UNIVERSITÉ DE STRASBOURG

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

IGBMC, UMR 7104, U 964

THÈSE

présentée par :

Anne-Sophie HOFFBECK

soutenue le : **25 Octobre 2013**

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

Discipline/ Spécialité : Aspect moléculaire et cellulaire de la biologie

Chromatin structure and DNA repair

THÈSE dirigée par :

Dr SOUTOGLOU Evi	CR, IGBMC, Strasbourg

RAPPORTEURS :

Dr REINA SAN MARTIN Bernardo	DR, IGBMC, Strasbourg
Dr PAPAMARCAKI Thomais	Assoc. Professor, University of Ioannina, Ioannina
Dr CAMMAS Florence	CR, Institut de recherche en cancérologie, Montpellier

AUTRES MEMBRES DU JURY :

Dr SARDET Claude	DR, Institut de génétique moléculaire, Montpellier
Dr SCHREIBER Valérie	DR, IREBS, Strasbourg

Acknowledgments

First of all I would like to thank a lot my thesis director Evi for giving me the opportunity of doing my PhD in her lab as well as for the patience, enthusiasm and motivation she put in supervising me these last years.

I would also like to express my gratitude to the member of my committee for taking the time to evaluate my work:

Dr Bernardo Reina, Dr Thomais Papamarkaki and Dr Florence Cammas as reviewers

Dr Claude Sardet, Dr Valérie Schreiber and Dr Jérôme Déjardin as examiners.

I thank more particularly Jérôme that is also a collaborator for the PICh project, for welcoming me for a week in his lab in order to learn about the technical tricks of this technique.

I would like to thank all the lab members past and present from the Soutoglou team, Alkmini, Anastazja, Audrey, Céline, Charlène, Katerina, Tibor and Zita for their support, discussions, advices, friendship and the very good atmosphere they put in the lab. I would like to add a special thanks to Alkmini, who I share the SET project with.

Thank you also to the members of the Reina lab: Anne-Sophie, Ebe, Isabelle, Léa, Sara and Vincent for discussions, FACS and cloning advices, PICh/BirA* collaborations (Sara and Léa) and sweet edibles I could find in their lab.

Thank you to the people of the common facilities of IGBMC that helped me for my different projects, Marc and Pascal from the imaging facility, Claudine from the flow cytometry facility and all the members from the cell culture and screening platforms.

Finally I would like to thank the "ministère de la recherche et de l'enseignement supérieur" as well as the "fondation pour la recherche médicale" as my two principal financial supports during my thesis.

Je souhaite dire un grand merci à mes amis qui m'ont soutenu et surtout supporté pendant ces années de thèse. Merci d'une part à mes amis « scientifiques » qui sont passé par là ou pas loin et qui connaissent bien le principe : Xavier (et Anne), Guillaume, Kasia, Sophie, Sercan, Thibaut, Barbara (et Laurent) et Maeva (et Rémi).

Toutefois je tiens également à remercier les autres qui n'ont pas un pied dans le milieu mais qui ont fait tout leur possible pour essayer de comprendre, Martin et Amélie, Thomas et Pauline, Clément et Justyna, Laetictia, Amandine et Laetitia. Enfin je tiens à remercier également mes amis de la danse et du cheval qui m'ont souvent posé la question : « C'est quand que tu la finis déjà ta thèse ? ».

Merci également à mes collègues d'OpenLAB : Antoine, Aurélie, Denis, Dimitri, François, Isabelle, Laura, Mélanie et Ricardos et nos superviseurs Laurence et Michel. OpenLAB a été une super expérience dont je garde un très bon souvenir. Je remercie aussi Gilliane, Pauline et Philippe mes tuteurs lors de mes deux années de monitorat en fac de pharmacie.

Un énorme merci à mon chéri Guillaume. Ces dernières années et particulièrement la période de rédaction n'ont pas été des plus simple. Je le remercie pour sa patience sans faille, son soutien, sa motivation dans mes moments de doute et son amour. Je tiens également à remercier sa famille Dany et Claude, pour le « support technique » et Lionel et Nathalie, pour les discussions et conseils.

Enfin je ne remercierais jamais assez mes parents Marie-Anne et Edy pour toute leur aide, autant personnelle que financière durant mes longues études. Ils m'ont soutenu, supporté, encouragé, aimé chaque jour et ont toujours cru en moi.

Table of content

Abbreviations

53BP1	p53 binding protein 1
9-1-1	Rad9-Rad1-Hus1
A	adenine
AD	activation domain
ADP	adenosine diphosphate
AID	activation induced deaminase
ALC1	amplified in liver cancer 1
ALDH2	aldehyde dehydrogenase 2
Alt-NHEJ	alternative NHEJ
APLF	aprataxin-PNK-like factor
APTX	aprataxin
ATM	ataxia telangiectasia mutated
ATR	ATM and Rad3 related
ATRIP	ATR interacting protein
BER	base excision repair
BLM	Bloom syndrome protein
bp	base pair
BRCA1	breast cancer 1
BRCA2	breast cancer 2
BRCT	BRCA1 C-terminal domain
BrdU	5-bromo-2'-deoxyuridine
C	cytosine
C-NHEJ	classical NHEJ
CAF1	chromatin assembly factor 1
CBP	CREB binding protein
CDK	cyclin dependent kinase
Chk1	checkpoint kinase 1
Chk2	checkpoint kinase 2
CK2	casein kinase 2
CtIP	CtBP-interacting protein
D-loop	Displacement loop
DAPI	4',6-diamidino-2-phenylindole
DDR	DNA damage response
DNA	deoxyribonucleic acid
DNA-PK	DNA-dependent protein kinase
DNA-PKcs	DNA-dependent protein kinase catalytic subunit
DNAse	deoxyribonuclease
DNMT1	DNA (cytosine-5)-methyltransferase 1
DNMT3a	DNA (cytosine-5)-methyltransferase 3a
dNTP	deoxynucleotide
Dox	doxycycline

DSB	double strand break
DSBR	double strand break repair
DSBR	DSB repair
dsDNA	double stranded DNA
ERα	estrogen receptor α
Exo1	exonuclease 1
FAT	FRAP ATM TTRAP
FATC	FAT C-terminal
FHA	fork head associated
G	guanine
G0 phase	gap 0 phase
G1 phase	gap 1 phase
G2 phase	gap 2 phase
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GFP	green fluorescent protein
GFP-SET	GFP fused to SET
GFP-SETΔNAP	GFP fused to SET mutant lacking the NAP domain
GR	glucocorticoid receptor
GSTP1	glutathione S-transferase pi 1
GzmA	granzyme A
GzmK	granzyme K
H1	histone 1
H2A	histone 2A
H2AX	histone 2AX
H2B	histone 2B
H3	histone 3
H4	histone 4
HDAC1	histone deacetylase 1
HDAC2	histone deacetylase 2
HEAT	huntingtin, elongation factor 3, PP2A and TOR1
HP1α	heterochromatin protein 1 α
HP1β	heterochromatin protein 1 β
HP1γ	heterochromatin protein 1 γ
HR	homologous recombination
HU	hydroxyurea
IGAAD	inhibitor of Granzyme A activated DNase
INHAT	inhibitor of histone acetyl transferase
K	lysine
KAP1	KRAB-associated protein 1
kb	kilobase
KRAB	Krueppel-associated box
lacI	lac repressor
LacO	lac operator
LIG3	ligase III

LIG4	ligase IV
LNA	locked nucleic acid
MCPH1	microcephalin
MDC1	mediator of DNA damage checkpoint 1
me	methylation
me2	dimethylation
me3	trimethylation
mH2A	macro H2A
MMR	mismatch repair
MNase	micrococcal nuclease
MRN	Mre11-Rad50-Nbs1
MS	Mass spectrometry
NAP1	nucleosome assembly protein 1
NCS	neocarzinostatin
NER	nucleotide excision repair
NHEJ	non homologous end joining
NLS	nulear localization sequence
NuRD	nucleosome remodeling deacetylase
PALB2	partner and localizer of BRCA2
PAR	poly(ADP)ribosylation
PARG	poly(ADP-ribose) glycohydrolase
PARP	poly(ADP-ribose) polymerase
PBS	phosphate buffered saline
PCAF	P300/CBP-associated factor
PCNA	proliferating cell nuclear antigen
PFA	paraformaldehyde
PHAPII	putative class II human histocompatibility leukocyte-associated protein II
PI3K-like	phosphatidylinositol 3-kinase like protein kinases
PIAS1	protein inhibitor of activated STAT-1
PIAS4	protein inhibitor of activated STAT-4
PICh	proteomics of isolated chromatin segments
PIKKs	phosphatidylinositol 3-kinase like protein kinases
PMSF	phenylmethylsulfonyl fluoride
PNKP	polynucleotide kinase/phosphatase
PP2A	protein phosphatase 2A
PP5	protein phosphatase 5
PRD	PIK-kinase regulatory domain
ProTα	prothymosin α
RAG1	recombination activating gene 1
RAG2	recombination activating gene 2
RBCC	RING finger B-box coiled-coil
RNAse	ribonuclease A
RNF168	RING finger protein 168
RNF8	RING finger protein 8

RPA	replication protein A
RT-qPCR	reverse transcription quantitative polymerase chain reaction
S	serine
S phase	synthesis phase
siRNA	small interfering RNA
SMC	structural maintenance of chromosome
Spo11	Sporulation protein 11
SSB	single strand break
SSBR	single strand break repair
SSC	saline-sodium citrate
ssDNA	single stranded DNA
SUMO	small ubiquitin-related modifier
Suv39H1	suppressor of variegation 3-9 homolog 1
SWI-SNF	SWItch/Sucrose NonFermentable
T	thymine
T	threonine
TAF-Iα	template activating factor 1 α
TAF-Iβ	template activating factor 1 β
TetO	tet operator
TIF1β	Transcription intermediary factor 1 β
Tip60	Tat interactive protein 60 kDa
TOPBP1	topoisomerase II Binding Protein 1
TREX1	three prime repair exonuclease 1
TRIM28	tripartite motif containing 28
TTRAP	TRAF and TNF receptor associated protein
UBC9	ubiquitin-conjugating enzyme 9
V(D)J	variable diverse joining
WIP1	wild-type p53-induced protein 1
WRN	Werner syndrome protein
XLF	XRCC4-like factor
XRCC1	x-ray cross-complementing gene 1
XRCC4	x-ray cross-complementing gene 4
Y	tyrosine
γH2AX	phosphorylated H2AX

A. INTRODUCTION

I. DNA and chromatin

Deoxyribonucleic acid (DNA), the macromolecule essential for life, is present in every cell of living organisms and encodes the genetic information. It is a polymer structured as a double helix and composed of a succession of nucleotides. Each nucleotide is formed by a nucleobase attached to a backbone made of a sugar (ribose) and a phosphate group. They are four types of nucleobases: guanine (G), adenine (A), thymine (T) and cytosine (C). The succession of these different bases encodes the genetic information. As a simple uncompacted double helix the DNA molecule present in a human cell is more than two meters long. That is why DNA is organized and compacted as chromatin, a DNA-protein association, in the nucleus. However, compaction is not the unique role of chromatin. It is also involved in strengthening the DNA for mitosis, preventing and repairing DNA damage and finally controlling gene expression and replication.

1. Chromatin structure

Chromatin is highly structured. They are three levels of chromatin organization: the nucleosome, the 30 nm fiber and higher-order folding levels corresponding to chromosome territories or metaphase chromosomes[1] (fig. 1A). In the nucleosome, DNA is wrapped around proteins named histones. A nucleosome is composed of an octamer of four canonical histones: H2A, H2B, H3 and H4, where each of them is present twice. DNA is wrapped around the core histones leading to an approximate six-fold length compaction[1,2] (fig. 1B). The DNA packaged around the octamer is 147 base pair (bp) long and is followed by a linker part of variable length (between 20 bp

and 90 bp in steps of 10 bp). The linker region is associated with H1, a histone not present in the core particle.[1] This structure is called "beads-on-the-string" can be only observed under non-physiological conditions. Under physiological conditions the string of nucleosome forms a 30 nm structure (fig. 1). The 30 nm fiber undergoes complex folding steps that are still under debate. The current view is that this fiber is arranged as a helix with a compaction of 70 to 160 kilobase per micrometer (kb/µm) and several structural models exist to explain this helical arrangement.[1] Finally the 30 nm fiber is organized as chromosome territories in interphase and compacted as extremely condensed chromosome during metaphase. A chromosome territory corresponds to the space that each chromosome occupies in the interphase nucleus under its decondensed form.[1]

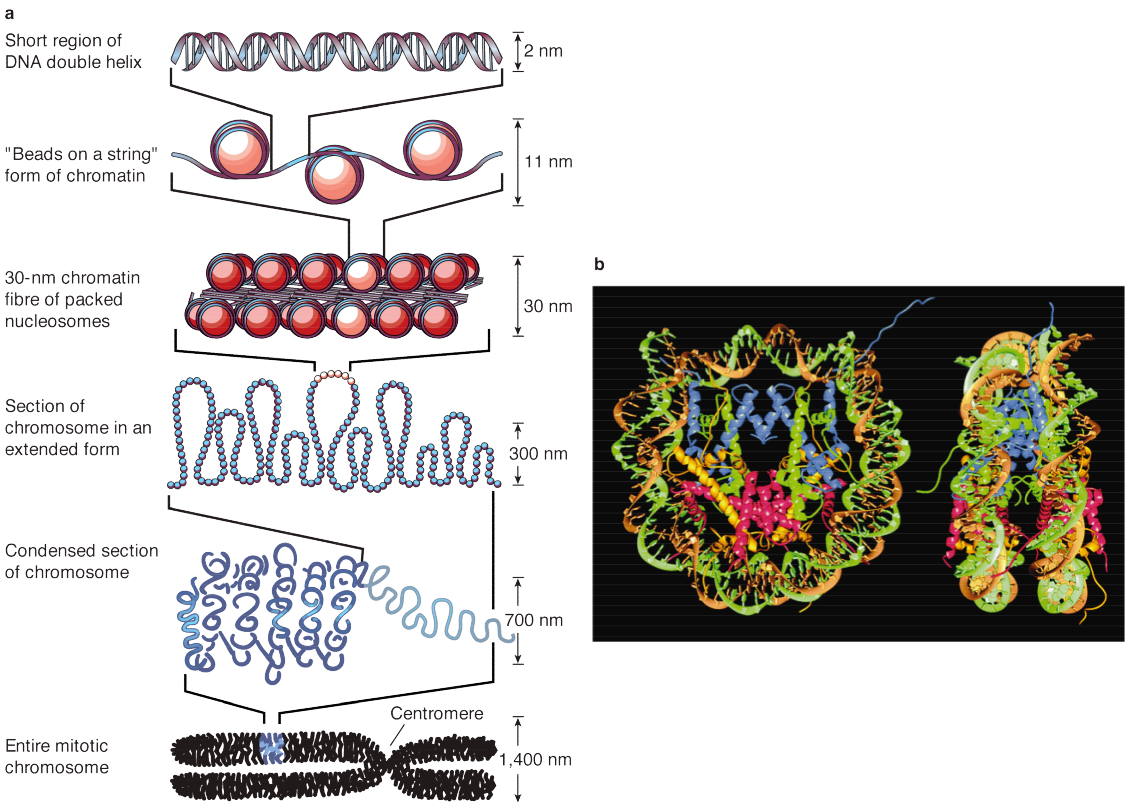

Figure 1: Packaging of DNA. [3]

(A) The organization of DNA within the chromatin structure. The lowest level of organization is the nucleosome, in which two superhelical turns of DNA are wound around the outside of a histone octamer. Nucleosomes are connected to one another by short stretches of linker DNA. At the next level of organization the string of nucleosomes is folded into a fibre about 30 nm in diameter, and these fibres are then further folded into higher-order structures. At levels of structure beyond the nucleosome the details of folding are still uncertain. (B) The structure of the nucleosome core particle was uncovered by X-ray diffraction, to a resolution of 2.8Å. It shows the DNA double helix wound around the central histone octamer. Hydrogen bonds and electrostatic interactions with the histones hold the DNA in place.

2. Heterochromatin and euchromatin

Heterochromatin and euchromatin are terms that refer more to states of chromatin compaction and transcriptional potential rather than categories of chromatin higher-order structure.[4] Heterochromatin was historically defined as region of nuclei that are stained strongly with basic dyes way before DNA discovery. On the contrary, euchromatin is defined as heterochromatin counterpart. Heterochromatin is visible in light and electron microscopy and is usually located at the nuclear periphery and as blocks surrounding the nucleolus[4] (fig. 3). Transcription mostly happens in euchromatin, which is more gene-rich than heterochromatin and the term heterochromatin is commonly applied to transcriptionally silent region of chromatin.[4] A distinction can be made between constitutive heterochromatin and facultative heterochromatin. Constitutive heterochromatin is always found condensed and is enriched in repetitive, gene-poor and late replicating DNA such as telomeres or centromeres. On the other hand, facultative heterochromatin can undergo reversible decompaction to allow gene transcription.[4]

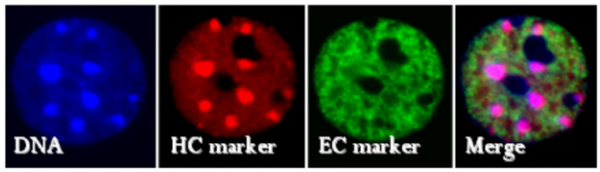

Figure 3: Chromocenters.[5]

NIH3T3 cells stained with DAPI (DNA, blue), KAP1 (HC marker, red) and Acetyl-K9 Histone H3 (EC marker, green). The densely compacted DNA regions overlapping with areas of strong KAP1 staining and weak Acetyl-K9 Histone H3 represent heterochromatic chromocentres.

Althought different types of heterochromatin exist, most share common molecular components. Heterochromatin is built thanks to five different elements: DNA sequence specific transcriptional repressors and associated co-repressors, regulators of histone acetylation and methylation, chromodomain adaptor proteins that frequently bind to methylated histones, proteins involved in DNA methylation and nucleosome remodeling enzymes which may be includes in complexes with the above components.[5]

To regulate which genomic regions are intended for silencing, the transcription repressors act at specific to DNA sequences. Usually they have little silencing activity and trigger the recruitment of co-repressors such as KRAB-associated protein 1 (KAP1)[5] (fig. 4A). As histone acetylation is a marker of transcriptionally active regions, histone deacetylases such as HDAC1 and HDAC2 are critical for heterochromatin formation.[5] Histone methylation is an essential mark for proper heterochromatinization. Specific histone methyltransferases involved in chromatin compaction are Suv39H1, Suv39H2 and SETDB1. Their role is increase mono-, di-

and tri-methylation of histones, as for instance on lysines 9 and 27 of histone 3 (H3K9me and H3K27me) and on lysine 20 of histone 4 (H4K20me)[5] (fig. 4B). These marks attract the chromodomain proteins HP1 (heterochromatin protein 1). HP1 proteins bind to the methylated tails of histones and form nucleosomal interactions compressing and stabilizing the silenced chromatin[5] (fig. 4C). DNA has been also showed to be "painted" with methyl groups on CpG islands in heterochromatic regions.[5] CpG islands are often positioned in the promoter region of genes and their methylation has been shown to cause gene silencing. Finally, ATP-dependent nucleosome remodeling enzymes, like CHD3/Mi-2α and CHD4/Mi-2β, provide the necessary energy for histone spacing rearrangements (fig. 4B). CHD3/Mi-2α or CHD4/Mi-2β together with HDAC1/2, MBD2/3, RbAp46/48, MTA1/2/3 and p66 corm the NuRD (nucleosome remodeling and deacetylase) complex. The NuRD complex has an helicase activity coupled with HDAC functions making it an essential component for generation of densely packed, hypoacetylated nucleosomes.[5]

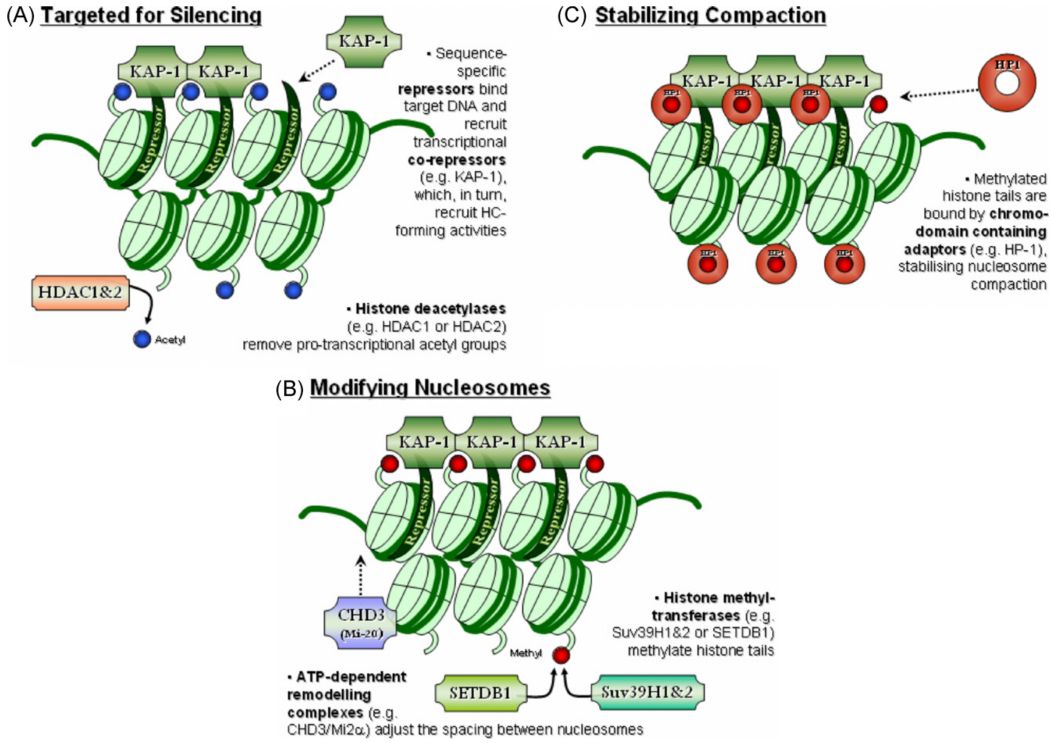

Figure 4: The foundations of heterochromatin.[5]

(A) Genomic regions targeted for heterochromatinization are first bound by sequence-specific repressors. Repressors initiate transcriptional silencing by recruiting co-repressors such as KAP1. Co-repressors promote the loss of pro-transcriptional histone modifications such as the acetylation of K9 of histone H3 by recruiting histone deacetylases (such as HDAC1 and HDAC2). (B) Deacetylated histones within nascent heterochromatin are methylated by histone methyltransferases such as SETDB1 or Suv39H1 and Suv39H2. Whilst this occurs, ATP-dependent remodeling enzymes increase the compaction of nucleosomes. (C) The methylated histone tails of compacted nucleosomes provide a platform for chromodomain protein (such as HP1) binding, which interact with one another across nucleosomes to stabilize the overall structure of the newly heterochromatinized DNA.

II. DNA damage

1. Types of damage

DNA is constantly assaulted by several types of agents causing DNA damage. They are various types of DNA damage. Besides single strand break (SSB) and double strand break (DSB) of the DNA filament, DNA can encounter a variety of base modifications such as, mismatch, intrastrand crosslink, pyrimidine dimers, bulky adducts, oxidation, alkylation, deamination, depurination or depyrimidation of bases.[6,7] Specific repair mechanisms for theses types of lesions evolved to counteract DNA damage. SSBs are repaired through single strand break repair (SSBR) and DSBs through either non-homologous end joining (NHEJ) or homologous recombination (HR). Mispaired bases are corrected by mismatch repair (MMR), small base modifications by base excision repair (BER) and more complex base alterations by nucleotide excision repair (NER).[6,7,8]

Among all these DNA damages, DSBs are the most hazardous type of damage as broken ends are able to dissociate and cause chromosomal translocation, leading to genomic instability and thus cancer.[9]

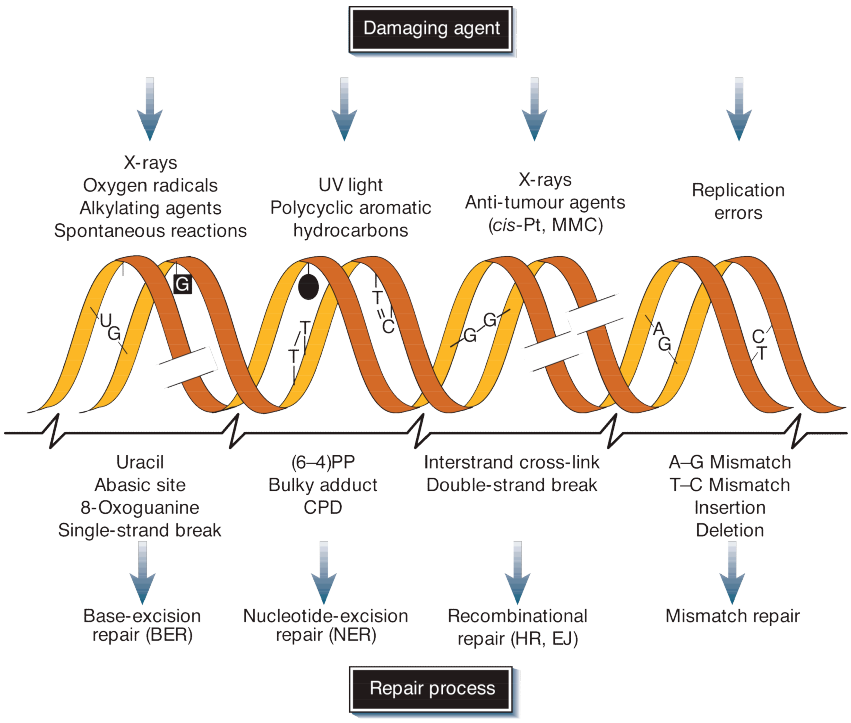

Figure 5: DNA damage and repair mechanisms.[6]

Common DNA damaging agents (top); examples of DNA lesions induced by these agents (middle); and most relevant DNA repair mechanism responsible for the removal of the lesions (bottom). Abbreviations: cis-Pt and MMC, cisplatin and mitomycin C, respectively (both DNA-crosslinking agents); (6–4)PP and CPD, 6–4 photoproduct and cyclobutane pyrimidine dimer, respectively (both induced by UV light); HR, homologous recombination; EJ, end joining.

2. DSBs

DSBs can be originated from exogenous sources such as ionizing radiation or radiomimetic drugs[8–10] (fig. 6A). Endogenous sources like reactive oxygen species from our own metabolism or replication stress can also produce DSBs.[8–10]

Replication stress is the most common endogenous damage as it results in the blockage of a replication fork. When a replication fork meets an obstacle it gets stalled. This results in a long stretch of single stranded DNA that is coated by RPA.

This accumulation of RPA triggers the recruitment of different proteins one of them being ATR that in turn phosphorylates Chk1 to block the cell cycle. If the obstacle cannot be bypassed the fork collapses and this results in a one end DSB that is repaired by homologous recombination[11] (fig. 6B).

Finally, DSBs are created voluntarily in several physiological processes such as meiotic recombination (fig. 6D), V(D)J recombination (fig. 6C) or class switch recombination (fig. 6E). In this case DSBs are generated through the action of endonucleases, Spo11, RAG1/2 and AID respectively.[8,10] In spite of the fact that these breaks are induced, the repair mechanisms used here are the same as for unpredicted DSBs. For V(D)J recombination and class switch recombination non-homologous end joining (NHEJ) is the mechanism repair the breaks. During meiosis, the DSBs are repaired through homologous recombination (HR) in order to allow crossing over thus promoting genomic diversity.[8,10]

As a matter of fact, DSBs have a key role in the balance between genome instability and genome diversity.

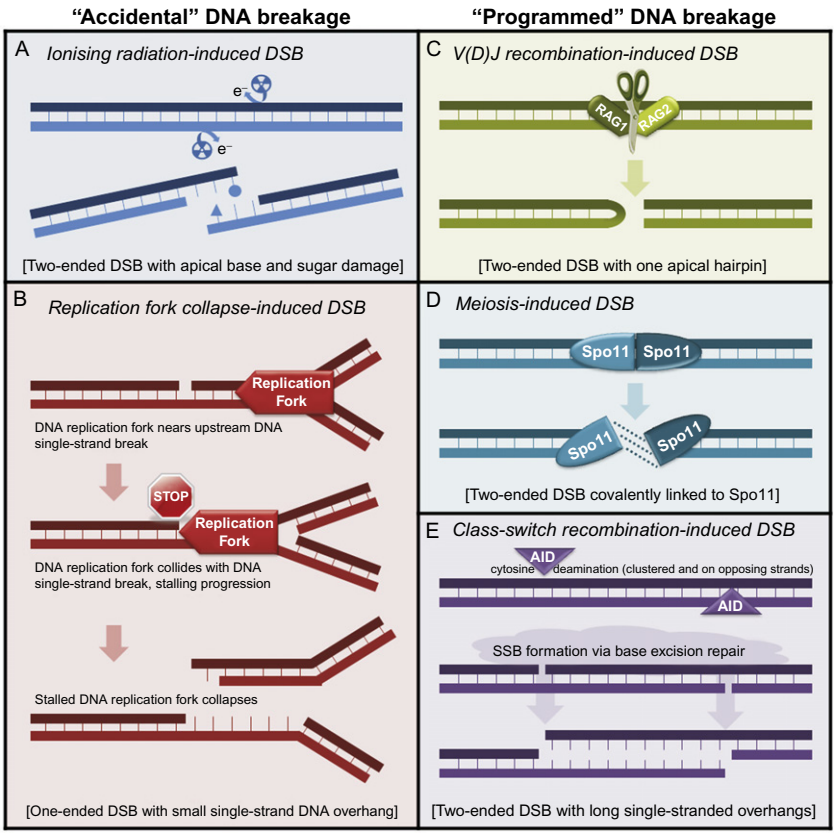

Figure 6: Variable routes of DSB formation. [10]

(A) Ionizing radiation can directly induce DSBs without replication or cellular processing. IR-induced DSBs have two termini (hence referred to as two-ended) and can have damaged bases or sugars at their termini and single-strand breaks in close proximity. (B) Replication forks can collapse when they encounter base damage or a single-strand break. Depending on whether the lesion is on the 50 or 30 strand, this can lead to single termini DSBs, that is, one ended. The ends can be a blunt end or, following resection, can have a 30 overhang; the lesion on the other strand is a single-strand gap or region, depending on polarity. (C) During V(D)J recombination, the RAG1/RAG2 endonucleases cleave DNA at recombination signal sequences. Breakage occurs via a transesterification reaction generating one blunt and one hairpin-ended DSB. (D) During meiosis, the topoisomerase II-like protein, Spo11, and the MRN complex generate protein-bridged DSBs. (E) During class-switch recombination (CSR), AID (activation-induced cytosine deaminase) generates uracil residues by deamination of cytosines at CSR sites at which SSBs can arise following their processing by uracil glycosylase. Two closely localized SSBs degenerate into DSBs with long single-strand overhangs.

III. DNA Damage Response, DDR

DNA damage response (DDR) is a signal transduction pathway triggered by DNA damage and replication stress.[7] It activates a signaling cascade to protect the cell and defend the organism against the threat of DNA damage. Proteins of the phosphatidylinositol 3-kinase like protein kinases (PI3K-like or PIKKs) family (ATM, ATR and DNA-PKcs) and of the poly(ADP)ribose polymerase (PARP) family are key players in this response.[7] ATM and DNA-PK are activated by the presence of double strand breaks. ATM has a large amount of substrates compared to DNA-PK that is involved in DSB end joining. Activation of ATR, in complex with its interacting partner ATRIP, is dependent on the presence of long single stranded DNA stretch coated with replication protein A (RPA). These ssDNA regions are often generated by replication fork staling or DSBs.[7] PARP proteins (especially PARP1 and PARP2) catalyze the addition of poly(ADP)ribose chains on proteins in the presence of single strand breaks or double strand breaks, in order to recruit DDR factors to the break.[7]

DDR factors are localized to the site of damage thanks to sensor proteins that directly recognize specific DNA lesions and activate DDR.[7] The DDR factors form discrete nuclear foci, which can be visualized by microscopy.[7,8] The factors recruitment of the DDR cascade is precisely spatiotemporally regulated and dependent on several types of post-translational modifications such as phosphorylation, ubiquitination, sumoylation, poly(ADP)ribosylation, acetylation and methylation.[7]

1. Sensing the break: the MRN complex

The Mre11-Rad50-Nbs1 (MRN) complex is the first player in the recognition of DSBs.[12] It is a key participant in the DDR as it is very conserved throughout evolution and lack of any of its component leads to embryonic lethality in mammals.[12] The MRN complex has three critical roles: DNA binding and processing, DNA tethering to bridge DNA over short and long distances, and finally activation of DDR and checkpoint signaling pathways.[13] MRN structure is composed of a heterohexamer: a Mre11 dimer, a Rad50 dimer and two Nbs1 proteins. These proteins form four distinct regions: the "head" built by the Mre11 dimer and two Rad50 N-terminal and C-terminal domains, the "coils" and "hook" made the region of Rad50 separating the N and C-terminal domains, and the Nbs1 "flexible adapter"[13] (fig. 7).

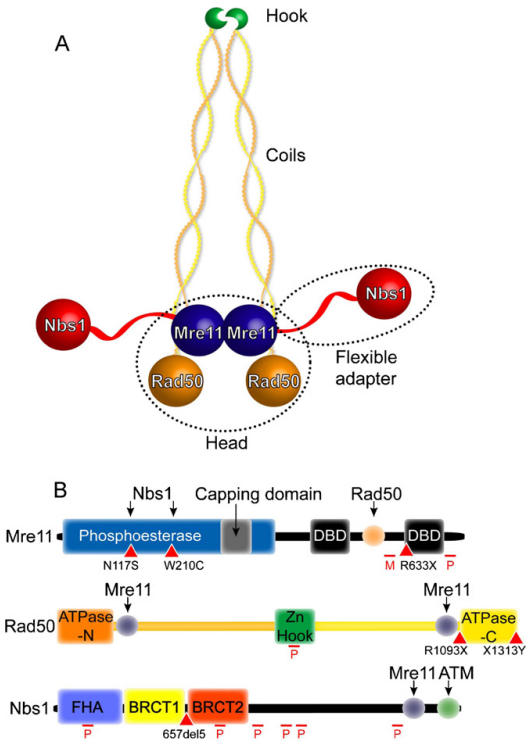

Figure 7: Overall MRN assembly and key domains.[13]

(A) MRN can assemble as a heterohexamer and consists of 4 key regions: the processing "head", formed by the Mre11 dimer and two Rad50 ABC ATPase domains (indicated by dotted line), the "coils" and "hook" encoded by the region of Rad50 separating the N- and C-terminal ABC ATPase halves, and the Nbs1 "flexible adapter" (indicated by dotted line) that provides the key link to signaling functions. (B) Schematic representations of the MRN subunits Mre11, Rad50 and Nbs1 showing key domains. The approximate location of reported methylation sites are indicated by M and DNA damage inducible phosphorylation sites by P. The major sites corresponding to inherited human disorders associated with each gene are indicated by a red triangle, with amino acid substitutions labeled for Mre11 and Rad50 (X is a stop codon) and 657del5 representing the major Nbs1 mutation responsible for >90% of Nijmegan breakage syndrome cases.

a. Mre11

The Mre11 protein exists as a U-shaped dimer. The homodimerization of Mre11 is necessary for MRN function and Mre11 dimerization domain is needed for its binding to DNA and repair activity. Mre11 interacts with Rad50 through a domain in the C-terminal part of the protein and with Nbs1 via a N-terminal located domain.[12,13] Mre11 has an endonuclease and an exonuclease activity. The nuclease activities of Mre11 are dependent of ATP, manganese and Nbs1 and stimulated by Rad50.[12] Mre11 nuclease activity is important for DSB repair, as it is involved in the early steps of resection.[14,15] Nevertheless, MRN role in signaling is independent of Mre11 nuclease activity.[12,14,15]

b. Rad50

Rad50 is part of the SMC (structural maintenance of chromosome) family of proteins, involved in chromosome repair and sister-chromatid cohesion. It has Walker A and B motifs situated in the two ends of the protein. These motifs are required for ATPase activity, DNA binding and partial unwinding of DNA. The

Mre11 protein interacts with Rad50 through a region located near the Walker motifs. The center of the protein is made of two coiled-coil domains separated by a hinge containing a zinc hook that allows dimerization of the protein.[12,13]

c. Nbs1

The Nbs1 protein does not have any known enzymatic activity but it regulates MRN functions. It is required for Mre11 and Rad50 nuclear localization, for rapid focal assembly of MRN complexes at sites of DNA damage and also stimulates the activity of Rad50 and Mre11. It contains a FHA (fork head-associated) domain flanked by two tandemly repeated BRCT (BRCA1 carboxyl-terminus) motifs. These domains are necessary for its interaction with γH2AX and phosphorylated MDC1. The FHA domain is also known to be required for interaction with ATR. Finally, the C-terminal part of Nbs1 contains short motifs responsible for its interaction with Mre11 and ATM.[12,13]

Lastly, it is important to note that as many other proteins, components of the MRN complex are modified by post-translational modifications. Several phosphorylation sites have been reported for Mre11, Rad50 and Nbs1. To date, the effects of post-translational modifications on MRN are poorly understood but are clearly involved in regulating the complex activities. Thus, these modifications aid MRN function in helping coordinate sensing, signaling, and effector functions in response to DNA damage.[12,13]

2. PIKKs kinases activation

As previously described, DDR triggers the activation of three kinases from the phosphatidylinositol 3-kinase like protein kinases family, ATM, ATR and DNA-PKcs. The DDR signalization cascade finally aims to phosphorylate two other kinases, Chk1 and Chk2. These activated checkpoint kinases lead to the arrest of the cell cycle until the lesion is repaired.[7]

a. Ataxia-Telangiectasia mutated, ATM

ATM is a kinase from the PI3K-like protein family. Mutation in the ATM gene provokes a human genomic instability disorder called Ataxia-telangiectasia. This disease is characterized by neurodegeneration that affects the cerebellum, dilatation of blood vessels, immunodeficiency, predisposition to malignancy and hypersensitivity to ionizing radiation.[16] ATM is a large protein of 3056 residues, and because of its size its structure has not been solved yet. ATM has an active domain containing a PI3K signature in its C-terminal part and numerous target sites for post-translational modifications[16,17] (fig. 8). ATM is known to be activated after DNA DSB induction, to in turn phosphorylate the histone variant H2AX on S139, called then γH2AX.[16] After DNA damage, ATM relocalizes to the break and gets catalytically active. Only a fraction of proteins concentrate to the lesion and stays there for several hours whereas the rest of the pool stays nucleoplasmic.[16] ATM spatial relocalization is responsible for many phosphorylation events happening in the repair focus. In undamaged cells, ATM exists as a quiescent homodimer. These dimers dissociate in active monomers in presence of DSBs.[16,17]

Several post-translational modifications, especially phosphorylations, are associated with ATM activation. The first and major one is ATM autophosphorylation on S1981, being a hallmark of human activated ATM. Three other autophosphorylation sites have been identified (S367, S1893 and S2996) (fig. 8) as well as an acetylation site by Tip60 (K3016). These modifications are crucial as their abolition impedes ATM function in DDR. The autophosphorylation sites are especially required for ATM retention at the DSB; however, they are not necessary for the initial ATM recruitment. Dephosphorylation events might also be important in ATM activation; indeed, PP2A and PP5 have been reported to be involved in the process.[16,17]

The exact stimulus that triggers ATM activation is not clearly established yet. The debate favors a conformational change that follows DSB induction rather than a direct contact of ATM with the break.[16] Nevertheless, it has been shown that MRN complex is essential for proper ATM activation. In fact, as discussed previously ATM is interacting with Nbs1 and this interaction is decisive to ATM recruitment and retention at damaged sites. 53BP1 and BRCA1, two mediator proteins, are also involved in optimizing this interaction.[16] ATM also interacts with another crucial mediator protein MDC1 (mediator of DNA damage checkpoint protein 1). MDC1 interacts with γH2AX in parallel of ATM; this double interaction permits the spreading of γH2AX as it allows ATM to phosphorylate additional H2AX histones, which in turn allow new MDC1 binding.[16]

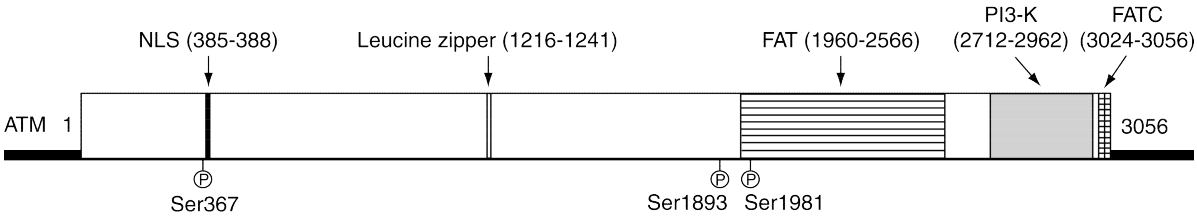

Figure 8: A schematic diagram of the ATM protein.[17]

ATM is a 350 kDa protein, consisting of 3056 amino acids. The phosphorylation sites (P) indicate the positions of serine residues that are autophosphorylated. The identified domains within ATM consist of the nuclear localization signal (NLS), the leucine zipper (LZ), the FRAP/ATM/TRRAP (FAT) domain, the Kinase domain (PI3-K), and the FAT c-terminal (FATC) domain.

b. ATM and Rad3 related, ATR

ATR shares many biochemical and functional similarities with ATM. As ATM, ATR is also a large serine/threonine kinase of 2644 residues. ATR is essential for human cell survival and mutations in ATR gene is compatible with viability only when heterozygous or hypomorphic. The only disease related with ATR mutation is the Seckel syndrome, a rare disorder characterized by growth retardation and microcephaly.[18]

ATR is activated in response to several types of DNA damage, as for example DSBs, bulky adducts, crosslinks or replication stress.[18] Though, a single DNA structure might be responsible of ATR activation. This structure includes single stranded DNA coated by RPA, that has affinity for most ssDNA in the cell. This structure can be formed when a replication fork gets stalled, during resection at DSBs or at uncapped telomeres.[18] ATR recognizes RPA coated ssDNA only through another protein called ATRIP (ATR-interacting protein) (fig. 9). ATR and ATRIP

interact with each other in an unregulated manner and their stability is interlinked. ATRIP is a crucial subunit of the ATR kinase since its loss is as detrimental as ATR depletion. ATRIP binds RPA directly via an acidic alpha helix that connects to a basic cleft in RPA n-terminal part.[18,19] RPA coated ssDNA allows ATR-ATRIP recruitment however this structure is not sufficient to promote ATR activation. The Rad9-Rad1-Hus1 (9-1-1) complex, a heterotrimeric ring shaped molecule related to the replicative sliding clamp PCNA, is necessary for ATR activation. The 9-1-1 complex localizes to ATR recruitment site as it recognizes a DNA end that is adjacent to a stretch of RPA coated ssDNA[18,19] (fig. 9). This complex allows the recruitment of TopBP1 that is responsible for ATR activation. TopBP1 is a BRCT domain containing protein that includes an ATR activation domain (AD). TOBP1 AD binds ATR within its ATRIP domain and ATR activation involves its PIK-kinase regulatory domain (PRD)[18–20] (fig. 9).

