

DNA damage in transcriptional regulation

A thesis submitted to the University of Kent for the degree of

PhD in Cell Biology at the Division of Natural Sciences

2022

Ane Stranger

School of Biosciences

I Declaration

No part of this thesis has been submitted in support of an application for any degree or qualification at the University of Kent or any other University or institute of learning.

Ane Stranger

December 2022

II Acknowledgements

I would like to thank my supervisor Dr Peter Ellis, for offering this exciting PhD project.

I would like to thank him for helping me develop as a scientist, helping me interpret the (often) weird and (sometimes) wonderful data I have generated, and always coming up with new ideas when things haven't gone to plan. I would like to thank Professor Darren Griffin, all my friends in the 'reprogroup', and in the School of Biosciences for all their support, much needed coffee breaks and many, many pints at Spoons. I would also like to thank my co-supervisor Dr Tim Fenton, for his support and guidance developing the computational aspects of this project, and the Fenton group for inviting me to be part of their lab meetings, and for all their kind help. I also owe a big thank you to Dr Gaëlle Legube, for letting me visit her lab, where I learned many skills which have been essential to this project.

I am eternally grateful to my family for believing in me, and always being so excited about even the smallest achievements. And to Joe, for celebrating every step along the long and winded road of this PhD, for supporting me and always having my back.

III Contents

I Declaration	2
II Acknowledgements.....	3
III Contents	4
IV Abbreviations	8
V Abstract.....	10
1.0 Chapter 1: Introduction	11
1.1 DNA organisation	11
1.1.1 Chromatin	11
1.1.2 3D chromatin organisation	13
1.1.3 Cellular organisation by phase separation.....	15
1.1.4 R-loops	15
1.1.5 Histones and histone modifications.....	17
1.1.6 Methods to investigate DNA-protein interactions	18
1.2 DNA damage	21
1.2.1 Causes of DNA breaks	21
1.2.2 DNA damage repair.....	22
1.2.2.1 DNA repair pathways	22
1.2.2.2 Chromatin remodelling in DNA repair	26
1.2.2.3 Chromatin state affecting choice between HR and NHEJ.....	27
1.3 H2A histone variant H2AX and its involvement in DNA damage repair	28
1.3.1 H2A variant H2AX is a minor constituent of chromatin.....	28
1.3.2 γ H2AX: H2AX is phosphorylated at ser139 in response to DNA damage	29
1.3.3 γ H2AX recruits DNA damage repair proteins.....	31
1.3.4 H2AX is constitutively phosphorylated at tyr142	32
1.3.5 H2AX tyr142 phosphorylation is lost concurrently with ser139 phosphorylation.....	33
1.3.6 Blocking H2AX tyr142 phosphorylation perturbs the DNA damage response	33
1.3.7 Di- γ H2AX interacts with pro-apoptotic factors.....	36
1.3.8 The di-phosphorylated form of H2AX is proposed to be present during early phases of the DNA damage response	38
1.4 Induction of DNA breaks.....	39
1.4.1 DNA break induction methods.....	39

1.4.2 AsiSI-ER DNA double-strand break induction	39
1.5 The nuclear hormone receptor estrogen receptor.....	40
1.5.1 The role of estrogen receptor in regulation of gene expression and in cancer....	40
1.5.2 Estrogen and ER lead to formation of DNA damage.....	44
1.6 APOBEC3B.....	45
1.6.1 APOBEC3B is a cytidine deaminase.....	45
1.6.2 APOBEC3B is involved in estrogen-target gene expression.....	46
1.6.3 APOBEC3B activity leads to formation of DSBs and γ H2AX.....	47
1.6.4 APOBEC3B binds R-loops to regulate ER activity	48
1.7 Key questions addressed in this thesis	52
2.0 Chapter 2: Methods	53
2.1 Cell culture	53
2.2 The novel γ H2AX and di- γ H2AX antibodies used in this study	54
2.3 DNA damage induction	55
2.4 Immunofluorescence (IF).....	55
2.5 Microscopy and imaging	56
2.6 Cell lysis and protein extraction.....	56
2.7 Protein quantitation.....	57
2.8 SDS-PAGE and western blot (WB).....	57
2.9 Dot blot	58
2.10 Chromatin Immunoprecipitation (ChIP)	59
2.11 Cleavage Under Targets and Tagmentation (CUT&Tag)	61
2.11.1 Preparing testis samples	62
2.11.2 Preparing DivA cells	62
2.11.3 Antibody recognition of targets and tagmentation.....	63
2.11.4 Library amplification	65
2.11.5 PCR product purification.....	65
2.12 Quantitative real-time Polymerase Chain Reaction (qPCR).....	66
2.13 Bioanalyzer.....	66
2.14 Next-generation sequencing.....	67
2.15 ChIP-seq and BLESS analysis	67
2.16 CUT&Tag-seq analysis.....	71
3.0 Chapter 3: CUT&Tag can be used to study γ H2AX formation upon DNA damage	77

