. Oncogene *18*, 2261-2271.

Polo, S.E., Blackford, A.N., Chapman, J.R., Baskcomb, L., Gravel, S., Rusch, A., Thomas, A., Blundred, R., Smith, P., Kzhyshkowska, J.*, et al.* (2012). Regulation of DNA-end resection by hnRNPU-like proteins promotes DNA double-strand break signaling and repair. Molecular cell *45*, 505-516.

Polo, S.E., and Jackson, S.P. (2011). Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. Genes Dev *25*, 409-433.

Polo, S.E., Kaidi, A., Baskcomb, L., Galanty, Y., and Jackson, S.P. (2010). Regulation of DNA-damage responses and cell-cycle progression by the chromatin remodelling factor CHD4. EMBO J *29*, 3130-3139.

Pope, B.D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D.L., Wang, Y., Hansen, R.S., Canfield, T.K.*, et al.* (2014). Topologically associating domains are stable units of replication-timing regulation. Nature *515*, 402-405.

Poulsen, M., Lukas, C., Lukas, J., Bekker-Jensen, S., and Mailand, N. (2012). Human RNF169 is a negative regulator of the ubiquitin-dependent response to DNA double-strand breaks. The Journal of cell biology *197*, 189-199.

Price, B.D., and D'Andrea, A.D. (2013). Chromatin remodeling at DNA double-strand breaks. Cell *152*, 1344-1354.

Ramirez-Lugo, J.S., Yoo, H.Y., Yoon, S.J., and Dunphy, W.G. (2011). CtIP interacts with TopBP1 and Nbs1 in the response to double-stranded DNA breaks (DSBs) in Xenopus egg extracts. Cell cycle (Georgetown, Tex *10*, 469-480.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S.*, et al.* (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665-1680.

Rastogi, R.P., Richa, Kumar, A., Tyagi, M.B., and Sinha, R.P. (2010). Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. J Nucleic Acids *2010*, 592980.

Reaper, P.M., Griffiths, M.R., Long, J.M., Charrier, J.D., Maccormick, S., Charlton, P.A., Golec, J.M., and Pollard, J.R. (2011). Selective killing of ATM- or p53-deficient cancer

cells through inhibition of ATR. Nat Chem Biol *7*, 428-430.

Riballo, E., Woodbine, L., Stiff, T., Walker, S.A., Goodarzi, A.A., and Jeggo, P.A. (2009). XLF-Cernunnos promotes DNA ligase IV-XRCC4 re-adenylation following ligation. Nucleic Acids Res *37*, 482-492.

Rich, T., Allen, R.L., and Wyllie, A.H. (2000). Defying death after DNA damage. Nature *407*, 777-783.

Ried, K., Finnis, M., Hobson, L., Mangelsdorf, M., Dayan, S., Nancarrow, J.K., Woollatt, E., Kremmidiotis, G., Gardner, A., Venter, D.*, et al.* (2000). Common chromosomal fragile site FRA16D sequence: identification of the FOR gene spanning FRA16D and homozygous deletions and translocation breakpoints in cancer cells. Human molecular genetics *9*, 1651-1663.

Rodriguez, H., Akman, S.A., Holmquist, G.P., Wilson, G.L., Driggers, W.J., and LeDoux, S.P. (2000). Mapping oxidative DNA damage using ligation-mediated polymerase chain reaction technology. Methods *22*, 148-156.

Rogakou, E.P., Boon, C., Redon, C., and Bonner, W.M. (1999). Megabase chromatin domains involved in DNA double-strand breaks in vivo. The Journal of cell biology *146*, 905-916.

Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., and Bonner, W.M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. The Journal of biological chemistry *273*, 5858-5868.

Roos, W.P., and Kaina, B. (2013). DNA damage-induced cell death: from specific DNA lesions to the DNA damage response and apoptosis. Cancer Lett *332*, 237-248.

Roux, K.J., Kim, D.I., Raida, M., and Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. The Journal of cell biology *196*, 801-810.

Rozier, L., El-Achkar, E., Apiou, F., and Debatisse, M. (2004). Characterization of a conserved aphidicolin-sensitive common fragile site at human 4q22 and mouse 6C1: possible association with an inherited disease and cancer. Oncogene *23*, 6872-6880.

Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., Schulz, T.C., Robins, A.J., Dalton, S., and Gilbert, D.M. (2010). Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. Genome research *20*, 761-770.

Saito, S., Goodarzi, A.A., Higashimoto, Y., Noda, Y., Lees-Miller, S.P., Appella, E., and Anderson, C.W. (2002). ATM mediates phosphorylation at multiple p53 sites, including Ser(46), in response to ionizing radiation. The Journal of biological chemistry *277*, 12491-12494.