As for ATM, post-translational modifications of ATR might regulate its activity and localization. Several phosphorylation sites have been identified on ATR and ATRIP and like ATM, ATR-ATRIP may form oligomeric complexes. Unfortunately, for now, none of these phosphorylation sites have been found responsible for ATR activation and there is no proof that ATR-ATRIP oligomerization is regulated.[18] Post-translationally modified kinases can be often purified in their activated state, but this is not the case for ATR. This indicates that the active form of ATR is dependent on the presence of a protein, like TopBP1, that might be needed to trigger its activation.[18–20]

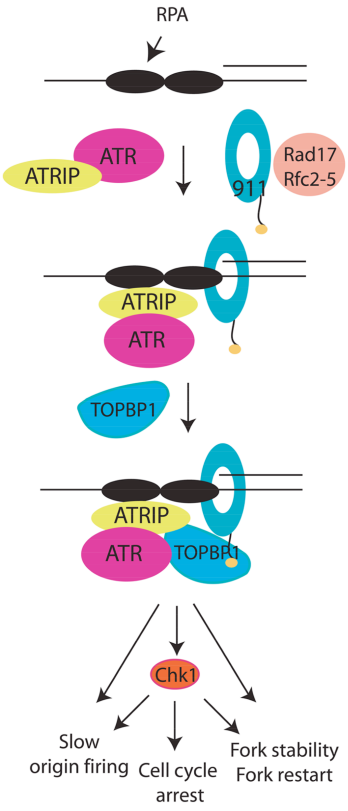

Figure 9: Simple models for ATR activation.18

The ATR-ATRIP complex and the 9-1-1 complex are recruited to the ssDNA-5′ primer junction independently. RPA binds ATRIP and directs the Rad17-RFC complex to load the 9-1-1 checkpoint clamp at the 5′ primer junction. Loading of 9-1-1 brings the ATR activator TopBP1 to the damage site. TopBP1 binds and activates ATR in an ATRIP-dependent manner, leading to phosphorylation of the downstream kinase Chk1 and other ATR effectors. In response to DNA damage or replication stress, ATR and its effectors ultimately slow origin firing and induce cell cycle arrest as well as stabilize and restart stalled replication forks.

c. DNA-dependent protein kinase complex, DNA-PK

DNA-PK, like ATM and ATR is a serine/threonine kinase from the PI3K-like family. It is composed of three proteins, Ku70, Ku80 and DNA-PKcs (DNA-PK catalytic subunit).[21] DNA-PKcs is a large protein of 4128 residues that contains a PI3K domain in its C-terminal part (fig. 10). Crystallography and electron

microscopy studies showed that DNA-PKcs encloses a channel able to hold double stranded DNA and that the whole DNA-PK complex forms an open ring where DNA can pass through.[21] Ku proteins bind DNA ends first and then recruit DNA-PKcs to stabilize the protein/DNA binding. Moreover two broken DNA ends can be stabilized in a complex containing two DNA-PKcs proteins in order to facilitate their rejoining.[21] In fact, DNA-PK principal role is to stabilize and help to process broken ends to allow their proper religation through non-homologous end joining (NHEJ). Nevertheless it has been shown that active DNA-PKcs is able to phosphorylate H2AX and initiate early step in the DNA damage response.[21]

As for ATM and ATR, post-translational modifications on DNA-PKcs are numerous and play an important role in its regulation. One of the major post-translational modifications of DNA-PKcs is its autophosphorylation. It results in kinase inactivation and dissociation of DNA-PKcs from the Ku70/80 heterodimer. They are two principal autophosphorylation clusters conserved through evolution in eukaryotes, one from residues 2023 to 2056 and the other from residues 2609 to 2647[21] (fig. 10). T2609 autophosphorylation is particularly interesting as it happens after IR and colocalizes with DNA damage. DNA-PKcs phosphorylation is very rapid after IR treatment, showing that DNA-PK is involved in the early response to damage. Intriguingly, phosphorylation of one of the clusters has opposite effects to the phosphorylation of the other cluster. Actually, phosphorylation of 2609-2647 promotes factor association at the DNA ends whereas phosphorylation of 2023-2056 results in DNA end joining with nucleotide deletion.[21]

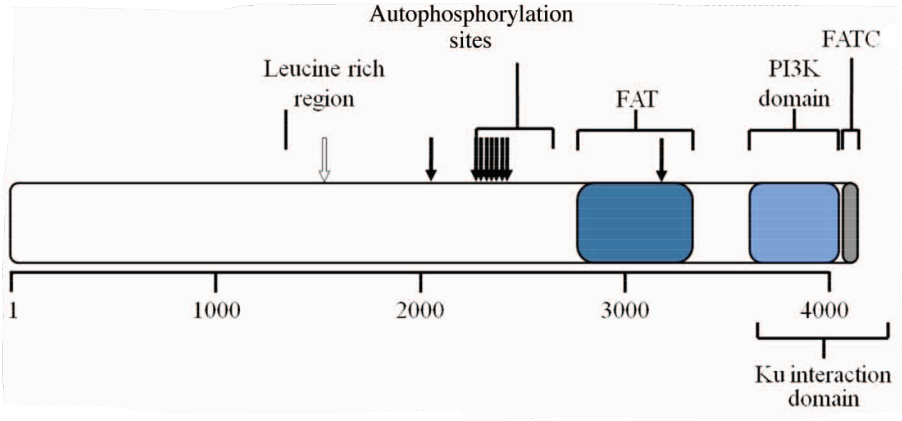

Figure 10: A schematic representation of human DNA-PKcs.[21]

DNA-PKcs is 4129 amino acid polypeptide and contains a FAT domain between 2908 and 3539 sharing sequence similarity between the related Pi3KK ATM and ATR. The FATC region spans 4096 and 4128. The Ku interaction domain is located between amino acids 3002 and 3850. Autophosphorylation sites (indicated by solid black arrows) have been identified at threonine 2609, 2638, 2647 and serine 2612, 2620, 2624 and 3205. A leucine zipper-like sequence (indicated by an open white arrow) between amino acid 1503 and 1538 has also be identified.

d. Checkpoint kinases, Chk1 and Chk2

Chk1 and Chk2 are activated through different mechanisms. Chk1 is primarily activated through ATR, in response to replication stress for instance.[18,22] The active ATR protein is able to phosphorylate Chk1 on at least two residues (S317 and S345). A mediator protein named Claspin is essential for this phosphorylation. Claspin binds to Chk1 in DNA damage dependent manner and is phosphorylated as well.[22] Following its activation, Chk1 is released from the chromatin in order to phosphorylate its substrates.[22] Activated ATM phosphorylates Chk2 on T68 but also on other residues of its N-terminal SCD domain. These last phosphorylations allow dimerization of the protein that promotes Chk2 autophosphorylation on several residues (T383, T387, S516, S379) promoting its activity.[22]

Activated Chk1 and Chk2 have essential role in regulating transcription, apoptosis, DNA repair and cell cycle progression. Concerning cell cycle, Chk1 is the primary effector of the intra-S and G2/M checkpoints, whereas Chk2 has a partial influence on the intra-S and G1/S checkpoints.[22]

3. DDR signaling

As described previously the first step of DDR is sensing of the break through the MRN complex (Mre11, Rad50, Nbs1). Then the ATM (Ataxia Telangiectasia mutated) kinase is recruited and activated (fig.11). This activation happens through the autophosphorylation of ATM (on S1981) that allows its monomerization. In fact, ATM is present in the undamaged cell as an inactive dimer. Activated ATM phosphorylates then the histone variant H2AX (then named γH2AX) in close proximity to the break. This allows the recruitment of the mediator protein MDC1 that has affinity with γH2AX. MDC1 acts as platform in order to recruit more MRN-ATM complexes to spread the signal along the DNA[9,10] (fig. 11). The second role of MDC1 is to recruit ubiquitin ligase such as RNF8 (RING finger protein 8) that is responsible of the recruitment of downstream factors like 53BP1 and BRCA1[9,10] (fig. 11). At the same time, DNA is resected and the single stranded DNA (ssDNA) coated with RPA (replication protein A). This results in the recruitment of the ATR kinase via its interacting partner ATRIP[9] (fig. 11). Both the ATM and ATR cascades leads to the activation of the checkpoint kinases Chk1 and Chk2 in order to block the cell cycle until the lesion is repaired or to direct the cell into senescence or apoptosis if the damage can not be overcome [9,10] (fig. 11).

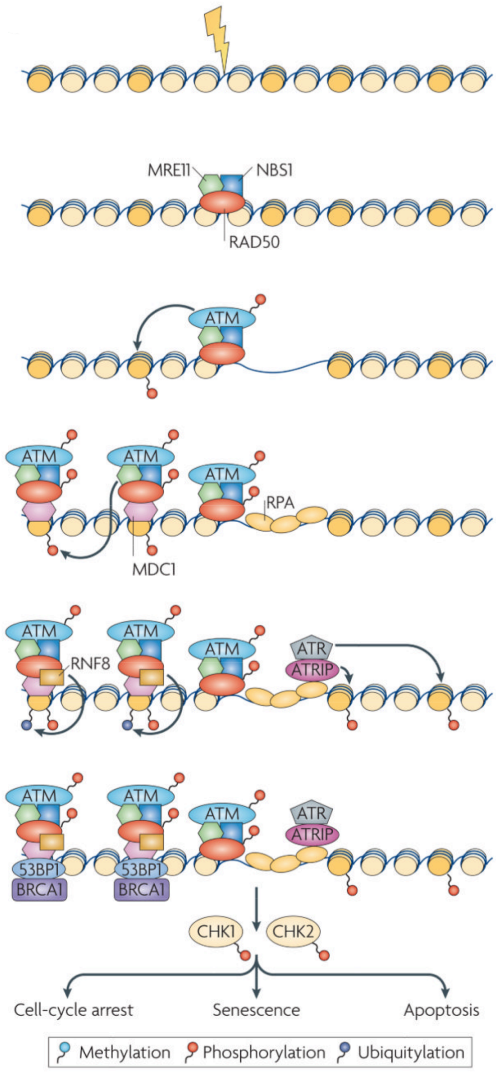

Figure 11: Assembly of DNA-damage response complexes.[9]

In mammalian cells, the presence of a double-stranded DNA break (DSB) is originally sensed by the MRN (Mre11–Rad50–Nbs1) sensor complex. The recruitment of MRN activates the transducer kinase ATM, which associates with DSBs and phosphorylates the histone variant H2AX (dark yellow) in DSB-flanking nucleosomes. The mediator MDC1 protein then binds to γH2AX and recruits additional copies of MRN and ATM, leading to spreading of the repair machinery along the chromosome. MDC1 also recruits ubiquitin ligase activities (for example, RNF8) that are responsible for the recruitment of downstream factors such as 53BP1 and BRAC1. DNA is then resected to single-stranded DNA (ssDNA) and recognized by replication protein A (RPA), which results in the recruitment of ATR through its interacting partner ATRIP. Both the ATM- and the ATR-dependent branches of the pathway, independently or in concert, lead to the activation of the checkpoint kinases Chk1 and Chk2.

4. Replication stress signaling

In case of a blocked replication fork, long stretch of ssDNA is formed. Single stranded DNA is coated by replication protein A (RPA). This will allow the recruitment of ATR via its interacting partner ATRIP. Activation of ATR requires the activator protein TopBP1. TopBP1 is implicated both in replication initiation as well as in DNA damage response. It is recruited to the stalled fork through the PCNA-like DNA clamp 9-1-1 complex (Rad9-Hus1-Rad1). Finally TopBP1 will activate the ATR-ATRIP complex, via its interaction with ATRIP, that will in turn induce phosphorylation of the Chk1 kinase[18] (fig. 12).

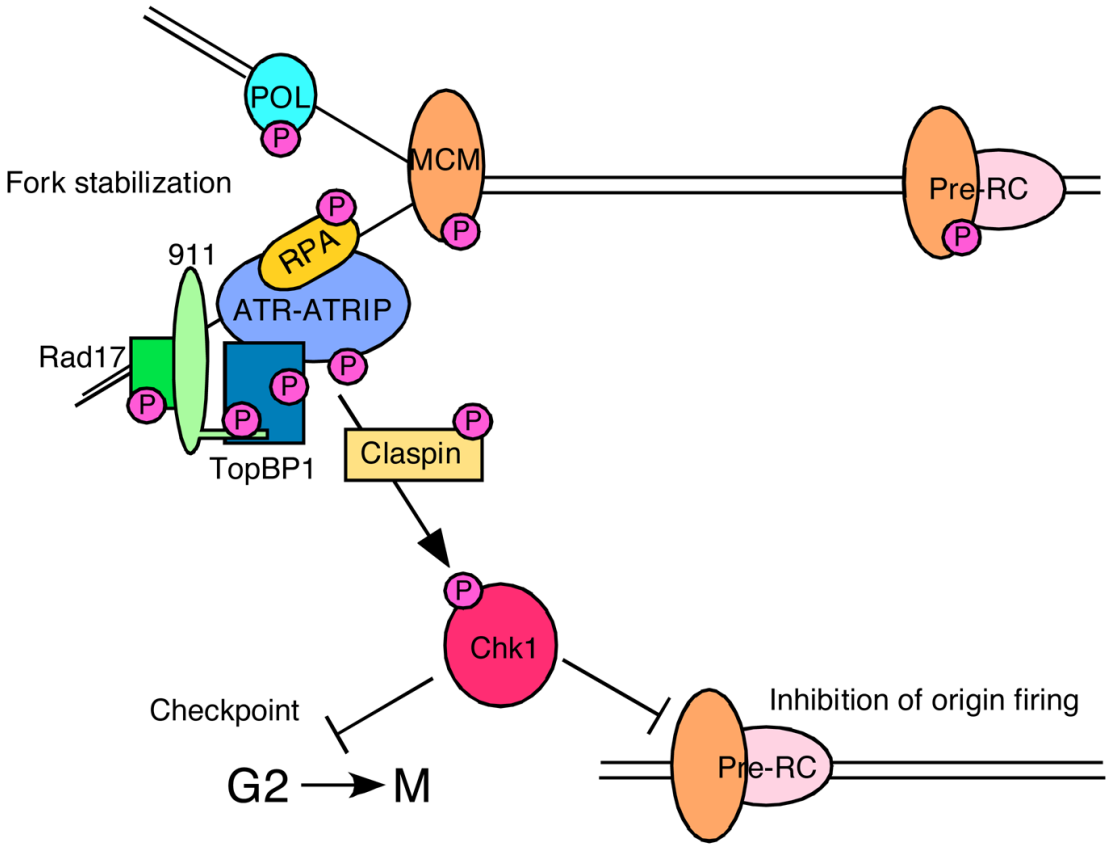

Figure 12: ATR phosphorylates numerous substrates to regulate replication and cell cycle transitions.[18]

A major ATR substrate is the checkpoint kinase Chk1. Chk1 phosphorylation releases it from chromatin and increases its kinase activity. Chk1 has numerous substrates some of which regulate cell cycle transitions and replication origin firing. Many ATR substrates are at the replication fork. In most cases, the consequences of phosphorylation remain unknown but likely contribute to fork stabilization. ATR-dependent MCM2 phosphorylation regulates binding to PLK1 which may promote the completion of DNA replication near the stalled fork.

5. DDR and chromatin

DNA damage response happens in the highly organized context of chromatin. Chromatin is highly dynamic in the presence of DSBs. Remodeling events happen in the structure of chromatin by opening it to allow better factor accessibility, or by closing it to silence transcription or prevent chromosomal translocations.[9,23] Emerging evidence suggests that the ability of repair factors to detect DNA lesions and to be retained efficiently at the site of damage is determined by histone modifications around the lesion and involves chromatin-remodeling events.[24,25] The most prominent DNA-damage-induced histone modification in DSB repair is the phosphorylation of the H2AX C-terminal tail, referred to as γH2AX.[23,26] Several other histone modifications, including acetylation, methylation, phosphorylation and ubiquitination of core histones have been linked to various aspects of DNA repair by facilitating access to the chromatin fiber, recruitment of repair factors and chromatin-remodeling components, as well as propagating cell-cycle checkpoint signals.[24] In response to DNA damage, chromatin undergoes rapid local and global decondensation[25], a process that is proposed to facilitate genome surveillance by enhancing access of DDR proteins to damage sites.

a. Histone modifications

i. Poly(ADP-ribosyl)ation

Poly(ADP-ribosyl)ation after induction of DSBs is mediated by three enzymes PARP1, PARP2 and PARP3 (poly(ADP-ribose) polymerases 1, 2 and 3). These enzymes add PAR chains on histones and this modification is one of the earliest detectable at the sites of DNA damage. The removal of PAR is achieved by PARG (poly(ADP-ribose) glycohydrolase).[23] PARP1 has a key role in DDR, it (ADP-ribosyl)ates lysines in the N-terminal tails of core histones. PAR chains are then bound by chromatin modifiers such as the NuRD (nucleosome remodeling and deacetylase) and polycomb complexes. These enzymes might silence transcription in the area near damage site, thus facilitating repair. The presence of these two complexes also enables the recruitment of RNF8 (RING finger protein 8) and RNF168 (RING finger protein 168), two ubiquitin ligases also implicated in transcriptional repression.[23] PARP3, on its side, triggers the assembly of APLF (aprataxin-PNK-like factor) at the break. APLF catalyzes the exchange of macroH2A (mH2A), a histone variant affecting gene silencing. This may further enhance the repressive chromatin produced by NuRD and polycomb complexes.[23]

Nevertheless it is also known that PAR can trigger chromatin relaxation. For instance, PAR dependent recruitment of ALC1 (amplified in liver cancer 1) promotes nucleosome sliding, therefore increasing chromatin accessibility at the break.[23]

ii. Phosphorylation

ATM phosphorylates H2AX on S139, forming γH2AX.[26] At the same time the neighbouring Y142, a residue constitutively phosphorylated in absence of DNA damage, is dephosphorylated. This dephosphorylation event is crucial as it is required for γH2AX recognition by MDC1. The phosphorylation resulting from this interaction coordinates all the downstream chromatin associated events.[23] MDC1 is constitutively phosphorylated by CK2 (casein kinase 2) allowing the anchorage of activated ATM on chromatin through Nbs1. Indeed, Nbs1 can simultaneously bind ATM and CK2 phosphorylated MDC1. This event permits the spreading of ATM and γH2AX expansion on the chromatin.[23] This γH2AX expansion also occurs in case of replication stress. In this situation, γH2AX spreading happens through recruitment of TopBP1, which activates ATR at stalled replication forks.[23] γH2AX doesn't spread linearly. High γH2AX density zones as well as low density γH2AX areas can be found around the break. To explain this, it was proposed that a fraction of activated ATM is not bound to chromatin and moves from the break. Additionally, this could explain why H2AX phosphorylation can "jump" over condensed heterochromatic regions.[23]

MDC1 is considered as the major γH2AX sensor. However MCPH1 (microcephalin) seems to bind γH2AX in a MDC1 independent manner. MCPH1 allows the recruitment of the SWI-SNF complex, a chromatin remodeling complex, thus increasing chromatin relaxation at the damage.[23]

Finally, γH2AX needs to be dephosphorylated to terminate the process and restore the chromatin to its initial state. They are several phosphatase implicated in

γH2AX dephosphorylation such as PP2A (protein phosphatase 2A) or WIP1 (wild-type p53-induced protein 1).[23] Another possibility proposed to remove γH2AX after DNA repair is replacement of histones. The FACT complex, known for its ability to facilitate transcription, is suggested to catalyze histone exchange of γH2AX with H2A in nucleosomes at DSB.[9]

iii. Acetylation

MDC1 is also a crucial player in acetylation induced by DNA damage. NuA4, a multisubunit complex is recruited in presence of MDC1. This complex contains the histone acetyltransferase Tip60. Once Tip60 recruited to the chromatin, a transient release of HP1 (heterochromatin protein 1) is needed to allow its activation. HP1 release is unmasking the constitutively tri-methylated lysine 9 of histone H3 (H3K9me3), this mark is then recognized by Tip60 chromodomain and permits the protein activation.[23] Tip60 is a tumor suppressor member of the MYST family of histone acetyltrasferases. It functions with its co-factor TTRAP (TRAF and TNF receptor associated protein).[27] Together these proteins are responsible for histone 4 hyperacetylation on lysines 5, 8, 12 and 16 after DNA damage, which promotes chromatin decondensation.[27,28]

Tip60 also acetylates H2AX, triggering its ubiquitination via the UC13 ubiquitin ligase. This ubiquitination event initiates γH2AX release from chromatin after DSBs in order to facilitate chromatin reorganization. Ubiquitination of H2A and H2AX is also important for the DNA damage signal amplification.[27]

Additionally Tip60 doesn't exclusively acetylate histones, as ATM is one of its targets too. ATM acetylation is required for its full activation after DSB induction (fig. 13). Tip60 also binds and acetylates the catalytic subunit of DNA-PK. Tip60 is recruited to break as a complex with ATM (fig. 13). Tip60 binding with H3K9me3 is not required for its recruitment to DSB but necessary for the acetyltransferase activity. The recruitment of Tip60 to breaks happens through its interaction with ATM and MRN.[27,29]

On the other side, HDAC1 and HDAC2 (histone deacetylase 1 and 2) assemble at DSBs too. These enzymes deacetylate lysine 56 of histone 3 (H3K56) and lysine 16 of histone 4 (H4K16) generating a more tight wrapping of DNA around nucleosomes. HDAC1/2 stimulate NHEJ by preventing aberrant spreading of the NHEJ protein Ku away from broken ends. They may also facilitate DSB tethering and diminish the risk of chromosomal translocations.[23]

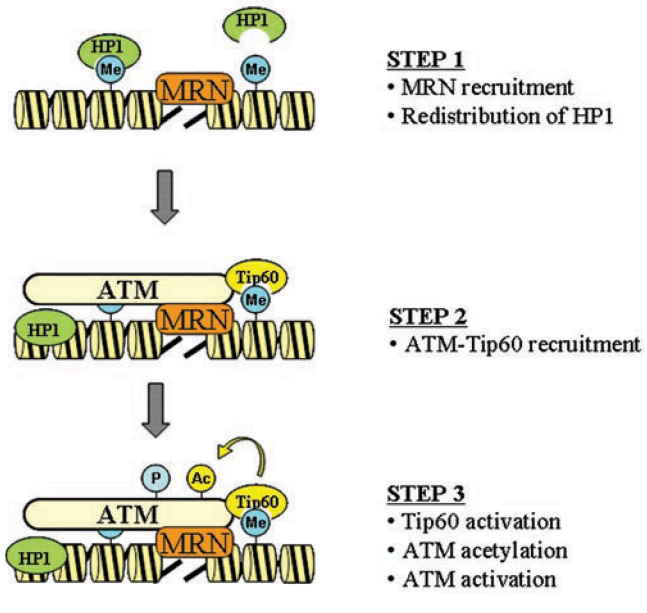

Figure 13: A mechanism for ATM activation.[29]

STEP 1: Following DSB production, MRN is recruited to DSB. In parallel, HP1 proteins are released from H3K9me3, and either retained on the chromatin, or released to the nucleoplasm by a process involving phosphorylation of HP1 by the CK2 kinase. It is not know if MRN participates in this process. STEP 2: inactive ATM-Tip60 complex is recruited to the DSB by MRN, facilitating interactions between Tip60's chromodomain and H3K9me3. STEP 3: Interaction between MRN and ATM, in combination with acetylation of ATM by Tip60, activates ATM's kinase activity. Me = methylation, ac = acetylation, p = phosphorylation.

iv. Ubiquitination and SUMOylation

Ubiquitin is a small regulatory protein of 76 residues, which can be attached to substrate protein via a three steps process. The first phase is activation done by ubiquitin-activating enzymes, then conjugation done by ubiquitin-conjugating enzymes and finally ligation performed by ubiquitin ligases. Ubiquitin can be added as a single molecule or as chains of several moieties linked by a specific amino acid, for instance lysine 48 or 63.[23,30]

Chromatin around breaks is enriched in K63-linked ubiquitin chains. This reaction is coordinated by MDC1 as MDC1 phosphorylation by ATM triggers the recruitment of RNF8 an ubiquitin ligase that initiates the ubiquitination reaction.[23] RNF168 another ubiquitin ligase completes the polymerization by binding the ubiquitin conjugate and amplifying it.[23] This cascade ultimately leads to the recruitment of chromatin associated genome caretakers as BRCA1, RAD18, 53BP1, TopBP1. It is also important to note that BRCA1 and RAD18 are also ubiquitin ligases and might further edit the ubiquitination initiated by DSBs.[23]

A "cousin" of ubiquitin, SUMO (small ubiquitin-related modifier) is implicated as well in DDR (fig. 14). DSBs trigger recruitment of SUMO ligases such as PIAS1

and PIAS4 (protein inhibitor of activated STAT-1 and 4) and SUMO conjugase such as UBC9 (ubiquitin-conjugating enzyme 9). These enzymes conjugate SUMO2/3 and SUMO1 respectively.[23,30] PIAS1/4 SUMOylations are necessary for proper formation of ubiquitin conjugates by RNF8, RNF168 and BRCA1. One example of SUMOylated factors is BRCA1 and 53BP1. RNF8/168 mediate the recruitment of BRCA1 and 53BP1 to the break that are then SUMOylated. Both BRCA1 and 53BP1 are SUMOylated by either PIAS1 or PIAS4, and BRCA1 SUMOylation enhances its ubiquitin ligase activity. In case of RNF8/RNF168 depletion SUMO1 and SUMO2/3 accrual is reduced, as SUMOylation targets are not recruited to the DSB.[30]

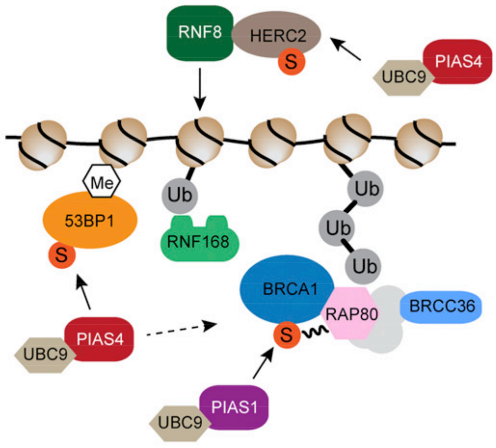

Figure 14: Examples of the role of SUMO in multiple DNA damage response pathways.[30]

Sumoylation acts at multiple steps during the chromatin-based DSB response.

v. Methylation

Methylation of lysine 79 of histone 3 (H3K79me) and lysine 20 of histone 4 (H4K20me) are constitutive marks in chromatin. Nevertheless, they promote 53BP1

accumulation via its Tudor domain after DNA damage. This suggests that DSBs provoke nucleosome repositioning thus exposing these methylated marks.[23] Furthermore H4K20me is found to be increased at damage site through the methyltransferase MMSET. MMSET phosphorylated by ATM binds to MDC1, involving this protein in another DNA damage activated chromatin organization mechanism.[23] Finally, SET8 (SET-domain-containing protein 8) a methyltransferase responsible for H4K20 monomethylation, accumulates at DSBs. Its activity seems necessary for 53BP1 recruitment in case of replication stress, as SET8 accumulation requires PCNA.[23]

b. DDR in heterochromatin

Heterochromatin, known to be more compacted and gene poor compared to euchromatin, has been shown to be refractory to γH2AX foci formation. Two hypotheses could explain this statement. First, heterochromatin might be less accessible to phosphorylation due to its compacted nature. Second, heterochromatin could be more resistant to DNA damage as the high concentration of protein bound to DNA could act as a protective layer.[27]

DNA damage and the mechanisms associated with it involve chromatin remodeling. The structure of the chromatin has to be modified via the manipulation of nucleosomes to allow access to the DNA.[27,31] The enzymes that coordinate this reorganization are ATP-dependent chromatin modifiers and histone modifiers. For instance, INO80, an ATP-dependent chromatin remodeler, is able to move or evict nucleosome to allow access to the DNA.[27] Chromatin remodeling happens both in

heterochromatin and euchromatin as both types of chromatin contain nucleosomes.[27,31] Howewer, DDR and DSB repair have a slower kinetic in heterochromatin compared to euchromatin. Heterochromatic DSBs might be more difficult to repair and require more processing than euchromatic breaks. Critical mediators have been identified in DSB induced heterochromatin decondensation: KAP1 and HP1. All these proteins function in close collaboration with the key DSB signaling kinase, ATM.[27]

i. KAP1

KAP1 (KRAB-associated protein 1) also known as TIF1β and TRIM28 is a corepressor for the Krüppel-associated box containing zinc-finger proteins (KRAB-ZFPs), a large class of transcription factors. KRAB-ZFPs bind to specific target sequences in chromatin and then interact with KAP1 through its RING finger B-box coiled-coil (RBCC) motif. KAP1 then recruits HP1, SETDB1 and Mi2-α (a compononent of the NuRD/Mi2 deacetylase complex).[27,32] SETDB1, that is a methyltransferase, tri-methylates histone 3 lysine 9 (H3K9me3), which is a marker of heterochromatin. The co-repressive activity of KAP1 and the recruitment of its partner proteins are also dependent on its auto-SUMOylation on at least three lysines (554, 779 and 804).[27]

KAP1 was found to be phosphorylated after DNA damage on serine 824. ATM is one of the kinases responsible for its phosphorylation but ATR and DNA-PK can also play this role. The phosphorylation of KAP1 helps to regulate its SUMOylation, and therefore its function in transcriptional repression and chromatin condensation.[27,32]

Phosphorylation of S824 causes decreased SUMOylation and S824 dephosphorylation triggers increased SUMOylation.[27] KAP1 can also be phosphorylated on serine 473 by either Chk1 or Chk2 after DNA damage. Phosphorylation of KAP1 on S473 attenuates its binding to the HP1 family proteins and inhibits its transcriptional repression of KRAB-ZFPs target genes.[27]

KAP1 plays also a role in ATM dependent chromatin relaxation after DNA damage. When KAP1 is phosphorylated at S824 by ATM, the chromatin becomes more sensitive to micrococcal nuclease after DSBs induction.[27] The proposed model is that, in presence of DSBs, activated ATM phosohorylates KAP1 throughout the nucleus, which in turn permits transcription of genes involved in cell cycle regulation and apoptosis. Then in a second step, the phosphorylated KAP1 (by ATM and Chk1/2) nearby the break initiates local chromatin decondensation by release of the HP1 proteins in order to allow access of the repair factors.[27]

ii. HP1

HP1 proteins are critical proteins in heterochromatin formation. They bind the heterochromatic histone mark H3K9me3 through their chromodomain, the DNA methyltransferases DNMT1 and DNMT3a and the histone methyltransferase Suv39H1. Suv39H1 is known to trimethylate H3K9, and as HP1 interacts with both the mark and the methyltransferase, it is proposed to mediate heterochromatin expansion.[27,33,34] In mammalian cells they are three HP1 isoforms: HP1α, HP1β and HP1γ. All three isoforms can form homo or heterodimers via their chromoshadow domain. The dimerization of the protein potentially helps the chromatin compaction but also creates a hydrophobic surface allowing the interaction with KAP1 and the

p150 subunit of CAF1 (chromatin assembly factor 1).[27,33,34] HP1 proteins are involved in assembly and maintenance of constitutive heterochromatin like at telomeres or centromeres but it is also implicated in regulation of transcription (facultative heterochromatin).[27,33,34]

In presence of DSBs HP1 can be released from the break, either via KAP1 phosphorylation or for HP1β via its phosphorylation by casein kinase 2 (CK2).[27] However HP1 proteins are also recruited to sites of DNA damage. This recruitment is dependent of the chromoshadow domain and independent of the interaction with H3K9me3 and chromatin. A two phases model could explain these opposed roles of HP1. In a first step HP1 is released from H3K9me3 at the break to help chromatin decondensation to allow a better access to repair factors. In a second step, HP1 is recruited to the damage via its chromoshadow domain and is proposed to help mediate and enhance the DNA damage signaling.[27]

Furthermore, a recent paper suggests that HP1 proteins are involved in resection. HP1α and HP1β depletion trigger a reduction in RPA phosphorylation contrary to HP1γ depletion that has only a mild effect on the phosphorylation status of RPA. Additionally, homologous recombination rates are decreased when HP1α and HP1β are silenced whereas HP1γ knockdown stimulates homologous recombination. These results suggest that HP1γ presence at DSB is inhibitory for their processing but that HP1α and HP1β could stimulate it.[35]

IV. DSB repair

In mammals the two major pathways known to repair DSBs are non-homologous end joining (NHEJ) and homologous recombination (HR)[10] (fig. 15). NHEJ consist in the direct religation of the break. This method can be prone to errors, as it requires DNA end processing that can introduce mutations such as deletion or insertions (fig. 15A). On the contrary, HR is thought to be an accurate repair mechanism, as it requests the presence of a homologous sequence that operates as a repair template. This template is often found on the sister chromatid; therefore HR happens mostly through S/G2 phases of the cell cycle whereas NHEJ occurs throughout the cell cycle[8,10](fig. 15B).

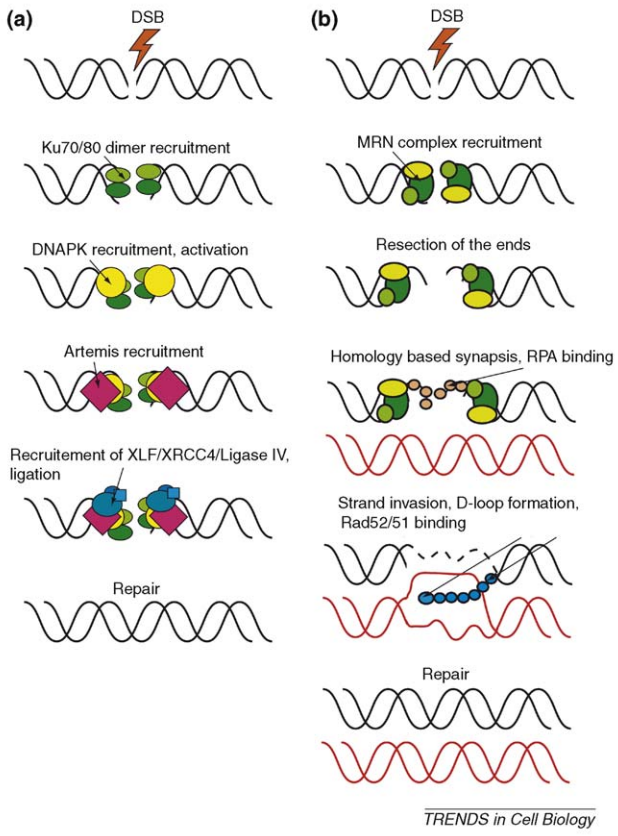

Figure 15: Double strand break repair pathways.[8]

(A) Non-homologous end joining (NHEJ): The first step of NHEJ is the recruitment of the Ku70/80 dimer; this is followed by recruitment of the DNAPK complex.