3.1 Introduction	77
3.2 Results.....	79
3.2.1 CUT&Tag-seq identifies chromosome-wide histone marks.....	79
3.2.2 OHT-induced γ H2AX formation in DivA cells	82
3.2.3 qPCR validation of CUT&Tag efficiency.....	84
3.2.4 Quality control and genome alignment of γ H2AX CUT&Tag-seq files.....	86
3.2.5 CUT&Tag-seq can be used to identify γ H2AX near DSBs.....	88
3.3 Discussion.....	95
4.0 Chapter 4: Investigating the role of doubly phosphorylated H2AX (serine 139 and tyrosine 142) in the DNA damage response.....	98
4.1 Introduction	98
4.2 Results.....	101
4.2.1 Testing specificity of H2AX-antibodies against differentially phosphorylated H2AX C-terminal peptides	101
4.2.2 Investigating di- γ H2AX and γ H2AX in DNA damaged cells by immunofluorescence staining.....	105
4.2.3 Chromatin Immunoprecipitation to investigate di- γ H2AX at DNA double-strand breaks in DivA cells	112
4.2.4 The di- γ H2AX and H2AX tyr142ph antibodies show signal via western blot in an H2AX knock-out cell line	117
4.3 Discussion.....	118
5.0 Chapter 5: APOBEC3B is recruited to promoters and enhancers in presence of estrogen, and binds estrogen response pathway genes	121
5.1 Introduction	121
5.1.1 APOBEC3B cytidine deamination regulates expression of estrogen target genes	121
5.1.2 APOBEC3B binds chromatin together with/without ER	122
5.2 Results.....	123
5.2.1 Quality control check of sequencing data by FastQC	123
5.2.2 Sequence alignment to the human genome	125
5.2.3 Calling ChIP-seq peaks using MACS2	126
5.2.4 Peak calling using the novel tool LanceOtron.....	131
5.2.5 A3B and ER binds estrogen-target genes in presence of estrogen.....	142
5.2.6 Peak region annotation of A3B/ER binding sites	147
5.2.7 A3B overlaps significantly with transcriptional start sites with/without ER.....	148

5.2.8 Gene ontology enrichment of peak regions	151
5.2.9 <i>De novo</i> motif analysis: A3B with/without ER binds the ERE	152
5.2.10 A3B and ER are enriched at promoters and enhancers in presence of estrogen	157
5.2.11 A3B-only/A3B & ER/ER-only bind genes belonging to similar biological pathways	160
5.2.12 Functional enrichment analysis on genes whose enhancers are bound by A3B and ER	164
5.3 Discussion.....	167
5.3.1 Re-analysis of A3B and ER CHIP-seq datasets previously analysed by Periyasamy <i>et al</i>	167
5.3.2 γ H2AX forms narrow peaks in response to A3B/estrogen-induced DNA damage	169
5.3.3 Disagreement in A3B binding regions discovered by the peak callers MACS2 and LanceOtron.....	170
5.3.4 A3B binds genes related to the estrogen pathway independent of ER.....	171
5.3.5 Further steps	172
6.0 Chapter 6: Discussion	174
6.1 Thesis outline	174
6.2 CUT&Tag detects γ H2AX formation in response to DNA damage using fewer cells and with less background signal than CHIP-seq.....	175
6.3 The novel di- γ H2AX antibody is not specific, and binds other targets than H2AX	178
6.4 Estrogen recruits APOBEC3B to enhancers and promoters, where it binds genes involved in the estrogen response pathway.....	181
7.0 References	186