San Filippo, J., Sung, P., and Klein, H. (2008). Mechanism of eukaryotic homologous recombination. Annu Rev Biochem *77*, 229-257.

Santoro, R., and Blandino, G. (2010). p53: The pivot between cell cycle arrest and

senescence. Cell cycle (Georgetown, Tex *9*, 4262-4263.

Saretzki, G., Armstrong, L., Leake, A., Lako, M., and von Zglinicki, T. (2004). Stress defense in murine embryonic stem cells is superior to that of various differentiated murine cells. Stem Cells *22*, 962-971.

Saretzki, G., Walter, T., Atkinson, S., Passos, J.F., Bareth, B., Keith, W.N., Stewart, R., Hoare, S., Stojkovic, M., Armstrong, L.*, et al.* (2008). Downregulation of multiple stress defense mechanisms during differentiation of human embryonic stem cells. Stem Cells *26*, 455-464.

Sartori, A.A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J., and Jackson, S.P. (2007). Human CtIP promotes DNA end resection. Nature *450*, 509-514.

Savatier, P., Huang, S., Szekely, L., Wiman, K.G., and Samarut, J. (1994). Contrasting patterns of retinoblastoma protein expression in mouse embryonic stem cells and embryonic fibroblasts. Oncogene *9*, 809-818.

Savic, V., Yin, B., Maas, N.L., Bredemeyer, A.L., Carpenter, A.C., Helmink, B.A., Yang-Iott, K.S., Sleckman, B.P., and Bassing, C.H. (2009). Formation of dynamic gamma-H2AX domains along broken DNA strands is distinctly regulated by ATM and MDC1 and dependent upon H2AX densities in chromatin. Molecular cell *34*, 298-310.

Schick, S., Fournier, D., Thakurela, S., Sahu, S.K., Garding, A., and Tiwari, V.K. (2015). Dynamics of chromatin accessibility and epigenetic state in response to UV damage. Journal of cell science *128*, 4380-4394.

Schipler, A., and Iliakis, G. (2013). DNA double-strand-break complexity levels and their possible contributions to the probability for error-prone processing and repair pathway choice. Nucleic Acids Res *41*, 7589-7605.

Schneiderman, J.I., Orsi, G.A., Hughes, K.T., Loppin, B., and Ahmad, K. (2012). Nucleosome-depleted chromatin gaps recruit assembly factors for the H3.3 histone variant. Proceedings of the National Academy of Sciences of the United States of America *109*, 19721-19726.

Schotta, G., Sengupta, R., Kubicek, S., Malin, S., Kauer, M., Callen, E., Celeste, A., Pagani, M., Opravil, S., De La Rosa-Velazquez, I.A.*, et al.* (2008). A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. Genes Dev *22*, 2048-2061.

Schultz, L.B., Chehab, N.H., Malikzay, A., and Halazonetis, T.D. (2000). p53 binding protein 1 (53BP1) is an early participant in the cellular response to DNA double-strand breaks. The Journal of cell biology *151*, 1381-1390.

Schwartz, D.C., and Cantor, C.R. (1984). Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell *37*, 67-75.

Schwartz, M., Zlotorynski, E., and Kerem, B. (2006). The molecular basis of common and rare fragile sites. Cancer Lett *232*, 13-26.

Schwer, B., Wei, P.C., Chang, A.N., Kao, J., Du, Z., Meyers, R.M., and Alt, F.W. (2016). Transcription-associated processes cause DNA double-strand breaks and translocations in neural stem/progenitor cells. Proceedings of the National Academy of Sciences of the United States of America *113*, 2258-2263.

Sedelnikova, O.A., Horikawa, I., Zimonjic, D.B., Popescu, N.C., Bonner, W.M., and Barrett, J.C. (2004). Senescing human cells and ageing mice accumulate DNA lesions with unrepairable double-strand breaks. Nature cell biology *6*, 168-170.

Serrano, M., Lin, A.W., McCurrach, M.E., Beach, D., and Lowe, S.W. (1997). Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. Cell *88*, 593-602.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458-472.

Shah, S.N., Opresko, P.L., Meng, X., Lee, M.Y., and Eckert, K.A. (2010). DNA structure and the Werner protein modulate human DNA polymerase delta-dependent replication dynamics within the common fragile site FRA16D. Nucleic Acids Res *38*, 1149-1162.

Shaltiel, I.A., Krenning, L., Bruinsma, W., and Medema, R.H. (2015). The same, only different - DNA damage checkpoints and their reversal throughout the cell cycle. Journal of cell science *128*, 607-620.