Activation and autophosphorylation of the DNAPK catalytic subunit leads to the recruitment of Artemis and the XLF/XRCC4/Ligase IV complex that perform the end-ligation step. (B) Homologous recombination: the MRN sensor complex composed of Mre11, Rad50 and Nbs1 binds to the DSB and resects the DNA ends to generate single-stranded DNA (ssDNA). The ssDNA bound by RPA and RAD proteins invades the homologous intact DNA template and DNA synthesis takes place to restore the integrity of the genetic information.

1. Non-homologous end joining, NHEJ

In mammalian cells NHEJ is the major mechanism for repairing DSBs. It is initiated when an intact DNA template cannot be found to perform homologous recombination. That is why NHEJ predominates in G_0 and G_1 but can function throughout the whole cell cycle. NHEJ reaction can ligate ends without any homologous nucleotides as well as ends with short nucleotidic homologies called microhomologies.[36]

a. Classical NHEJ, C-NHEJ

i. C-NHEJ model

Classical NHEJ can be divided into 3 different phases: end detection and tethering, processing and ligation[37] (fig. 16).

During the first step, the Ku70/80 heterodimer recognizes the DNA ends of the DSB and bind to them. Ku70/80 binding triggers the recruitment of all the other factors involved in the NHEJ repair. DNA-PKcs is another critical factor involved in the initial step of C-NHEJ, it interacts with Ku at the break and tethers the DNA ends

in a synaptic complex (fig. 16). The interaction between Ku and DNA-PKcs at the DSB enhances its kinase activity, which is required for NHEJ.[37]

Through the second step, DNA ends are enzymatically processed in order to produce DNA termini suitable for ligation. The enzymes handling this step are numerous. Some of them remove the non-ligatable chemical groups that can be present at the DNA ends (fig. 16). Examples of these enzymes are PNKP (polynucleotide kinase/phosphatase), a DNA 3'-phosphatase/5'-kinase or aprataxin (APTX) that may remove abortive ligation products.[37] Nucleases such as exonuclease 1 (Exo1), Mre11, Artemis or even the helicase/exonuclease WRN (Werner syndrome protein) might be also involved in DNA end processing by revealing DNA microhomology sequences that help position ends for ligation. Finally DNA polymerases such as polymerase μ or λ fill nucleotide gaps or deletions. This step of processing is often leading to nucleotides insertion or deletion rendering NHEJ prone to errors.[37]

The third and final step of NHEJ consists in ligation of the two ends by DNA ligase IV (LIG4). LIG4 interacts with x-ray cross-complementing gene 4 (XRCC4) and XRCC4-like factor (XLF or Cernunnos) and forms a complex with them[37] (fig. 16).

The organization in time of NHEJ is poorly understood. Besides the initial Ku binding and the final LIG4 ligation very little is known about the timing of the intermediate steps. NHEJ might be a dynamic multicomponent process in which specific factors operate in accordance with the chromatin environment, the cell cycle stage and the conformation type of the DSB. For instance, DNA-PKcs might be

required only for slowly repaired, complex lesions, whereas ATM and Artemis are required for DSB repair in heterochromatin.[37]

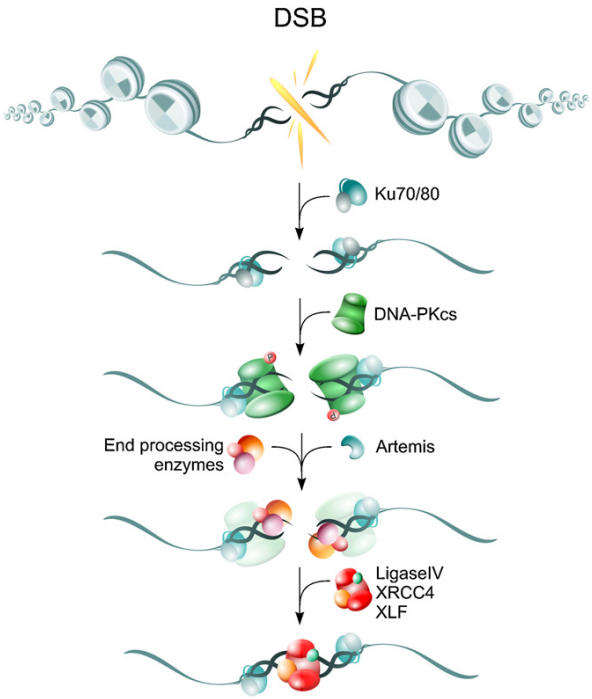

Figure 16: Outline of the main players and of the key steps of DNA-PKcs-dependent NHEJ (C-NHEJ).[38]

C-NHEJ efficiently restores genomic integrity without ensuring sequence restoration. Association of Ku to DNA ends facilitates the recruitment of DNA-PKcs, a major kinase activated by DNA ends and contributing to the efficiency of this repair pathway. DNA-PKcs promotes end processing by the Artemis nuclease and subsequent rejoining of broken DNA ends by the LIG4/XRCC4/XLF complex.

ii. Ku

The Ku70/80 heterodimer is a critical component of NHEJ. It forms a basket shaped structure that encircles dsDNA and that acts as a platform to recruit other proteins, such as DNA-PKcs. Indeed the C-terminal region of Ku80 is structured as a long "arm" that extends from the DNA binding core, allowing protein-protein

interaction.[37] Unfortunately very little is known on how DNA-PKcs is recruited to Ku at breaks. Ku also interacts with WRN as well as XRCC4 and is required for XLF recruitment to DSBs. Furthermore Ku is required for the recruitment of APTX and PNK-like factor (APLF), which stabilizes the repair complex and promotes NHEJ.[37]

iii. DNA-PKcs

DNA-PKcs function was described earlier. Here I describe in more detail the structure of the kinase.

The different domains of DNA-PKcs are ordered as followed: a large N-terminal domain composed of helical elements and HEAT (Huntingtin, Elongation factor 3, PP2A and TOR1) repeats, a FAT (FRAP, ATM, TRRAP) domain, a PI3K kinase domain and a FACT domain in the very C-termini[37] (fig. 17A). These regions organize in a 3D structure composed of a head or crown and a palm constituted of two arms. The FAT-PI3K-FACT domains are located in the head/crown region whereas the N-terminal domain forms the arms (fig. 17B). The protein binds dsDNA through the channel formed by the opening between the arms. Furthermore, the regions at the apex of the arms are predicted to be highly flexible, suggesting that the arms might open and close. An hypothesis is that phosphorylation of DNA-PKcs might trigger this conformational change and thus regulate its interaction with Ku and DNA[37](fig. 17C).

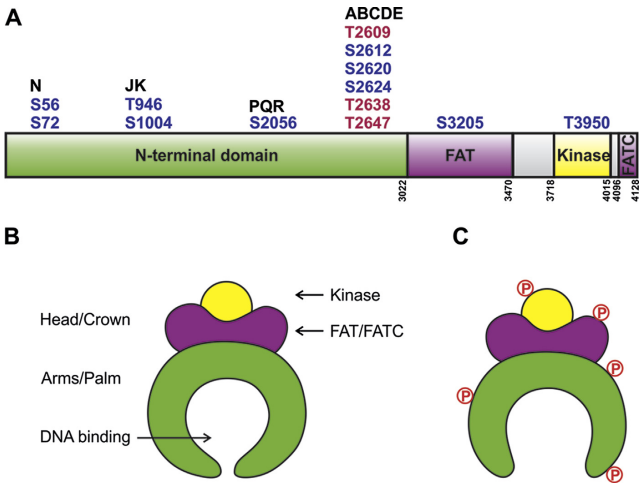

Figure 17: Structural organization of DNA-PKcs kinase.[37]

(A) Schematic of DNA-dependent protein kinase catalytic subunit (DNA-PKcs) showing major in vivo phosphorylation sites. Domain boundaries are shown by vertical numbers. (B) Model for structure of DNA-PKcs (C) Model for putative effects of phosphorylation on DNA-PKcs structure.

iv. XRCC4 and XLF

XRCC4 has no catalytic activity but plays an important role in NHEJ. Indeed it is able to interact and stabilize LIG4. XRCC4 is constitutively phosphorylated by CK2 on threonine 233. This mark allows its interaction with PNKP, APTX and APLF via their fork head associated (FHA) domains suggesting that XRCC4 is triggering their recruitment to damage.[37] XLF has no catalytic activity either but is able to facilitate XRCC4-LIG4 mediated end joining. It also promotes LIG4 readenylation, making it ready for the next ligation event. XLF and XRCC4 form dimers that interact with each other forming long helical filaments. These filaments are proposed to bridge or align DNA ends for ligation.[37]

v. DNA ligase IV, LIG4

DNA ligase IV structure is composed of a N-terminal DNA binding and catalytic domain followed by a C-terminal tandem BRCT domain. The N-terminal domain interacts with Artemis to regulate NHEJ in V(D)J recombination whereas the C-terminal region interacts with XRCC4, which is important for LIG4 stability. The XLF-XRCC4 filaments are destabilized by LIG4, which suggest that the ligase regulates their formation.[37]

b. Alternative NHEJ, Alt-NHEJ

In addition to the classical NHEJ pathway, another much less characterized alternative NHEJ pathway (Alt-NHEJ) has been identified. This pathway has been discovered in cells lacking crucial C-NHEJ factors that were still able to ligate DNA broken ends. Sequence analysis of these junctions showed that the frequency of microhomologies is clearly increased in the Alt-NHEJ pathway compared to C-NHEJ.[36,38] Practically, cells lacking DNA-PKcs itself or its activity are able to repair almost all IR-induced DSBs, although with a slower kinetic. This is also the case when Ku70/80, LIG4 or other factors of C-NHEJ are missing. It is important to note that Alt-NHEJ seems particularly error prone as cells defective for DNA-PKcs or other C-NHEJ components show increased levels of chromosomal aberrations after exposure to IR.[38]

The recognition of the break involves PARP1 as an early step as well as the MRN complex (fig. 18). There might be a competition between PARP1 and Ku when both

proteins are present. However, Ku has much greater affinity for DNA ends than PARP1, enhancing the idea that Alt-NHEJ is a backup repair mechanism. Furthermore, PARP1 is involved in repair of other types of damage (SSBs, base damages…) present in excess compared to DSBs; this is further contributing to the backup character of Alt-NHEJ. DNA end processing might be handled by the MRN complex and especially via Mre11 that has a nuclease activity (fig. 18). Finally the ligation step might be performed by DNA ligase III (LIG3) in this pathway (fig. 18). LIG3 functions in a complex with XRCC1 and is regulated by PARP1 but this LIG3/XRCC1/PARP1 complex is also involved in other repair mechanism such as repair of single stranded breaks or base damages. Moreover, Werner syndrome proteins seem to be involved as well in Alt-NHEJ and recruited to DSBs in this mechanism.[38]

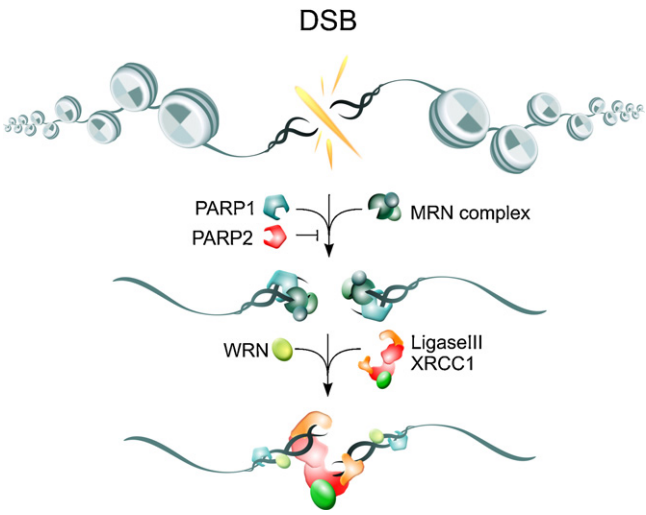

Figure 18: Outline of the main players and of the key steps of an alternative form of NHEJ functioning as a backup to D-NHEJ (Alt-NHEJ).[38]

The enzymatic activities implicated in Alt-NHEJ are shown together with their possible contributions in the process.o

2. Homologous recombination, HR

Homologous recombination in opposition with non-homologous end joining requires the presence of an undamaged homologous DNA sequence to operate as repair template. HR could be thought to function throughout the cell cycle if the homologous chromosome can be used as a template. However this not the case as polymorphisms or homologue sequence differences are too great to allow efficient heteroduplex formation and also as homologous chromosomes are often located far away from their partners. That is why HR happens in the S/G2 phases of the cell cycle when a sister chromatid is available.[10]

The first step of HR is DNA end resection at the DSB site in order to produce 3' single stranded DNA overhangs that are able to invade a DNA duplex containing a homologous sequence. In mammals, the resection happens through the collaboration of the MRN complex, CtIP (CtBP-interacting protein) and BRCA1 (breast cancer 1) (fig. 19). Furthermore EXO1 has been shown to resect DNA *in vitro* and is activated by BLM (Bloom syndrome protein), so these proteins might be implicated in DNA end resection too.[39] The ssDNA created during resection is bound by RPA, which is required for subsequent checkpoint and HR protein recruitment like RAD51. RAD51 is a DNA dependent ATPase that forms nucleoproteins filaments with DNA and operates as a strand exchange protein. RAD51 recruitment is also helped by BRCA2 (breast cancer 2) that bind RAD51 via its eight BRC repeats (fig. 19). The presence of these repeats stimulates RAD51 filament formation and BRCA2 is known to interact with ssDNA. Furthermore BRCA1 and BRCA2 interact through BRCA2 interacting partner PALB2. The disruption of this interaction results in decreased levels of HR.[39]

RAD51 has a crucial role during the three phases of HR: presynapsis, synapsis and post-synapsis. During the presynapsis RAD51 is loaded onto the 3′ single stranded overhang forming the presynaptic filament (fig. 19). This filament has a right handed helical structure composed of six RAD51 moieties and eighteen nucleotides per turn. The ssDNA within the filament is stretched and this stretching is essential for fast and efficient homology search. In the synaptic phase a physical connection between the invading DNA substrate and homologous substrate is formed, generating a heteroduplex DNA (D-loop) (fig. 19). RAD51 helps the formation of this heteroduplex. Finally during the post-synapsis DNA is synthetized using the invading 3′ stand as a primer, at this time RAD51 dissociates from DNA in order to expose 3′-OH required for the synthesis[40] (fig. 19).

After synthesis initiation, the D-loop can go through several paths. In the synthesis dependent strand annealing pathway happening in mitotic cells, the 3′ end in the D-loop is extended by repair synthesis, and then the newly synthesized DNA strand dissociates to anneal to the other DNA end to complete the reaction. In the double strand break repair model, the second end of DSB can be "captured" by the D-loop to stabilize its structure forming a double Holliday junction. Several different proteins such as GEN1 and SLX1/SLX4 can resolve the double Holliday junction (fig. 19). There are two ways to resolve this junction, leading to crossover and non-crossover products.[39,40] During meiotic recombination crossover play an important role in facilitating chromosome segregation whereas during mitotic recombination crossovers may have serious deleterious effects as for instance, loss of heterozygosity. Some proteins such as BLM are capable of dissolving D-loops suppressing then mitotic crossovers thus decreasing the risk of genomic instability.[39]

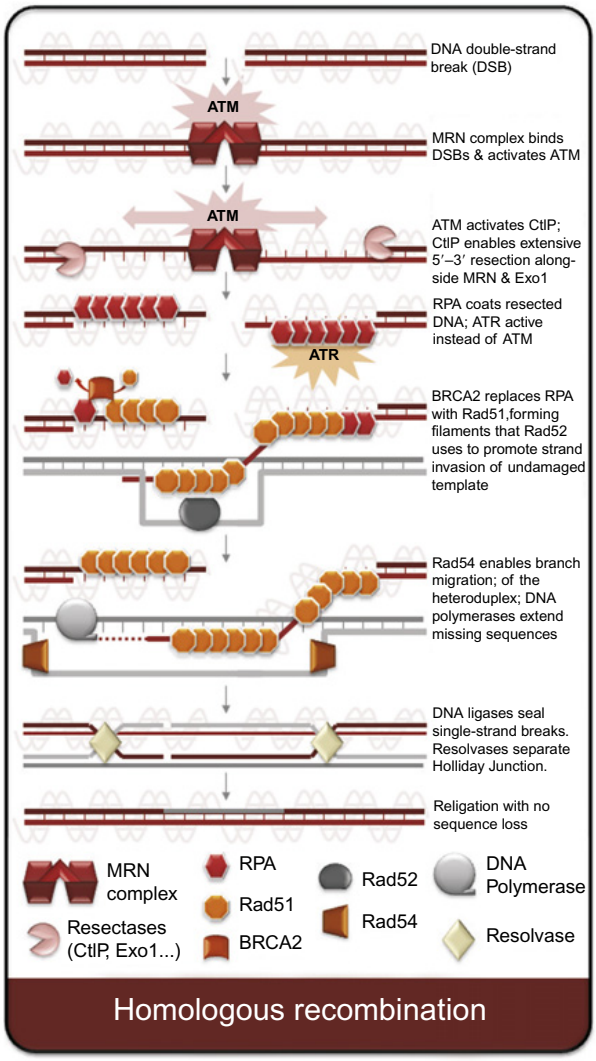

Figure 19: Homologous recombination repair.[10]

During HR, resection is initiated by MRN and CtIP and progressed by exonucleases such as EXOI and BLM/DNA2. The single-strand DNA generated is coated with RPA, which can activate ATR. Via BRCA2 and RAD52, RPA is displaced by RAD51 to generate DNA filaments, which are capable of invading the undamaged strand to generate D-loops. The hybrid DNA generated is called heteroduplex DNA with junctions formed by DNA-strand crossing over referred to as Holliday junctions. The cross-over point can move by branch migration mediated by Rad54. Finally, resolvases separate Holliday junctions and generate products with or without crossovers, depending on the relative direction of resolution of the two Holliday junctions.

V. SET

As introduced previously, DNA damage happens in the context of the highly structured chromatin. It is now established that chromatin remodelers such as histone chaperones have a critical role in DSB signaling and repair. Histone chaperones are defined as factors having a histone binding activity as well as a histone dependent ATP independent nucleosome assembly activity.[41] Here is introduced SET, a histone chaperone belonging to the NAP1 family that was identified in our lab through a siRNA screen targeting DNA remodeler involved in DNA damage. SET has been referred under several names such as PHAPII (putative class II human histocompatibility leukocyte-associated protein II), TAF-Iβ, I2PP2A, INHAT, IGAAD or StF-IT-1 according to its function when identified.[41]

1. SET identification

SET was first identified as a translocated gene in acute undifferentiated leukemia. It was found in fusion with the can/NUP214 gene.[42] SET was also independently identified as template activating factor I (TAF-I) from a HeLa S3 cell extract as a stimulator of replication of adenoviral core DNA.[43] Nucleotide sequence analysis revealed afterwards that SET and TAF-Iβ were the same proteins. Two isoforms of SET originated from alternative splicing do exist. These isoforms are called TAF-Iα and TAF-Iβ. TAF-Iα differs from TAF-Iβ by an additional short N-terminal sequence, 13 extra amino acids are present on TAF-Iα.[41] SET/TAF-Iβ is a nuclear protein ubiquitously expressed in different cell types whereas TAF-Iα expression level differs in different cell types.[44] SET is highly conserved in several vertebrates (frog,

mouse, rat and human) but homologs are also found in *Ceanorhabditis elegans* (Spr-2) and *Saccharomyces cerevisiae* (Vps75).[41] Vps75 is very similar to human SET. It has the same structure and preferential binding for H3-H4. Vps75 interacts with the histone acetyltransferase Rtt109 and promotes its activity.[45]

2. SET structure

SET is composed of three domains: a N-terminal domain, a central NAP domain and an acidic C-terminal domain. The NAP domain can be divided in two parts, a coiled coil region in N-terminal side of the NAP and an "earmuff" second domain located on the C-terminal side of NAP. A nuclear localization sequence is can be also found in the NAP domain. SET is present in the cell as a homodimer where the proteins interact with each other via their coiled coil domain. The whole structure forms a 'headphone" like configuration with the dimerization domains on the top and the two earmuffs domains on each side[46] (fig. 20). The N-terminal domain forms a α-helix that is not involved in protein dimerization and is supposed to be highly mobile. The structure of SET was determined without the presence of the acidic C-terminal domain as this type of regions is an obstacle for proper crystallization. Nevertheless, this part of the protein is an intrinsically disordered region. The region of SET involved in histone chaperone activity is positioned at the bottom surface of the earmuff domain as it can bind core histones and double stranded DNA.[46]

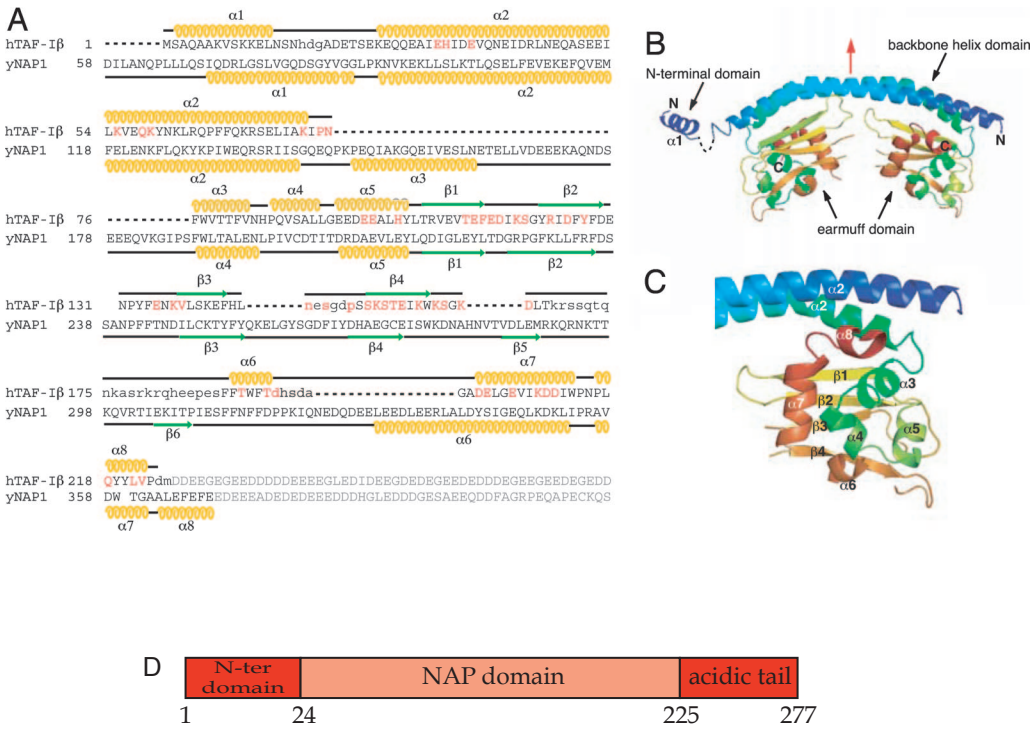

Figure 20: Structure of SET/TAF-Iβ/INHATΔC.[46]

(A) Amino acid sequences of SET/TAF-Iβ/INHAT and NAP-1. The α-helices (yellow) and β-strands (green) in the sequences are indicated. Residues with no observable electron density are not capitalized. Residues in the acidic stretch are gray. (B) Overall structure of the SET/TAF-Iβ/INHATΔC dimer, with the pseudo twofold axis highlighted in red. (C) The structure of the earmuff domain. (D) Different domains of the SET protein with amino acid number.

3. SET functions

SET is one of the best-characterized NAP family members and is known to be a multi-tasking proteins. Indeed, it has several functions and is implicated in various pathways.

a. SET is a histone chaperone part of the INHAT complex

First of all SET is a histone chaperone that has a preferential binding for H3 and H4 even if it is able to bind all four core histones.[46] The C-terminal region is thought be involved in histone interaction and due to its acidic nature, favors its binding with basic histones. This is the case only for H2A and H2B as a mutant lacking this domain is still able to bind H3 and H4.[46] Indeed the H3-H4 interacting region is found in the NAP domain (bottom surface of the earmuff domain).[46] SET interacts with unacetylated, hypoacetylated or repressively marked histones but not with hyperacetylated histones.[47,48]

SET has also more recently been identified as a H1 chaperone.[49] It was found to interact with linker histone H1 and to regulate its incorporation to chromatin. Fluorescence recovery after photobleaching analysis revealed that H1 mobility is increased when SET is overexpressed whereas H1 mobility is reduced in case of SET depletion, suggesting that SET is removing H1 from chromatin.[49]

SET was also established being part of the INHAT (inhibitor of histone acetyltransferase) complex.[50] INHAT is a multiprotein complex composed of TAF-Iα, SET/TAF-Iβ and pp32, a member of the family of leucine rich acidic nuclear phosphoproteins. INHAT strongly inhibits the histone acetyltransferase activity of p300/CBP and PCAF by masking histones from being acetylated thus decreasing transcription.[50]

b. SET is a transcription factor

SET also acts as a transcription factor and more precisely as a transcription cofactor. SET interacts with transcriptions factors KLF5 and Sp1, negatively regulating their activity thus causing transcription inhibition.[51,52] SET also interacts with nuclear receptor type transcription factors such as ERα (estrogen receptor α)[53] or GR (glucocorticoid receptor).[54] In absence of estrogen, SET binds to ERα to inhibit promoter acetylation thus preventing its activation.[53] In the case of GR, SET acts also as GR-responsive transcriptional repressor.[54]

c. SET is a chromatin remodeler

SET has been reported to remodel chromatin in either a direct or indirect way and with the ability to condense or decondense chromatin.

SET is more often described as a protein that provokes chromatin decondensation. First of all SET is known to be a stimulator of adenoviral genome replication and transcription.[43,55–58] This is achieved by increased accessibility to the transcription and replication machinery to the adenoviral core DNA.[43,55–58] It was also shown that SET is able to decondense demembraned *Xenopus laevis* sperm *in vitro* via direct interaction with basic proteins.[59] SET was also found to interact with Prothymosin α (ProTα) and that highlights another way of SET to indirectly decondense the chromatin as it was shown that that ProTα is a histone 1 binding protein that efficiently relaxes human sperm chromatin.[60] Finally SET interacts as well with microcephalin (MCPH1) via the N-terminal BRCT domain of MCPH1. MCPH1 ensures the proper temporal activation of chromosome condensation during

mitosis and SET depletion caused abnormal chromosome condensation similar to MCPH1 deficient phenotype.[61]

On the other hand, SET acting as a transcription cofactor or as member of INHAT in inhibiting histone acetylation (as described earlier), inhibits transcription.[50] This suggests that in this case SET originates chromatin condensation. Furthermore SET overexpression was also shown to inhibit DNA methylation causing gene silencing and thus chromatin compaction.[62]

d. SET is an inhibitor of PP2A phosphatase

SET has been identified as a potent inhibitor of protein phosphatase 2A (PP2A).[63] PP2A is a ubiquitous and conserved serine/threonine phosphatase that regulates diverse cellular processes, such as cell metabolism, cell cycle, DNA replication, transcription and translation, signal transduction, cell proliferation, cytosqueleton dynamics, cell mobility and apoptosis. The majority of soluble phosphatases activity at phospho serine and phospho threonine is catalyzed by PP2A. Interestingly, PP2A is known to be implicated in γH2AX de-phosphorylation. PP2A is considered as a tumor suppressor and is suggested to be functionally inactivated in cancer. Several mutations impairing PP2A activity have been identified in different types of cancers. That is why over inhibition of PP2A could lead to cancerogenesis.[64]

e. SET is involved in cell cycle regulation

SET is implicated in cell cycle regulation. SET was found interacting with $p21^{Cip1}$ an inhibitor cyclins A-CDK2 and E-CDK2.[65] SET is a modulator of $p21^{Cip1}$ inhibitory function to E-CDK2. Consequently SET can regulate G_1/S checkpoint by modulating the activity of of E-CDK2 (but not A-CDK2) by reversing inhibitory effect of $p21^{Cip1}$ and thus pushing cells into S phase.[65] SET was also shown to cooperate with $p21^{Cip1}$ to inhibit B-CDK1 in order to regulate the G2/M transition of the cell cycle.[66] Finally, SET interacts with GAPDH (glyceraldehyde-3-phosphate dehydrogenase) that reverses in a dose dependent manner, the inhibition of cyclin B-CDK1 activity provoked by SET.[67]

f. SET is involved in apoptosis

SET has a role in apoptosis. The tumor suppressor NM23-H1 is a Granzyme A activated DNase (GAAD) that can induce a caspase-independent cell death by creating DNA nicks in the cell genome. SET is NM23-H1 specific inhibitor and is thus termed IGAAD (inhibitor of GAAD). In case of apoptosis, Granzyme A (GzmA) cleaves SET releasing it from its interaction with NM23-H1. SET and NM23-H1 translocate to the nucleus and SET is degraded, allowing NM23-H1 to nick chromosomal DNA.[68] In addition to this mechanism, an exonuclease TREX1 is present in the SET complex. After GzmA NM23-H1 activation, TREX1 removes nucleotides from the nicked 3' to reduce the risk of possible end religation.[69] Granzyme K (GzmK) can also activate NM23-H1 by cleaving SET to release it from its interaction with NM23-H1.[70] This phenomenon happens in the context of natural killer (NK) cells from the immune system in order to kill viruses, intracellular

bacteria or transformed cells. When a killer lymphocyte recognizes its target cytotoxic granules containing granzymes are liberated by exocytosis and fuse their membranes with the target cell in order to release their content by endocytosis.[68,70]

g. SET has a specific role in neurons

In neurons, SET binds $p35^{nck5a}$ a neuronal Cdk5 activator and enhances its activity.[71] Besides of its role on Cdk5 kinase activation SET has been shown to have a role in neuronal development. Indeed in a study conducted in Drosophila melanogaster, it has been noticed that SET is highly expressed in cells before neuronal differentiation and then its expression decreases through development. SET is thus acting as a negative regulator of neuronal development.[72] Furthermore in a recent study SET has been shown to be highly expressed and translocated to the cytoplasm in neurons of Alzheimer's disease brains.[73] Indeed, in these cells, SET is phosphorylated on serine 9, disrupting what is thought to be a secondary NLS causing cytoplasmic translocation. Consequently PP2A is highly inhibited in the cytoplasm of these neurons, thus leading to hyperphosphorylation of tau, which is one of the hallmarks of Alzheimer's disease.[73]

h. SET's role in meiosis

SET plays a role into sister chromatid segregation in meiosis and more precisely in meiosis II. During meiosis I chromosome pairs separate whereas chromatid sisters are dissociated over meiosis II. Centromeric cohesin is protected from separase-

dependent removal in meiosis I through the activity of PP2A-B56 phosphatase, which is recruited to centromeres by shugoshin (Sgo). In meiosis II, SET as an inhibitor of PP2A, is positioned at centromeres colocalizing with PP2A and the cohesin subunit Rec8. Inhibition of PP2A by SET allows chromatid segregation during meiosis II and its depletion leads to failure of sister chromatid separation. SET's role is essential for faithful sister chromatid segregation by mediating deprotection of centromeric cohesion in meiosis II.[74]

i. SET's role in detoxification and DDR

A recent study shows that SET overexpression decreases cell detoxification efficiency.[75] This study is based on two detoxification enzymes, ALDH2 (aldehyde dehydrogenase 2), that catalyzes the oxidation of acetaldehyde to acetic acid and other reactive aldehydes, and GSTP1 (glutathione S-transferase pi 1), an enzyme for xenobiotic elimination through conjugation reactions that detoxifies cigarette-derived xenobiotics. ALDH2 and GSTP1 mRNA levels are measured in cells overexpressing or depleted for SET. When SET is overexpressed both ALDH2 and GSTP1 levels are downregulated. Treatment with trichostatin A (TSA), an inhibitor of HDAC1/2 causing global chromatin decondensation, even increased the phenotype.[75] Ethanol and cisplatin are classical compouds used to assess ALDH2 and GSTP1 activities. Ethanol treatment upregulates both ALDH2 and GSTP1, whereas cisplatin treatment upregulates only GSTP1. SET seems to act as a transcription repressor on ALDH2 and GSTP1 promoter regions in absence of ethanol or cisplatin.[75] At the same time SET overexpression leads to downregulation of ATM, BRCA1 and Chk2 causing

DDR impairment. This is leading to DNA damage accumulation in the cells when treated with cisplatin. Cell overexpressing SET exhibit also a G_2/M stalling instead of a G_0/G_1 arrest when treated with cisplatin.[75]

j. SET is overexpressed in various types of cancer

As described previously SET has been reported to be implicated in cancer, notably, as a translocation partner with the Can/NUP214 gene in acute undifferentiated leukemia.[42] Since this first implication of SET with cancer many more have been described. In the last years numerous studies came out showing that SET is overexpressed in various types of cancer. Indeed SET proteins levels are upregulated in serous epithelial ovarian cancer[76], human colorectal adenocarcinoma[77], oral squamous cell carcinoma[77], B-cell chronic lymphocytic leukemia[78], non-Hodgkin lymphoma[78], hepatocellular carcinoma[79], acute myeloid leukemia[80] and head and neck squamous cell carcinoma[81]. In some of these cancers the level of SET overexpression could even be correlated with disease severity and agressivity.[78,80]

SET might be used as a therapeutic target. One strategy is to target the inhibition activity of SET to PP2A and NM23-H1 that are respectively tumor and metastasis suppressors. COG112 is a peptide interacting with SET that disrupt its binding with PP2A and NM23-H1, thus suppressing its inhibitory effect on these proteins.[82]

B. THESIS OBJECTIVES

The goal of my thesis entitled “Chromatin structure and DNA repair” is to identify novel chromatin related proteins involved in DNA damage response and repair. To achieve this goal, we utilized two different strategies:

A. First we attempted a direct purification of DNA repair foci using proteomics of isolated chromatin segments (PICh) that allows purification of proteins associated with a specific genomic locus. Due to technical limitations this project didn't give any convincing results.

B. We performed a siRNA screen using a library with chromatin modifiers, remodelles and associated proteins. The goal of this part was to identify novel factors that their dowregulation leads to persistent DNA damage, suggesting a role in DNA repair. One of the hits of the screen is the nuclear oncogene SET and our goal was to address the role of SET in the process.

Below I describe the two different objectives in a format of two separate research manuscripts for training purposes.

C. RESULTS

Identification of the composition of DNA repair foci by proteomics of isolated chromatin segments (PICh)

Anne-Sophie Hoffbeck[1], Sara Milosevic[1], Bernardo Reina[1], Jérôme Déjardin[2] and Evi Soutoglou[1]

[1]Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), UMR 7104 CNRS, UdS, INSERM U964, BP 10142, F-67404 Illkirch Cedex, CU de Strasbourg, France

[2]INSERM AVENIR Team, Institute of Human Genetics, CNRS UPR 1142, Montpellier, France

The project had technical limitations and had to be discontinued

Abstract

Our genome is constantly assaulted by various DNA damaging agents (from exogenous or endogenous sources) and must be efficiently repaired to maintain its integrity. The most deleterious and difficult to repair lesions are the double strand breaks (DSBs). Inefficient repair of DSBs can lead to aberrations that may induce cancer. DSBs are repaired through, Non-homologous end joining (NHEJ) and homologous recombination (HR). The DNA damage response is a complex cascade of events that includes sensing of DNA damage followed by activation and recruitment of a series of proteins. At the end this cascade leads to activation of cell-cycle checkpoint kinases, which stall damaged cells in their cell cycle until the lesion is repaired at the broken DNA. If unrepaired DSBs persist, the DDR activates apoptotic pathways or induces cell senescence. A huge amount of DDR proteins are present around the DSB and spread among the DNA, they are visible in the cell as DNA damage foci. Our objective is to identify unknown proteins involved in the DDR that are present at DNA repair foci. We utilized a technique called proteomics of isolated chromatin segments (PICh) that permits purification of proteins associated with specific genomic loci. The PICh technique was applied to a cellular system developed in our lab in which we can induce a DSB at a specific genomic site. Unfortunately, we have faced many technical limitations that made us conclude that this method might not be the most suitable for our purpose.

Introduction

Double strand breaks (DSBs) are hazardous because interaction between DNA ends from different DSBs can produce tumorigenic chromosome translocations. DSBs are repaired by two major pathways, homologous recombination (HR) and non-homologous end joining (NHEJ).[8–10]

Cells response to DSBs doesn't only involve the repair machinery; double strand breaks trigger a complex cascade called DNA damage response (DDR). DDR implies sensing of DNA damage followed by activation of phosphatidylinositol 3-kinases ATM/DNAPK/ATR (ataxia telangiectasia mutated/DNA-dependent protein kinase/ataxia telangiectasia and Rad3-related) and recruitment of a series of proteins as mediator proteins MDC1 and 53BP1. This cascade leads to activation of cell-cycle checkpoint kinases, which pauses the cell cycle until the lesion is repaired.[7–10]

The DDR manifests itself morphologically in the form of DNA repair foci, structures formed by the recruitment and accumulation of DNA repair factors at a site of DNA damage. As a consequence, a single double stranded break nucleates a macroscopically discernible focus, which spreads up to one megabase in the surrounding chromatin.[7,8] Although today we know several factors that are recruited to DSBs, the exact composition of the repair foci is rather unknown.

Biochemical fractionation of γ-irradiation induced repair foci has been technically very challenging. Evi Soutoglou's laboratory has developed a cell system in which a DSB can specifically be induced at a defined genomic site. NIH 3T3 stable cell lines (NIH 2/4) containing a copy of the unique 18 nucleotide I-SceI restriction site that does not exist in the mammalian genome, flanked by an array of 256 copies of the

lac-repressor binding site (lac operator: LacO) and by 96 copies of the tetracycline response element (TetO) have been generated[83] (fig. 1A). Expression of I-SceI induces a unique DSB in the genome. This system offers the opportunity for probing the repair foci composition and purification of novel factors that interact with DSBs because these can be induced in a specific genomic locus of known sequence.

In this study, we have used an unbiased approach, the Proteomics of Isolated Chromatin Segments (PICh) method[84] (fig. 1C), to identify proteins that accumulate into nuclear foci in response to DSBs. Unfortunately, we have faced many technical limitations that made us conclude that this method might not be the most suitable for our purpose. Alternative methods that can be designed to answer our question using our unique experimental system are discussed.