IV Abbreviations

53BP1	p53 binding protein 1
AP site	Apurinic/apurimidinic site
APOBEC3 A3(B)	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide 3(B)
AR	Androgen receptor
ATM	Ataxia-telangiectasia mutated
ATR	Ataxia telangiectasia and Rad3-related
BER	Base excision repair
BLESS	Breaks labelling, enrichment on streptavidin, and next- generation sequencing
BRCA1/BRCA2	Breast cancer gene 1/2
BRG1	Brahma-related gene-1
ChIP(-seq)	Chromatin immunoprecipitation (-sequencing)
CHPT1	Choline phosphotransferase 1
CTCF	CCCTC-binding factor
CTSD	Cathepsin D
CUT&Tag	Cleavage under targets and tagmentation
DAPI	4',6-diamidino-2-phenylindole
DICER1	Dicer 1, ribonuclease III
DlvA	DSB inducible via AsiSI
DNA-PKcs	DNA-dependent protein kinase, catalytic subunit
DSB	Double-strand break
EGR3	Early growth response 3
ER, ER α , ER β	Estrogen receptor, estrogen receptor α/β
ERE	Estrogen response element
EYA	Eyes-absent
FDR	False discovery rate
FOS	Fos proto-oncogene, AP-1 transcription factor subunit
FOXA1	Forkhead box A1
GATA3	GATA binding protein 3
GREB1	Growth regulating estrogen receptor binding 1
HR	Homologous recombination
IF	Immunofluorescence
IP	Immunoprecipitation

ISWI	Imitation switch
JNK1	Jun N-terminal protein kinase 1
KO	Knock-out
LINE	Long interspersed nuclear elements
LTR	Long terminal repeat
MAX	MYC associated factor X
MCPH1	Microcephalin 1
MDC1	Mediator of DNA damage checkpoint 1
MRN	Mre11, Rad50 and Nbs1 complex
NHEJ	Non-homologous end joining
OHT	4-hydroxytamoxifen
(q)PCR	(Quantitative) polymerase chain reaction
PDZK1	PDZ domain containing 1
RARA	Retinoic acid receptor alpha
RNA pol II	RNA polymerase II
RNF8	Ring finger protein 8
SINE	Short interspersed nuclear elements
SSB	Single strand break
ssDNA	Single-strand DNA
SWI/SNF	Swlthc/sucrose non-fermentable
TAD	Topologically associated domain
TF	Transcription factor
TFF1	Trefoil factor 1
TSS	Transcriptional start site
UCSC	University of California Santa Cruz
UNG	Uracil DNA glycosylase
WB	Western blot
WICH complex	WSTF–ISWI ATP-dependent chromatin-remodelling complex
WSTF	Williams syndrome transcription factor
WT	Wild type
XRCC4	X-ray repair cross-complementing protein 4
ZNF506	Zinc finger protein 506

V Abstract

The genome is continuously exposed to DNA damage, the most genotoxic of which are double-strand DNA breaks (DSBs). Their presence is signalled to the cell through phosphorylation of the histone variant H2AX at ser139, forming γ H2AX, which recruits DNA repair factors. γ H2AX has been extensively studied using ChIP-seq; here, a novel ChIP-based technique, CUT&Tag, was used to detect γ H2AX formation in DNA damage-inducible cells and the results were analysed alongside publicly available γ H2AX ChIP-seq datasets. CUT&Tag required fewer cells and sequencing reads than ChIP-seq, and yet exhibited improved signal:noise. H2AX can also be phosphorylated at tyr142, which is lost upon DNA damage. It is thought that H2AX phosphorylated at ser139 and tyr142 (di- γ H2AX) exists transiently in cells, however, that a sustained di- γ H2AX signal is pro-apoptotic. Di- γ H2AX was investigated using a newly developed antibody by immunofluorescence, western blot and ChIP-qPCR; however, it was later discovered that the antibody was non-specific. DNA damage also plays a role in gene regulation. Estrogen receptor (ER) is bound by its ligand, estrogen, upon which it activates its target genes. In the ER+ breast cancer cell line MCF7, ER recruits the cytidine deaminase APOBEC3B (A3B), which deaminates cytosine, leading to DSBs, repair of which facilitates activation of target genes. To investigate A3B binding further, in the presence and absence of ER, existing ChIP-seq datasets were re-analysed; sites of recruitment were categorised based on ER coincidence and each category characterised in detail, based on the enriched motifs, targeted genes and gene ontology. Sites occupied by both A3B and ER were strongly enriched for the ER motif and estrogen-regulated genes. Interestingly, sites at which A3B binds without ER were also enriched for the ER motif and estrogen-associated genes, suggesting that even sites at which A3B binds independently of ER are associated with the estrogen response.