Sharma, G.G., So, S., Gupta, A., Kumar, R., Cayrou, C., Avvakumov, N., Bhadra, U., Pandita, R.K., Porteus, M.H., Chen, D.J.*, et al.* (2010). MOF and histone H4 acetylation at lysine 16 are critical for DNA damage response and double-strand break repair. Molecular and cellular biology *30*, 3582-3595.

Shi, W., Feng, Z., Zhang, J., Gonzalez-Suarez, I., Vanderwaal, R.P., Wu, X., Powell, S.N., Roti Roti, J.L., Gonzalo, S., and Zhang, J. (2010). The role of RPA2 phosphorylation in homologous recombination in response to replication arrest. Carcinogenesis *31*, 994-1002.

Shieh, S.Y., Ahn, J., Tamai, K., Taya, Y., and Prives, C. (2000). The human homologs of checkpoint kinases Chk1 and Cds1 (Chk2) phosphorylate p53 at multiple DNA damage-inducible sites. Genes Dev *14*, 289-300.

Shiloh, Y. (2001). ATM and ATR: networking cellular responses to DNA damage. Curr Opin Genet Dev *11*, 71-77.

Shiloh, Y. (2003). ATM and related protein kinases: safeguarding genome integrity. Nat Rev Cancer *3*, 155-168.

Shiloh, Y., and Ziv, Y. (2013). The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. Nature reviews *14*, 197-210.

Shim, E.Y., Ma, J.L., Oum, J.H., Yanez, Y., and Lee, S.E. (2005). The yeast chromatin remodeler RSC complex facilitates end joining repair of DNA double-strand breaks. Molecular and cellular biology *25*, 3934-3944.

Shimura, T., Martin, M.M., Torres, M.J., Gu, C., Pluth, J.M., DeBernardi, M.A., McDonald, J.S., and Aladjem, M.I. (2007). DNA-PK is involved in repairing a transient surge of DNA breaks induced by deceleration of DNA replication. Journal of molecular biology *367*, 665-680.

Shridhar, R., Shridhar, V., Wang, X., Paradee, W., Dugan, M., Sarkar, F., Wilke, C., Glover, T.W., Vaitkevicius, V.K., and Smith, D.I. (1996). Frequent breakpoints in the 3p14.2 fragile site, FRA3B, in pancreatic tumors. Cancer research *56*, 4347-4350.

Shrivastav, M., De Haro, L.P., and Nickoloff, J.A. (2008). Regulation of DNA double-strand break repair pathway choice. Cell Res *18*, 134-147.

Singh, I., Ozturk, N., Cordero, J., Mehta, A., Hasan, D., Cosentino, C., Sebastian, C., Kruger, M., Looso, M., Carraro, G.*, et al.* (2016). High mobility group protein-mediated transcription requires DNA damage marker gamma-H2AX. Cell Res *25*, 837-850.

Singh, N.P., McCoy, M.T., Tice, R.R., and Schneider, E.L. (1988). A simple technique for quantitation of low levels of DNA damage in individual cells. Exp Cell Res *175*, 184-191.

Smerdon, M.J., and Thoma, F. (1990). Site-specific DNA repair at the nucleosome level in a yeast minichromosome. Cell *61*, 675-684.

Smith, B.L., Bauer, G.B., and Povirk, L.F. (1994). DNA damage induced by bleomycin, neocarzinostatin, and melphalan in a precisely positioned nucleosome. Asymmetry in protection at the periphery of nucleosome-bound DNA. The Journal of biological chemistry *269*, 30587-30594.

Smith, J., Tho, L.M., Xu, N., and Gillespie, D.A. (2010). The ATM-Chk2 and ATR-Chk1 pathways in DNA damage signaling and cancer. Adv Cancer Res *108*, 73-112.

Smogorzewska, A., and de Lange, T. (2002). Different telomere damage signaling pathways in human and mouse cells. EMBO J *21*, 4338-4348.

Solinger, J.A., Kiianitsa, K., and Heyer, W.D. (2002). Rad54, a Swi2/Snf2-like recombinational repair protein, disassembles Rad51:dsDNA filaments. Molecular cell *10*, 1175-1188.

Sone, K., Piao, L., Nakakido, M., Ueda, K., Jenuwein, T., Nakamura, Y., and Hamamoto, R. (2014). Critical role of lysine 134 methylation on histone H2AX for gamma-H2AX production and DNA repair. Nature communications *5*, 5691.

Soria, G., Polo, S.E., and Almouzni, G. (2012). Prime, repair, restore: the active role of chromatin in the DNA damage response. Molecular cell *46*, 722-734.

Soutoglou, E., Dorn, J.F., Sengupta, K., Jasin, M., Nussenzweig, A., Ried, T., Danuser, G., and Misteli, T. (2007). Positional stability of single double-strand breaks in mammalian cells. Nature cell biology *9*, 675-682.