Results

PICh is based on formaldehyde crosslinking of interacting proteins bound to DNA, followed by the precipitation of specific genomic loci using desthiobiotin-tagged oligonucleotides containing locked nucleic acids (LNAs) (fig. 1B). The DNA/probe hybrids are isolated using magnetically labeled avidin beads. The crosslink is then reversed and co-precipitating proteins are resolved by SDS-PAGE and identified by mass spectrometry (MS) (fig. 1C). This method is suitable for analyzing abundant genomic loci (i.e. telomeres, centromeres, lac repeats etc) and it is not readily applicable to the analysis of single copy loci.

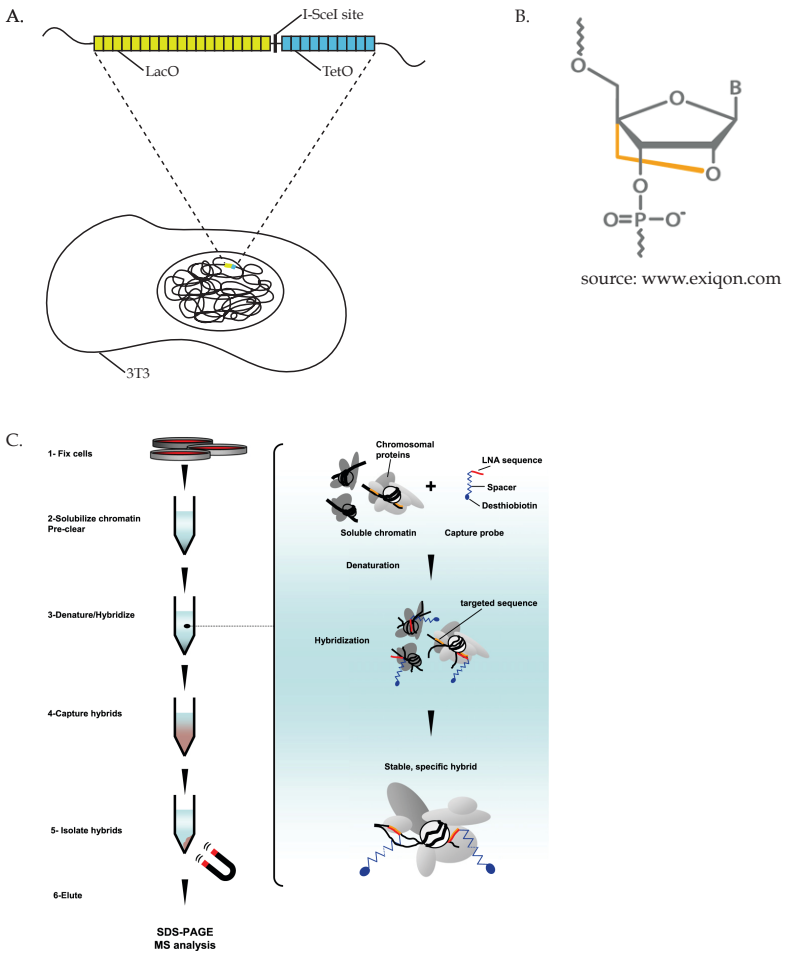

Figure 1: LacO/TetO system in 3T3 and PICh outline.[84]

(A) LacTet system in the 3T3 cell line. An array composed of an I-SceI site flanked by 256 LacO sequence repeats on one side and 96 tetO sequence repeats on the other side is stably integrated in a NIH 3T3 cell line. (B) The structure of LNA™. The ribose ring is connected by a methylene bridge (orange) between the 2′-O and 4′-C atoms thus "locking" the ribose ring in the ideal conformation for Watson-Crick binding. When incorporated into a DNA or RNA oligonucleotide LNA™ makes the pairing with a complementary nucleotide strand more rapid and increases the stability of the resulting duplex. (C) Outline of PICh protocol. Cells are crosslinked and soluble chromatin is extracted. Desthiobiotin tagged LNA probes are hybridized to DNA and stable specific DNA/proteins hybrids are captured. Proteins are eluted, migrated on SDS-PAGE gel and identified via mass spectrometry.[84]

Based on the published PICh protocol[84], each individual experiment requires the production of $1.8x10^{10}$ cells (18 liters at $1x10^{6}$ cells/mL). To obtain this number, we need to grow cells in suspension. Since our original cell line (NIH 2/4) didn't grow in suspension, we generated HeLa cell lines where the LacO/I-SceI array is stably integrated in the genome (HeLa 111) (fig. 2A left and upper right panels). Southern blot analysis revealed that HeLa 111 contain 10-20 tandem copies of the LacO-I-SceI unit (fig. 2A right bottom panel).

Since transient transfection in the mount of cells referred above is not feasible, we needed to establish a system to inflict DSBs in an inducible manner. To this end, we generated stable cell lines containing the array and expressing the I-SceI under the p-tight doxycycline induced promoter (HeLa 111 p-tight) (fig. 2B). In this way, I-SceI is only expressed and has access to the array after the addition of doxycycline (Dox) (fig. 2B, C).

To determine the time after doxycycline addition where the majority of DSBs are induced, we performed a time course (fig. 2D). To measure the cutting efficiency, we

performed an immunoFISH and counted the colocalization between the array and γH2AX foci (fig. 2C). We observed that the cutting efficiency was highest 12h after Dox addition (fig. 2D).

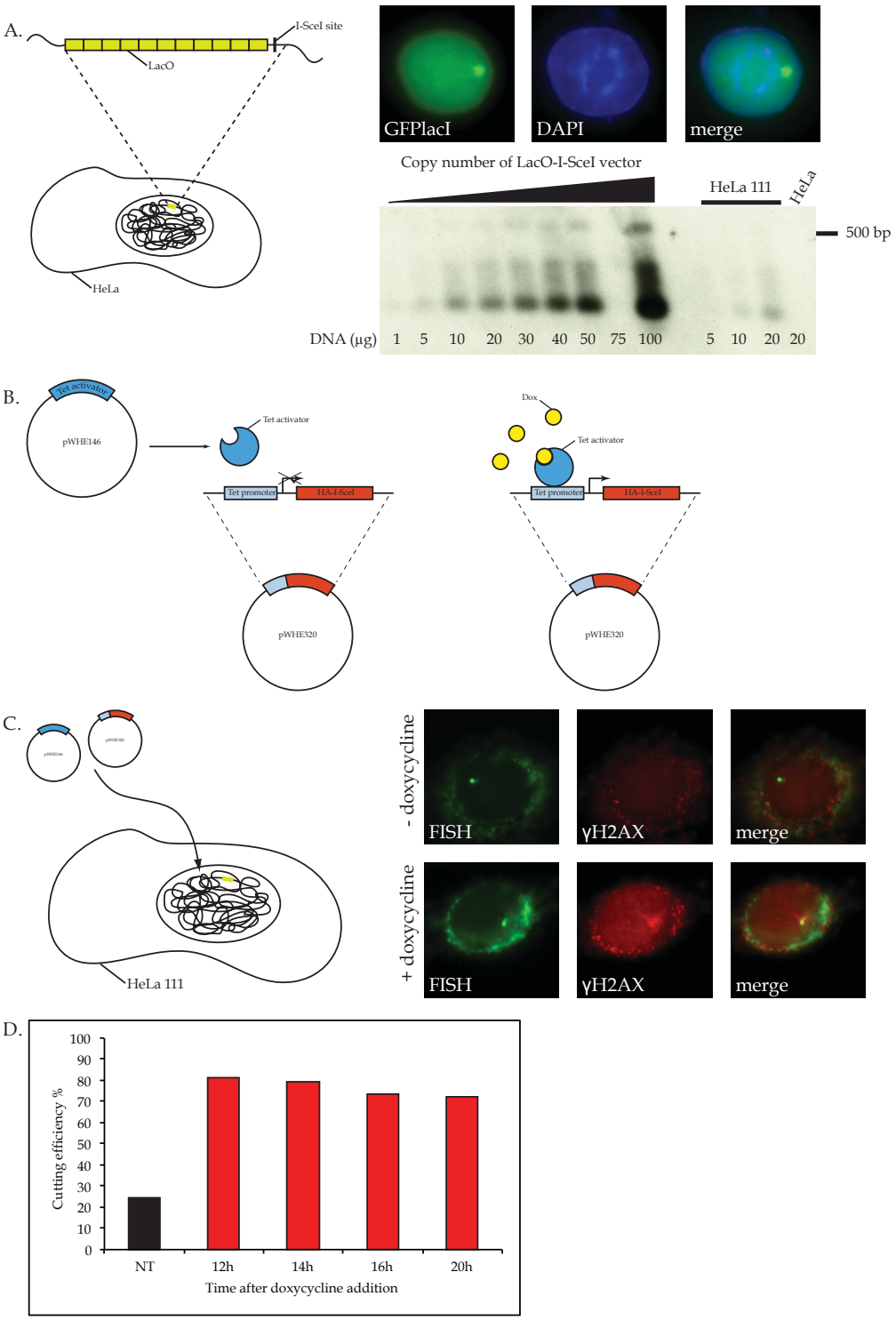

Figure 2: Establishment of the HeLa 111 p-tight cell line.

(A) Left panel: Lac system in the HeLa cell line. An array composed of 256 repeats of the LacO sequence flanked by an I-SceI restriction site is tandemly inserted in the genome of a HeLa cell line. Right upper panel: If the cells are transfected with GFP fused to the lac repressor lacI (GFPlacI), the array is visible as green focus. Right lower panel: The number of tandem integrations of the array is determined via Southern blot. A calibration range is done thanks to the LacO-I-SceI vector used to create the cell line and the number of repeats is determined with HeLa 111 genomic DNA. (B) p-tight system. The system is composed of two vectors. A helping vector named pWHE146 codes for an activator and an other vector (pWHE320) contains the I-SceI gene preceded by a Tet promoter. In presence of doxycycline (Dox) the activator gets functional, bind to the Tet promoter sequence in order to allow I-SceI expression. (C) Left panel: The p-tight system is stably integrated into the genome of the HeLa 111 cell line (HeLa 111 p-tight). Right panel: When doxycycline is added to HeLa 111 p-tight, I-SceI cuts the array. The array is visible here by FISH and γH2AX accumulates on the array when Dox is added. (D) The cutting efficiency of the array is determined by counting colocalizations of the array visualized by FISH and γH2AX foci. The cutting efficiency has been determined 12, 14, 16 and 20 h after Dox addition.

The LacO-I-SceI array is composed of 256 repeats of the lac operator sequence (292bp) followed by an I-SceI site. The LacO repeats are organized in 32 concatamers of a cassette that contains 8 LacO repeats (fig. 3A). We designed two LNA probes; one that is overlapping between two LacO repeats (LNA probe 2) and one that is not (LNA probe 1) (fig. 3B). Each probe is composed of 25 nucleotides where one third of which are locked nucleic acids (LNA). They are designed to have a low secondary structure and their melting temperature is comprised between 85 and 92°C.

To assess the specificity of these probes for the LacO array, we tested them by FISH (fig. 3C). We used as a positive control the telomeric probe provided by the group of Jérôme Déjardin. Although, we didn't detect specific FISH signal that corresponds to the lac array, we decided to undertake the PICh and test directly whether our probes were specific.

We performed the PICh protocol in the absence of doxycycline (no breaks) and after 12 hours of doxycycline addition (breaks). In this way we should be able to distinguish the proteins recruited to the broken array from the ones that are normally bound to it. We then separated the precipitated proteins by SDS page electrophoresis (fig. 3D).

As observed in figure 3D the background was very high and we couldn't distinguish differences between the – dox and + dox conditions (fig. 3D). Nevertheless, to further test whether we have precipitated proteins related to DNA repair foci when the break is induced, we performed Western blot analysis for γH2AX or 53BP1. Regrettably no signal was detected for any of these DDR markers (fig. 3E). These results suggest that our probes are not specific.

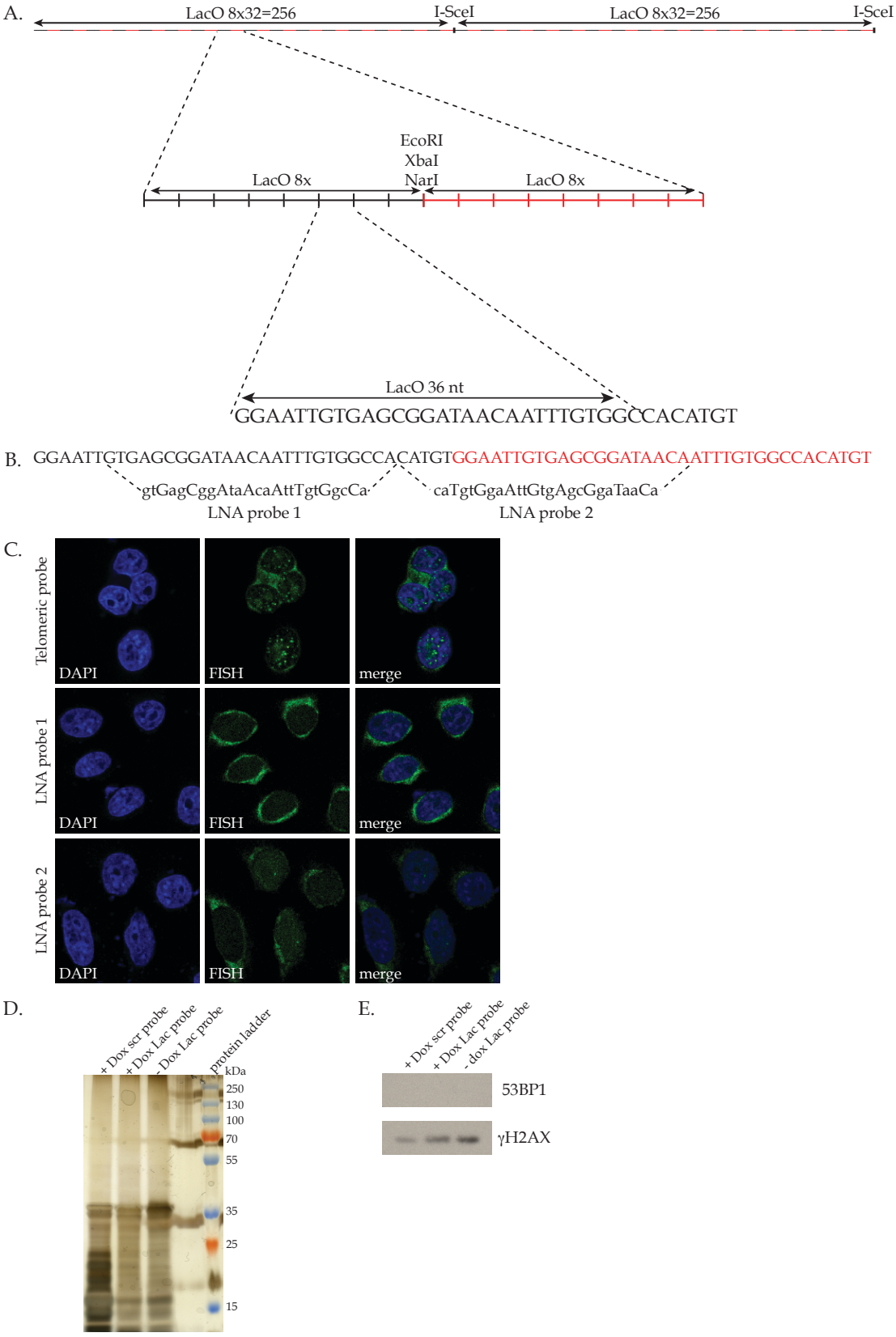

Figure 3: LacO-I-SceI array, LNA probes and PICh.

(A) The LacO-I-SceI array is composed of 256 repeats of the lac operator sequence (292bp) followed by an I-SceI site. The LacO repeats are organized in 32 concatamers of a 8 repeats cassette separated by restriction sites (EcoRI, XbaI and NarI). Each LacO repeat is 36 nucleotides long. (B) LNA probe 1 is designed not overlapping over repeats and LNA probe 2 is designed overlapping over two LacO repeats. (C.) FISH is performed with LNA probe 1, LNA probe 2 and a LNA telomeric probe provided by Jérome Déjardin (control) on HeLa 111 to verify their specificity. (D) PICh is performed in three different conditions: with doxycycline (Dox) addition but with a probe of scrambled sequence (scr probe), with Dox and with LNA probe 1 and 2 (Lac probe) and finally without doxycycline and with LNA probe 1 and 2. Eluted proteins from PICh are separated by SDS-PAGE electrophoresis and silver stained. (E) Eluted proteins from PICh are immunobloted against γH2AX and 53BP1.

We subsequently repeated the PICh procedure with newly designed probes (LNA probes 3, 4 and 5) (fig. 4A). Surprisingly, although these probes were more specific and gave a clear signal at a FISH experiment (fig. 4B), the pull down was extremely inefficient as no protein was detected on the silver stained SDS-PAGE gel (data not shown). In line with these observation Western blot analysis didn't reveal any specific recruitment of DDR factors at the LacO array after DSB induction (data not shown).

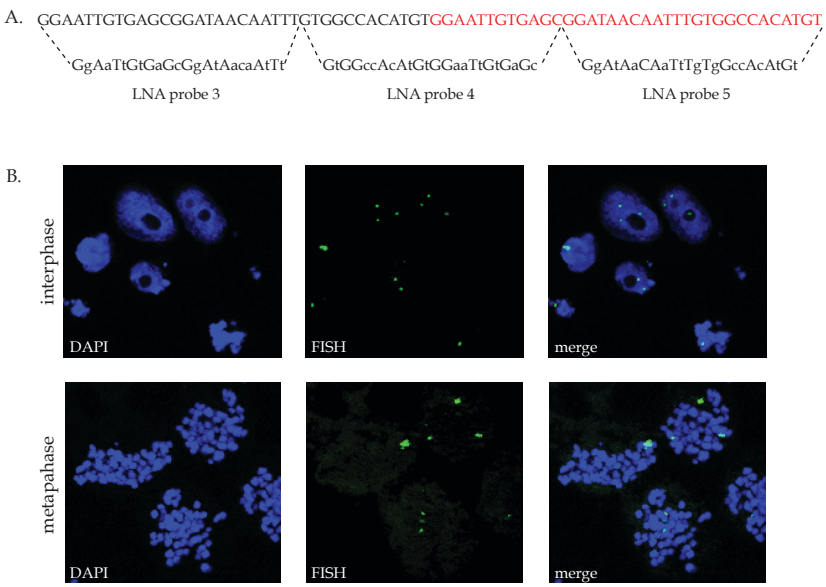

Figure 4: Specific LNA probes.

(A) LNA probes 3, 4 and 5 are designed covering the sequence of two full LacO repeats. (B) FISH is performed with a mix of LNA probes 3, 4 and 5 to verify their specificity.

Discussion

Here, we tried to apply an unbiased approach to identify novel proteins that are recruited to a DNA repair foci. In spite of our several attempts, the PICh technique doesn't seem to be the appropriate technique to achieve this goal.

One explanation could be that the number of repeated sequences at the LacO locus is not high enough. In our cell line the LacO array is tandemly integrated around 20 times and the Lac operator is thus present approximately 5120 times at this locus. This number of repetitions might be not sufficient to have a good pull down of the proteins. If we compare our system to the initial PICh that was performed with telomeric sequences, we note that telomere repeated sequences are present in numerous different positions in the cells, unlike our array that is located at a single locus. Therefore, even if the accumulation of the DDR proteins at a DSB is immense, it happens only at one unique place in the cell with this system.

As the lac array is composed of repetitive sequences, one could think that it might be heterochromatinized, consequently preventing the probe access. The chromatin status of the array is not determined yet for this cell line, but as the original PICh has been performed in telomeres, known to be hetrochromatinized, this shouldn't be an issue. The chromosomal localization of the array might be also a line to follow, to understand why the precipitation of the proteins is not efficient.

Finally, even if we tried to design the probes as accurate as possible and that we could determine their proper interaction by FISH, the probes we chose might not hybridize stably enough to pull down the DNA/proteins hybrids.

To conclude, we haven't determined yet why the PICh technique is not appropriate to pull down proteins located at the LacO array. One of the above hypotheses could be true as well as a combination of them.

Perspectives

We will develop an alternative approach to isolate proteins that are recruited to the LacO array upon DNA damage. This project will be conducted in collaboration with the laboratory of Bernardo Reina and it will be part of the thesis project of Léa Gaudot, a student in his lab. We will use a new technique for proximity-dependent labeling of proteins in eukaryotic cells called BioID[85].

In this assay, we will fuse the Lac repressor (LacI) to a promiscuous *Escherichia coli* biotin protein ligase (BirA*). We will generate stable HeLa 111 cells expressing LacI-BirA* and tethering of BirA* at the LacO array will result in non-biased proximity-dependent biotinylation of all proteins that are near-neighbors of the fusion protein.

Proteins will be extracted from cells treated or not with Dox and biotinylated proteins will be isolated by affinity capture and identified by mass spectrometry. Furthermore, to obtain a comparative and quantitative profile of the proteins present in the different conditions, we will use the Stable Isotope Labeling by Amino acids in Cell Culture (SILAC) procedure. To this end, we will label the cells with different amino acid isotopes and compare samples with and without Dox addition.

The approach is relatively straightforward; it doesn't rely in protein crosslinking or DNA hybridization and thus promises to yield good results.

Over expression of the nuclear oncogene SET impedes replication stress signaling and Homologous Recombination by KAP1 and HP1 retention to chromatin

Anne-Sophie Hoffbeck[1*], Alkmini Kalousi[1*], Jordan Pinder[2], Benoît Fisher [1], Laurent Brino [1], Thomais Papamarcaki[3], Kum Kum Khana[4], Graham Dellaire[2] and Evi Soutoglou[1#]

[1]Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), UMR 7104 CNRS, UdS, INSERM U964, BP 10142, F-67404 Illkirch Cedex, CU de Strasbourg, France

[2]Departments of Pathology and Biochemistry & Molecular Biology, Dalhousie University, Halifax, NS, Canada; Beatrice Hunter Cancer Research Institute, Halifax, NS, Canada.

[3]Laboratory of Biological Chemistry, Medical School, University of Ioannina, 451 10 Ioannina, Greece

[4]Queensland Institute of Medical Research, Herston, Australia

Corresponding author

* Equal contribution

Manuscript in preparation

Abstract

Cells continuously experience stress and damage from exogenous sources, such as UV light or irradiation, and endogenous sources, such as oxidative by-products of cellular metabolism that endanger genome stability. Several pathways have been evolved to detect DNA damage, signal its presence and mediate its repair. Double Strand Breaks (DSBs) are the most harmful DNA breaks because their unfaithful repair can lead to chromosomal translocations. Cells respond to DSBs by initiating a signaling cascade called DNA damage response (DDR) that activates checkpoints and promotes repair of the broken ends. Although, there has been much progress in identifying key factors of DDR and DNA repair, only recent cell -biological approaches have started to reveal how they function in the context of local high-ordered chromatin structure. Here we performed a siRNA screen to identify chromatin related proteins that are involved in repair of DSBs. We indentified SET as a moderator of DDR in chromatin surrounding DSBs. We show that depletion of SET increases DDR and survival in the presence of radiomimetic drugs. On the other hand, we find that that overexpression of SET impairs the DNA damage response and cell cycle progression upon replication stress, and the DNA repair by HR. Moreover cells with high levels of SET are hypersentitive to agents that cause replication stress. All the above functions of SET depend on its NAP domain. Furthermore we find that SET interacts with the corepressor KAP1 and its overexpression results to the sustained retention of KAP1 and HP1 in chromatin. Our results suggest a model in which this retention triggers an inhibition of resection impairing HR regulation.

Introduction

Various types of agents from either exogenous or endogenous sources constantly assault DNA[8]. DNA Double strand breaks (DSBs) are the less frequent but more toxic lesions because interaction between DNA ends from different DSBs can produce tumorigenic chromosome translocations[9]. DSBs are repaired by two main pathways: Non Homologous End Joining (NHEJ) and Homologous Recombination (HR)[10]. NHEJ is used by cells to join broken ends by simple re-ligation and although this pathway is active throughout the cell cycle, it mainly occurs during G1[36]. The NHEJ pathway is intrinsically error prone and can drive chromosome translocations by joining distal DSBs from different parts of the genome[9]. HR functions predominantly when pairing of sister chromatids occurs during S/G2 and takes advantage of the information encoded by the homologous template to eliminate the DSB in an error-free manner[39]. During HR, DNA is processed to generate single stranded ends that are covered by RPA and later exchanged to RAD51. These nucleoprotein filaments are then prone to invade the homologous sequence so that subsequent repair can take place [39].

One of the most common DNA damage is replication stress, which results in the blockage of a replication fork. When a replication fork meets an obstacle it gets stalled[11]. This will result in a long stretch of ssDNA that will be coated by RPA[18]. This accumulation of RPA will trigger the recruitment of different proteins like ATR that will in turn phosphorylate Chk1 to block the cell cycle[22]. If the obstacle cannot be bypassed, the fork will collapse and this will result in a DSB that will be repaired by HR[18].

Cells' response to DSBs does not only involve the repair machinery; double strand breaks trigger a complex cascade called DNA damage response (DDR)[9,10]. DDR implies sensing of DNA damage followed by activation of phosphatidylinositol 3-kinases ATM/DNAPK/ATR (ataxia telangiectasia mutated/DNA-dependent protein kinase/ataxia telangiectasia and Rad3-related) and recruitment of a series of mediator proteins such as MDC1 and 53BP1[9,10]. This cascade leads to activation of cell-cycle checkpoint kinases, which pauses the cell cycle until the lesion is repaired[22].

All types of DNA repair occur in the context of highly structured chromatin[9,23,25]. Emerging evidence suggests that the ability of repair factors to detect DNA lesions is determined by histone modifications around the lesion and involves chromatin-remodeling events[9,23,25]. The most prominent DNA-damage-induced histone modification in double strand break repair (DSBR) is the phosphorylation of the C-terminal tail of H2AX, referred to as γH2AX[26]. Several other histone modifications, including acetylation, methylation, phosphorylation and ubiquitination of most core histones, have been linked to various aspects of DNA repair by facilitating access to the chromatin fiber, recruiting repair factors and chromatin-remodeling components, and propagating cell-cycle checkpoint signals[23]. Although there is increasing evidence that chromatin alterations are essential for efficient DSBR, the complete picture about the mechanisms ruling these chromatin changes are far from being understood.

Here we performed a siRNA screen to identify chromatin related proteins that are involved in repair of DSBs. We indentified SET as a moderator of DDR in chromatin surrounding DSBs. We show that depletion of SET increases DDR and

survival in the presence of radiomimetic drugs. On the other hand, we find that SET overexpression impairs replication stress signaling, DNA breaks processing and consequent repair by Homologous Recombination. SET interacts with the corepressor KAP1 and its overexpression results to the sustained retention of KAP1 and HP1 in chromatin. Our results suggest a model in which this retention triggers an inhibition of resection impairing HR regulation.

Results

The nuclear oncogene SET is a modulator of DDR in chromatin surrounding DSBs.

To evaluate the impact of high ordered chromatin structure in DNA damage response and DNA repair in an unbiased way, we performed a siRNA screen using a library with chromatin related proteins and their interaction partners (fig. 1A). HeLa cells were transfected in 96 well plates with a library consisting of 1184 siRNAs and they were subsequently challenged with the radiomimetic drug Neocarzinostatin (NCS) (fig. 1A). 16hours after damage, cells were fixed and immunostained for γH2AX and 53BP1 (fig. 1A). As positive control, a siRNA targeting XRCC4 was used since it is known that downregulation of XRCC4 leads to persistent DNA damage[86]. A non-targeting siRNA was used as negative control. Among other hits, we identified the nuclear oncogene SET as a modulator of DDR and downregulation of SET resulted in significant increase in γH2AX intensity and 53BP1 foci formation, similar to that observed in cells depleted for XRCC4 (fig. 1B). Interestingly, depletion of SET by 4 different siRNAs validated the phenotype observed at the primary screen (fig. 1C).

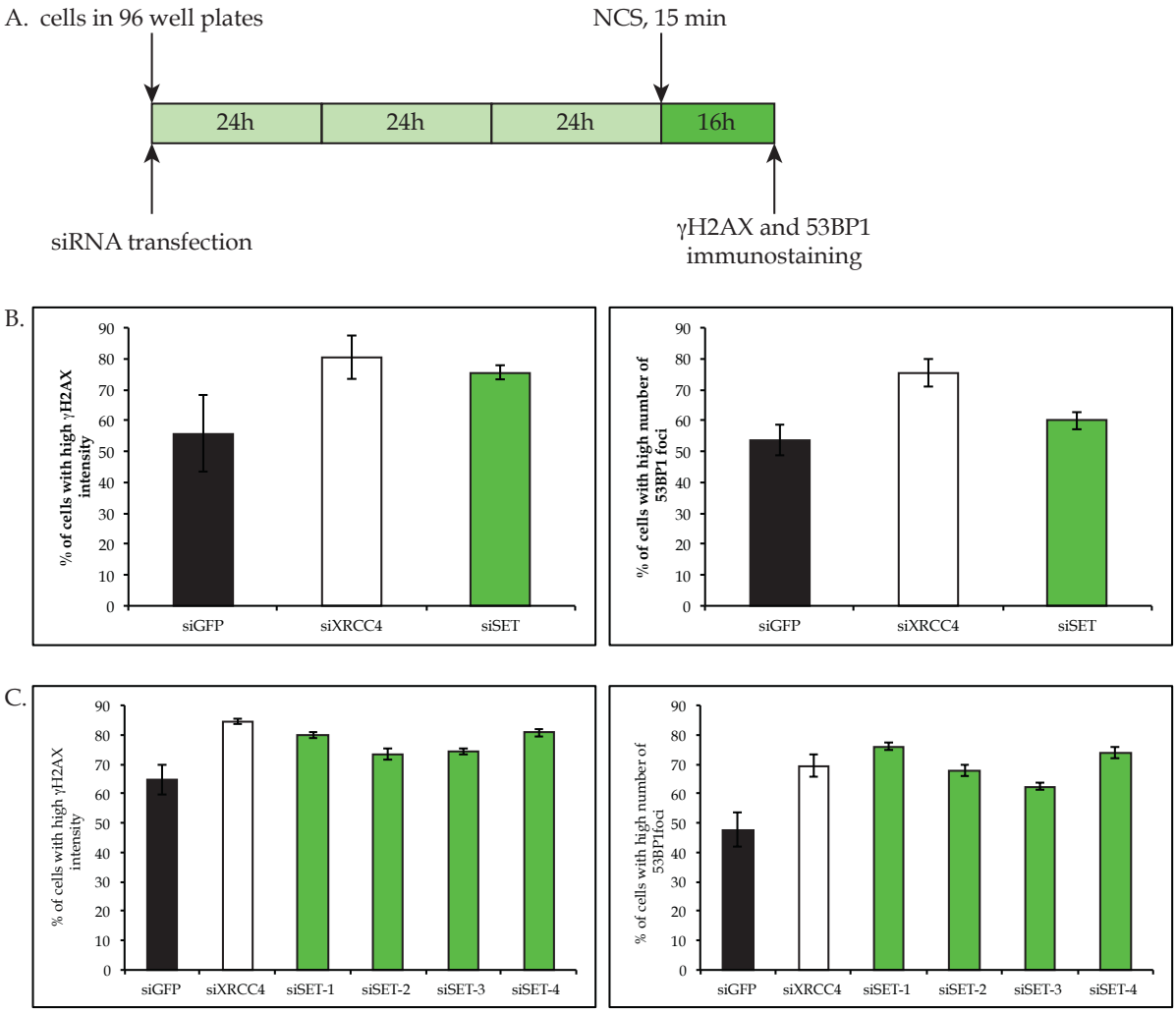

Figure 1: SET identification through a siRNA screen

(A) Outline of siRNA screen experiment. HeLa cells are transfected with siRNAs in 96 well plates. 72 h after transfection cells are treated with 50 ng/mL NCS for 15 min before being released for 16 h. Cells are then fixed and immunostained for γH2AX and 53BP1. (B) HeLa cells are transfected with siGFP, siXRCC4 or siSET (pool) 72 h after transfection cells are treated with 50 ng/mL NCS for 15 min before being released for 16 h. Cells are then fixed and immunostained for γH2AX and 53BP1. Level of γH2AX intensity (left histogram) or number of 53BP1 foci (right histogram) are measured. siSET-1, 2, 3 and 4 are four different siRNAs targeting SET. (C) Same as (B) but with individual siRNAs for SET (siSET-1, 2, 3 and 4).

To investigate whether SET has a direct role in DDR, and to exclude the possibility that the phenotype described above is due to global effect in gene expression that might occur upon SET depletion, we tested the recruitment of SET fused to GFP in laser-induced breaks. Notably, SET was recruited to lesions instantly upon laser induction (fig. 2A, second panel) and it was released immediately after, pointing to a transient association of SET with DNA breaks (fig. 2A, second panel). Interestingly, a deletion mutant of SET that lacks the NAP domain failed to accumulate in laser induced lesions (fig. 2A, third panel) suggesting a NAP domain dependent association of SET with DNA breaks. The proper induction of DNA damage by the 405 laser after Hoechst 33342 sensitization was monitored by the recruitment of mRuby2-Ku80 (fig. 2A bottom panel).

To validate the screen results in another cell type, we performed Western blot analysis in control U2OS cells and cells depleted for SET. We indeed observed higher γH2AX levels in cells where SET is downregulated (fig. 2B). Interestingly, γH2AX was increased immediately upon DNA damage and was sustained higher until 8 hours after (fig. 2B). These results point to two different possibilities: First that depletion of SET results in higher number of DNA breaks pointing to a role of SET in facilitating repair of DNA lesions. Another possible explanation is that downregulation of SET results in higher accessibility of H2AX to DDR kinases due to altered chromatin conformation, without affecting the number of unrepaired DNA breaks.

To decide between the two different possibilities, we first measured the number of DNA lesions in control and SET depleted cells, by comet assays. Interestingly, we

do not observe a significant increase of breaks in cells lacking SET (fig. 2C).These results suggest that depletion of SET is not affecting repair efficiency and point to the second possibility.

To further investigate the role of SET in checkpoint activation, we tested the phosphorylation of Chk1 and Chk2 in controls cells and SET depleted cells. Interestingly, depletion of SET results does not alter check-point proteins' phosphorylation (fig. 2B) further strengthening the point that SET depletion doesn't lead to persistent DNA damage.

To further evaluate the involvement of SET in genome integrity, we performed clonogenic survival assays in U2OS cells treated with the radiomimetic drug phleomycin. We observed that SET depleted cells survive better in the presence of DNA damage than control cells (fig. 2D), suggesting that downregulation of SET renders DDR more efficient. All the above results suggest that SET play an inhibitory role in DDR.

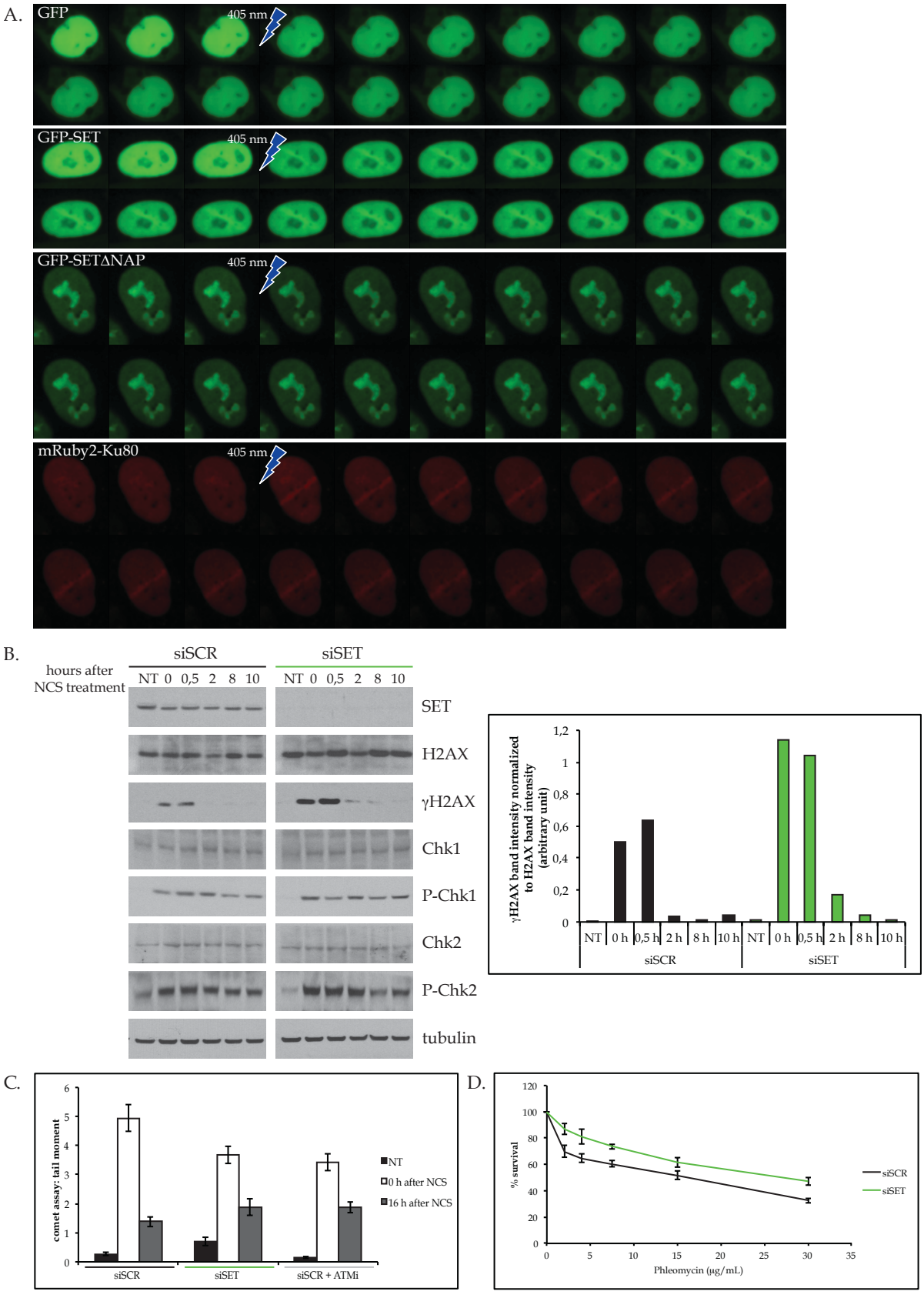

Figure 2: SET recruitment at laser induced breaks and effect of SET depletion on DDR and survival.