1.0 Chapter 1: Introduction

There are two main aspects to this thesis. In the first I tested the use of novel methodology for detecting histone modifications associated with DNA damage, and searched for the presence of a particular modification hypothesised to be present during the early stage of the DNA damage response. In the second part, following COVID19-related interruption to lab work, I used bioinformatic methods to study the location of APOBEC3B cytidine deaminase binding to chromatin, which induce programmed DNA double-strand breaks during activation of estrogen-responsive promoters. Collectively my work addresses the chromatin events occurring in response to DNA damage, and how these may be invoked deliberately as part of transcriptional regulation.

1.1 DNA organisation

1.1.1 Chromatin

The average mammalian cell has a nucleus of 5-10µm in diameter, whereas the linear, uncondensed human genome is approximately 2 metres long. DNA therefore has to be extensively packaged and condensed in order to fit inside the relatively small nucleus (Li and Zhu, 2015). To condense the genome, DNA molecules are wrapped around histone proteins, together forming nucleosomes, which represent the fundamental unit of chromatin (figure 1.1) (Li and Zhu, 2015). A single nucleosome consists of around 147 bp of DNA wrapped 1.7 times around a histone octamer (figure 1.1). The histone octamer is made up of two copies each of H2A, H2B, H3 and H4 (Li and Zhu, 2015). Nucleosomes are interspersed by linker-DNA, giving it an appearance of – and its name – ‘beads on a string’ (Bednar *et al.*, 1998). This chromatin structure has a diameter of 11 nm. Linker histone H1 proteins also bind the nucleosomes, further aiding packaging (Bednar *et al.*, 1998; Li and Zhu, 2015).

In addition to packaging the DNA so that it fits inside the nucleus, chromatin has two main purposes. 1) To control the accessibility of DNA to other freely-diffusible binding factors. The precise location of nucleosomes will determine whether specific DNA sequences are exposed for binding, or not. This is controlled by nucleosome remodelling factors that can deposit, remove or move nucleosomes. 2) To act as a signalling platform regulating aspects of DNA metabolism including transcription and DNA repair. This is achieved by covalent modification of the histone tails, which protrude from the core of the nucleosome and carry modifiable residues (figure 1.1, discussed below in section 1.1.5).