Sozzi, G., Veronese, M.L., Negrini, M., Baffa, R., Cotticelli, M.G., Inoue, H., Tornielli, S., Pilotti, S., De Gregorio, L., Pastorino, U.*, et al.* (1996). The FHIT gene 3p14.2 is abnormal in lung cancer. Cell *85*, 17-26.

Sterner, D.E., and Berger, S.L. (2000). Acetylation of histones and transcription-related factors. Microbiol Mol Biol Rev *64*, 435-459.

Stewart, G.S., Panier, S., Townsend, K., Al-Hakim, A.K., Kolas, N.K., Miller, E.S., Nakada, S., Ylanko, J., Olivarius, S., Mendez, M.*, et al.* (2009). The RIDDLE syndrome protein mediates a ubiquitin-dependent signaling cascade at sites of DNA damage. Cell *136*, 420-434.

Stewart, G.S., Wang, B., Bignell, C.R., Taylor, A.M., and Elledge, S.J. (2003). MDC1 is a mediator of the mammalian DNA damage checkpoint. Nature *421*, 961-966.

Stiff, T., O'Driscoll, M., Rief, N., Iwabuchi, K., Lobrich, M., and Jeggo, P.A. (2004). ATM and DNA-PK function redundantly to phosphorylate H2AX after exposure to ionizing radiation. Cancer research *64*, 2390-2396.

Stiff, T., Walker, S.A., Cerosaletti, K., Goodarzi, A.A., Petermann, E., Concannon, P., O'Driscoll, M., and Jeggo, P.A. (2006). ATR-dependent phosphorylation and activation of ATM in response to UV treatment or replication fork stalling. EMBO J *25*, 5775-5782.

Stucki, M., Clapperton, J.A., Mohammad, D., Yaffe, M.B., Smerdon, S.J., and Jackson, S.P. (2005). MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. Cell *123*, 1213-1226.

Stucki, M., and Jackson, S.P. (2006). gammaH2AX and MDC1: anchoring the DNA-damage-response machinery to broken chromosomes. DNA Repair (Amst) *5*, 534-543.

Subba Rao, K. (2007). Mechanisms of disease: DNA repair defects and neurological disease. Nat Clin Pract Neurol *3*, 162-172.

Subramanian, P.S., Nelson, D.L., and Chinault, A.C. (1996). Large domains of apparent delayed replication timing associated with triplet repeat expansion at FRAXA and FRAXE. Am J Hum Genet *59*, 407-416.

Sugiyama, T., and Kowalczykowski, S.C. (2002). Rad52 protein associates with replication protein A (RPA)-single-stranded DNA to accelerate Rad51-mediated displacement of RPA and presynaptic complex formation. The Journal of biological chemistry *277*, 31663-31672.

Sung, P., Krejci, L., Van Komen, S., and Sehorn, M.G. (2003). Rad51 recombinase and recombination mediators. The Journal of biological chemistry *278*, 42729-42732.

Sutherland, J.C., Lin, B., Monteleone, D.C., Mugavero, J., Sutherland, B.M., and Trunk, J. (1987). Electronic imaging system for direct and rapid quantitation of fluorescence from electrophoretic gels: application to ethidium bromide-stained DNA. Anal Biochem *163*, 446-457.

Svetlova, M.P., Solovjeva, L.V., and Tomilin, N.V. (2010). Mechanism of elimination of phosphorylated histone H2AX from chromatin after repair of DNA double-strand breaks. Mutation research *685*, 54-60.

Swuec, P., and Costa, A. (2014). Molecular mechanism of double Holliday junction

dissolution. Cell Biosci *4*, 36.

Symington, L.S., and Gautier, J. (2011). Double-strand break end resection and repair pathway choice. Annual review of genetics *45*, 247-271.

Tamm, C., Pijuan Galito, S., and Anneren, C. (2013). A comparative study of protocols for mouse embryonic stem cell culturing. PLoS One *8*, e81156.

te Riele, H., Maandag, E.R., and Berns, A. (1992). Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs. Proceedings of the National Academy of Sciences of the United States of America *89*, 5128-5132.

Thoma, F. (1999). Light and dark in chromatin repair: repair of UV-induced DNA lesions by photolyase and nucleotide excision repair. EMBO J *18*, 6585-6598.

Thompson, L.H. (2012). Recognition, signaling, and repair of DNA double-strand breaks produced by ionizing radiation in mammalian cells: the molecular choreography. Mutation research *751*, 158-246.