(A) Laser induced DNA damage. Live cell imaging is performed after laser irradiation of cells expressing either GFP, GFP-SET, GFP-SETΔNAP or mRuby2Ku80. Cells are pre-sensitized with Hoechst 33342 (2 μM in PBS for 5 min, then washed 3 times in PBS) and irradiated via a 405 nm laser. Time between each picture is 1 second. (B) Cells are transfected with siSCR or siSET (on-target-plus smart pool) for 72H before being treated with 50 ng/mL NCS for 15 and released for different time points. Immunoblotting is performed on total cell extract against SET, H2AX, γH2AX, Chk1, P-Chk1, Chk2, P-Chk2 and tubulin. Left graph: quantification of band intensity for γH2AX. (C) Comet assay. Cells are transfected with siSCR (negative control) or with siSET (on-target-plus smart pool) with or without addition of ATM inhibitor. The cells are then treated with NCS and released for 0 or 16h before being subjected to comet assay. (D) U2OS cells transfected with siSCR or siSET (on-target-plus smart pool) are treated with different concentrations of phleomycine for 1 h. Their survival is then measured through counting of colony formation.

SET overexpression impedes replication stress signaling and repair by HR

The above findings fit with a deleterious effect of SET in cases of its overexpression. Indeed, SET is found highly overexpressed in a variety of cancers[76–81]. To assess the impact of SET overexpression in genomic instability, we established a cellular model for SET overexpression in U2OS cells (fig. 3A, B). We generated U2OS stable cell lines overexpressing SET or a truncated form of SET lacking the NAP domain fused to GFP (GFP-SET, GFP-SETΔNAP). A U2OS stable cell line overexpressing GFP was used as negative control (fig. 3A). The level of overexpression of the SET protein is 3-4 times higher than the endogenous level of SET (fig. 3C) and the GFP-SET full length exerted similar nuclear localization with the endogenous protein (fig. 3D). Moreover, biochemical fractionation revealed that GFP-SET was partitioned to the different cellular compartments in a similar manner as the endogenous protein (fig. 3D). Surprisingly, although by Immunofluoresence, SET is localized exclusively in the nucleus (fig. 3E), the majority of the protein is

detected in the cytoplasm after biochemical fractionation (fig. 3D) suggesting that the protein is shuttling between the different cellular compartments.

To investigate the impact of SET overexpression in genomic instability, we performed clonogenic survival in U2OS GFP, GFP-SET and GFP-SETΔNAP cells in the presence of different radiomimetic drugs (fig. 3F, G). We observed that cells overexpressing SET are particularly sensitive to the replication stress agent camptothecin, which is an inhibitor of Topoisomerase I (fig. 3G). Interestingly, deletion of the NAP domain abrogates the phenotype (fig. 3G). Treatment of cells with camptothecin results in stalled and collapsed replication forks and thus DNA lesions exclusively in S cell cycle phase. Therefore, our results suggest that SET overexpression affects the DDR or repair of lesions occurring in S phase.

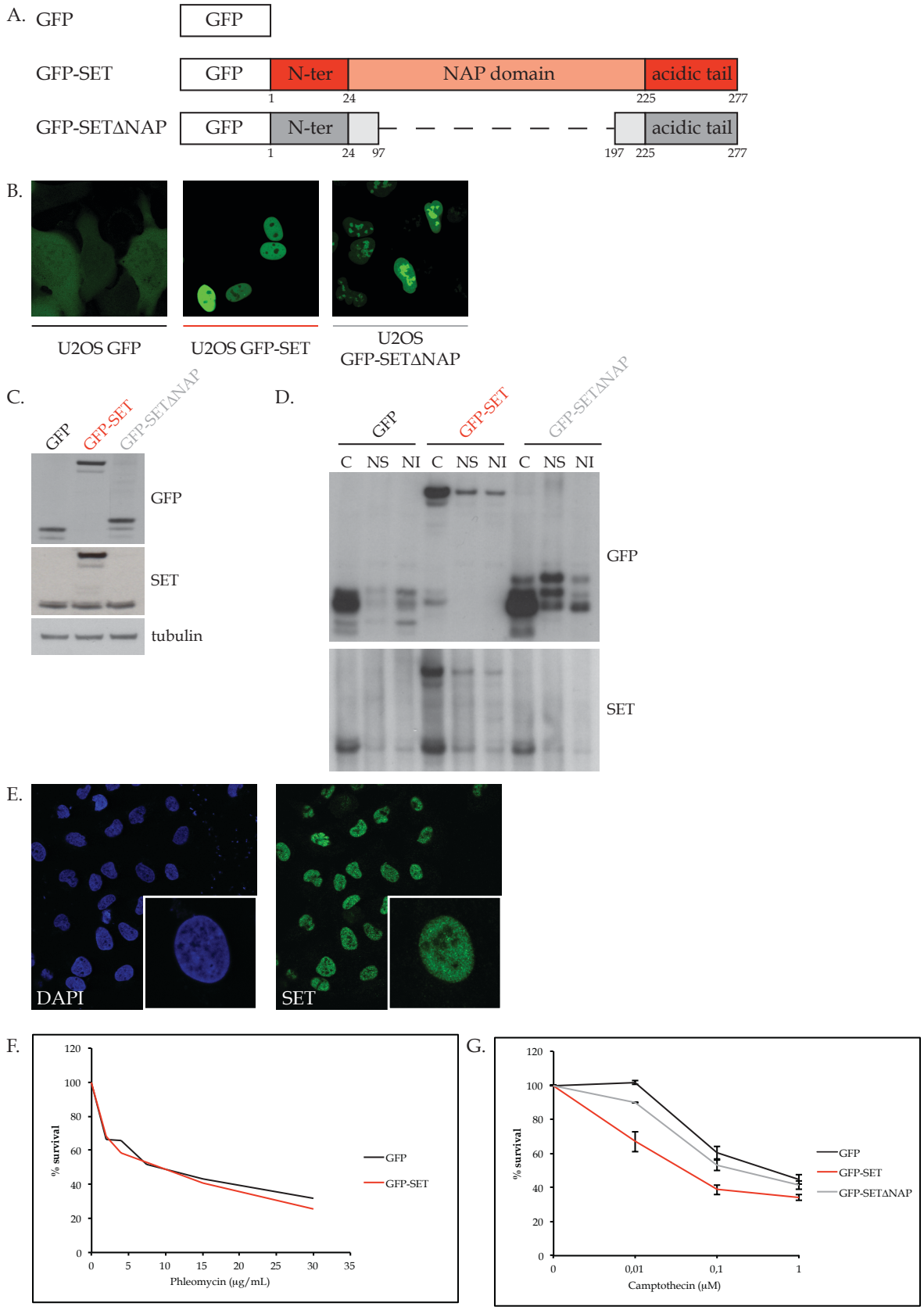

Figure 3: SET overexpression, effect on survival.

(A) Schematic representation of the fusion proteins stably overexpressed in U2OS cells (GFP, GFP-SET and GFP-SETΔNAP). (B) Stable cell lines are generated containing either GFP, GFP-SET or GFP-SETΔNAP. Cell lines are imaged through confocal microscopy. (C) Total cell extract is collected from U2OS GFP, GFP-SET and GFP-SETΔNAP and immublotting is performed against GFP, SET and tubulin. (D) Cytoplasmic (C), nuclear soluble (NS) and nuclear insoluble (NI) protein fractions are collected from U2OS GFP, GFP-SET and GFP-SETΔNAP and immunoblotted for GFP and SET. (E) Immunostaining of endogenous SET in U2OS cells. Pictures are taken via confocal microscopy. (F) U2OS GFP, GFP-SET and GFP-SETΔNAP cells are treated with different concentrations of phleomycine for 1 h. Their survival is then measured through counting of colony formation. (G) Same as (F) with treatment with different concentrations of camptothecin for 1 h.

In line with these results, cells overexpressing SET exerted slower progression through S phase compared to GFP cells (fig. 4A) when challenged with the replication stress agent hydroxyurea (HU), further suggesting a delay in repairing lesions occurring during S phase. To further analyze this phenomenon with a specific S phase marker, we performed BrdU staining after treatment with HU in U2OS cells overexpresing GFP-SET or GFP (fig. 4B). In agreement with the previous observation, we observed a dramatic increase of cells in S when SET is overexpressed (fig. 4B).

To investigate whether the delay in S-phase progression is due to a defect in replication stress signaling caused by SET overexpression, we assessed the levels of RPA phosphorylation after HU treatment in GFP-SET, GFP-SETΔNAP and GFP cells. Interestingly, RPA phosphorylation is delayed in U2OS GFP-SET compared to GFP-SETΔNAP and GFP cells (fig. 4C). Along the same lines, SET overexpression impedes the phosphorylation of H2AX and this phenotype is abolished upon deletion of the NAP domain (fig. 4C). These results suggest an inhibitory role of SET in replication stress signaling that is dependent on the NAP domain.

When stalled replication forks eventually collapse give rise to double strand breaks that are mainly repaired by homologous recombination (HR)[18]. To test whether the defect in replication stress signaling observed when SET is overexpressed is leading to defect in HR, we used the DR-GFP system (fig. 4D). This recombination assay relies on the two inactivated tandem repeated (DR)-GFP plasmid developed by M. Jasin[87], which contains two mutated GFP genes oriented as direct repeats and separated by a drug selection marker (fig. 4D). The upstream (5′) GFP gene (cassette I) carries a recognition site for I-SceI, a rare-cutting endonuclease that does not cleave several eukaryotic genomes tested. The downstream (3′) GFP (cassette II) is inactivated by upstream and downstream truncations, leaving only ~502 bp of GFP. Expression of the meganuclease I-SceI leads to formation of DSBs and GFP reconstitution is a marker of HR efficiency (fig. 4D). As expected, SET overexpression decreased the efficiency of HR (fig. 4E). This decrease was abrogated when the SETΔNAP protein was overexpressed. (fig. 4E).

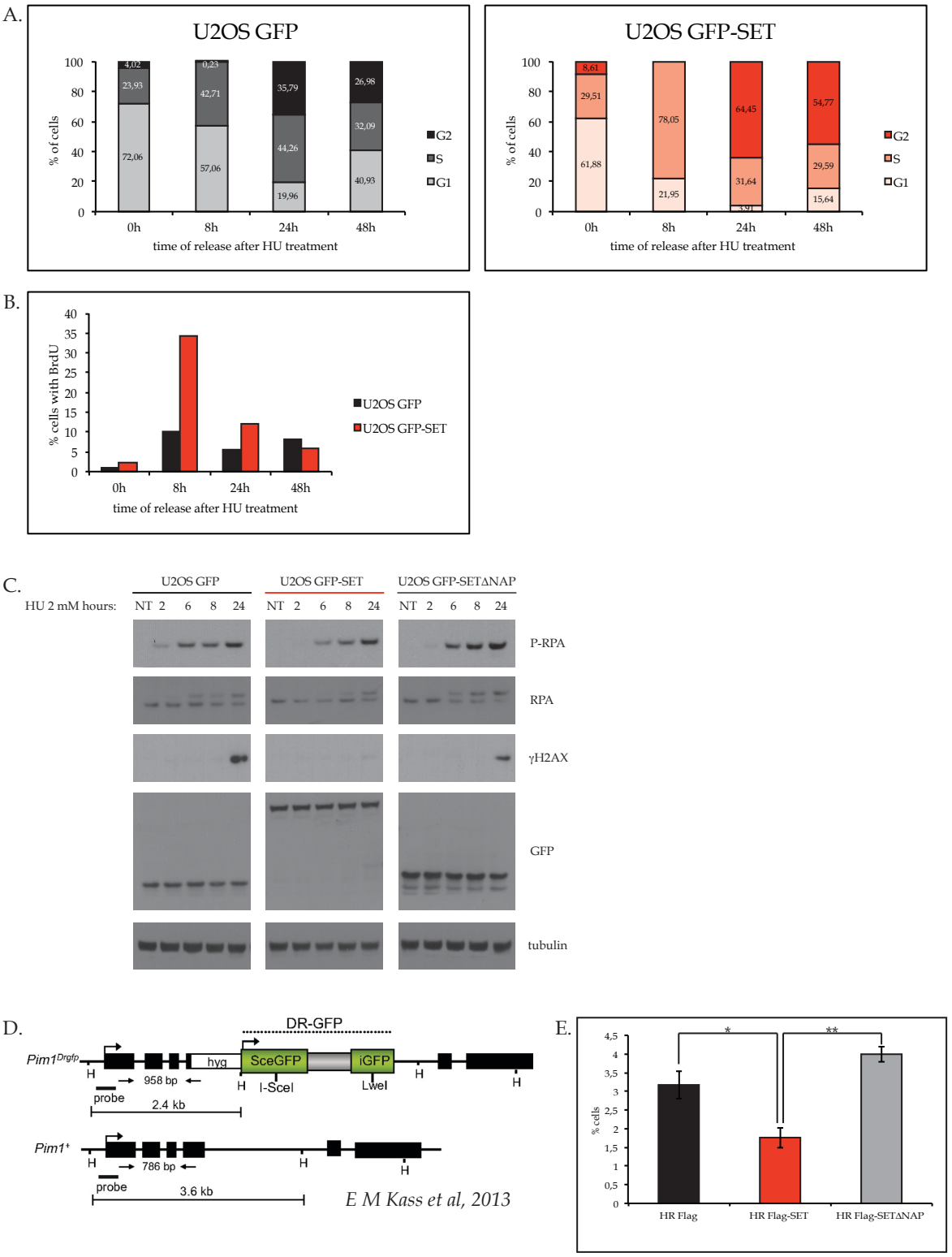

Figure 4: Repercussion of SET overexpression on cell cyle, replication stress signaling and HR.

(A) U2OS GFP (left graph) and GFP-SET (right graph) cell lines are treated with 2 mM hydroxyurea (HU) for 24 h and then released for 6, 8, 24 or 48 h before being

submitted to propidium iodide cell cycle analysis. (B) U2OS GFP (black bars) and GFP-SET (red bars) cell lines are treated with 2 mM hydroxyurea (HU) for 24 h and then released for 0, 8, 24 or 48 h before being submitted to BrdU cell cycle analysis. Bars display percentage of cells in S phase. (C) U2OS GFP, GFP-SET and GFP-SETΔNAP cells are treated with 2 mM HU for 2, 6, 8 and 24h before collecting protein extracts. Immunoblotting is performed against P-RPA, RPA, γH2AX, GFP and tubulin. (D) Schematic representation of GFP reconstitution array integrated in DR-GFP cells. The upstream (5′) GFP gene (cassette I) carries a recognition site for I-SceI, a rare-cutting endonuclease that does not cleave several eukaryotic genomes tested. The downstream (3′) GFP (cassette II) is inactivated by upstream and downstream truncations, leaving only ~502 bp of GFP. Expression of the meganuclease I-SceI leads to formation of DSBs and GFP reconstitution is a marker of HR efficiency. (E) DR-GFP cells are transfected with Flag, Flag-SET or Flag-SETΔNAP and number of cells repairing DSBs through homologous recombination are assessed.

SET overexpression or tethering to chromatin affects the recruitment of HR factors in DSBs

To get more insights into the mechanism that leads to the impairment of HR, we asked whether DNA repair proteins involved in HR are properly recruited at DSBs in SET overexpressing conditions. To tackle this question, we used a cellular system in which a single DSB can be created at a specific genomic site. This system consists of an array composed of 256 repeats of the lac operator flanked by an I-SceI restriction enzyme site stable integrated in U2OS cells. Visualization of the break is achieved by the expression of lac repressor (lacI) fused to GFP and expression of I-SceI induces a DSB that is exemplified by the early DDR marker γH2AX (fig. 5A, B). Interestingly, when SET is overexpressed the phosphorylation of RPA as well as the recruitment of BRCA1 and RAD51 at the LacO array upon DNA damage is impaired (fig. 5C). Moreover, the NAP domain seemed crucial for this phenotype as

overexpression of SETΔNAP did not impair P-RPA, BRCA1 and RAD51 recruitment at the LacO array (fig. 5C).

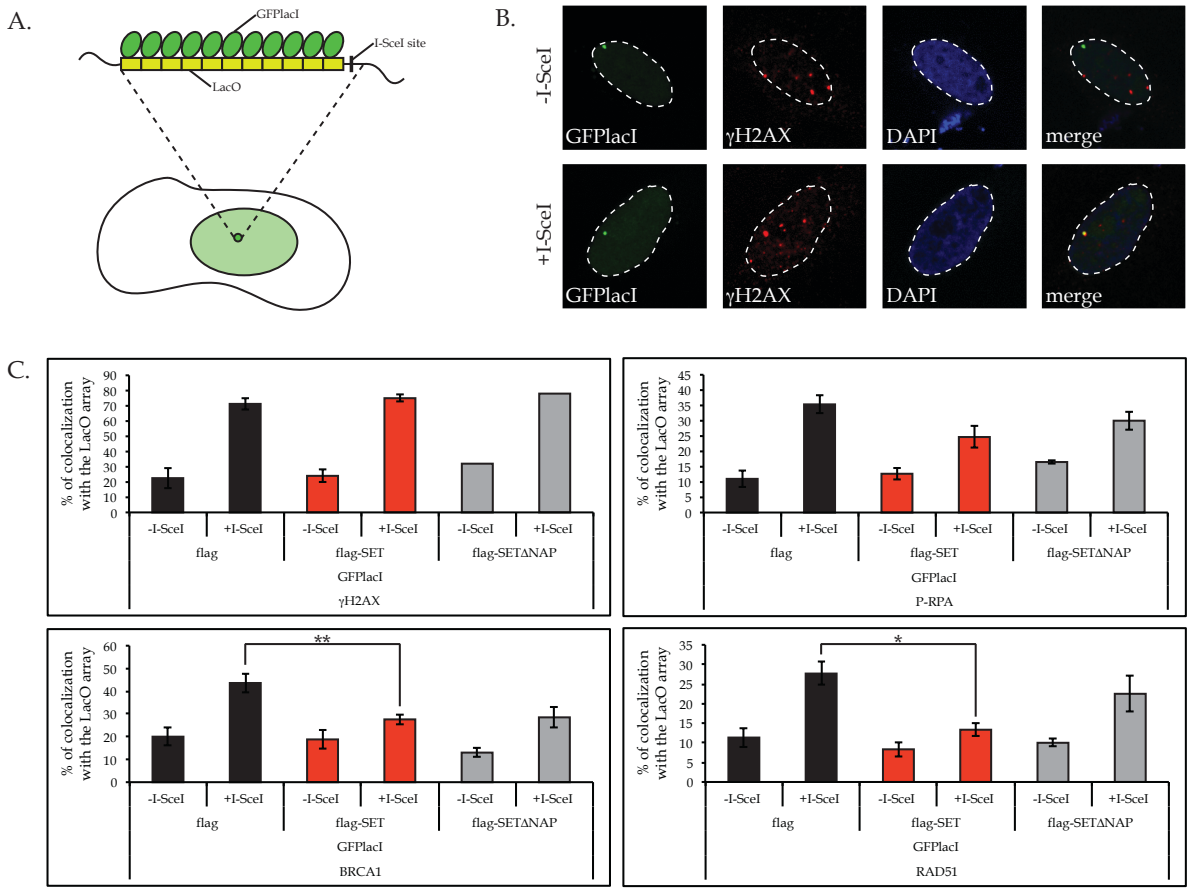

Figure 5: Effect of SET overexpression on recruitment of DNA repair factor at a single DSB.

(A) LacO-I-SceI system description. An array of 256 repeats of the LacO sequence flanked by an I-SceI site is stably integrated in an U2OS cell line. Transfection of GFPlacI (GFP fused to the lac repressor lacI) induces a green focus at the site of the array via the binding of the lac repressor on the LacO sequences. (B) If the cells are transfected with I-SceI and GFPlacI at the same time, a DSB is created in close location to the array inducing the apparition of a γH2AX focus colocalizing with the GFPlacI focus. (C) U2OS stably containing the lac system and the GFPlacI are transfected with Flag, Flag-SET or Flag-SETΔNAP. In each cases the cells are transfected or not with I-SceI. Number of colocalizations between the array and γH2AX, P-RPA, BRCA1 or RAD51 are counted.

We next took advantage of the LacO-lacI/I-SceI system to evaluate the impact of direct SET tethering to chromatin on the recruitment of HR proteins to DSBs. To this end, SET protein was fused to lacI and mCherry to generate mCherrylacISET (fig. 6A, B). Interestingly, stable association of SET on chromatin had a dramatic effect in H2AX and RPA phosphorylation as well as in the recruitment of ATR, CtIP, BRCA1 and RAD51 (fig. 6C). Tethering of the NAP deletion mutant of SET mimicked the phenotype of the control (fig. 6C).

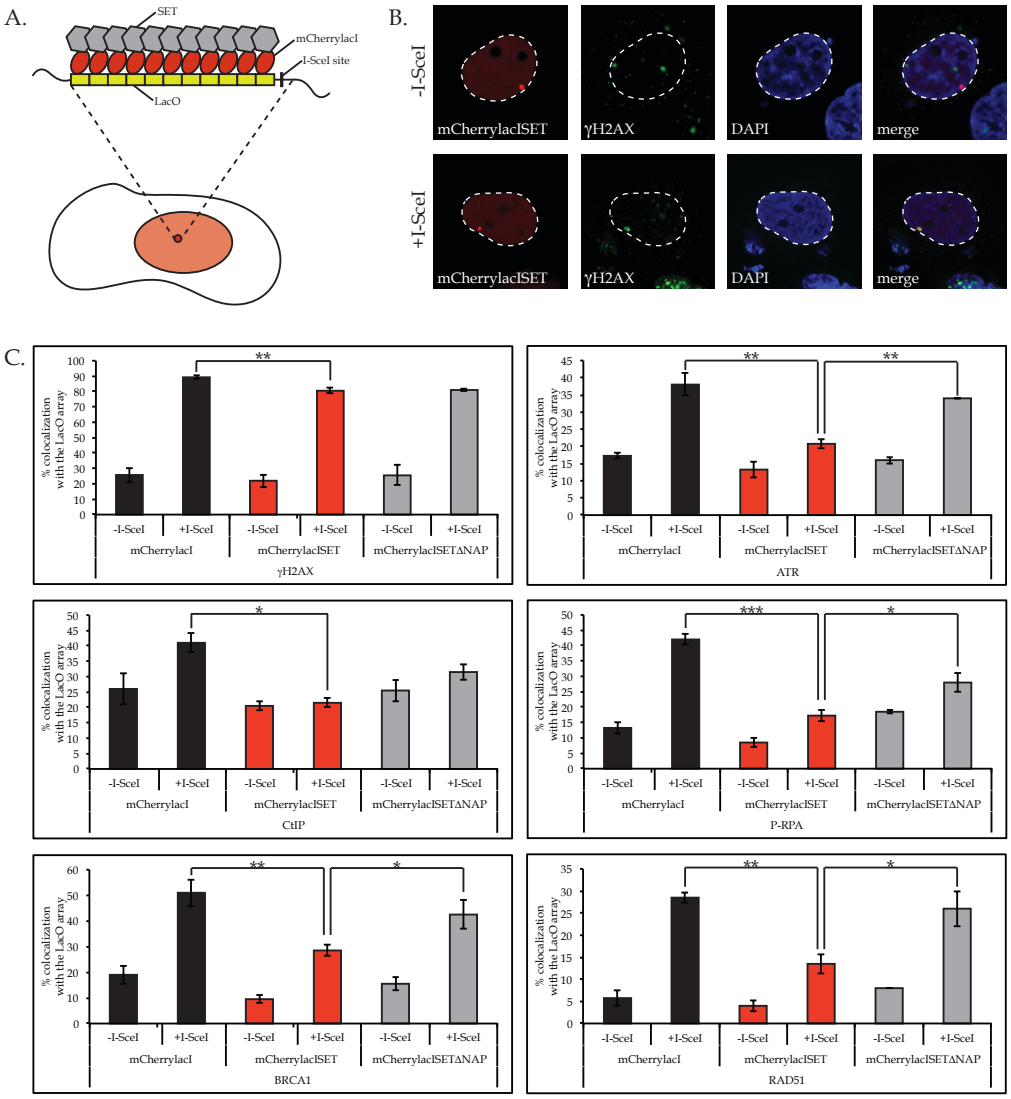

Figure 6: Effect of SET tethering on chromatin on DNA damage factor recruitment.

(A) Lac system description. An array of 256 repeats of the LacO sequence flanked by an I-SceI site is stably integrated in an U2OS cell line. Transfection of mCherrylacISET (mCherry fused to the lac repressor lacI and SET) induces the tethering of SET on the chromatin. This is visible as a red focus at the site of the array via the binding of the lac repressor on the LacO sequences. (B) If the cells are transfected with I-SceI and mCherrylacISET at the same time, a DSB is created in close location to the array where SET is tethered inducing the apparition of a γH2AX focus colocalizing with the red focus. (C) U2OS stably containing the lac system are transfected with mCherrylacI, mCherrylacISET or mCherrylacISETΔNAP. In each cases the cells are transfected or not with I-SceI. Number of colocalizations between the array and γH2AX, ATR, CtIP, P-RPA, BRCA1 or RAD51 are counted.

SET interacts with KAP1 and facilitates KAP1 and HP1 retention to chromatin

To investigate the mechanism of action of SET, we searched for potential interaction partners. SET Immunoprecipitation followed by mass spectrometry, revealed corepressor KAP1 (KRAB-associated protein-1) as an interacting partner of SET (thesis "BRCA1, KAP1 and the DNA Damage Response" from Kienan Savage, Queensland Institute of Medical Research, Herston, Australia). To verify this interaction, we performed GFP-Trap immunoprecipitation experiments using the previously described U2OS GFP and GFP-SET cell lines. We observed an interaction between SET and KAP1 that was occurring in the presence or absence of DNA damage (fig. 7A).

To further explore the functional significance of SET and KAP1 interaction, and to test whether KAP1 is involved in any of the phenotypes related to SET overexpression, we examined KAP1 expression and localization in cells overexpressing SET. RT-qPCR and Western blot analysis revealed that the expression levels of KAP1 in GFP-SET overexpressing cells are slightly higher than in GFP

control cells (less than 2 fold) (fig. 7B, C). Nevertheless, regular IF experiments showed similar KAP1 nuclear localization in U2OS cells overexpressing GFP-SET or GFP (fig. 7C). On the other hand, when IF is performed in conditions that all soluble proteins are pre-extracted before fixation, we observed a dramatic increase of KAP1 retention to chromatin when SET is overexpressed (fig. 7C, IF second lane). Interestingly, KAP1 retention was not observed in cells overexpressing GFP-SETΔNAP. These results point to a role of SET in KAP1 chromatin retention that depends on the NAP domain.

KAP1 is known to mediate gene silencing by recruiting CHD3, a subunit of the nucleosome remodeling and deacetylation (NuRD) complex, and SETDB1 a methyltransferase which specifically tri-methylates histone H3 at 'Lys-9' (H3K9me3)[27,32]. To examine whether KAP1 chromatin retention in SET overexpression was leading to SETDB1 chromatin retention and increased H3 K9 methylation, we performed Immunofluoresence in U2OS GFP, GFP-SET and GFP-SETΔNAP cells after pre-extraction of soluble proteins. Following the pattern of KAP1, SETDB1 retention to chromatin is increased when SET is overexpressed (fig. 7D). As expected, H3K9me3 levels are also higher in SET overexpressing conditions (fig. 7E). Both phenotypes were abrogated in cells expressing the ΔNAP deletion (fig. 7D, E), pointing to a role of NAP domain in retention of KAP1 its partners to chromatin.

KAP1 mediated gene silencing also involves the recruitment of Heterochromatin protein 1 (HP1s: HP1α, β, γ) through direct protein-protein interaction with KAP1, or through binding to H3K9me3 mark[27,32]. In line with the previous observations, SET

overexpression leads to higher retention of HP1s in chromatin (fig. 7F, G, H). The phenomenon was particularly pronounced in HP1β and HP1γ (fig. 7G, H). Concerning HP1γ, its localization was globally perturbed, being more in the nucleoplasm than in the heterochromatic foci. (fig. 7H).

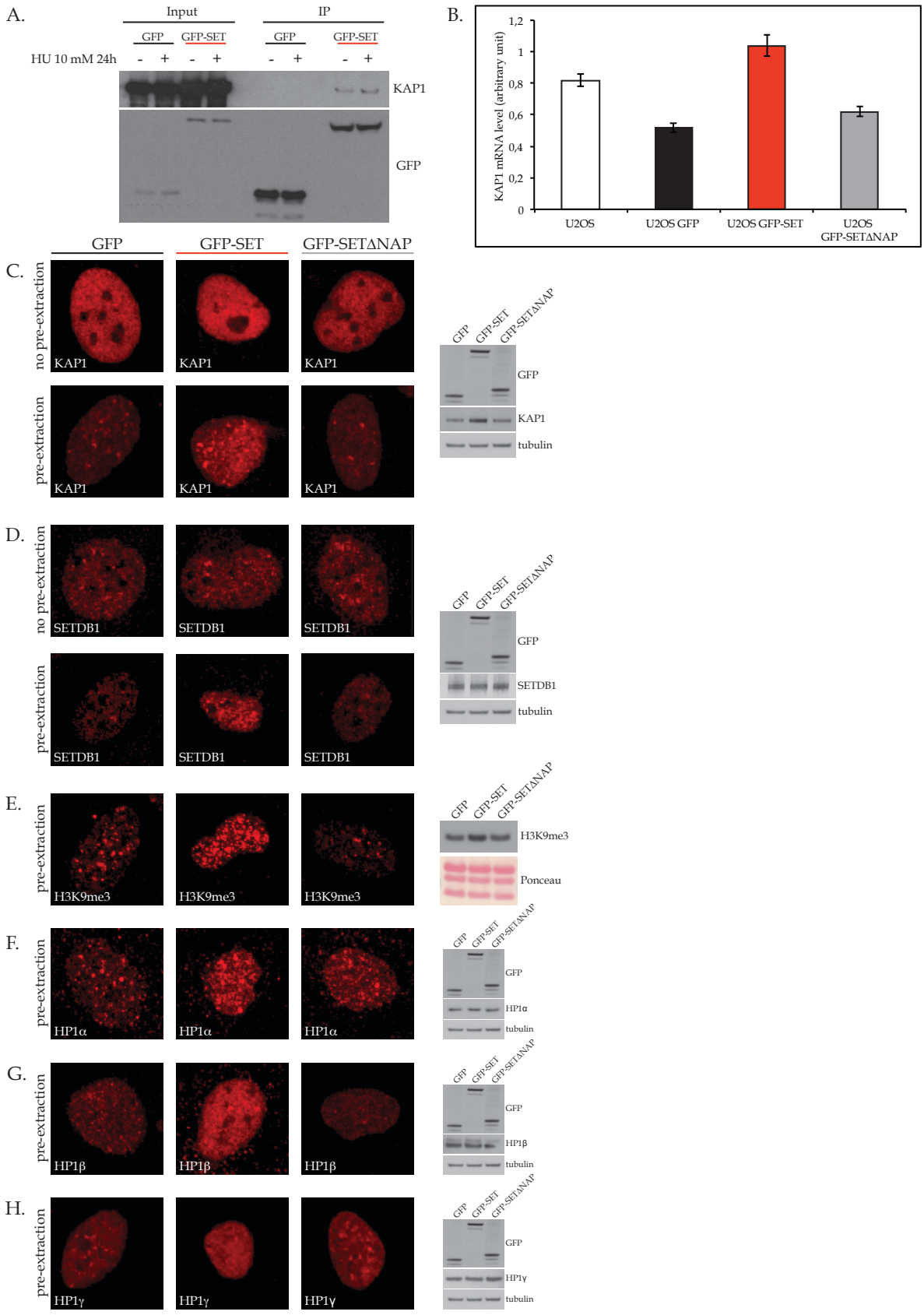

Figure 7: SET overexpression leads to retention of KAP1, SETDB1 and HP1s in chromatin

(A) U2OS GFP and GFP-SET cells are used to perform immunoprecipitation (IP) with beads labeled with antibodies against GFP. Each IP is performed with or without pre-treatment with 10 mM HU for 24 h. Input and elution are immunoblotted against KAP1 and GFP. (B) KAP1 mRNA levels are evaluated through RT-qPCR in U2OS, U2OS GFP, U2OS GFP-SET and U2OS GFP-SETΔNAP. (C) Left panel: Nuclear soluble proteins of U2OS GFP, GFP-SET and GFP-SETΔNAP cells are pre-extracted or not prior to PFA fixation. Immunostaining is performed against KAP1 and cells are imaged through confocal microscopy. Right panel: Total protein extract from U2OS GFP, GFP-SET and GFP-SETΔNAP cells are immunoblotted against GFP, KAP1 and tubulin. (D) Left panel: Nuclear soluble proteins of U2OS GFP, GFP-SET and GFP-SETΔNAP cells are pre-extracted or not prior to PFA fixation. Immunostaining is performed against SETDB1 and cells are imaged through confocal microscopy. Right panel: Total protein extract from U2OS GFP, GFP-SET and GFP-SETΔNAP cells are immunoblotted against GFP, SETDB1 and tubulin. (E) Left panel: Nuclear soluble proteins of U2OS GFP, GFP-SET and GFP-SETΔNAP cells are pre-extracted prior to PFA fixation. Immunostaining is performed against H3K9me3 and cells are imaged through confocal microscopy. Right panel: Histone extract from U2OS GFP, GFP-SET and GFP-SETΔNAP cells are immunoblotted against H3K9me3. Ponceau staining is used as a loading control. (F) Left panel: Nuclear soluble proteins of U2OS GFP, GFP-SET and GFP-SETΔNAP cells are pre-extracted prior to PFA fixation. Immunostaining is performed against HP1α and cells are imaged through confocal microscopy. Right panel: Total protein extract from U2OS GFP, GFP-SET and GFP-SETΔNAP cells are immunoblotted against GFP, HP1α and tubulin. (G) Left panel: Nuclear soluble proteins of U2OS GFP, GFP-SET and GFP-SETΔNAP cells are pre-extracted prior to PFA fixation. Immunostaining is performed against HP1β and cells are imaged through confocal microscopy. Right panel: Total protein extract from U2OS GFP, GFP-SET and GFP-SETΔNAP cells are immunoblotted against GFP, HP1β and tubulin. (H) Left panel: Nuclear soluble proteins of U2OS GFP, GFP-SET and GFP-SETΔNAP cells are pre-extracted prior to PFA fixation. Immunostaining is performed against HP1γ and cells are imaged through confocal microscopy. Right panel: Total protein extract from U2OS GFP, GFP-SET and GFP-SETΔNAP cells are immunoblotted against GFP, HP1γ and tubulin.

To verify the above results and to test whether SET tethering to chromatin is enhancing the retention of KAP1 and its interacting partners at a specific genomic location, we used the LacO-lacI/I-SceI tethering system. To this end, we tethered SET to the LacO array chromatin and we assessed the recruitment of KAP1, HP1α, HP1β

and HP1γ (fig. 8). In accordance with the previous observations, tethering of SET and not its deletion mutant at the LacO locus leads to the retention of KAP1 and HP1s at the locus (fig. 8A, B, C, D). Among the HP1s, the most significant effect was observed with Hp1γ (fig. 8D). Similar results were obtained when an I-SceI DSB was induced adjacent to the LacO locus. It is noteworthy that Hp1γ colocalization with the array is reduced to 50% upon induction of the I-SceI break in cells that lac repressor is expressed (fig. 8D), showing that HP1γ is evicted from the LacO chromatin upon DNA damage. Interestingly, upon SET tethering eviction of HP1γ is not observed, suggesting that SET inhibits the eviction of HP1γ and retains it stably bound to chromatin (fig. 8D).

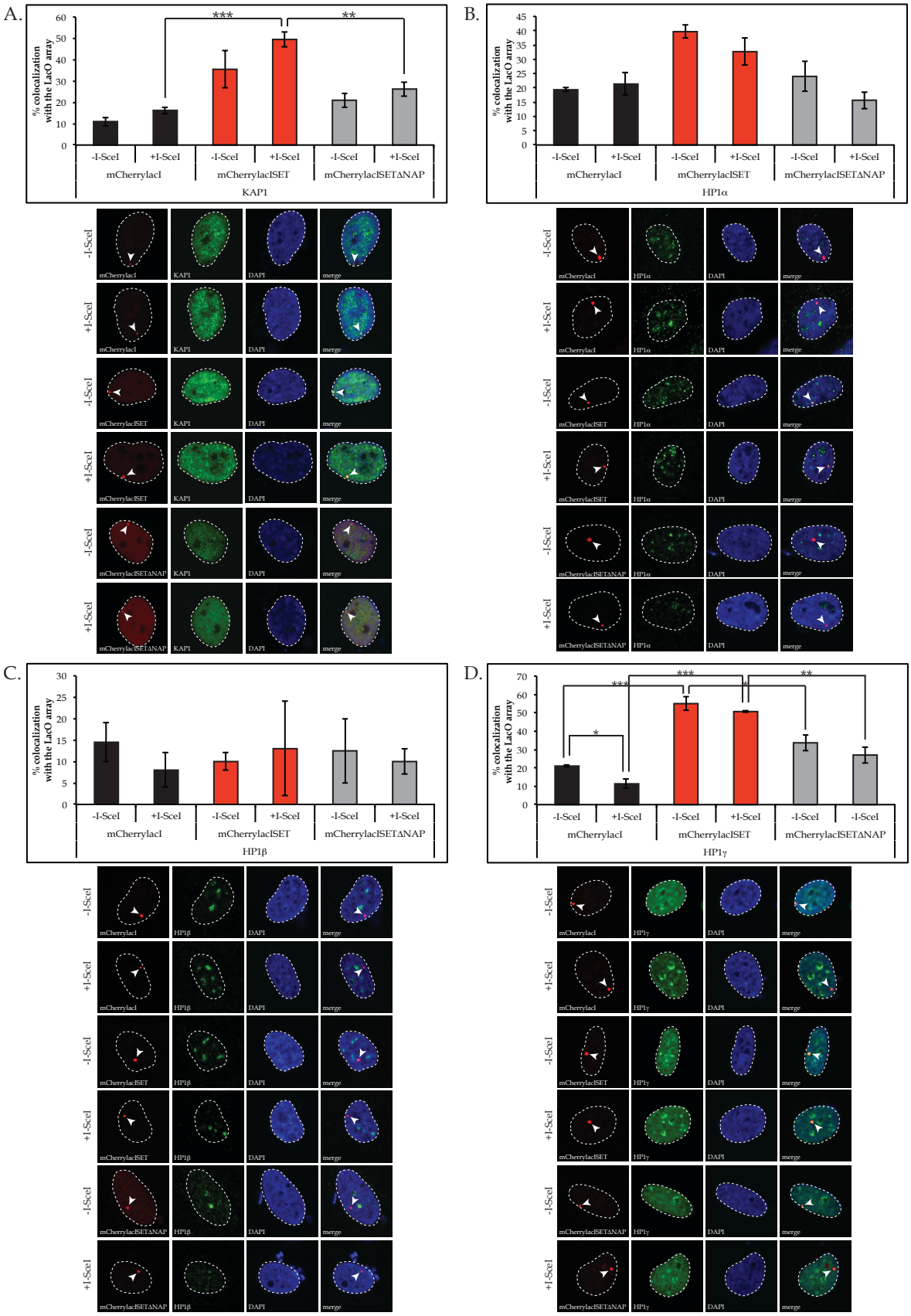

Figure 8: Effect of SET tethering on chromatin on KAP1 and HP1s.