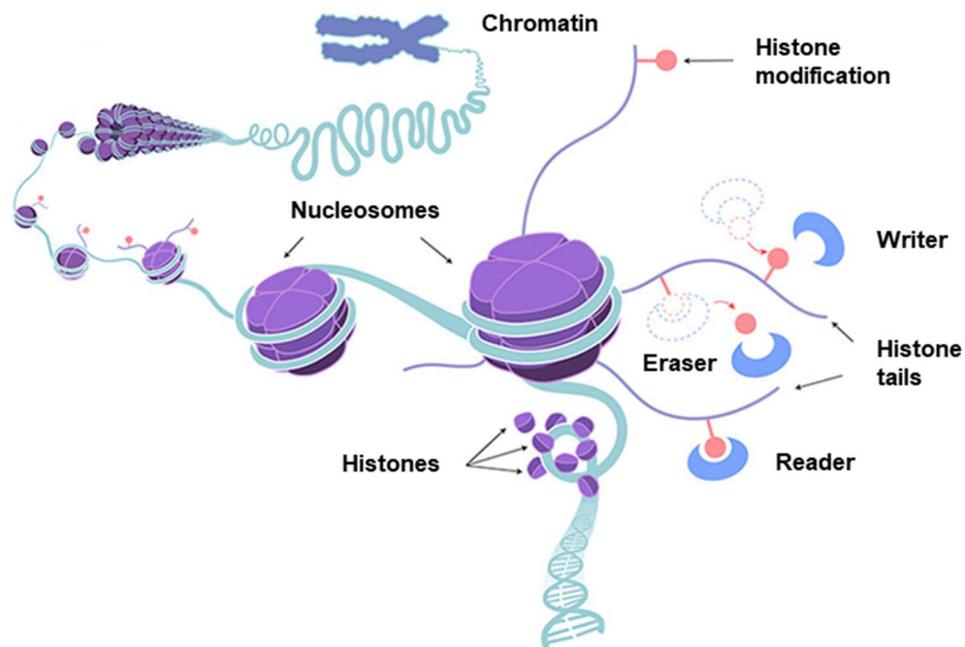


Figure 1.1. Histones have protruding histone tails where histone modifications can be added (Li and Li, 2021). Nucleosomes are the basic subunits of chromatin, and are made up of around 147 bp of DNA wrapped 1.7 times around two of each of the core histones H2A, H2B, H3 and H4. The histone tails are a common site for post-translational modifications, which are added by «writer» proteins, lead to interactions with «reader» proteins and are removed by «eraser» proteins.

1.1.2 3D chromatin organisation

In the interphase nucleus, chromosomes occupy distinct chromosome territories (figure 1.2 A). The chromatin is organised into two compartments, largely distinguished by the state of the chromatin they contain. Actively transcribed chromatin/euchromatin is less compact, and make up compartment A, and inactive heterochromatin is firmly packed and make up compartment B (figure 1.2 B) (Lieberman-Aiden *et al.*, 2009). These are further organised into topologically associated domains (TADs), which are domains up to 1Mb wide, and made up of sequences that interact within the TAD (figure 1.2 C) (Rowley and Corces, 2018). TADs can be further organised into subTADs. The TAD boundaries are mostly defined by convergently oriented CCCTC-binding factor (CTCF) binding sites, on either side of the sequence contained within the TAD (figure 1.2 C). These CTCF binding sites interact, organising the TAD chromatin into a loop (Rowley and Corces, 2018). The loops are formed by a cohesin or condensin complex binding chromatin at two close sites, and chromatin extruding bi-directionally until CTCF boundary elements are reached (Sanborn *et al.*, 2015). The CTCF boundary elements are occupied by CTCF, which anchor the loops (Sanborn *et al.*, 2015).

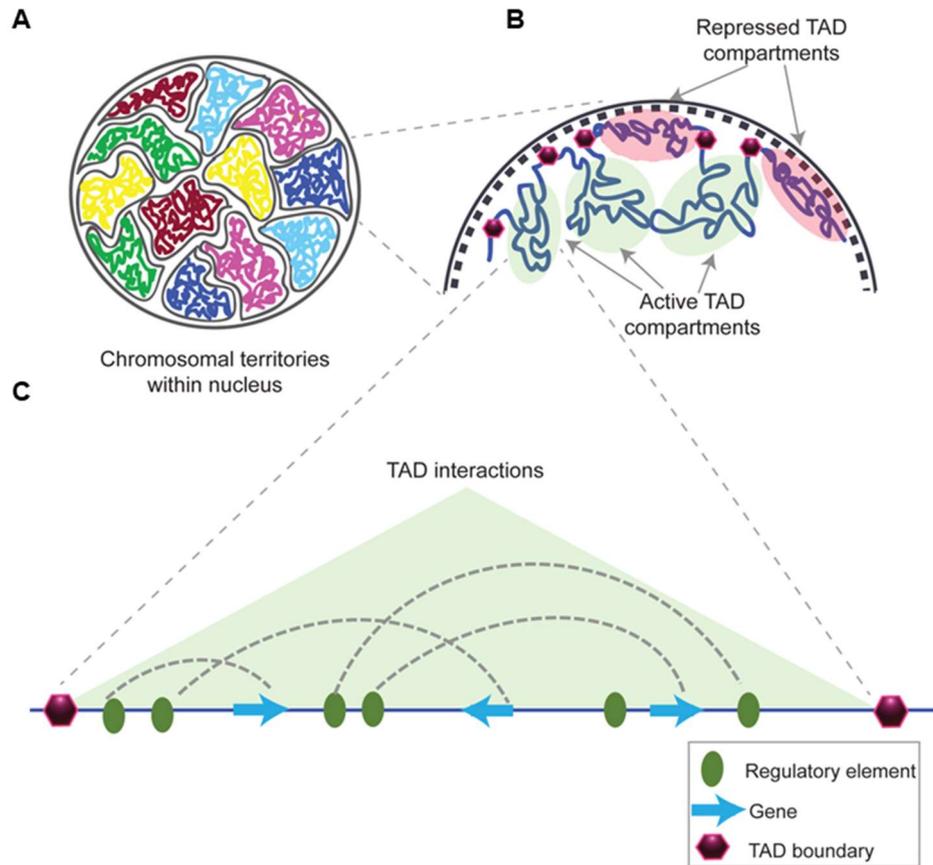


Figure 1.2. Chromatin is organised at various levels, from chromatin loops to chromosomal territories (Matharu and Ahituv, 2015).