Thoms, K.M., Baesecke, J., Emmert, B., Hermann, J., Roedling, T., Laspe, P., Leibeling, D., Truemper, L., and Emmert, S. (2007). Functional DNA repair system analysis in haematopoietic progenitor cells using host cell reactivation. Scand J Clin Lab Invest *67*, 580-588.

Thorslund, T., Ripplinger, A., Hoffmann, S., Wild, T., Uckelmann, M., Villumsen, B., Narita, T., Sixma, T.K., Choudhary, C., Bekker-Jensen, S.*, et al.* (2015). Histone H1 couples initiation and amplification of ubiquitin signalling after DNA damage. Nature *527*, 389-393.

Tichy, E.D., Pillai, R., Deng, L., Tischfield, J.A., Hexley, P., Babcock, G.F., and Stambrook, P.J. (2012). The abundance of Rad51 protein in mouse embryonic stem cells is regulated at multiple levels. Stem Cell Res *9*, 124-134.

Tomimatsu, N., Mukherjee, B., Deland, K., Kurimasa, A., Bolderson, E., Khanna, K.K., and Burma, S. (2012). Exo1 plays a major role in DNA end resection in humans and influences double-strand break repair and damage signaling decisions. DNA Repair (Amst) *11*, 441-448.

Trojer, P., and Reinberg, D. (2007). Facultative heterochromatin: is there a distinctive molecular signature? Molecular cell *28*, 1-13.

Tsai, S.Q., and Joung, J.K. (2016). Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. Nat Rev Genet *17*, 300-312.

Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P.*, et al.* (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat Biotechnol *33*, 187-197.

Tsouroula, K., Furst, A., Rogier, M., Heyer, V., Maglott-Roth, A., Ferrand, A., Reina-San-Martin, B., and Soutoglou, E. (2016). Temporal and Spatial Uncoupling of DNA Double Strand Break Repair Pathways within Mammalian Heterochromatin.

Molecular cell *63*, 293-305.

Tuduri, S., Crabbe, L., Conti, C., Tourriere, H., Holtgreve-Grez, H., Jauch, A., Pantesco, V., De Vos, J., Thomas, A., Theillet, C.*, et al.* (2009). Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. Nature cell biology *11*, 1315-1324.

Turenne, G.A., Paul, P., Laflair, L., and Price, B.D. (2001). Activation of p53 transcriptional activity requires ATM's kinase domain and multiple N-terminal serine residues of p53. Oncogene *20*, 5100-5110.

Turinetto, V., Orlando, L., Sanchez-Ripoll, Y., Kumpfmueller, B., Storm, M.P., Porcedda, P., Minieri, V., Saviozzi, S., Accomasso, L., Cibrario Rocchietti, E.*, et al.* (2012). High basal gammaH2AX levels sustain self-renewal of mouse embryonic and induced pluripotent stem cells. Stem Cells *30*, 1414-1423.

Uziel, T., Lerenthal, Y., Moyal, L., Andegeko, Y., Mittelman, L., and Shiloh, Y. (2003). Requirement of the MRN complex for ATM activation by DNA damage. EMBO J *22*, 5612-5621.

Valerie, K., and Povirk, L.F. (2003). Regulation and mechanisms of mammalian double-strand break repair. Oncogene *22*, 5792-5812.

Van Sloun, P.P., Jansen, J.G., Weeda, G., Mullenders, L.H., van Zeeland, A.A., Lohman, P.H., and Vrieling, H. (1999). The role of nucleotide excision repair in protecting embryonic stem cells from genotoxic effects of UV-induced DNA damage. Nucleic Acids Res *27*, 3276-3282.

van Steensel, B., and Belmont, A.S. (2017). Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. Cell *169*, 780-791.

Vempati, R.K., Jayani, R.S., Notani, D., Sengupta, A., Galande, S., and Haldar, D. (2010). p300-mediated acetylation of histone H3 lysine 56 functions in DNA damage response in mammals. The Journal of biological chemistry *285*, 28553-28564.

Veres, A., Gosis, B.S., Ding, Q., Collins, R., Ragavendran, A., Brand, H., Erdin, S., Cowan, C.A., Talkowski, M.E., and Musunuru, K. (2014). Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. Cell Stem Cell *15*, 27-30.

Verkerk, A.J., Pieretti, M., Sutcliffe, J.S., Fu, Y.H., Kuhl, D.P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., Zhang, F.P.*, et al.* (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell *65*, 905-914.

Wang, B., Matsuoka, S., Carpenter, P.B., and Elledge, S.J. (2002). 53BP1, a mediator of the DNA damage checkpoint. Science (New York, NY *298*, 1435-1438.

Wang, L., Darling, J., Zhang, J.S., Huang, H., Liu, W., and Smith, D.I. (1999). Allele-specific late replication and fragility of the most active common fragile site, FRA3B. Human molecular genetics *8*, 431-437.