U2OS stably containing the lac system are transfected with mCherrylacI, mCherrylacISET or mCherrylacISETΔNAP. In each cases the cells are transfected or not with I-SceI. Number of colocalizations between the array and KAP1 (A), HP1α (B), HP1β (C) or HP1γ (D) are counted and cells are imaged via confocal microscopy.

As HP1γ is not released from the break when SET is tethered, we investigated whether HP1γ retention could be part of the mechanism leading to resection impairment observed in SET overexpressing conditions. Therefore, we asked whether HP1γ tethering to LacO array by fusion with lac repressor and with GFP could recapitulate the SET tethering phenotype (fig. 9A, B). Interestingly, tethering of HP1γ at the LacO array resulted in a substantial decrease in P-RPA and BRCA1, RAD51 and CtIP recruitment upon I-SceI break induction compared to the lac repressor alone (fig. 9C, D, E, F). This phenotype was specific to HP1γ, since it was not observed upon tethering of HP1α (fig. 9C, D, E, F).

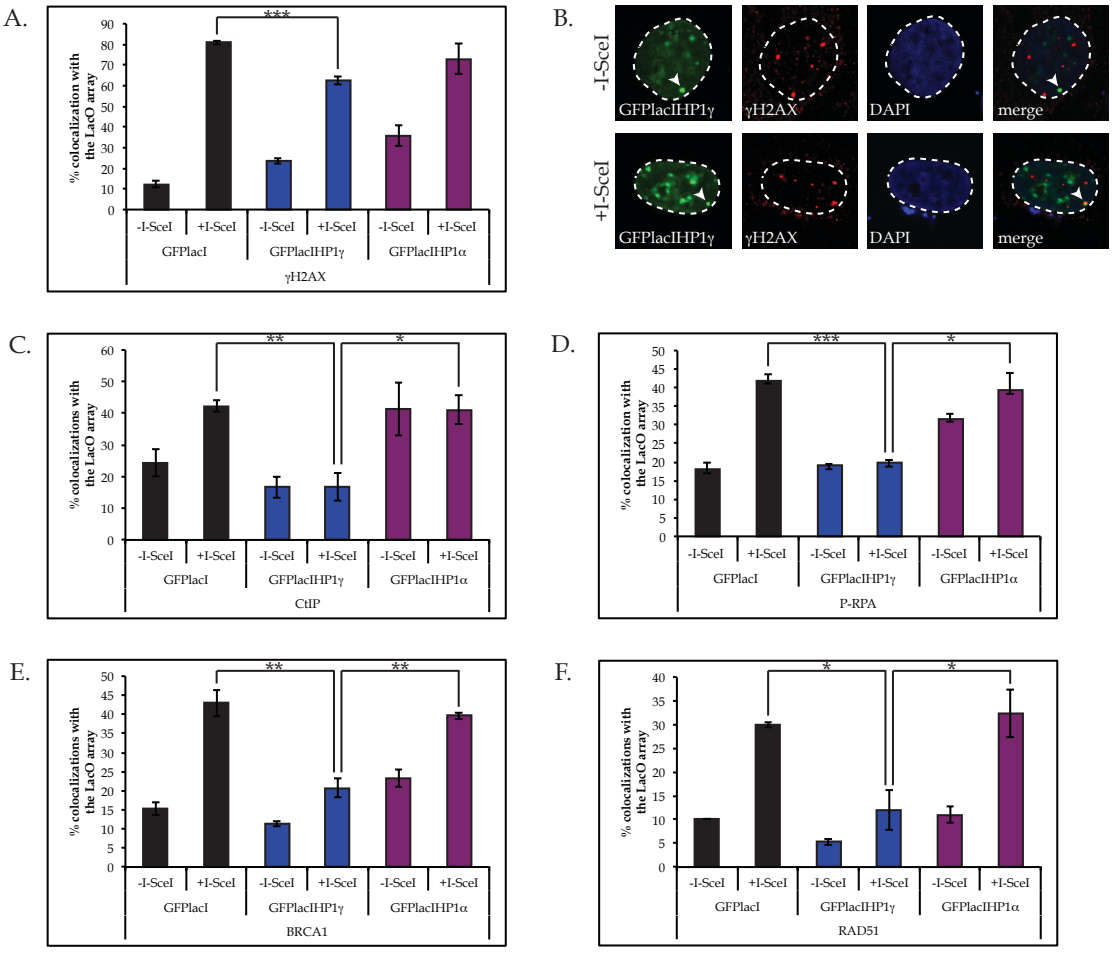

Figure 9: Effect of HP1γ tethering on chromatin on DNA damage factor recruitment.

U2OS stably containing the lac system are transfected with GFPlacI, GFPlacIHP1γ and GFPlacIHP1α. In each case the cells are transfected or not with I-SceI. Number of colocalizations between the array and γH2AX (A), P-RPA (C), BRCA1 (D), RAD51 (E) or CtIP (F) are counted and γH2AX colocalizations are imaged by confocal microscopy (B).

Our results point to a role of SET in KAP1 and HP1 retention to chromatin and when this is exaggerated, results in replication signaling and homologous recombination defects.

Discussion

SET/TAF-Iβ, also known as I2PP2A and INHAT, was originally identified as a translocated gene in acute undifferentiated leukemia[42]. It is a multi-tasking protein and it was shown to be a potent inhibitor of phosphatase 2A (PP2A)[63]. It belongs to the NAP1 family of histone chaperones[55]. Other studies have shown that SET/TAF-Iβ binds to nucleosomal histones and inhibits histone acetylation by masking histone tails as a component of the INHAT complex. Here, we describe a novel function of SET in DDR and DNA repair. We have found that SET is an endogenous modulator of DDR and when depleted enhances DDR and survival in radiomimetic drugs. In addition, we show that SET overexpression impairs DDR and homologous recombination and reduces survival in replication stress agents.

A recent study documented that SET overexpression leads to increased DNA damage, due to the down regulation of Aldehyde dehydrogenase 2 (ALDH2) and glutathione S-transferase P 1 (GSTP1), enzymes that play essential role in cellular detoxification[75]. Moreover, this study showed that SET overexpression leads to decrease of ATM, BRCA1 and Chk2 protein levels due to increased histone deacetylation at their promoters induced by SET[75]. In contrary to this observation, we do not detect changes in protein levels of all the DNA repair proteins we have tested including BRCA1, ATM and Chk2 (fig 2B and not shown). However, we observe a decrease in recruitment of HR factors like BRCA1, RAD51, CtIP, etc when SET is overexpressed or tethered to the chromatin (fig. 5C), which is in line with the decreased recruitment of BRCA2 in DSBs revealed by the same study[75].

In search of potential mechanism of action of SET, we found that it interacts with KAP1 and mediates its retention to chromatin. Our results altogether suggest a model in which SET-dependent KAP1 chromatin retention leads to the retention of its interaction partner, the methyltransferase SETDB1 and increase in its target modification tri-methylate histone 3 lysine 9 (fig. 10). Consequently, this heterochromatic mark triggers increased retention of the HP1 proteins to chromatin. In the presence of DNA damage, KAP1 and HP1s are not properly released from chromatin leading to inaccessibility to DNA repair factors and subsequent repair defect (fig. 10).

Recent studies have highlighted the importance of KAP1 and HP1 release from heterochromatin, to allow chromatin relaxation and access to DNA repair factors leading to efficient repair of heterochromatic lesions[31,35]. For HP1β, this release was dependent in its phosphorylation by Casein Kinase II[88]. On the other hand, all three HP1 isoforms are shown to accumulate in DNA lesions, through their chromoshadow domain[31,89]. These contradictory findings can be reconciled to a model in which HP1 mobilization from DSBs is followed by its accumulation at these or other sites. Our results reveal distinct behavior of HP1 isoforms in response to DNA damage. First we show that exclusively HP1γ and not α and β is released from I-SceI induced breaks. Moreover, persistent binding of HP1γ and not α to chromatin inhibits resection and subsequent strand invasion exemplified by the defective recruitment of RAD51 and BRCA1. Our observations are in line with recent data that reveal differences in how the different HP1 isoforms regulate Homologous recombination. Although, HP1α and HP1β promote RPA phosphorylation,

recruitment of RAD51 and stimulate HR, HP1γ plays an inhibitory role suggesting that its release is necessary for efficient repair by Homologous recombination.

All the above-described effects observed under SET overexpression depend on the NAP domain. This can be explained by the fact that the NAP domain is important for SET interaction with chromatin. Indeed it was shown in vitro, that NAP interacts with histone H3[46]. This idea is in line with our finding that the SETΔNAP deletion mutant partitions more in the nuclear soluble fraction than in the chromatin after fractionation. Another possible explanation is that the NAP domain of SET mediates its interaction with KAP1. This idea is supported by the observation that cells overexpressing SETΔNAP do not exert increased chromatin retention of KAP1.

The NAP domain is key for the chaperone activity of SET. It was recently reported that SET is a histone 1 chaperone and that overexpression of SET favors H1 mobility[49]. Although, increased H1 mobility and thus eviction from chromatin is associated with chromatin decondensation, we cannot exclude the possibility that KAP1 retention functions as a compensation mechanism to reverse this phenotype.

SET has been described to have two opposite roles in the chromatin organization, the ability to compact or relax the chromatin[43,50,55–62]. SET dependent chromatin compaction is described in its role as transcription cofactor and as a member of the inhibitor of histone acetyltransferase (INHAT) complex[50]. SET was also shown to inhibit DNA methylation causing chromatin condensation[62]. On the other hand, SET induces chromatin relaxation during adenoviral core DNA replication, and through its interaction with prothymosine α and microcephalin[61]. Our results show that SET overexpression leads to chromatin condensation. The presence of specific

heterochromatic marks like H3K9me3 and proteins such as, HP1 proteins and KAP1 on the chromatin of cells overexpressing SET corroborates this point. Furthermore, area measurements of DAPI stained nucleus of U2OS GFP-SET and U2OS GFP cells showed a significant size of nucleus reduction in U2OS GFP-SET cells (data not shown) further pointing to a chromatin compaction induced by SET overexpression.

The contradictory evidence about SET's role in chromatin structure, suggest that SET may have a different roles in different chromatin transactions. As described in the literature SET may open chromatin in certain chromatin context to allow adenoviral replication[43,58] or transcription[56,57] but at the same time SET might also have a role in condensing the chromatin to modulate access of DNA repair factors to chromatin and moderate DDR and DNA repair.

SET is highly overexpressed in various types of cancers[42,76–81] and in certain cases SET levels are correlating with disease severity[78,80]. Although, it was proposed that SET leads to tumorigenesis because it inhibits the tumor suppressor PP2A[82] or metastasis suppressor NM23-H1[82], our results suggest that defective DNA repair by HR that are observed when SET is overexpressed, might have contributed to the initiation of carcinogenesis.

Moreover, SET has been shown to interact with the tumor suppressor p53[90]. SET inhibits p53 acetylation thus repressing transcription of its target genes[90] leading to impairment of p53 dependent cell cycle arrest and apoptosis[90]. These results are fitting with the observation that SET is overexpressed in cancer. Indeed, if p53 signaling is deregulated by SET overexpression, cell cycle arrest and apoptosis are not mounted, leading the cells to tumorigenesis. The impairment of p53 signaling in SET overexpressing conditions fit with our results that show that cells

overexpressing SET are hypersensitive to replication stress agents. Although, we haven't directly tested the activation of p53 target genes in our system, decrease in cell survival can be explained by impairment of apoptosis and cell cycle arrest.

Therefore, SET is an obvious therapeutic target for cancer therapy. Recent studies have reported the development of peptides like COG112[82] that inhibit the binding of SET with PP2A or NM23-H1 and release them from the SET inhibitory effect. Our results demonstrated that SET overexpressing cells are hypersensitive to replication stress agents such as camptothecin (fig. 3B). This suggests that SET overexpression although might have contributed in the initiation or progression of cancer, is at the same time the Achilles' heel of these tumors and that the sensitivity of SET overexpressing cells in replication stress agents that are widely used in clinic, can be exploited to target these tumors.

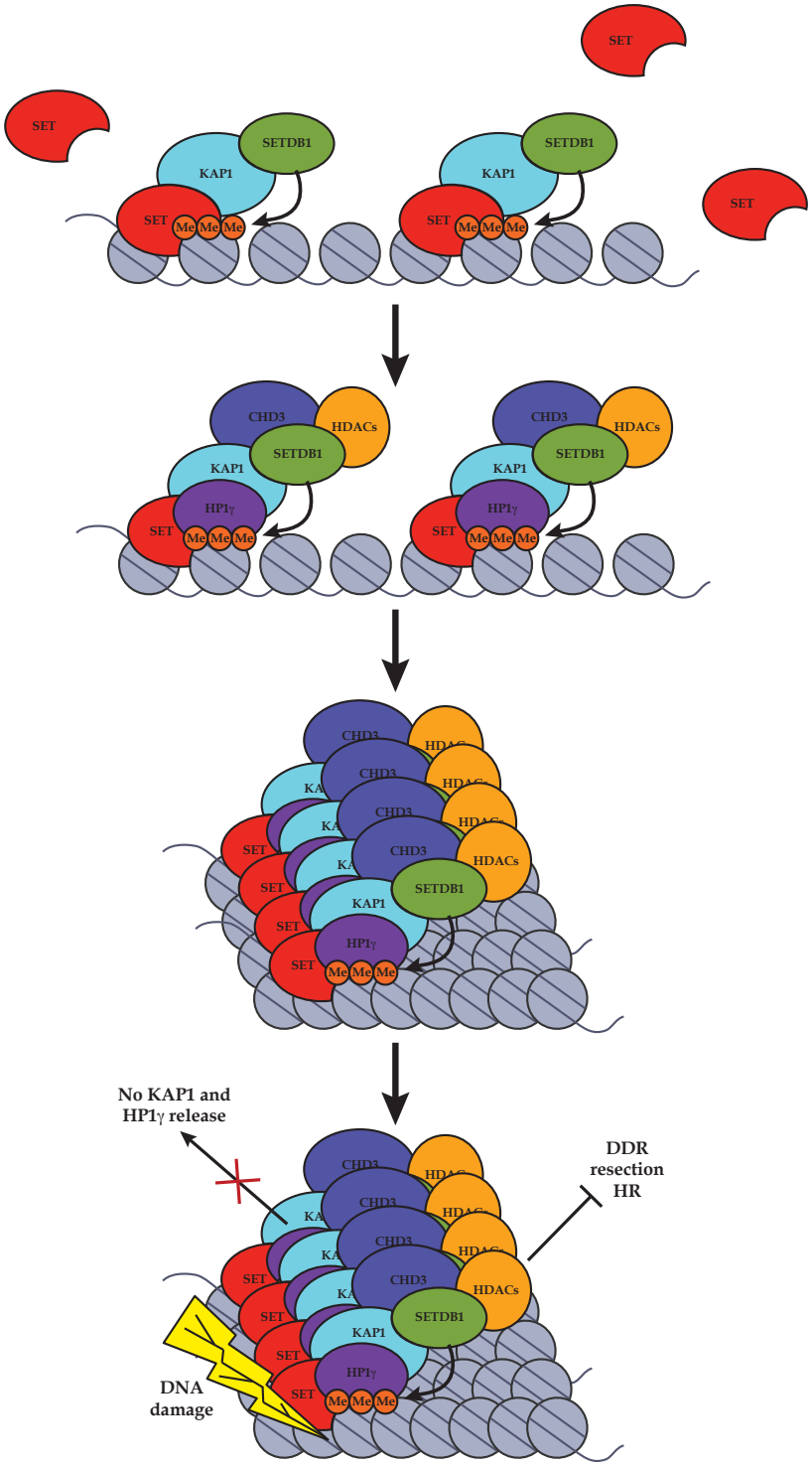

Figure 10: SET model.

Normally a fraction of SET is found on the chromatin. When SET is in excess it can cause retention of KAP1 on chromatin. KAP1 then recruits the methyltransferase SETDB1, which trimethylates H3K9. This marks allows the recruitment of the HP1 proteins as well as the components of the NuRD complex and causes DNA condensation. If a DSB happens in this context KAP1 and the HP1 proteins are

retained on the chromatin and cannot be released to allow chromatin opening, thus creating a repressive environment for DNA signaling and repair factor recruitment. Furthermore, the retention of HP1γ on chromatin inhibits resection that is the first step of homologous recombination repair.

Perspectives

To complete the characterization of the SET overexpression phenotype and to get more insights into the mechanism of action of SET in DDR and DNA repair, we plan to perform the following experiments:

-Fluorescence recovery after photobleaching (FRAP) analysis on KAP1 and HP1α, β and γ to determine if their mobility is altered after SET overexpression.

-To assess directly the inhibitory role of KAP1 and HP1γ retention on chromatin in DDR and HR, we will deplete KAP1 and HP1 by siRNA and we will test whether it alleviates the effects of SET overexpression, namely the HR decrease and HR factors binding in DSBs.

-To validate that SETDB1 retention to chromatin induces the increased for H3K9 methylation in case of SET overexpression, we will assess H3K9me3 levels in SETDB1 siRNA conditions in cells overexpressing SET.

-Finally to investigate whether the SET overexpressing phenotype observed in a cellular model is also apparent in cancers overexpressing SET, we will analyze and correlate SET expression levels with H3K9me3 levels in T-cells derived from leukemia patients where SET is very often highly overexpressed.

D. Conclusion

During my thesis entitled "Chromatin structure and DNA repair" I aimed to get a better understanding of the mechanisms involved in DNA repair in the context of chromatin.

In my first project, I attempted to identify in an unbiased way novel factors involved in DNA damage signaling and repair. I took advantage of the PICh (proteomics of isolated chromatin segments) technique established by Jérôme Déjardin that allows purification of proteins associated with specific genomic loci. I combined this technique with the LacO-I-SceI system developed by Evi Soutoglou, which permits the induction of a double strand break at a precise genomic site of known sequence. In this way I would have been able to purify proteins found in close proximity of a DSB. Unfortunately PICh was not feasible in our system

During my second project, I sought to understand the role of SET in DNA damage response and repair. SET was identified in the lab via a siRNA screen done in order to identify chromatin related proteins involves in DSB signaling. When SET was depleted, the cells showed a delayed DDR. I demonstrated in my thesis that SET overexpression impairs replication stress signaling and homologous recombination repair. Moreover cells overexpressing SET are more sensitive to replication stress agents such as camptothecin. SET interacts with KAP1 and in context of SET overexpression it retains KAP1 and its binding partner HP1γ to the chromatin thus creating an inhibitory context for homologous recombination repair. As SET is overexpressed in various types of cancers their defect in DNA repair could be exploited to treat these cancers.

E. Materials and methods

I. Materiel and methods: PICh

Cell culture and transfection

The cells are cultured in dulbecco's modified eagle medium containing 4,5 g/L glucose supplemented with 10% fetal calf serum and 40 µg/mL gentamycine.

In order to establish stable cell lines the plasmids (table ???) are transfected with FuGene 6 (Promega) according to the manufacturer protocol and then single cell colony isolated under the selection of the corresponding antibiotic.

Determination of transgene copy number by Southern blot

Different amounts of genomic DNA (5, 10 and 20 µg) were digested overnight with 50 units of EcoRI. In parallel, increasing quantities of LacO-I-SceI plasmid (corresponding to 1 to 150 copies of the transgene) were digested with 20 units of EcoRI. Electrophoretic migration of the samples was performed on a 0.8 % agarose gel in TAE for 16 h (60 V). The gel was transferred on a positively charged Nylon membrane (Genescreen plus, Perkin Elmer) in 0.4 N NaOH (overnight capillarity transfer). The membrane was dried and prehybridized overnight at 60°C with 20 ml of Church buffer supplemented with sonicated and heat-denatured salmon sperm DNA (100 µg/ml). A probe corresponding to the LacO-I-SceI plasmid digested with EcoRI was radiolabelled with adCTP by random priming (using the Prime-It II Random Primer Labeling Kit from Agilent). The membrane was hybridized in 20 ml of Church buffer for 16 h at 60°C, under constant agitation (15 rpm). Washes were performed as described, 2 X 15 min washes with 2X SSC, 0,1% SDS at 60°C, followed by 2 more stringent washes, (2 X 15 min) with 1X SSC, 0,1% SDS at 60°C, and 2 x 15

min washes with 0,5X SSC, 0,1% SDS at 60°C. The membrane was exposed for autoradiography at -80°C with intensifying screens for several days.

FISH LNA probes

Cells are grown on glass coverslips and fixed in 4% paraformaldehyde in PBS for 10 min at room temperature. Cells are then permeabalized with 0.1 % Triton-X100 in PBS for 10 min at room temperature. Following permabalization cells are blocked with 1 % BSA in PBS for 30 min at room temperature. Cells are then washed with 2X SSC 5 min at RT and then 45 min with temperature increasing progressively from RT to 72°C. Cells are then quickly whashed with 70 % ethanol once and 100 % ethanol twice before letting them air dry. The coverslips are then incubated in 0.1N NaOH for 10 min at RT before being washed with 2X SSC for 5 min at RT. Once again, cells are quickly washed with 70 % ethanol once and 100 % ethanol twice before letting them air dry.

The hybridization mix (10 μM LNA probe in LNA hybridization buffer) is then added to the slide under a glass coverslip. The whole slide is placed on a heat block for 30 sec at 85 °C and then incubated overnight at 37 °C in a humid chamber. The coverslip is then removed and the slide washed with 2X SSC two times 20 min at 42 °C. The cells are then stained with anti-biotin-FITC (antibody table) at RT for 45 to 60 min. The cells are then washed with PBS for 15 min at RT before being stained with 40 ng/mL DAPI in PBS for 5 min. Finally the coverslips are mounted using ProLong Gold mounting medium (Invitrogen).

ImmunoFISH

Immunostaining

Cells are cytospined on glass microscopic slides and fixed in 4% paraformaldehyde in PBS for 10 min at room temperature. Cells are then permeabilized with 0.1 % Triton-X100 in PBS for 10 min at room temperature. Following permabilization cells are blocked with 1 % BSA in PBS for 30 min at room temperature. Blocked cells are then stained with primary antibody (antibody table) diluted in 1 % BSA in PBS, for 1 hour at room temperature. Cells are then washed 3 times 10 min with PBS. Cells are then stained with fluorescent secondary antibody (species specific for primary) (Alexa Fluor, Molecular Probes) for 1 hour at room temperature. Cells are then washed 3 times 10 min with PBS after which fixed for 10 min with 4 % paraformaldehyde.

Probe synthesis: nick translation

The following mix is prepared: 2 µg of template DNA, 10 µL of 10X NT-Buffer, 10 µL of 10X dNTP, 10 µL of 0.1 M β-mercaptoethanol, 3 µL of BIO-16-dUTP (Roche), 4 µL of 1 mg/mL DNAse (Roche), 2 µL *E. coli* DNA polymerase I (New England Biolabs) and water qsp 100 µL. the mix is then incubated between 1 h to 2 h at 15 °C. The DNA samples are run on a 1 % agarose gel to check the probe length that should be between 250 to 500 bp. When the probe size is right the reaction is stopped with 1 µL of 0.5 M EDTA and the tube stored at -20 °C before use.

FISH

The probe is prepared as following; quantities are given for one cell sample (one slide). 0.3 µg of DNA of nick translation is mixed with 3 µg human Cot-I DNA (Invitrogen), 9 µg of salmon sperm DNA, 1/10 of the volume of 3 M sodium acetate pH 5 and 2.5 volume of 100 % ethanol. After vortexing the mix is incubated at least 30 min at -80°C. The tube is then centrifuged 20min, 14 000 rpm, 4°C and the supernatant discarded. One mL 70 % ethanol is added to the pellet and the sample is centrifuged once more 5 min, 14 000 rpm, 4°C. The supernatant is discarded and the pellet let to dry to air for 30 to 60 min. The pellet is then resuspended in 20 µL hybridization buffer and vortexed for 1 h at room temperature.

At the same time cells are prepared. Cells on slide are washed with 2X SSC 5 min at RT and then 45 min with temperature increasing progressively from RT to 72°C. Cells are then quickly washed with 70 % ethanol once and 100 % ethanol twice before letting them air dry. The slides are then incubated in 0.1N NaOH for 10 min at RT before being washed with 2X SSC for 5 min at RT. Once again, cells are quickly washed with 70 % ethanol once and 100 % ethanol twice before letting them air dry.

The hybridization mix (precipitated probe in hybridization buffer) is then added to the slide under a glass coverslip. The whole slide is placed on a heat block for 30 sec at 85 °C and then incubated overnight at 37 °C in a humid chamber. The coverslip is then removed and the slide washed with 2X SSC two times 20 min at 42 °C. The cells are then stained with anti-biotin-FITC (antibody table) and fluorescent secondary antibody (Alexa Fluor, Molecular Probes) specific for the species of the

primary antibody used in the pre-FISH immunostaining at RT for 45 to 60 min. The cells are then washed with PBS for 15 min at RT before being stained with 40 ng/mL DAPI in PBS for 5 min. Finally the slide is mounted using ProLong Gold mounting medium (Invitrogen).

PICh

From 18 L HeLa 111-ptight-D1 cell equivalent (around 10^6 cells/mL): cells are crosslinked in PBS–3% formaldehyde for 30 min at room temperature (RT) and washed four times in PBS. Cells are then equilibrated in sucrose buffer and dounced 20 times with a tight pestle. Cells are equilibrated in glycerol buffer and pellet is collected following centrifugation at 3 200 g for 10 min at 4 °C. The pellet is resuspended into the same volume of glycerol buffer. The pellet is frozen into liquid nitrogen and stored at -80 °C.

The following volumes and numbers are given for one purification (that is around 6 x 10^9 cell equivalent): the material is centrifuged at 2 000 g for 2 min at RT and the pellet is resuspended into the same pellet volume of 1x PBS-0.5%Triton X-100 and 90 µL of RNaseA (Qiagen 100 mg/ml) is added. The mixture is incubated for 4 h at RT with shaking. The samples are then centrifuged at 3 200 g for 10 min at 4 °C, the supernatant discarded and the pellet resuspended in MNAse Buffer and 2 000 U MNAse (Roche) is added before being incubated 25 min at 37 °C. The reaction is stopped with addition of EGTA to a final concentration of 20 mM and the mixture is washed four times in PBS-5 mM EGTA. The material is equilibrated in LB3JD solution and the pellet is resuspended into 55% pellet volume of LB3JD solution.

Samples are sonicated (Micro-tip, Misonix 3000) using the following parameters: Power setting 7 (36–45 Watts), 15 s constant pulse, and 45 s pause for a 7 min total process time. Sample is collected by centrifugation at 16000 g for 15 min at RT (supernatant) and incubated at 58 °C for 5 min. LB3JD pre-equilibrated streptavidin beads are added (Pierce Ultralink streptavidin, 0.5 mL) and the sample is incubated for 16 h at 4 °C. Beads are discarded and supernatant is saved. Chromatin sample is then applied to Sephacryl S-400-HR (GE Healthcare) spin columns. 1/100 final volume of 20 % SDS is added together with the LNA probe (1 mM final concentration). Hybridization is conducted as follows (25 °C for 3 min, 71 °C for 7 min, 37 °C for 3 h, 25 °C final temperature). The sample is centrifuged at 16 000 g for 15 min and the supernatant is diluted twice with milliQ water and the LBJD pre-equilibrated MyONE C1 beads (Invitrogen) solution is added (typically 750 µL). The sample is incubated for 16 h at RT before increasing the volume to 10 ml with LBJD. Beads are washed seven times with 10 ml with LBJD at RT and once with 1 mL with LB3JDLS at RT. Beads are resuspended into 750 µL of elution buffer. The sample is incubated for 30 min at RT with shaking and temperature is raised to 65 °C for 10 min. The eluate is precipitated using tri-chloroacetic acid (18% final) and the pellet is resuspended into 50 ml of crosslinking reversal solution. The sample is incubated at 99 °C for 25 min.

SDS-PAGE and silver staining

4x loading buffer is added to the eluted PICh samples that are then boiled 15 min at 95°C. The samples are loaded on 12% Bis-Tris acrylamide pre-cast gel (Invitrogen)

and migrated at 150V for 45 min. The proteins are silver stained with the Invitrogen Silver Quest silver staining kit according to the manufacturer instructions.

Western blot

4x loading buffer is added to the eluted PICh samples that are then boiled 15 min at 95°C. The samples are loaded on 12% Bis-Tris acrylamide pre-cast gel (Invitrogen) and migrated at 150V for 45 min. The proteins are then transferred on a nitrocellulose membrane through wet transfer, 1h at 150V. The membrane is blocked in a solution of 5% non-fat dry milk in PBS for 30 mins at room temperature. Primary detection antibodies are diluted in 1% non-fat dry milk in PBS. Dilutions used for each primary antibody are shown in table ???. The membranes are incubated with primary antibodies overnight at 4°C. The membrane is washed 3 times for 10 min each in PBS-0.1% Tween 20. Secondary antibody (Jackson Immunosearch) diluted in 1% non-fat dry milk in PBS, is added for at least 1 h at room temperature. The membrane is washed 3 times for 10 min each in PBS-0.1% Tween 20 before incubating it 1 min in chemilunescence reagent (ECL, Fisher Scientific). Finally the membrane is exposed to photographic film (Kodak) that is then developed.

II. Materiel and methods: SET

Cell culture and transfections

The cells are cultured in dulbecco's modified eagle medium containing 4,5 g/L glucose supplemented with 10% fetal calf serum and 40 µg/mL gentamycine.

Transient transfections of plamsids are performed using FuGene 6 (Promega) according to the manufacturer protocol. In order to establish stable cell lines the plasmids are transfected with FuGene 6 according to the manufacturer protocol and then single cell colony isolated under the selection of the corresponding antibiotic.

siRNA transfections are performed thanks to Interferin (Polyplus transfection) according to the manufacturer protocol with a final concentration of 20 nM.

Cell cycle analysis

Cells are cultured in 10 cm round dish and treated either with phleomycin or hydroxyurea for the different time points. 10 min before collection, BrdU (Sigma) is added at a final concentration of 10 µM. The cells are detached from the plates using trypsin and resuspended after centrifugation (400g, 10 min, 4°C) in 1 mL ice-cold 70 % ethanol under mild vortexing. The cells are then incubated at least overnight at -20°C for fixation. Fixed cells are incubated in pepsine buffer for 10min at 37°C before being incubated in HCl 2N for 20 min at room temperature. Anti-BrdU antibody (antibody table) is then added in Bu buffer for 45 min. Cells are washed with PBS before addition of the secondary antibody (Alexa Fluor, Molecular Probes)in Bu buffer for 30 min. Finally cells are washed and incubated in PI buffer for 30 min just

before FACS analysis. Acquisition of the data is performed through Cell Quest software, and analysis through FlowJo and ModFit.

Comet assay

Comet assay was performed according to the protocol of the Trevigen Kit FOR Comet assay. Briefly, U2OS cells were treated with NCS (200 ng/ml) for 15 minutes and were collected either right after the treatment or 16 hours after the treatment. 5000 cells from each condition were combined with Low Melting Agarose (LMAgarose) at a ratio of 1:5 and were placed onto a CometSlide. The slides were treated according to the Trevigen kit protocol and were finally electrophorized for 15 min at 21 V at 4 oC. The slides were subsequently stained with SYBR green and let to dry in the dark. The Leica epifluorescence microscope DM600B was used to view the CometSlides. Photos of the cells were analyzed by CometScore and the average Comet Tail Moment was measured.

Fractionation

Cells are collected, resuspended in hypotonic buffer, dounced 10 times with a loose pestle and centrifuged 10 min, 10 000 rpm, 4°C. The supernatant forms the cytoplasmic extract. The pellet is resuspended in sucrose buffer and high salt buffer is added drop wise under mild vortex agitation. The samples are then incubated on ice for 30 min and centrifuged 10 min, 10 000 rpm, 4°C. The supernatant forms the nuclear soluble extract. The pellet is then subjected to micrococcal digestion (MNAse,

Roche) for 10 min and sonicated 3 times 1 min. The samples are centrifuged 10 min, 10 000 rpm, 4°C and the supernatant forms the nuclear insoluble extract.

GFP-trap

Cells are collected in PBS-PMSF, lysed in lysis buffer 20 min on ice and centrifuged 20 min, 14 000 rpm, 4°C. Pellet is discarded and supernatant is incubated with GFP-trap beads (Chromotek) for 2 h at 4°C under rotation. Beads are washed two times with Buffer 150 and one time with buffer 400 and bound proteins are analysed by western blot.

Histone extraction

Cells are harvested and washed twice with ice cold PBS. The cells are then resuspended in triton extraction buffer (TEB) at cell density of 10^7 cell/mL. Cells are lysed on ice for 10 min with gentle stirring before being centrifuged at 2 000 rpm for 10 min at 4 °C. Supernatant is discarded. The cells are then washed in half the volume of TEB and centrifuged as before. The pellet is resuspended in 0.2 N HCl at a cell density of $4x10^7$ cell/mL and incubated overnight at 4 °C. Samples are centrifuged at 2 000 rpm for 10 min at 4 °C and the supernatant collected. The quantity of protein is determined using Bradford (Bio-Rad) measurement according to the manufacturer instructions before subjecting the samples to western blot analysis.

Homologous recombination assay

DR-GFP cells are transfected using Fugene 6 (Promega) according to the manufacturer instructions with the following constructs:

- pcDNA Flag 3 µg + salmon sperm DNA 1 µg
- pcDNA Flag 3 µg + HA-ISceI 1 µg
- pcDNA Flag-SET 3 µg + salmon sperm DNA 1 µg
- pcDNA Flag-SET 3 µg + HA-ISceI 1 µg
- pcDNA Flag-SETΔ 3 µg + salmon sperm DNA 1 µg
- pcDNA Flag-SET 3 µg + HA-ISceI 1 µg

72h after transfection the cells are collected with trypsin and fixed with 4 % paraformaldehyde in PBS for 10 min at room temperature. Samples are then submitted to FACS analysis. Acquisition of the data is performed through Cell Quest software, and analysis through FlowJo.

Immunofluorescence

Cells are grown on glass coverslips and fixed in 4% paraformaldehyde in PBS for 10 min at room temperature. Cells are then permeabalized with 0.1 % Triton-X100 in PBS for 10 min at room temperature. Following permeabalization cells are blocked with 1 % BSA in PBS for 30 min at room temperature. Blocked cells are then stained with primary antibody (antibody table) diluted in 1 % BSA in PBS, for 1 hour at room temperature. Cells are then washed 3 times 10 min with PBS. Cells are then stained with fluorescent secondary antibody (species specific for primary) (Alexa Fluor,

Molecular Probes) for 1 hour at room temperature. Cells are then washed 3 times 10 min with PBS after which cellular DNA is stained with 40 ng/mL DAPI in PBS. Finally cells are washed once with PBS and the coverslips mounted using ProLong Gold (Invitrogen).

In case of pre-extraction of the soluble protein, the cells are incubated with CSK buffer 30 sec, on ice prior to fixation.

Laser irradiation

Live cell imaging is performed after laser irradiation of cells expressing either GFP, GFP-SET, GFP-SETΔNAP or mRuby2Ku80. Cells are pre-sensitized with Hoechst 33342 (2 μM in PBS for 5 min, then washed 3 times in PBS) and irradiated via a 405 nm laser. Pictures are taken by confocal microscopy every second.

RNA extraction and RT-qPCR

RNA is extracted from cells thanks to the RNeasy kit (Qiagen) according to the manufacturer instructions. RT-qPCR was then performed with the Quantitect SYBR Green RT-PCR Kit (Qiagen) in the LightCycler 480 system (Roche).

siRNA screen

Screening was performed at the high throughput screening facility of the IGBMC, using a custom siRNA library (Dharmacon) of chromatin related and phosphatome siRNA smartpool (4 different siRNA/pool). Controls are performed with smartpool siRNAs from Dharmacon (Thermo Scientific): siRNAs against human XRCC4 (positive control) and siGFP (negative control). For each target, 25 nM final concentration of siRNA is reverse transfected in 5,000 HeLa cells per 0.3 cm^2 by using Interferin (Polyplus).

All screens are performed in 96-well cell culture microplates with a particular focus on avoiding microplate edge effects. γH2AX and 53BP1 foci are analyzed by immunofluorescence 3 days after siRNA transfection and 16 h after NCS treatment (100 ng/mL, 15 min). Cells are washed, fixed with 3% paraformaldehyde, permeabilized with 0.1% Triton X-100, blocked with 2% BSA, and incubated with γH2AX and 53BP1 antibodies followed by Alexa Fluor 488 and 568 conjugated second antibodies (Invitrogen). Cell nucleuses are counterstained with 1 µg/mL DAPI. The screens are achieved owing to a TECAN robotic station (for cell transfection, staining, and immunocytochemistry) and to a Caliper Twister II robotic arm coupling microplate stacks to the InCELL1000 analyzer microscope (GE LifeSciences).

Images are acquired with the InCELL1000 analyzer microscope (GE LifeSciences) and analyzed with the Multi Target Analyzer from GE Healthcare (γH2AX intensity and 53BP1 foci counting). Datasets are analyzed via RReportGenerator software (Wolfgang Rafelsberger, IGBMC).

Survival assay

U2OS cells were plated in 6 well plates in triplicates (500 cells per well) and were subsequently treated with increasing concentrations of either phleomycin (Sigma) or camptothecin (Sigma) for 1 hour. The cells were then released with the addition of fresh medium and were let to grow in colonies for 10 days. After 10 days, the cells were washed with PBS, fixed with 4% Formaldehyde for 30 minutes and stained with Crystal Violet (0,1% w/v) for 1 hour. The number of colonies per well was measured by ImageJ.

Western blot

The cells are harvested in RIPA buffer supplemented with protease inhibitor cocktail and phosSTOP (Roche). Lysis is done 15 min on ice before centrifugation (15 min, 14 000 rpm, 4°C). The quantity of protein is the determined by Bradford (Bio-Rad) measurement according to the manufacturer instructions. 4x loading buffer is then added to the extracted proteins that are then boiled 15 min at 95°C. The samples are loaded on Invitrogen pre-casted gradient gel and migrated at 150V for 45 min. The proteins are then transferred on a nitrocellulose membrane through wet transfer, 1h at 150V. The membrane is blocked in a solution of 5% non-fat dry milk in PBS for 30 mins at room temperature. Primary detection antibodies are diluted in 1% non-fat dry milk in PBS. Dilutions used for each primary antibody are shown in table 1. The membranes are incubated with primary antibodies overnight at 4°C. The membrane is washed 3 times for 10 min each in PBS-0.1% Tween 20. Secondary antibody (Jackson Immunoresearch) diluted in 1% non-fat dry milk in PBS, is added for at

least 1 h at room temperature. The membrane is washed 3 times for 10 min each in PBS-0.1% Tween 20 before incubating it 1 min in chemilunescence reagent (ECL, Fisher Scientific). Finally the membrane is exposed to photographic film (Kodak) that is then developed.