- (A) Chromosomes occupy distinct regions within the nucleus, known as chromosomal territories.
- (B) Within individual chromosomes, chromatin is organised into “A” and “B” compartments, distinguished by their propensity to contain mostly active or inactive DNA, respectively. The compartments themselves contain topologically associating domains (TADs), which are regions of chromatin whose probability of interacting with each other is higher than the probability of interactions with DNA outside of the TAD.
- (C) TADs are demarcated by boundary elements, most frequently CTCF binding sites. Within a TAD, interactions between genes and regulatory elements e.g., enhancers can be increased; however, whether this increase in interaction is based on a higher frequency of short-lived interactions or stabilised interactions (or both) remains unclear.

1.1.3 Cellular organisation by phase separation

Phase separation is a biophysical phenomenon in which specific molecules e.g., lipids or particular proteins form membraneless compartments called biomolecular condensates. These condensates are thought to play a crucial role in organising cellular components and regulating various biological processes. Recently, phase separation has been hypothesised to play a particularly important role in DNA organisation and regulation, although this is an area of intense of debate. Certain proteins and nucleic acids can undergo phase separation, leading to the formation of liquid-like droplets or compartments (Shin and Brangwynne, 2017). These dynamic structures may facilitate the spatial organization of DNA, bringing together specific genomic regions, regulatory elements, and transcriptional machinery. By influencing chromatin accessibility and genome architecture, phase separation is hypothesised to contribute significantly to the control of DNA-related processes, such as gene expression (Shin and Brangwynne, 2017).

1.1.4 R-loops

R-loops are RNA-DNA hybrid structures which can form throughout the genome. They primarily form during DNA replication or at the active site of RNA polymerase II (RNA pol II) as a gene is being transcribed (figure 1.3) (Santos-Pereira and Aguilera, 2015). R-loops are more stable than double-strand DNA, and they preferentially form at particular DNA structural features, such as regions enriched for DNA nicks, or regions at which the free single-strand DNA can form a G-quadruplex (i.e. high GC content can increase the likelihood of R-loop formation).

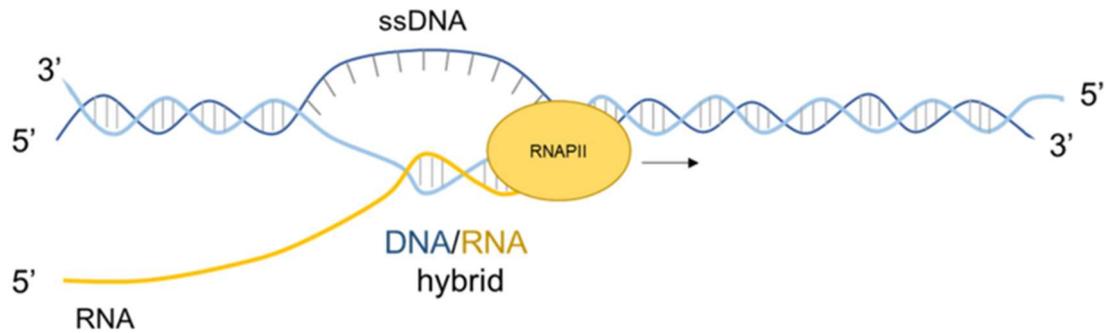


Figure 1.3. R-loops can be formed during transcription (Marabitti *et al.*, 2022). As actively transcribing RNA polymerase II (RNAPII) moves along DNA, the nascent RNA transcript is extruded behind and initially forms a DNA-RNA hybrid molecule with the transcribed DNA strand. This structure along with the naked, displaced DNA strand is known as an R-loop.