Wang, M., Wu, W., Wu, W., Rosidi, B., Zhang, L., Wang, H., and Iliakis, G. (2006). PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. Nucleic Acids Res *34*, 6170-6182.

Wang, X., Wang, Y., Wu, X., Wang, J., Wang, Y., Qiu, Z., Chang, T., Huang, H., Lin, R.J., and Yee, J.K. (2015). Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. Nat Biotechnol *33*, 175-178.

Ward, I.M., Minn, K., Jorda, K.G., and Chen, J. (2003a). Accumulation of checkpoint protein 53BP1 at DNA breaks involves its binding to phosphorylated histone H2AX. The Journal of biological chemistry *278*, 19579-19582.

Ward, I.M., Minn, K., van Deursen, J., and Chen, J. (2003b). p53 Binding protein 53BP1 is required for DNA damage responses and tumor suppression in mice. Molecular and cellular biology *23*, 2556-2563.

Ward, J.F. (1988). DNA damage produced by ionizing radiation in mammalian cells: identities, mechanisms of formation, and reparability. Prog Nucleic Acid Res Mol Biol *35*, 95-125.

Weinfeld, M., Chaudhry, M.A., D'Amours, D., Pelletier, J.D., Poirier, G.G., Povirk, L.F., and Lees-Miller, S.P. (1997). Interaction of DNA-dependent protein kinase and poly(ADP-ribose) polymerase with radiation-induced DNA strand breaks. Radiat Res *148*, 22-28.

Wellinger, R.E., and Thoma, F. (1997). Nucleosome structure and positioning modulate nucleotide excision repair in the non-transcribed strand of an active gene. EMBO J *16*, 5046-5056.

Wen, B., Wu, H., Shinkai, Y., Irizarry, R.A., and Feinberg, A.P. (2009). Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. Nature genetics *41*, 246-250.

Whalen, S., Truty, R.M., and Pollard, K.S. (2016). Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. Nature genetics *48*, 488-496.

White, J., Stead, E., Faast, R., Conn, S., Cartwright, P., and Dalton, S. (2005). Developmental activation of the Rb-E2F pathway and establishment of cell cycle-regulated cyclin-dependent kinase activity during embryonic stem cell differentiation. Molecular biology of the cell *16*, 2018-2027.

Wikonkal, N.M., and Brash, D.E. (1999). Ultraviolet radiation induced signature mutations in photocarcinogenesis. J Investig Dermatol Symp Proc *4*, 6-10.

Wold, M.S. (1997). Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. Annu Rev Biochem *66*, 61-92.

Wood, R.D. (2010). Mammalian nucleotide excision repair proteins and interstrand crosslink repair. Environ Mol Mutagen *51*, 520-526.

Wray, J., Kalkan, T., and Smith, A.G. (2010). The ground state of pluripotency. Biochem

Soc Trans *38*, 1027-1032.

Wright, J.A., Keegan, K.S., Herendeen, D.R., Bentley, N.J., Carr, A.M., Hoekstra, M.F., and Concannon, P. (1998). Protein kinase mutants of human ATR increase sensitivity to UV and ionizing radiation and abrogate cell cycle checkpoint control. Proceedings of the National Academy of Sciences of the United States of America *95*, 7445-7450.

Wu, J., Chen, Y., Lu, L.Y., Wu, Y., Paulsen, M.T., Ljungman, M., Ferguson, D.O., and Yu, X. (2011). Chfr and RNF8 synergistically regulate ATM activation. Nature structural & molecular biology *18*, 761-768.

Wyles, S.P., Brandt, E.B., and Nelson, T.J. (2014). Stem cells: the pursuit of genomic stability. Int J Mol Sci *15*, 20948-20967.

Xiao, A., Li, H., Shechter, D., Ahn, S.H., Fabrizio, L.A., Erdjument-Bromage, H., Ishibe-Murakami, S., Wang, B., Tempst, P., Hofmann, K.*, et al.* (2009). WSTF regulates the H2A.X DNA damage response via a novel tyrosine kinase activity. Nature *457*, 57-62.

Xie, A., Hartlerode, A., Stucki, M., Odate, S., Puget, N., Kwok, A., Nagaraju, G., Yan, C., Alt, F.W., Chen, J.*, et al.* (2007). Distinct roles of chromatin-associated proteins MDC1 and 53BP1 in mammalian double-strand break repair. Molecular cell *28*, 1045-1057.

Xu, D., Muniandy, P., Leo, E., Yin, J., Thangavel, S., Shen, X., Ii, M., Agama, K., Guo, R., Fox, D., 3rd*, et al.* (2010). Rif1 provides a new DNA-binding interface for the Bloom syndrome complex to maintain normal replication. EMBO J *29*, 3140-3155.