III. Buffers and antibodies

Buffer table

Buffer name	Composition
10X dNTP	0.5 mM dATP 0.5 mM dCTP 0.5 mM dGTP 0.1 mM dTTP
10X NT-Buffer	0.5 M Tris-HCl pH 8 50 mM $MgCl_2$ 0,5 mg/mL BSA
4x loading buffer	40% Glycerol 240 mM Tris/HCl pH 6.8 8% SDS 0.04% bromophenol blue 5% β-mercaptoethanol
Bu buffer	0,5 % fetal calf serum 0,5 % Tween20 20 mM Hepes in PBS 1x
Buffer 150	25 mM Hepes 7,5 150 mM NaCl 0,1 % triton X-100 2 mM $MgCl_2$ 1 mM PMSF 1x PIC 1x phoSTOP
Buffer 400	25 mM Hepes 7,5 400 mM NaCl 0,1 % triton X-100 2 mM $MgCl_2$ 1 mM PMSF 1x PIC 1x phoSTOP

Church buffer	0,25 M Sodium Phosphate Buffer (0,5M pH 7.2) 1 mM EDTA 1 % BSA 7 % SDS
Crosslinking reversal solution	250 mM Tris-HCl pH 8.8 2 % SDS 0.5 M 2-mercaptoethanol
CSK buffer	0,5 % triton-100 50 mM Hepes pH 7 150 mM NaCl 10 mM EGTA 2mM $MgCl_2$
Elution buffer	75 % LB3JD 25 % biotin 50 mM
Fractionation Sucrose buffer	20 mM Tris-HCl pH 7,65 15 mM KCl 60 mM NaCl 0,34 M Sucrose 0,15 mM Spermin 0,5 mM Spermidin
Glycerol Buffer	20 % glycerol 10 mM Hepes-NaOH pH 7.9 0.1 mM EDTA 0.1 mM EGTA 5 mM magnesium acetate (MgOAc)
High salt buffer	20 mM Tris-HCl pH 7,65 25 % glycerol 1,5 mM $MgCl_2$ 0,2 mM EDTA 900 mM NaCl
Hybridization buffer	50 % formamide 10 % dextran sulfate 4X SSC
Hypotonic buffer	10 mM Tris-HCl pH 7,65 1,5 mM $MgCl_2$ 10 mM KCl
LB3JD	10 mM Hepes-NaOH pH7.9 100 mM NaCl

	2 mM EDTA pH8 2 mM EGTA pH8 0.2 % SDS 0.1 % sodium sarkosyl
LB3JDLS	10 mM Hepes-NaOH pH7.9 30 mM NaCl 2 mM EDTA pH8 2 mM EGTA pH8 0.2 % SDS 0.1 % sodium sarkosyl
LNA hybridization buffer	50 % formamide 10 % dextran sulfate 0.08 % salmon sperm DNA 2XSSC
Lysis buffer	25 mM Hepes 7,5 150 mM NaCl 1 % triton X-100 2 mM $MgCl_2$ 1 mM PMSF 1x PIC 1x phoSTOP
MNAse Buffer	20 mM Hepes-NaOH pH 7.9 5 mM $MgCl_2$ 5 mM $CaCl_2$ 70 mM KCl
PBS-PMSF	1x PBS 1 mM PMSF
Pepsine buffer	0,5 mg/mL pepsine 30 mM HCl
PI buffer	25 µg/mL propidium iodide 50 µg/mL RNAse A in PBS 1x
PICh Sucrose buffer	0.3 M sucrose 10 mM Hepes-NaOH pH 7.9 1 % triton X-100 2 mM magnesium acetate (MgOAc)
RIPA buffer	25 mM Tris-HCl pH 7,6 150 mM NaCl

	1 % NP-40 1 % sodium deoxycholate 0,1 % SDS
Triton extraction buffer (TEB)	0.5 % triton X-100 2 mM PMSF 0.02 % NaN_3 1X PBS

Antibody table

Name	Company	Reference	Specie	dilution WB	dilution IF	dilution FISH	dilution FACS
anti-biotin FITC	Vectorlab	SP-3040	goat			1/50	
ATR (N-19)	Santa Cruz	sc-1887	goat		1/300		
BRCA1	Calbiochem	OP92+OP93	rabbit		1/200		
BrdU	Roche	11170376001	mouse				1/50
CtIP	Active Motif	61141 cl. 14-1	mouse		1/50		
GFP	Abcam	ab 6673-100	goat	1/1000			
H3K9me3	Abcam	ab8898	rabbit	1/1000	1/500		
HP1α	Euromedex	2HP2G9	mouse	1/500			
HP1α	Euromedex	2HP1H5	mouse		1/500		
HP1β	Euromedex	1MOD1A9	mouse	1/500	1/1000		
HP1γ	Euromedex	2MOD1G6	mouse	1/500	1/1000		
KAP1	Euromedex	1tb1a9	mouse	1/2000	1/2000		
phospho RPA32	Bethyl	A300-245A	rabbit	1/500	1/500		
RAD51	Calbiochem	PC130	rabbit		1/200		
RPA2	Novus	NB600-565	mouse	1/1000	1/500		
SET (I2PP2A)	Globozymes	GLO150-100	rabbit	1/1000			
SETDB1	Sigma	WH0009869M7	rabbit	1/500	1/500		
αtubulin	Sigma	T9026	mouse	1/10000			
γH2AX	Abcam	ab22551	mouse		1/1000		
γH2AX	Millipore	07-627	rabbit	1/500			

Bibliography

1. Fransz, P. & de Jong, H. From nucleosome to chromosome: a dynamic organization of genetic information. *Plant J.* **66,** 4–17 (2011).

2. Luger, K., Dechassa, M. L. & Tremethick, D. J. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat. Rev. Mol. Cell Biol.* **13,** 436–447 (2012).

3. Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* **421,** 448–453 (2003).

4. Woodcock, C. L. & Ghosh, R. P. Chromatin Higher-order Structure and Dynamics. *Cold Spring Harb. Perspect. Biol.* **2,** (2010).

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

6. Hoeijmakers, J. H. J. Genome maintenance mechanisms for preventing cancer. *Nature* **411,** 366–374 (2001).

7. Ciccia, A. & Elledge, S. J. The DNA damage response: making it safe to play with knives. *Mol. Cell* **40,** 179–204 (2010).

8. Nagy, Z. & Soutoglou, E. DNA repair: easy to visualize, difficult to elucidate. *Trends Cell Biol.* **19,** 617–629 (2009).

9. Misteli, T. & Soutoglou, E. The emerging role of nuclear architecture in DNA repair and genome maintenance. *Nat. Rev. Mol. Cell Biol.* **10,** 243–254 (2009).

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

11. Allen, C., Ashley, A. K., Hromas, R. & Nickoloff, J. A. More forks on the road to replication stress recovery. *J. Mol. Cell Biol.* **3,** 4–12 (2011).

12. Rupnik, A., Lowndes, N. F. & Grenon, M. MRN and the race to the break. *Chromosoma* **119,** 115–135 (2010).

13. Williams, G. J., Lees-Miller, S. P. & Tainer, J. A. Mre11-Rad50-Nbs1 conformations and the control of sensing, signaling, and effector responses at DNA double-strand breaks. *DNA Repair* **9,** 1299–1306 (2010).

14. Buis, J. *et al.* Mre11 Nuclease Activity Has Essential Roles in DNA Repair and Genomic Stability Distinct from ATM Activation. *Cell* **135,** 85–96 (2008).

15. Rass, E. *et al.* Role of Mre11 in chromosomal nonhomologous end joining in mammalian cells. *Nat. Struct. Mol. Biol.* **16,** 819–824 (2009).

16. Shiloh, Y. & Ziv, Y. The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. *Nat. Rev. Mol. Cell Biol.* **14,** 197–210 (2013).

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

18. Cimprich, K. A. & Cortez, D. ATR: an essential regulator of genome integrity. *Nat. Rev. Mol. Cell Biol.* **9,** 616–627 (2008).

19. Mordes, D. A. & Cortez, D. Activation of ATR and related PIKKs. *Cell Cycle Georget. Tex* **7,** 2809–2812 (2008).

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

21. Hill, R. & Lee, P. W. K. The DNA-dependent protein kinase (DNA-PK): More than just a case of making ends meet? *Cell Cycle Georget. Tex* **9,** 3460–3469 (2010).

22. Stracker, T. H., Usui, T. & Petrini, J. H. J. Taking the time to make important decisions: the checkpoint effector kinases Chk1 and Chk2 and the DNA damage response. *DNA Repair* **8,** 1047–1054 (2009).

23. Lukas, J., Lukas, C. & Bartek, J. More than just a focus: The chromatin response to DNA damage and its role in genome integrity maintenance. *Nat. Cell Biol.* **13,** 1161–1169 (2011).

24. Polo, S. E. & Jackson, S. P. Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. *Genes Dev.* **25,** 409–433 (2011).

25. Price, B. D. & D'Andrea, A. D. Chromatin remodeling at DNA double-strand breaks. *Cell* **152,** 1344–1354 (2013).

26. Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. M. DNA Double-stranded Breaks Induce Histone H2AX Phosphorylation on Serine 139. *J. Biol. Chem.* **273,** 5858–5868 (1998).

27. Cann, K. L. & Dellaire, G. Heterochromatin and the DNA damage response: the need to relaxThis paper is one of a selection of papers in a Special Issue entitled 31st Annual International Asilomar Chromatin and Chromosomes Conference, and has undergone the Journal's usual peer review process. *Biochem. Cell Biol.* **89,** 45–60 (2010).

28. Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **128,** 693–705 (2007).

29. Sun, Y., Jiang, X. & Price, B. D. Tip60: Connecting chromatin to DNA damage signaling. *Cell Cycle* **9,** 930–936 (2010).

30. Jackson, S. P. & Durocher, D. Regulation of DNA Damage Responses by Ubiquitin and SUMO. *Mol. Cell* **49,** 795–807 (2013).

31. Soria, G., Polo, S. E. & Almouzni, G. Prime, repair, restore: the active role of chromatin in the DNA damage response. *Mol. Cell* **46,** 722–734 (2012).

32. Iyengar, S. & Farnham, P. J. KAP1 Protein: An Enigmatic Master Regulator of the Genome. *J. Biol. Chem.* **286,** 26267–26276 (2011).

33. Zeng, W., Ball, A. R. & Yokomori, K. HP1: Heterochromatin binding proteins working the genome. *Epigenetics* **5,** 287–292 (2010).

34. Maison, C. & Almouzni, G. HP1 and the dynamics of heterochromatin maintenance.

Nat. Rev. Mol. Cell Biol. **5,** 296–305 (2004).

35. Soria, G. & Almouzni, G. Differential contribution of HP1 proteins to DNA end resection and homology-directed repair. *Cell Cycle* **12,** 422–429 (2013).

36. Boboila, C., Alt, F. W. & Schwer, B. Classical and alternative end-joining pathways for repair of lymphocyte-specific and general DNA double-strand breaks. *Adv. Immunol.* **116,** 1–49 (2012).

37. Wang, C. & Lees-Miller, S. P. Detection and repair of ionizing radiation-induced DNA double strand breaks: new developments in nonhomologous end joining. *Int. J. Radiat. Oncol. Biol. Phys.* **86,** 440–449 (2013).

38. Mladenov, E. & Iliakis, G. Induction and repair of DNA double strand breaks: the increasing spectrum of non-homologous end joining pathways. *Mutat. Res.* **711,** 61–72 (2011).

39. Kass, E. M. & Jasin, M. Collaboration and competition between DNA double-strand break repair pathways. *FEBS Lett.* **584,** 3703–3708 (2010).

40. Krejci, L., Altmannova, V., Spirek, M. & Zhao, X. Homologous recombination and its regulation. *Nucleic Acids Res.* **40,** 5795–5818 (2012).

41. Eitoku, M., Sato, L., Senda, T. & Horikoshi, M. Histone chaperones: 30 years from isolation to elucidation of the mechanisms of nucleosome assembly and disassembly. *Cell. Mol. Life Sci.* **65,** 414–444 (2008).

42. Adachi, Y., Pavlakis, G. N. & Copeland, T. D. Identification and characterization of SET, a nuclear phosphoprotein encoded by the translocation break point in acute undifferentiated leukemia. *J. Biol. Chem.* **269,** 2258–2262 (1994).

43. Matsumoto, K., Nagata, K., Ui, M. & Hanaoka, F. Template activating factor I, a novel host factor required to stimulate the adenovirus core DNA replication. *J. Biol. Chem.* **268,** 10582–10587 (1993).

44. Asaka, M. N., Murano, K. & Nagata, K. Sp1-mediated transcription regulation of TAF-Ialpha gene encoding a histone chaperone. *Biochem. Biophys. Res. Commun.* **376,** 665–670 (2008).

45. Selth, L. & Svejstrup, J. Q. Vps75, A New Yeast Member of the NAP Histone Chaperone Family. *J. Biol. Chem.* **282,** 12358–12362 (2007).

46. Muto, S. *et al.* Relationship between the structure of SET/TAF-Iβ/INHAT and its histone chaperone activity. *Proc. Natl. Acad. Sci.* **104,** 4285–4290 (2007).

47. Kutney, S. N., Hong, R., Macfarlan, T. & Chakravarti, D. A Signaling Role of Histone-binding Proteins and INHAT Subunits pp32 and Set/TAF-Iβ in Integrating Chromatin Hypoacetylation and Transcriptional Repression. *J. Biol. Chem.* **279,** 30850–30855 (2004).

48. Schneider, R., Bannister, A. J., Weise, C. & Kouzarides, T. Direct Binding of INHAT to H3 Tails Disrupted by Modifications. *J. Biol. Chem.* **279,** 23859–23862 (2004).

49. Kato, K., Okuwaki, M. & Nagata, K. Role of Template Activating Factor-I as a chaperone in linker histone dynamics. *J. Cell Sci.* **124,** 3254–3265 (2011).

50. Seo, S. B. *et al.* Regulation of histone acetylation and transcription by INHAT, a human cellular complex containing the set oncoprotein. *Cell* **104,** 119–130 (2001).

51. Miyamoto, S. *et al.* Positive and Negative Regulation of the Cardiovascular Transcription Factor KLF5 by p300 and the Oncogenic Regulator SET through Interaction and Acetylation on the DNA-Binding Domain. *Mol. Cell. Biol.* **23,** 8528–8541 (2003).

52. Suzuki, T. *et al.* Functional Interaction of the DNA-binding Transcription Factor Sp1 through Its DNA-binding Domain with the Histone Chaperone TAF-I. *J. Biol. Chem.* **278,** 28758–28764 (2003).

53. Wagner, S., Weber, S., Kleinschmidt, M. A., Nagata, K. & Bauer, U.-M. SET-mediated Promoter Hypoacetylation Is a Prerequisite for Coactivation of the Estrogen-responsive pS2 Gene by PRMT1. *J. Biol. Chem.* **281,** 27242–27250 (2006).

54. Ichijo, T., Chrousos, G. P. & Kino, T. Activated Glucocorticoid Receptor Interacts with the INHAT Component Set/TAF-I? and Releases it from a Glucocorticoid-responsive Gene Promoter, Relieving Repression: Implications for the Pathogenesis of Glucocorticoid Resistance in Acute Undifferentiated Leukemia with Set-Can Translocation. *Mol. Cell. Endocrinol.* **283,** 19–31 (2008).

55. Kawase, H. *et al.* NAP-I is a functional homologue of TAF-I that is required for replication and transcription of the adenovirus genome in a chromatin-like structure. *Genes Cells Devoted Mol. Cell. Mech.* **1,** 1045–1056 (1996).

56. Matsumoto, K. *et al.* Stimulation of DNA Transcription by the Replication Factor from the Adenovirus Genome in a Chromatin-like Structure. *J. Biol. Chem.* **270,** 9645–9650 (1995).

57. Okuwaki, M. & Nagata, K. Template Activating Factor-I Remodels the Chromatin Structure and Stimulates Transcription from the Chromatin Template. *J. Biol. Chem.* **273,** 34511–34518 (1998).

58. Pegoraro, G., Marcello, A., Myers, M. P. & Giacca, M. Regulation of Adeno-Associated Virus DNA Replication by the Cellular TAF-I/Set Complex. *J. Virol.* **80,** 6855–6864 (2006).

59. Matsumoto, K., Nagata, K., Miyaji-Yamaguchi, M., Kikuchi, A. & Tsujimoto, M. Sperm Chromatin Decondensation by Template Activating Factor I through Direct Interaction with Basic Proteins. *Mol. Cell. Biol.* **19,** 6940–6952 (1999).

60. Karetsou, Z. *et al.* Prothymosin alpha associates with the oncoprotein SET and is involved in chromatin decondensation. *FEBS Lett.* **577,** 496–500 (2004).

61. Leung, J. W. *et al.* SET Nuclear Oncogene Associates with Microcephalin/MCPH1 and Regulates Chromosome Condensation. *J. Biol. Chem.* **286,** 21393–21400 (2011).

62. Cervoni, N., Detich, N., Seo, S., Chakravarti, D. & Szyf, M. The Oncoprotein Set/TAF-1β, an Inhibitor of Histone Acetyltransferase, Inhibits Active Demethylation of DNA, Integrating DNA Methylation and Transcriptional Silencing. *J. Biol. Chem.* **277,** 25026–25031 (2002).

63. Li, M., Makkinje, A. & Damuni, Z. The Myeloid Leukemia-associated Protein SET Is a Potent Inhibitor of Protein Phosphatase 2A. *J. Biol. Chem.* **271,** 11059–11062 (1996).

64. Seshacharyulu, P., Pandey, P., Datta, K. & Batra, S. K. Phosphatase: PP2A structural importance, regulation and its aberrant expression in cancer. *Cancer Lett.* **335,** 9–18 (2013).

65. Estanyol, J. M. *et al.* The Protein SET Regulates the Inhibitory Effect of p21Cip1 on Cyclin E-Cyclin-dependent Kinase 2 Activity. *J. Biol. Chem.* **274,** 33161–33165 (1999).

66. Canela, N. *et al.* The SET Protein Regulates G2/M Transition by Modulating Cyclin B-Cyclin-dependent Kinase 1 Activity. *J. Biol. Chem.* **278,** 1158–1164 (2003).

67. Carujo, S. *et al.* Glyceraldehyde 3-phosphate dehydrogenase is a SET-binding protein and regulates cyclin B-cdk1 activity. *Oncogene* **25,** 4033–4042 (2006).

68. Fan, Z., Beresford, P. J., Oh, D. Y., Zhang, D. & Lieberman, J. Tumor suppressor NM23-H1 is a granzyme A-activated DNase during CTL-mediated apoptosis, and the nucleosome assembly protein SET is its inhibitor. *Cell* **112,** 659–672 (2003).

69. Chowdhury, D. *et al.* The exonuclease TREX1 is in the SET complex and acts in concert with NM23-H1 to degrade DNA during granzyme A-mediated cell death. *Mol. Cell* **23,** 133–142 (2006).

70. Zhao, T. *et al.* Granzyme K cleaves the nucleosome assembly protein SET to induce single-stranded DNA nicks of target cells. *Cell Death Differ.* **14,** 489–499 (2006).

71. Qu, D. *et al.* The Protein SET Binds the Neuronal Cdk5 Activator p35 nck5a and Modulates Cdk5/p35 nck5a Activity. *J. Biol. Chem.* **277,** 7324–7332 (2002).

72. Kim, D.-W., Kim, K.-B., Kim, J.-Y., Lee, K.-S. & Seo, S.-B. Negative regulation of neuronal cell differentiation by INHAT subunit SET/TAF-Iβ. *Biochem. Biophys. Res. Commun.* **400,** 419–425 (2010).

73. Yu, G. *et al.* Ser9 phosphorylation causes cytoplasmic detention of I2PP2A/SET in Alzheimer disease. *Neurobiol. Aging* **34,** 1748–1758 (2013).

74. Chambon, J.-P. *et al.* The PP2A Inhibitor I2PP2A Is Essential for Sister Chromatid Segregation in Oocyte Meiosis II. *Curr. Biol.* **23,** 485–490 (2013).

75. Almeida, L. O. *et al.* SET overexpression decreases cell detoxification efficiency: ALDH2 and GSTP1 are downregulated, DDR is impaired and DNA damage accumulates. *FEBS J.* **279,** 4615–4628 (2012).

76. Ouellet, V. *et al.* SET complex in serous epithelial ovarian cancer. *Int. J. Cancer J. Int. Cancer* **119,** 2119–2126 (2006).

77. Jiang, Q., Zhang, C., Zhu, J., Chen, Q. & Chen, Y. The set gene is a potential oncogene in human colorectal adenocarcinoma and oral squamous cell carcinoma. *Mol. Med. Reports* **4,** 993–999 (2011).

78. Christensen, D. J. *et al.* SET oncoprotein overexpression in B-cell chronic lymphocytic leukemia and non-Hodgkin lymphoma: a predictor of aggressive disease and a new treatment target. *Blood* **118,** 4150–4158 (2011).

79. Li, C. *et al.* Quantitative Proteomics Reveal up-regulated Protein Expression of the SET Complex Associated with Hepatocellular Carcinoma. *J. Proteome Res.* **11,** 871–885 (2012).

80. Cristóbal, I. *et al.* Overexpression of SET is a recurrent event associated with poor outcome and contributes to protein phosphatase 2A inhibition in acute myeloid leukemia. *Haematologica* **97,** 543–550 (2012).

81. Leopoldino, A. M. *et al.* SET protein accumulates in HNSCC and contributes to cell survival: antioxidant defense, Akt phosphorylation and AVOs acidification. *Oral Oncol.* **48,** 1106–1113 (2012).

82. Switzer, C. H. *et al.* Targeting SET/I2PP2A Oncoprotein Functions as a Multi-pathway Strategy for Cancer Therapy. *Oncogene* **30,** 2504–2513 (2011).

83. Soutoglou, E. *et al.* Positional stability of single double-strand breaks in mammalian cells. *Nat. Cell Biol.* **9,** 675–682 (2007).

84. Déjardin, J. & Kingston, R. E. Purification of Proteins Associated with Specific Genomic Loci. *Cell* **136,** 175–186 (2009).

85. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* **196,** 801–810 (2012).

86. Ahnesorg, P., Smith, P. & Jackson, S. P. XLF Interacts with the XRCC4-DNA Ligase IV Complex to Promote DNA Nonhomologous End-Joining. *Cell* **124,** 301–313 (2006).

87. Pierce, A. J., Johnson, R. D., Thompson, L. H. & Jasin, M. XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes Dev.* **13,** 2633–2638 (1999).

88. Ayoub, N., Jeyasekharan, A. D., Bernal, J. A. & Venkitaraman, A. R. HP1-beta mobilization promotes chromatin changes that initiate the DNA damage response. *Nature* **453,** 682–686 (2008).

89. Luijsterburg, M. S. *et al.* Heterochromatin protein 1 is recruited to various types of DNA damage. *J. Cell Biol.* **185,** 577–586 (2009).

90. Kim, J.-Y. *et al.* Inhibition of p53 acetylation by INHAT subunit SET/TAF-Iβ represses p53 activity. *Nucleic Acids Res.* **40,** 75–87 (2012).

Résumé en français : Etude de la structure de la chromatine dans la réparation de l'ADN.

Les cassures double brin (CDBs) de l'ADN sont dangereuses car une religature aberrante peut produire des translocations chromosomiques tumorigènes[1]. Les CDBs sont réparées par deux voies principales, la recombinaison homologue (RH) et la ligature d'extrémités non-homologues (NHEJ)[2] (fig. 1). Par ailleurs, les cassures double brin déclenchent une voie de signalisation complexe appelée la réponse aux dommages de l'ADN (RDA). La RDA implique la reconnaissance de l'ADN endommagé suivi de l'activation des phosphatidylinositol 3 kinases ATM/DNAPK/ATR (ataxia telangiectasia mutated/DNA-dependent protein kinase/ataxia telangiectasia and Rad3-related) suivie du recrutement d'une cascade de protéines, comme par exemple, les protéines médiatrices MDC1 et 53BP1. Cette cascade de signalisation mène à l'activation des kinases point de contrôle du cycle cellulaire (Chk1 et Chk2) qui arrêtent le cycle jusqu'à ce que la lésion soit réparée[3–5] (fig. 2). La RDA se manifeste morphologiquement sous forme de foyers de réparation de l'ADN, des structures formées par le recrutement et l'accumulation des facteurs de réparation de l'ADN au niveau du dommage. De ce fait, une cassure double brin unique forme un foyer, visible en microscopie, qui s'étend jusqu'à une mégabase dans la chromatine environnante[6]. Bien qu'aujourd'hui nous connaissions un certain nombre de facteurs recrutés aux CDBs, la composition exacte des foyers de réparation n'est pas élucidée.

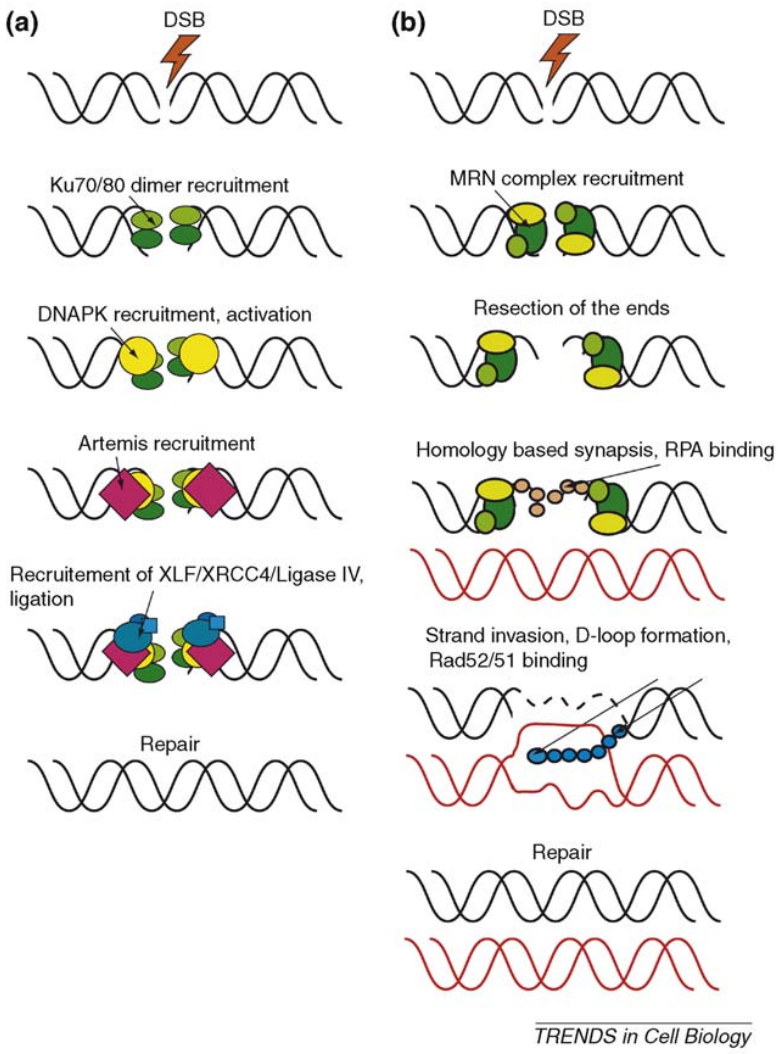

Figure 1 : Mécanismes de réparation des cassures double brin[6].

(A) Ligature d'extrémités non-homologues (NHEJ) et (B) recombinaison homologue (RH).

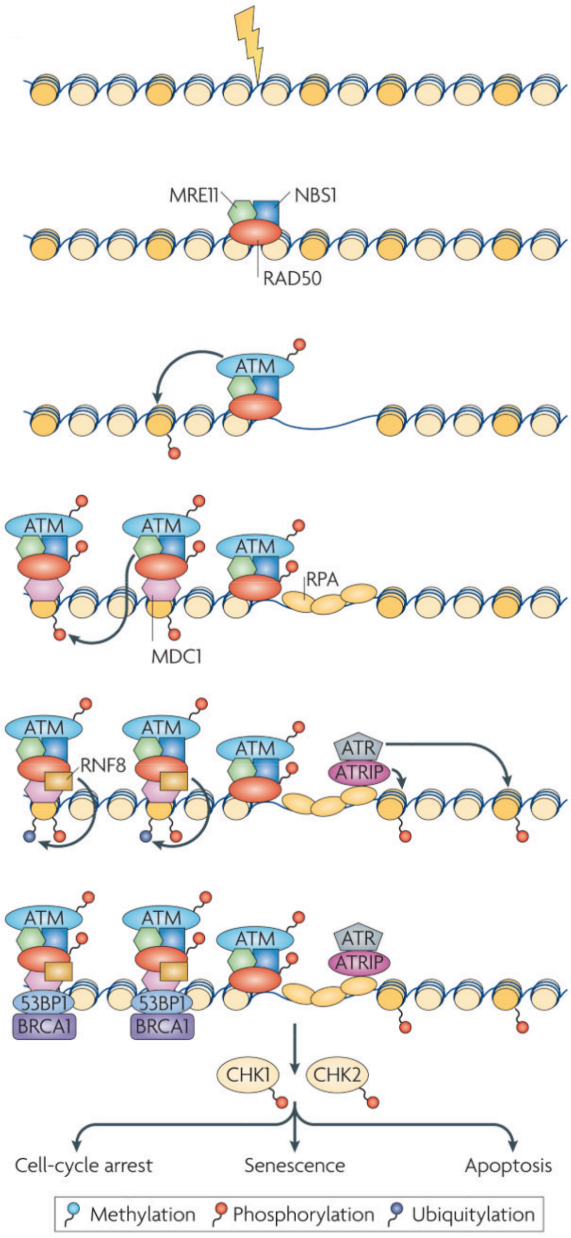

Figure 2 : Réponse aux dommages de l'ADN (RDA)[7].

Tous les types de réparations de l'ADN ont lieu dans un contexte hautement structuré, la chromatine. La capacité de certains facteurs de réparation à détecter l'ADN endommagé est déterminée par les modifications des histones autour de la lésion, et implique des événements de remodelage de la chromatine[8,9]. Lors de la

réparation des CDBs (RCDB), une des principale modifications d'histone est la phosphorylation de la queue C-terminale d'H2AX, appelée ainsi γH2AX[10]. D'autres modifications telles que l'acétylation, la méthylation, la phosphorylation et l'ubiquitination des histones du cœur du nucléosome, ont été liées à différents aspects de la réparation de l'ADN. En effet, elles permettent notamment de faciliter l'accessibilité à la fibre chromatinienne, de recruter des facteurs de réparation ainsi que des complexes protéiques permettant le remodelage de la chromatine et enfin de propager le signal de la RDA[11]. Les modifications de la chromatine sont essentielles pour une réparation efficace des CDBs, cependant les mécanismes précis à l'origine de ces changements sont encore partiellement méconnus. L'objectif de ma thèse est d'étudier la structure de la chromatine lors de la réparation des CDBs.

Pendant ma thèse, j'ai travaillé sur deux projets principaux : le développement d'une approche protéomique non biaisée afin d'identifier de nouvelles protéines recrutées aux CDBs et impliquées dans la modification de la chromatine ainsi que l'étude du rôle d'une protéine spécifique de la chromatine SET/TAF-1β dans la RCDB.

Projet A : Développement d'une approche protéomique pour l'identification de nouveaux modificateurs de la chromatine aux CDBs.

Le fractionnement biochimique de foyers de réparations obtenus par radiations ionisantes est très ardu techniquement. Le laboratoire d'Evi Soutoglou a développé un système cellulaire (système LacO-I-SceI) dans lequel une CDB peut être spécifiquement induite à un site génomique défini. Il est composé d'un site de restriction de 18 nucléotides pour l'enzyme I-SceI qui n'existe pas de façon endogène chez les mammifères. Ce site est bordé de part et d'autre de 256 copies du site de liaison du répresseur de l'opéron lactose (lac opérateur : LacO) et de 96 copies de la séquence opératrice de l'opéron tétracycline (TetO). Grâce à l'expression d'I-SceI, une CDB unique dans le génome peut être induite[12] (fig. 3A). Une lignée cellulaire NIH 3T3 stable pour ce système a été générée (NIH 2/4). Ce système offre l'opportunité d'isoler par sonde un foyer de réparation de l'ADN et de purifier des nouveaux facteurs protéiques interagissant avec les CDBs. Pour identifier les protéines qui s'accumulent dans les foyers en réponse aux CDBs, nous avons utilisé une méthode décrite récemment appelée PICh (Proteomics of Isolated Chromatin Segments)[13] (fig. 3B). Cette technique est basée sur une fixation au formaldéhyde des protéines liées à l'ADN, suivie par la précipitation des séquences génomiques spécifiques en utilisant des oligonucléotides contenant des LNA (locked nucleic acids) fusionnés à une desthiobiotine. Les hybrides ADN/sonde sont isolés en utilisant des billes magnétiques recouvertes d'avidine. La fixation est ensuite inversée et les protéines co-précipitées sont séparées sur gel SDS-PAGE avant d'être identifiées par spectrométrie de masse (SM). Cette méthode convient pour analyser la composition de loci génomiques abondants (i.e. télomères, centromères, répétitions lacO, etc) et

n'est pas pour le moment applicable à l'analyse de loci ne comportant qu'une seule copie.

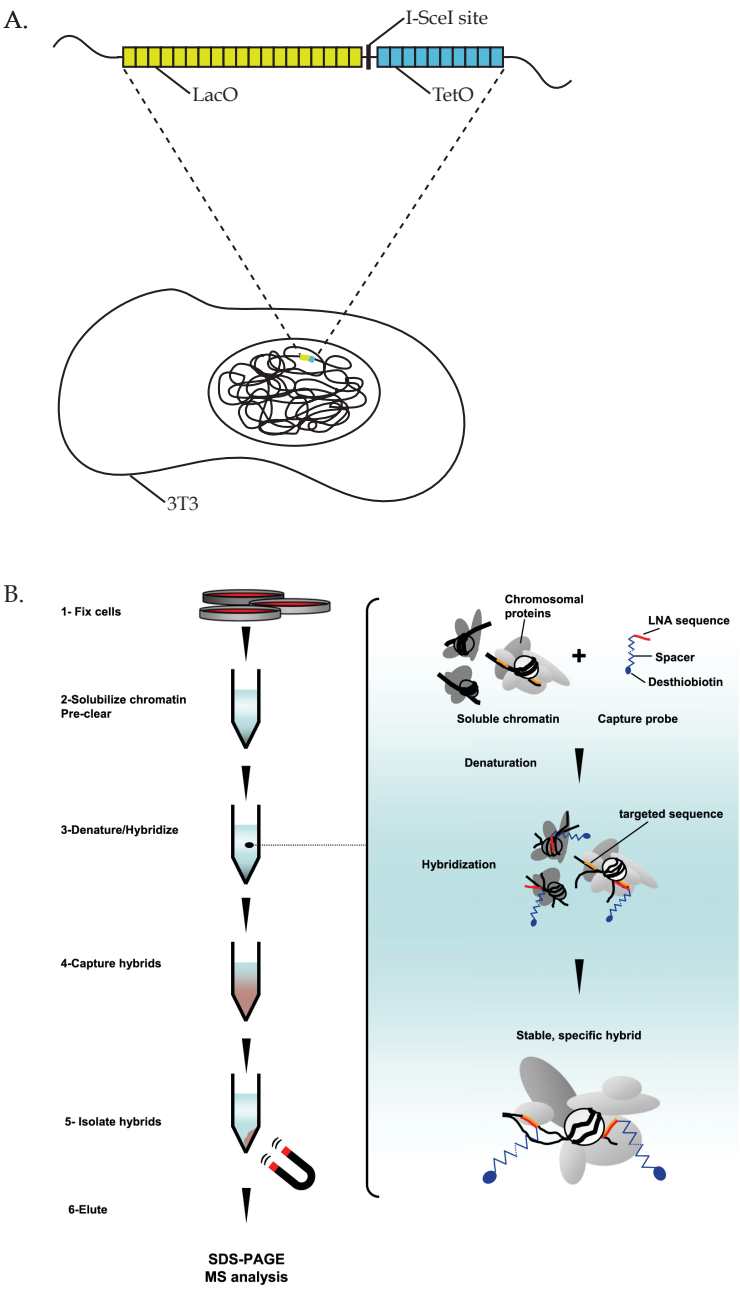

Figure 3 : Système LacO-I-SceI et protocole du PICh[13].

D'après le protocole publié du PICh, chaque expérience requiert la production de $1{,}8 \times 10^{10}$ cellules (18 litres à 1×10^{6} cellules/mL). Pour obtenir cette quantité de cellules,

il est nécessaire de les cultiver en suspension. Notre lignée originelle (NIH 2/4) ne poussant pas en suspension, j'ai généré une lignée HeLa contenant au moins 50 répétitions en tandem de l'unité LacO-I-SceI. Pour créer des CDBs inductibles, nous avons établi une lignée stable contenant la cassette LacO-I-SceI et exprimant l'enzyme I-SceI sous contrôle d'un promoteur inductible à la doxycyline. De cette manière, I-SceI est exprimé et à accès au site uniquement après addition de doxycycline. Nous avons réalisé tout d'abord un PICh à petite échelle où nous avons détecté la présence de γ-H2AX au niveau du site I-SceI coupé, par western blot. Nous avons ensuite effectué des PIChs à grande échelle, malheureusement, dû à un bruit de fond trop important, nous n'avons pas pu détecter de protéines de la RDA ou de la RCDB. Nous avons tenté de contourner le problème en modifiants différents paramètres et en concevant des sondes plus spécifiques. Mais malgré tous nos efforts aucune de nos tentatives n'a été probante, nous avons donc été amené à abandonner ce projet.

Projet B : Le rôle de SET/TAF-1β dans la régulation de la chromatine et la réparation des cassures double brin.

Afin d'identifier de nouvelles protéines de la chromatine impliquées dans la réparation nous avons utilisé une approche méthodique et non biaisée. Nous avons réalisé un criblage grâce à des siRNA de manière à identifier de nouvelles protéines chromatiniennes dont la sous-expression conduit à des CDBs persistantes. Dans ce but, une bibliothèque sur mesure couvrant un grand nombre de protéines en relation avec la chromatine (1 184 gènes) a été utilisée. Des cellules HeLa ont été transfectées avec ces siRNA et des CDBs ont été induites par l'addition d'une drogue radiomimétique, la néocarzinostatine (NCS). Seize heures après induction des dommages, les cellules ont été marquées par immunofluorescence avec les anticorps γH2AX et 53BP1 (fig. 4A). Les protéines dont la sous-expression provoque un signal γH2AX et 53BP1 persistant sont donc impliquées dans la RCDB. Le but de ce projet est d'examiner le rôle de SET, un des résultats positif parmi les plus robustes du criblage, dans la réponse aux dommages de l'ADN et la réparation de l'ADN. En effet, la diminution de l'expression de SET retarde la réparation, ce qui est illustré par un nombre de foyers de 53BP1 augmenté et une intensité de γ-H2AX plus importante (fig. 4B).