Xu, Y., Ayrapetov, M.K., Xu, C., Gursoy-Yuzugullu, O., Hu, Y., and Price, B.D. (2012). Histone H2A.Z controls a critical chromatin remodeling step required for DNA double-strand break repair. Molecular cell *48*, 723-733.

Yajima, H., Lee, K.J., and Chen, B.P. (2006). ATR-dependent phosphorylation of DNA-dependent protein kinase catalytic subunit in response to UV-induced replication stress. Molecular and cellular biology *26*, 7520-7528.

Yan, H., Toczylowski, T., McCane, J., Chen, C., and Liao, S. (2011). Replication protein A promotes 5'-->3' end processing during homology-dependent DNA double-strand break repair. The Journal of cell biology *192*, 251-261.

Yan, W.X., Mirzazadeh, R., Garnerone, S., Scott, D., Schneider, M.W., Kallas, T., Custodio, J., Wernersson, E., Li, Y., Gao, L.*, et al.* (2017). BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. Nature communications *8*, 15058.

Yang, H., Jeffrey, P.D., Miller, J., Kinnucan, E., Sun, Y., Thoma, N.H., Zheng, N., Chen, P.L., Lee, W.H., and Pavletich, N.P. (2002). BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. Science (New York, NY *297*, 1837-1848.

Yang, Y.G., Cortes, U., Patnaik, S., Jasin, M., and Wang, Z.Q. (2004). Ablation of PARP-1 does not interfere with the repair of DNA double-strand breaks, but

compromises the reactivation of stalled replication forks. Oncogene *23*, 3872-3882.

Ying, S., Minocherhomji, S., Chan, K.L., Palmai-Pallag, T., Chu, W.K., Wass, T., Mankouri, H.W., Liu, Y., and Hickson, I.D. (2013). MUS81 promotes common fragile site expression. Nature cell biology *15*, 1001-1007.

You, Y.H., Lee, D.H., Yoon, J.H., Nakajima, S., Yasui, A., and Pfeifer, G.P. (2001). Cyclobutane pyrimidine dimers are responsible for the vast majority of mutations induced by UVB irradiation in mammalian cells. The Journal of biological chemistry *276*, 44688-44694.

Yuan, J., Pu, M., Zhang, Z., and Lou, Z. (2009). Histone H3-K56 acetylation is important for genomic stability in mammals. Cell cycle (Georgetown, Tex *8*, 1747-1753.

Yuzhakov, A., Kelman, Z., Hurwitz, J., and O'Donnell, M. (1999). Multiple competition reactions for RPA order the assembly of the DNA polymerase delta holoenzyme. EMBO J *18*, 6189-6199.

Zeman, M.K., and Cimprich, K.A. (2014). Causes and consequences of replication stress. Nature cell biology *16*, 2-9.

Zgheib, O., Pataky, K., Brugger, J., and Halazonetis, T.D. (2009). An oligomerized 53BP1 tudor domain suffices for recognition of DNA double-strand breaks. Molecular and cellular biology *29*, 1050-1058.

Zhang, H., and Freudenreich, C.H. (2007). An AT-rich sequence in human common fragile site FRA16D causes fork stalling and chromosome breakage in S. cerevisiae. Molecular cell *27*, 367-379.

Zhang, J., Willers, H., Feng, Z., Ghosh, J.C., Kim, S., Weaver, D.T., Chung, J.H., Powell, S.N., and Xia, F. (2004). Chk2 phosphorylation of BRCA1 regulates DNA double-strand break repair. Molecular and cellular biology *24*, 708-718.

Zhang, Y., McCord, R.P., Ho, Y.J., Lajoie, B.R., Hildebrand, D.G., Simon, A.C., Becker, M.S., Alt, F.W., and Dekker, J. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. Cell *148*, 908-921.

Zhao, H., and Piwnica-Worms, H. (2001). ATR-mediated checkpoint pathways regulate phosphorylation and activation of human Chk1. Molecular and cellular biology *21*, 4129-4139.

Zhou, B.B., and Elledge, S.J. (2000). The DNA damage response: putting checkpoints in perspective. Nature *408*, 433-439.

Zhu, Q., and Wani, A.A. (2010). Histone modifications: crucial elements for damage response and chromatin restoration. J Cell Physiol *223*, 283-288.

Zickler, D., and Kleckner, N. (1999). Meiotic chromosomes: integrating structure and function. Annual review of genetics *33*, 603-754.

Ziegler-Birling, C., Helmrich, A., Tora, L., and Torres-Padilla, M.E. (2009). Distribution of p53 binding protein 1 (53BP1) and phosphorylated H2A.X during mouse preimplantation

development in the absence of DNA damage. Int J Dev Biol *53*, 1003-1011.

Zimmermann, M., Lottersberger, F., Buonomo, S.B., Sfeir, A., and de Lange, T. (2013). 53BP1 regulates DSB repair using Rif1 to control 5' end resection. Science (New York, NY *339*, 700-704.

Zink, D., Bornfleth, H., Visser, A., Cremer, C., and Cremer, T. (1999). Organization of early and late replicating DNA in human chromosome territories. Exp Cell Res *247*, 176-188.

Ziv, Y., Bielopolski, D., Galanty, Y., Lukas, C., Taya, Y., Schultz, D.C., Lukas, J., Bekker-Jensen, S., Bartek, J., and Shiloh, Y. (2006). Chromatin relaxation in response to DNA double-strand breaks is modulated by a novel ATM- and KAP-1 dependent pathway. Nature cell biology *8*, 870-876.

Zlotorynski, E., Rahat, A., Skaug, J., Ben-Porat, N., Ozeri, E., Hershberg, R., Levi, A., Scherer, S.W., Margalit, H., and Kerem, B. (2003). Molecular basis for expression of common and rare fragile sites. Molecular and cellular biology *23*, 7143-7151.

Zou, L., and Elledge, S.J. (2003). Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. Science (New York, NY *300*, 1542-1548.

Ujjwal Kumar BANERJEE

3-D Genome organization of DNA damage repair

Résumé

Notre génome est constamment attaqué par des facteurs endogènes et exogènes qui menacent son intégrité et conduisent à différents types de dommages. Les cassures double brins (CDBs) font partie des dommages les plus nuisibles car elles peuvent entraîner la perte d'information génétique, des translocations chromosomiques et la mort cellulaire. Tous les processus de réparation se déroulent dans le cadre d'une chromatine hautement organisée et compartimentée. Cette chromatine peut être divisée en un compartiment ouvert transcriptionnellement actif (euchromatine) et un compartiment compacté transcriptionnellement inactif (hétérochromatine). Ces différents degrés de compaction jouent un rôle dans la régulation de la réponse aux dommages à l'ADN.

L'objectif de mon premier projet était de comprendre l'influence de l'organisation 3D du génome sur la réparation de l'ADN. Pour cela, j'ai utilisé deux approches complémentaires dans le but d'induire et de cartographier les CDBs dans le génome de souris. Mes résultats ont mis en évidence un enrichissement de γH2AX, facteur de réparation des dommages à l'ADN, sur différentes régions du génome de cellules souches embryonnaires de souris, et ont également montré que les dommages persistent dans l'hétérochromatine, contrairement à l'euchromatine qui est protégée des dommages. Pour mon deuxième projet, j'ai cartographié l'empreinte génomique de 53BP1, facteur impliqué dans la réparation des CDBs, dans des cellules U2OS asynchrones et des cellules bloquées en G1 afin d'identifier de nouveaux sites de liaison de 53BP1. Mes résultats ont permis d'identifier de nouveaux domaines de liaison de 53BP1 couvrant de larges régions du génome, et ont montré que ces domaines de liaison apparaissent dans des régions de réplication moyenne et tardive.

Mots clés : réponse aux dommages à l'ADN, Cassures Double Brin, ChIP-Seq, cellules souches embryonnaire, organisation de la chromatine, γH2AX, 53BP1

Abstract

Our genome is constantly under attack by endogenous and exogenous factors which challenge its integrity and lead to different types of damages. Double strand breaks (DSBs) constitute the most deleterious type of damage since they may lead to loss of genetic information, translocations and cell death. All the repair processes happen in the context of a highly organized and compartmentalized chromatin. Chromatin can be divided into an open transcriptionally active compartment (euchromatin) and a compacted transcriptionally inactive compartment (heterochromatin). These different degrees of compaction play important roles in regulating the DNA damage response.

The goal of my first project was to understand the influence of 3D genome organization on DNA repair. I used two complementary approaches to induce and map DSBs in the mouse genome. My results have shown that enrichment of the DNA damage repair factor γH2AX occurs at distinct loci in the mouse embryonic stem cell genome and that the damage persists in the heterochromatin compartment while the euchromatin compartment is protected from DNA damage. For my second project, I mapped the genomic footprint of 53BP1, a factor involved in DSBs repair, in asynchronous and G1 arrested U2OS cells to identify novel 53BP1 binding sites. My results have identified novel 53BP1 binding domains which cover broad regions of the genome and occur in mid to late replicating regions of the genome.

Keywords: DNA damage response, Double strand breaks, ChIP-Seq, embryonic stem cells, chromatin organization, γH2AX, 53BP1.