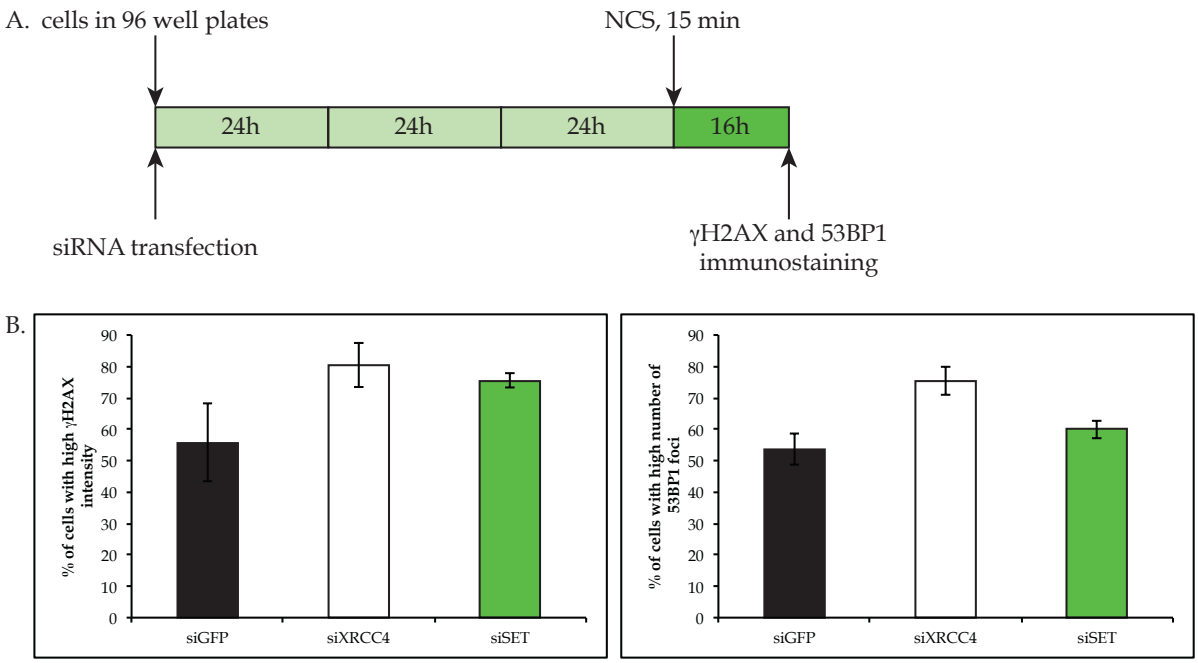

Figure 4 : Identification de SET par un criblage siRNA.

SET/TAF-1β : SET/TAF-1β aussi connue sous le nom de I2PP2A[14] et INHAT[15] a été identifiée à l'origine comme un gène ayant subi une translocation lors de leucémie aiguë indifférenciée[16–18]. SET est une protéine multitâche et a été démontrée comme un puissant inhibiteur de la phosphatase 2A (PP2A)[14], elle possède un domaine NAP (nucleosome assembly protein) et appartient à la famille NAP1 des chaperons d'histone[19] (fig. 5). Par ailleurs elle induirait potentiellement une décondensation de la chromatine[20]. SET/TAF-1β inhibe l'acétylation des histones en masquant leur queue N-terminale[21]. Il n'existe qu'une référence bibliographique connectant SET et les voies de la RCDB[22].

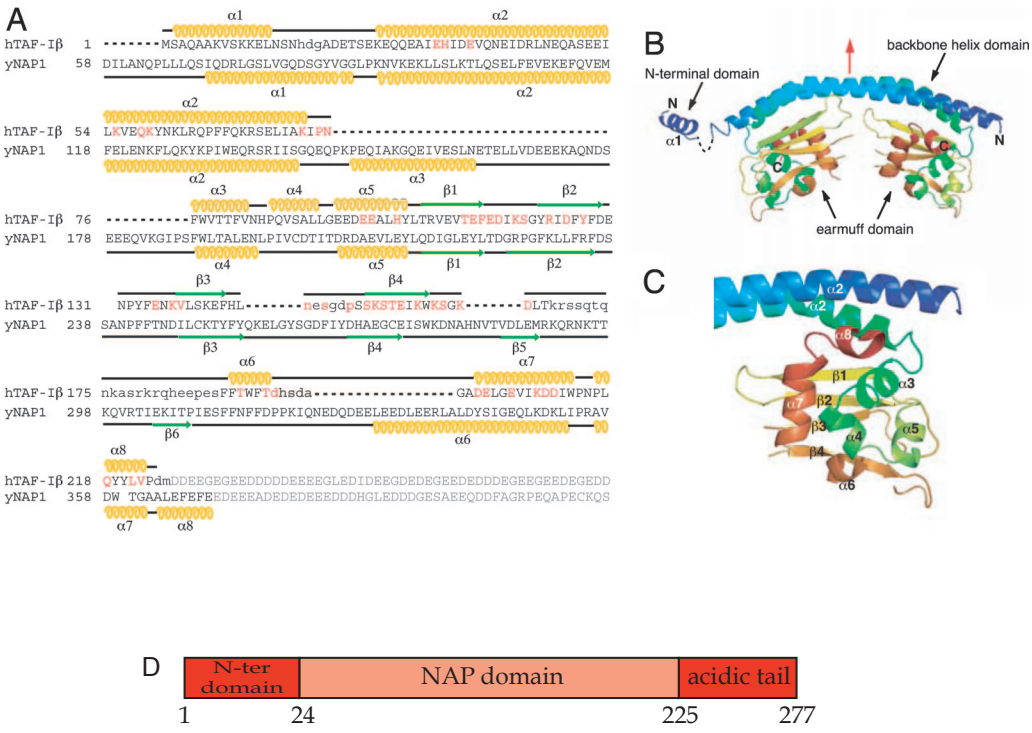

Figure 5 : Structure de la protéine SET[23]

Nous avons validé le résultat du criblage par western blot où le niveau de la protéine γ-H2AX est plus élevé dans les cellules où l'expression de SET est réduite (fig. 6A). Cela peut signifier deux choses, soit il existe des cassures persistantes qui activeraient alors les kinases Chk1 et Chk2, soit le nombre de CDBs est le même. Dans cette hypothèse, la RDA serait plus importante, car la chromatine plus accessible. Nous avons pu observer par western blot que Chk1 et Chk2 ont un niveau équivalent que la protéine SET soit réprimée ou non (fig. 6A), ce qui privilégie la deuxième hypothèse. De plus, les cellules sous-exprimant SET traitées avec une drogue provoquant des cassures double brin survivent mieux que leurs homologues possédant un taux normal de la protéine (fig. 6B). Ces résultats suggèrent une réparation plus efficace due à une structure plus ouverte de la chromatine.

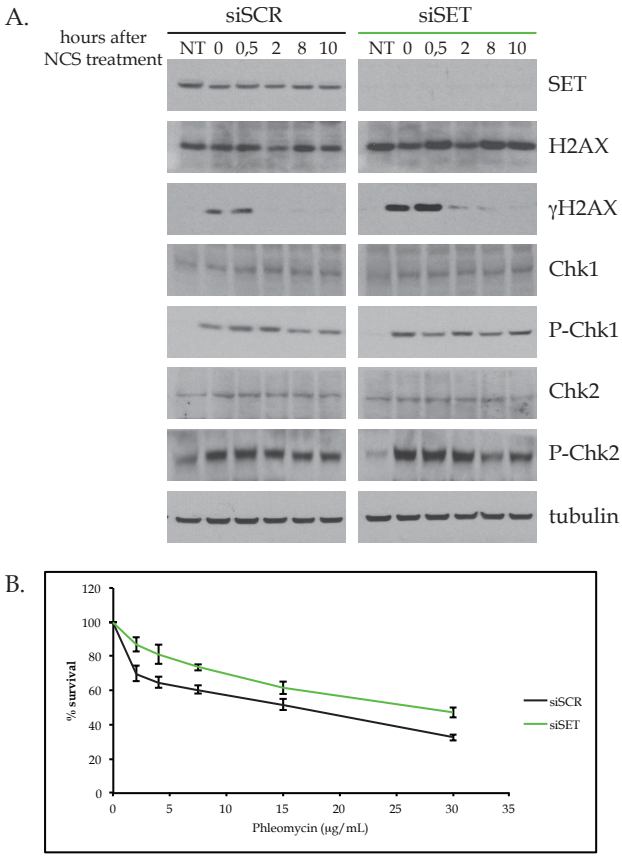

Figure 6 : Effet de la déplétion de SET sur la RDA et la survie.

SET/TAF-1β étant surexprimé dans différents types de cancer[24–28], nous avons donc décidé d'étudier les effets de la surexpression de cette protéine. Pour cela nous avons crée des lignées stables de cellules U2OS surexprimant soit la GFP (U2OS GFP) soit la protéine SET fusionnée à la GFP (U2OS GFP-SET). En premier lieu nous avons analysé le cycle cellulaire de ces cellules lorsqu'elles sont traitées par la NCS et nous avons remarqué que la surexpression de SET retarde la transition des cellules de la phase S à G2 ce qui suggère un problème de stress réplicatif. Le stress réplicatif correspond au blocage de la fourche de réplication, ce qui va entrainer le recrutement de la kinase ATR. Si la réplication ne peut pas redémarrer, la fourche peut évoluer en

cassure double brin qui va être réparée par la RH[29]. Nous avons donc traité ces cellules avec une drogue provoquant le blocage des fourches de réplication par déplétion des nucléotides, l'hydroxyurée (HU). Lorsque SET est surexprimée en présence de cette drogue, la phosphorylation de la protéine RPA (pRPA), essentielle pour la RDA et la RH, est retardée (fig. 7). La surexpression de SET provoque donc un défaut dans la cascade de signalisation du stress réplicatif.

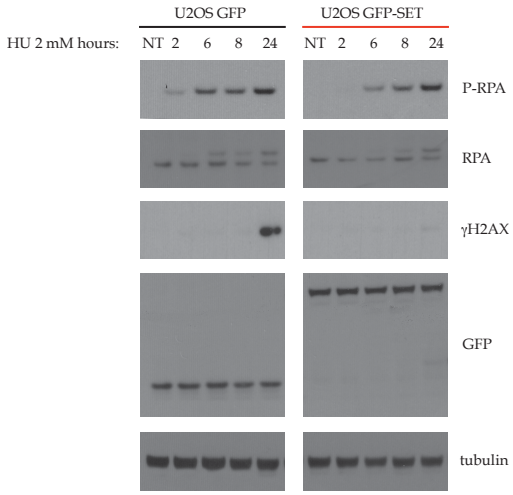

Figure 7 : Effet de la déplétion de SET sur la signalisation du stress réplicatif.

Pour déterminer l'effet de la protéine SET lorsqu'elle est liée à la chromatine, nous avons profité du système LacO-I-SceI décrit ci-dessus permettant de créer une cassure double brin à un locus génomique spécifique[12]. Nous avons créé une protéine de fusion (mCherrylacSET) contenant la protéine fluorescente mCherry, le répresseur de l'opéron lactose (LacI) et la protéine SET, que nous avons transfecté dans des cellules U2OS contenant la cassette LacO-I-SceI (U2OS 19). De cette manière, il est possible d'induire une CDB au niveau de l'accumulation de SET sur l'ADN en co-

transfectant l'enzyme de restriction I-SceI (fig. 8A). Nous avons ensuite marqué ces cellules par immunofluorescence avec des anticorps dirigés contre certaines protéines de la RH (ATR, γH2AX, P-RPA, BRCA1, RAD51 et CtIP) (fig. 8B). Puis, les colocalisations entre les foyers fluorescents formés par l'accumulation de SET et les foyers fluorescents correspondants aux protéines de la RH ont été dénombrées (fig. 8C). Nous utilisons comme contrôle une protéine de fusion mCherrylac ne comportant pas la protéine SET. Cette expérience nous a permis de montrer que lors de l'accumulation de SET près d'une CDB le recrutement des protéines ATR, γH2AX, P-RPA, BRCA1, RAD51 et CtIP est moins efficace (fig. 8C).

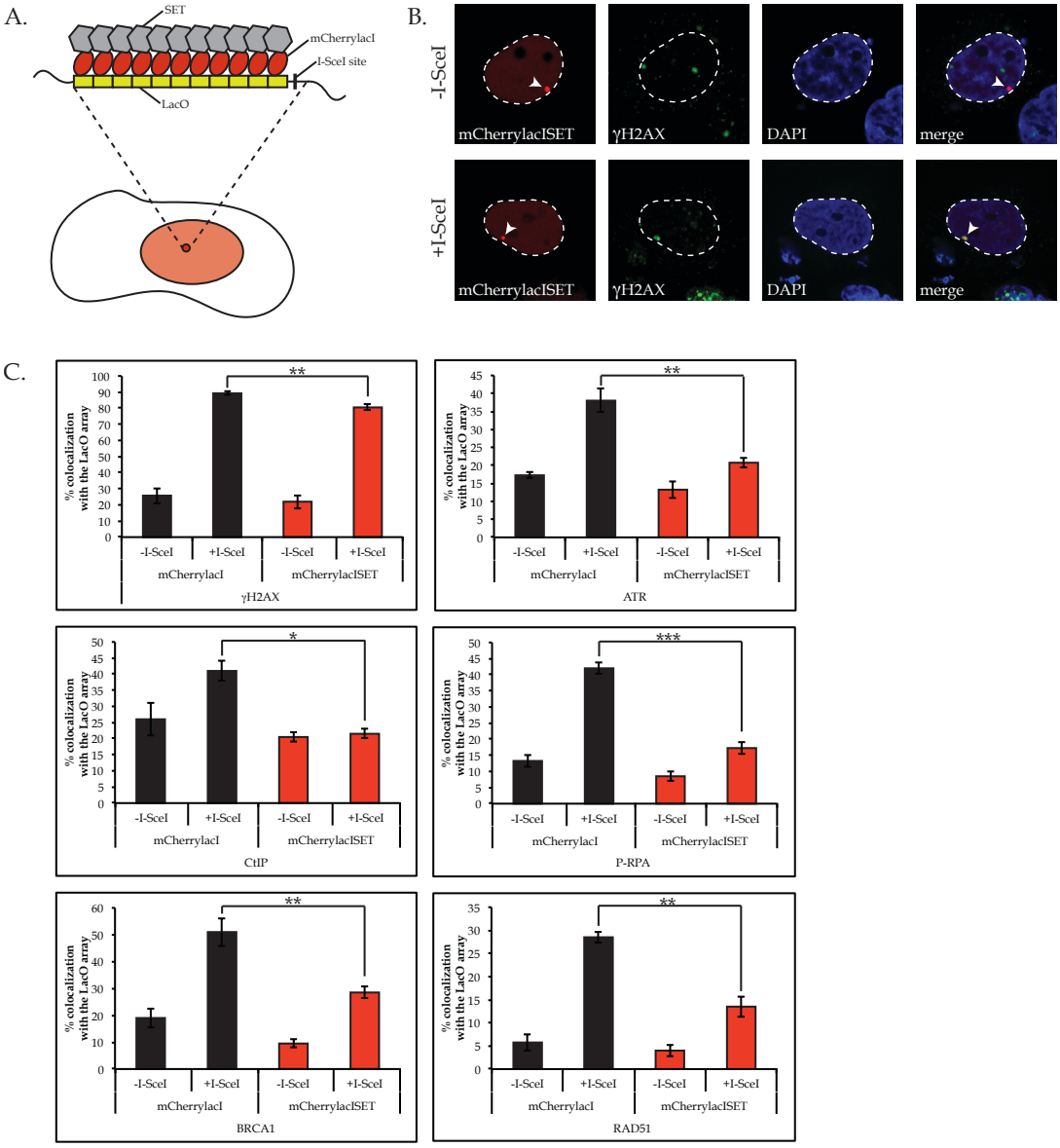

Figure 8 : Effet de l'ancrage de SET à la chromatine sur le recrutement aux dommages de l'ADN des protéines impliquées dans la réparation.

Comme nous avons observé un défaut des protéines de la RH, nous avons ensuite mesuré le taux de recombinaison homologue grâce un système cellulaire basé sur la reconstitution d'une séquence GFP (U2OS DR-GFP). Ces cellules contiennent

une séquence d'ADN comportant le gène de la GFP muté par un site I-SceI suivi d'une séquence du même gène tronqué mais correspondant à la partie mutée. Si un événement de recombinaison a eu lieu, la séquence du gène va être reconstituée et la cellule deviendra GFP positive[30], ce que nous pouvons quantifier par cytométrie en flux (fig. 9A). Lors de la surexpression de SET, le taux de RH est bien plus faible, ce qui est en accord avec les résultats précédents (fig. 9B). Par ailleurs, la survie des cellules surexprimant SET est réduite de façon dramatique en présence de camptothécine, une drogue bloquant la topoisomérase I et créant ainsi du stress réplicatif (fig. 9C). D'après ces résultats, nous pouvons conclure que, la surexpression de SET provoque une anomalie dans la cascade de signalisation du stress réplicatif et une inhibition de la recombinaison homologue.

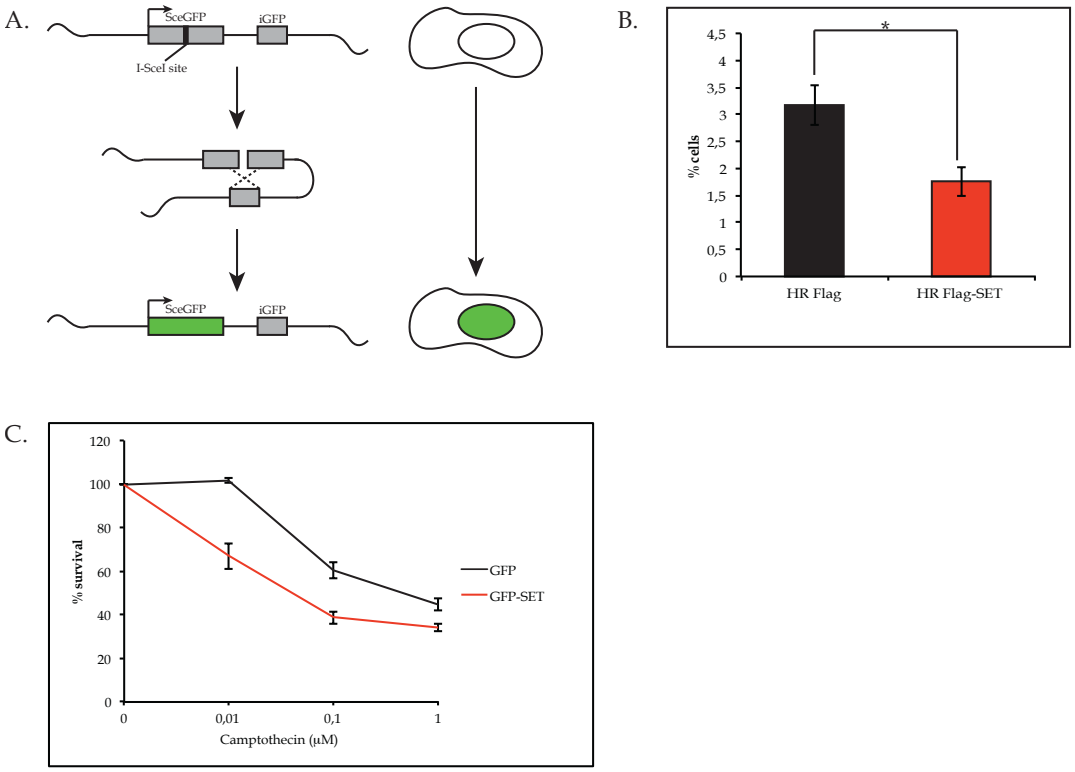

Figure 9 : Effet de la surexpression de SET sur l'efficacité de la recombinaison homologue et sur la survie cellulaire.

Afin de déterminer le mécanisme d'action de la protéine SET/TAF-1β, nous nous sommes intéressés aux protéines interagissant avec celle-ci. Des données de spectrométrie de masse obtenue par le Dr. Kum Kum Khanna montrent que SET interagit avec la protéine KAP1 (KRAB associated protein 1). KAP1 est une protéine hétérochromatique qui a un rôle important dans la RDA. En effet, lors d'une CDB dans l'hétérochromatine, KAP1 doit être phosphorylée par ATM afin d'être libérée de la chromatine et permettre la décondensation de celle-ci pour permettre aux protéines de la RDA d'accéder à la CDB[31]. Par ailleurs, il a été démontré que cette protéine est recrutée de façon transitoire aux CDBs afin de faciliter la RH[32]. Nous avons validé l'interaction de SET et KAP1 par co-immunoprecipitation (fig. 10A). Dans le but de déterminer le rôle de l'interaction de KAP1 et SET plus spécifiquement au niveau de la chromatine, nous avons éliminé les protéines nucléaires solubles dans des cellules surexprimant ou non la protéine SET. Puis, nous avons observé par immunofluorescence la localisation et l'intensité de la protéine KAP1. KAP1 est bien plus présente dans la chromatine des cellules surexprimant SET (fig. 10B). De plus KAP1 est recruté de façon plus abondante au niveau des CDBs lorsque SET est ancré à la chromatine à proximité de celles-ci (système LacO-I-SceI) (fig. 10C). Enfin, les mêmes phénomènes sont observés pour la marque hétérochromatique H3K9me3 et la protéine HP1 (hétérochromatin protein 1), ce qui suggère que la surexpression de SET provoque une hétérochromatinisation de l'ensemble de la chromatine.

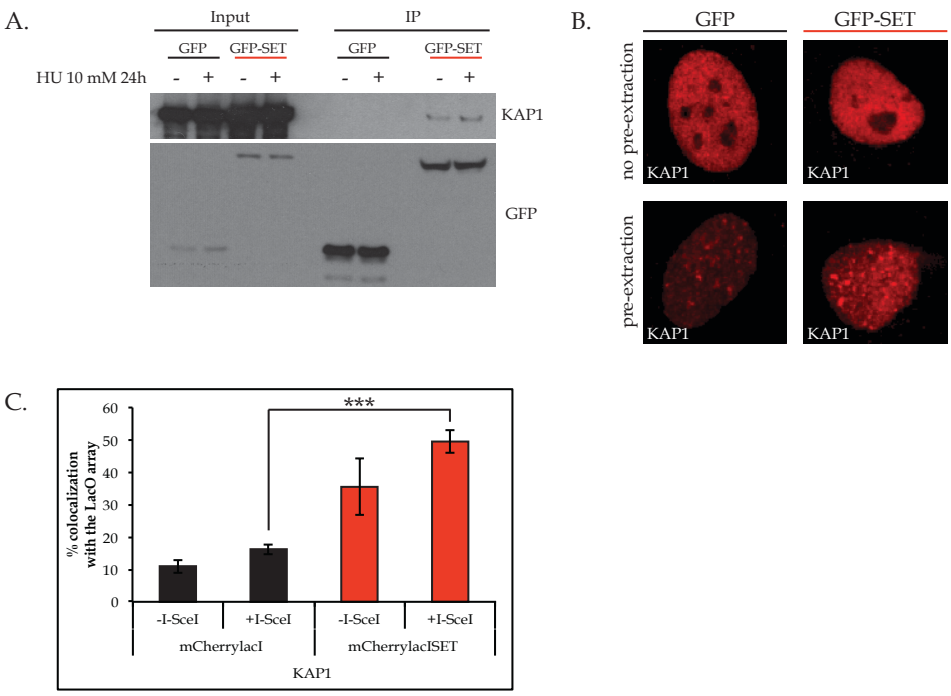

Figure 10 : Effet de la surexpression de SET sur la rétention de KAP1 à la chromatine.

Nous proposons le modèle suivant. En temps normal, une fraction de l'ensemble des protéines SET présentes dans le noyau se trouve liée à la chromatine. Suite à une surexpression de SET, une quantité plus importante se retrouve fixée sur la chromatine, résultant dans une rétention massive de KAP1. KAP1 engendre ensuite le recrutement de la lysine methyltransferase SETDB1, qui trimethyle H3 sur la lysine 9 (H3K9me3), aboutissant ainsi à une hétérochromatinisation générale. Dans ce contexte, suite à l'induction d'une CDB, la rétention de KAP1 ne permet pas l'accès des protéines de la RCDB au site du dommage (fig. 11).

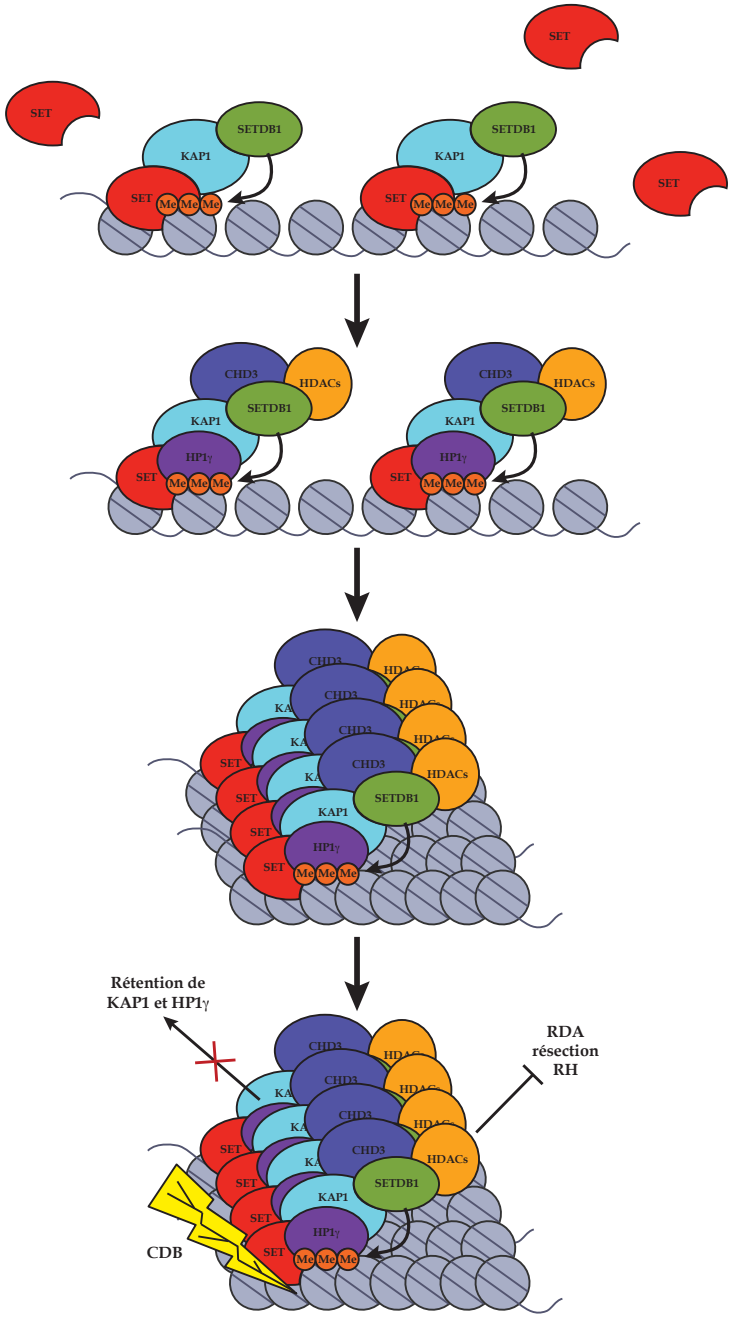

Figure 11 : Modèle d'action de la protéine SET.

Nous souhaitons par la suite étudier de manière plus approfondie la mobilité de KAP1 par des expériences de FRAP (fluorescence recovery after photobleaching). Nous aimerions déterminer la distribution globale de KAP1 et H3K9me3 sur l'ADN

quand SET est surexprimée par ChIP-seq (chromatin immunoprecipitation sequencing). Nous chercherons également à déterminer si SETDB1 est bien responsable des niveaux plus élevés de triméthylation de H3K9. Afin de vérifier que la chromatine est plus compactée quand SET est surexprimée nous souhaitons effectuer une analyse de la relaxation de la chromatine par digestion à la nucléase micrococcale (MNase). Nous souhaitons aussi déterminer le taux de NHEJ grâce un système cellulaire spécifique[33]. Enfin, puisque SET est surexprimée dans certains cancers, nous aimerions examiner si une corrélation existe entre les niveaux de SET et H3K9me3 dans des lymphocytes T de patients atteint de leucémie. Par ailleurs, l'ancrage de KAP1 à la chromatine par SET ne permet pas son recrutement aux CDBs. En effet, des résultats préliminaires de micro-irradiation laser permettant de créer des CDBs montrent un recrutment moins dynamique de KAP1 aux cassures double brin.

Les résultats obtenus à propos de l'implication de SET dans la réponse aux dommages de l'ADN et leur réparation sont particulièrement intéressants car cette protéine est surexprimée dans certains cancers. SET pourrait être le talon d'Achille de ce type de tumeurs et être la cible de chimiothérapies utilisant le stress réplicatif comme arme contre ces cellules cancéreuses.

Références

1. Elliott, B. & Jasin, M. Double-strand breaks and translocations in cancer. *Cell. Mol. Life Sci. CMLS* **59,** 373–385 (2002).

2. Weterings, E. & van Gent, D. C. The mechanism of non-homologous end-joining: a synopsis of synapsis. *DNA Repair* **3,** 1425–1435 (2004).

3. Bartek, J. & Lukas, J. DNA damage checkpoints: from initiation to recovery or adaptation. *Curr. Opin. Cell Biol.* **19,** 238–245 (2007).

4. Rouse, J. & Jackson, S. P. Interfaces between the detection, signaling, and repair of DNA damage. *Science* **297,** 547–551 (2002).

5. Lukas, J. & Bartek, J. Watching the DNA repair ensemble dance. *Cell* **118,** 666–668 (2004).

6. Nagy, Z. & Soutoglou, E. DNA repair: easy to visualize, difficult to elucidate. *Trends Cell Biol.* **19,** 617–629 (2009).

7. Misteli, T. & Soutoglou, E. The emerging role of nuclear architecture in DNA repair and genome maintenance. *Nat. Rev. Mol. Cell Biol.* **10,** 243–254 (2009).

8. Altaf, M., Saksouk, N. & Côté, J. Histone modifications in response to DNA damage. *Mutat. Res.* **618,** 81–90 (2007).

9. Kruhlak, M. J., Celeste, A. & Nussenzweig, A. Spatio-temporal dynamics of chromatin containing DNA breaks. *Cell Cycle Georget. Tex* **5,** 1910–1912 (2006).

10. Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. M. DNA Double-stranded Breaks Induce Histone H2AX Phosphorylation on Serine 139. *J. Biol. Chem.* **273,** 5858–5868 (1998).

11. Price, B. D. & D'Andrea, A. D. Chromatin remodeling at DNA double-strand breaks. *Cell* **152,** 1344–1354 (2013).

12. Soutoglou, E. *et al.* Positional stability of single double-strand breaks in mammalian cells. *Nat. Cell Biol.* **9,** 675–682 (2007).

13. Déjardin, J. & Kingston, R. E. Purification of Proteins Associated with Specific Genomic Loci. *Cell* **136,** 175–186 (2009).

14. Li, M., Guo, H. & Damuni, Z. Purification and characterization of two potent heat-stable protein inhibitors of protein phosphatase 2A from bovine kidney. *Biochemistry (Mosc.)* **34,** 1988–1996 (1995).

15. Gamble, M. J., Erdjument-Bromage, H., Tempst, P., Freedman, L. P. & Fisher, R. P. The Histone Chaperone TAF-I/SET/INHAT Is Required for Transcription In Vitro of Chromatin Templates. *Mol. Cell. Biol.* **25,** 797–807 (2005).

16. Von Lindern, M. *et al.* Can, a putative oncogene associated with myeloid leukemogenesis, may be activated by fusion of its 3' half to different genes:

characterization of the set gene. *Mol. Cell. Biol.* **12,** 3346–3355 (1992).

17. Adachi, Y., Pavlakis, G. N. & Copeland, T. D. Identification and characterization of SET, a nuclear phosphoprotein encoded by the translocation break point in acute undifferentiated leukemia. *J. Biol. Chem.* **269,** 2258–2262 (1994).

18. Nagata, K. *et al.* Cellular Localization and Expression of Template-Activating Factor I in Different Cell Types. *Exp. Cell Res.* **240,** 274–281 (1998).

19. Kawase, H. *et al.* NAP-I is a functional homologue of TAF-I that is required for replication and transcription of the adenovirus genome in a chromatin-like structure. *Genes Cells Devoted Mol. Cell. Mech.* **1,** 1045–1056 (1996).

20. Matsumoto, K., Nagata, K., Miyaji-Yamaguchi, M., Kikuchi, A. & Tsujimoto, M. Sperm Chromatin Decondensation by Template Activating Factor I through Direct Interaction with Basic Proteins. *Mol. Cell. Biol.* **19,** 6940–6952 (1999).

21. Seo, S. B. *et al.* Regulation of histone acetylation and transcription by INHAT, a human cellular complex containing the set oncoprotein. *Cell* **104,** 119–130 (2001).

22. Almeida, L. O. *et al.* SET overexpression decreases cell detoxification efficiency: ALDH2 and GSTP1 are downregulated, DDR is impaired and DNA damage accumulates. *FEBS J.* **279,** 4615–4628 (2012).

23. Muto, S. *et al.* Relationship between the structure of SET/TAF-Iβ/INHAT and its histone chaperone activity. *Proc. Natl. Acad. Sci.* **104,** 4285–4290 (2007).

24. Sirma Ekmekci, S. *et al.* SET oncogene is upregulated in pediatric acute lymphoblastic leukemia. *Tumori* **98,** 252–256 (2012).

25. Cristóbal, I. *et al.* Overexpression of SET is a recurrent event associated with poor outcome and contributes to protein phosphatase 2A inhibition in acute myeloid leukemia. *Haematologica* **97,** 543–550 (2012).

26. Jiang, Q., Zhang, C., Zhu, J., Chen, Q. & Chen, Y. The set gene is a potential oncogene in human colorectal adenocarcinoma and oral squamous cell carcinoma. *Mol. Med. Reports* **4,** 993–999 (2011).

27. Christensen, D. J. *et al.* SET oncoprotein overexpression in B-cell chronic lymphocytic leukemia and non-Hodgkin lymphoma: a predictor of aggressive disease and a new treatment target. *Blood* **118,** 4150–4158 (2011).

28. Li, C. *et al.* Quantitative Proteomics Reveal up-regulated Protein Expression of the SET Complex Associated with Hepatocellular Carcinoma. *J. Proteome Res.* **11,** 871–885 (2012).

29. Allen, C., Ashley, A. K., Hromas, R. & Nickoloff, J. A. More forks on the road to replication stress recovery. *J. Mol. Cell Biol.* **3,** 4–12 (2011).

30. Miller, K. M. *et al.* Human HDAC1 and HDAC2 function in the DNA-damage response to promote DNA nonhomologous end-joining. *Nat. Struct. Mol. Biol.* **17,** 1144–1151 (2010).

31. Cann, K. L. & Dellaire, G. Heterochromatin and the DNA damage response: the need to relax. *Biochem. Cell Biol. Biochim. Biol. Cell.* **89,** 45–60 (2011).

32. Baldeyron, C., Soria, G., Roche, D., Cook, A. J. L. & Almouzni, G. HP1alpha recruitment to DNA damage by p150CAF-1 promotes homologous recombination repair. *J. Cell Biol.* **193,** 81–95 (2011).

33. Rass, E. *et al.* Role of Mre11 in chromosomal nonhomologous end joining in mammalian cells. *Nat. Struct. Mol. Biol.* **16,** 819–824 (2009).

Anne-Sophie HOFFBECK

Chromatin structure and DNA repair

Résumé

Notre génome est continuellement endommagé par des agents provoquant des lésions de l'ADN. Les cassures doubles brins de l'ADN (CDBs) sont les lésions les plus dangereuses. En effet, une CDB mal réparée peut mener à des aberrations de l'ADN pouvant conduire à l'apparition d'un cancer. Dans le but d'éviter les effets délétères des CDBs, nos cellules ont développé une voie de signalisation, nommée réponse aux dommages de l'ADN (RDA), permettant la détection des cassures et l'activation des points de contrôle du cycle cellulaire afin d'arrêter le cycle pendant la réparation des CDBs. Une des caractéristiques principales de la RDA est l'accumulation d'un grand nombre de facteurs sur l'ADN autour de la cassure, formant un foyer visible en microscopie. Cependant, l'efficacité de réparation de l'ADN est entravée par la structure condensée de la chromatine environnante. Les mécanismes de réparation de l'ADN surmontent ce problème en recrutant de nombreuses protéines permettant le réarrangement de la chromatine afin de faciliter la réparation. Le but de mon travail de thèse est d'identifier de nouvelles protéines impliquées dans le remodelage de la chromatine autour des CDBs. D'une part nous avons pour but d'identifier le protéome complet d'un foyer de réparation de l'ADN grâce à la technique PICh (Proteomics of Isolated Chromatin loci). D'autre part, nous étudions le rôle de l'oncoprotéine SET/TAF-1β, que nous avons identifié lors d'un criblage siRNA réalisé dans le but de découvrir de nouveaux facteurs chromatiniens impliqués dans la réparation des CDBs.

Mots clés: Réponse aux dommages de l'ADN, Cassure double brin, Réparation de l'ADN, PICh, Recombinaison homologue, Stress réplicatif, SET, TAF-Iβ

Abstract

Various DNA damaging agents, that can cause DNA lesions, assault constantly our genome. The most deleterious DNA lesions are the breaks occurring in both strands of DNA (Double stand breaks: DSBs). Inefficient repair of DSBs can lead to aberrations that may induce cancer. To avoid these deleterious effects of DSBs, cells have developed signalling cascades which entail detection of the lesions and spreading of the signal that leads to arrest in cell cycle progression and efficient repair. A major characteristic of DNA damage response (DDR) is the accumulation of a vast amount of proteins around the DSBs that are visible in the cell as DNA damage foci. However, efficient DNA repair is hampered by the fact that genomic DNA is packaged into chromatin. The DNA repair machinery overcomes this condensed structure to access damaged DNA by recruiting many proteins that remodel chromatin to facilitate efficient repair. The aim of my PhD work is to identify novel proteins involved in the DDR and/or the remodelling of chromatin surrounding DSBs. On one hand, we take advantage of the PICh (Proteomics of Isolated Chromatin loci) technique and we aim to identify the entire proteome of DNA repair foci. On the other hand, we study the role of the oncogene SET/TAF-Iβ, a major hit of a siRNA screen performed to identify novel chromatin related proteins that play role in repair of DSBs.

Keywords: DNA damage response, Double strand break, DNA repair, PICh, Homologous recombination, Replication stress, SET, TAF-Iβ