R-loops can aid in both gene activation and silencing. R-loops are often enriched within CpG islands and at the transcriptional start site of genes; it is thought that their formation protects against recruitment and activity of DNA methyltransferase 3B1 (DNMT3B1), which silences genes, and R-loops can thereby aid transcription initiation (Ginno *et al.*, 2012; Ginno *et al.*, 2013). In some instances, non-coding RNAs can form R-loops at/near gene promoters, leading to chromatin opening and binding of transcription factors (R-loops can therefore act similarly to pioneer factors) (Boque-Sastre *et al.*, 2015). Conversely, in other cases, R-loops can cause DNA compaction and formation of heterochromatin, which can lead to gene silencing (Groh *et al.*, 2014).

R-loops are associated with genome instability and cancer. There is increased risk of DNA single strand and double-strand breaks at R-loops; transcription-associated recombination can then contribute to genome instability through chromosomal rearrangement, or in extreme cases, complete loss of the affected chromosome(s). R-loops can also stall replication-fork progression, which can lead to double-strand break formation, recombination and chromosomal rearrangement (Aguilera and

García-Muse, 2012). Interestingly, the breast cancer type 1/2 (BRCA1 and BRCA2) tumour suppressor/DNA repair proteins play roles in preventing and/or resolving R-loops, although it remains unclear how they mechanistically do this (Bhatia *et al.*, 2014). Mutations in BRCA1 or BRCA2 could therefore lead to R-loop accumulation, increased DNA damage, and increased chromosome rearrangements, all of which are associated with cancer.

1.1.5 Histones and histone modifications

In addition to the four core histones H2A, H2B, H3 and H4, and linker histone H1, other histone variants exist. These include the H3 variant H3.3 and H2A variant H2AZ, which are enriched together at promoters of actively transcribed genes (Jin *et al.*, 2009). Histone proteins, and their post-translational modifications, play a fundamental role in processes such as regulation of gene expression, chromatin packaging and the DNA damage response (Lawrence, Daujat and Schneider, 2016). Histone modifications include acetylation, methylation, ubiquitylation and phosphorylation. Most known modifications are located at the N-terminal histone tails, which protrude from the nucleosome, although residues of the histone core can also be modified (figure 1.1) (Lawrence, Daujat and Schneider, 2016). Together, histone modifications make up the histone code, which impacts the state of chromatin, and influences which protein factors are recruited (Bannister and Kouzarides, 2011).

All four core histones can be modified, but the exact position, and type of modification, determines the effect these modifications have on the surrounding chromatin. The factors which interact with histones can be categorised as “writers”, “readers” and “erasers” of histone modifications (figure 1.1). Writer proteins deposit histone modifications *de novo*, reader proteins bind selectively to modified histones and thereby interpret the histone modifications, while eraser proteins remove histone

marks. For example, histone 3 lysine 4 tri-methylation (H3K4me3) marks actively transcribed genes (Santos-Rosa *et al.*, 2002). H3K4 is methylated by a histone methyltransferase complex, which enables interactions with readers of this histone marks, including transcriptional regulators and chromatin remodellers, leading to transcription of the activated gene (Beacon *et al.*, 2021; Wysocka *et al.*, 2006). The H3K4me3 mark is removed by H3K4 demethylases (Beacon *et al.*, 2021). Another transcription-promoting histone mark is H3K27 acetylation (H3K27ac), Histone 3 can also be trimethylated by Polycomb repressive complex 2 at its lysine 27 residue, forming H3K27me3, which marks facultative heterochromatin and transcriptionally silenced regions (Saksouk, Simboeck and Déjardin, 2015).

1.1.6 Methods to investigate DNA-protein interactions

The investigations of chromatin state, temporal and spatial genomic presence of histones, histone modifications and transcriptional regulators have been facilitated by techniques allowing investigations of protein-DNA interactions. Several such techniques exist, but they commonly involve immunoprecipitating the protein of interest, and if followed by next-generation sequencing, can give a genome-wide view of the occupancy of the protein or protein-modification of interest.

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) is a powerful and widely-used technique used for studying protein-DNA interactions. It involves crosslinking proteins to DNA, isolating the protein-DNA complexes, fragmenting the DNA, immunoprecipitating the protein of interest, sequencing the associated DNA fragments, and analysing the resulting data to identify binding sites (figure 1.4) (Park, 2009). ChIP-seq is often used to probe transcription factor (TF) binding patterns and can be used to identify enrichment of the studied protein around particular chromatin

features e.g., enhancers or promoters and to identify DNA motifs associated with protein occupancy (Park, 2009).

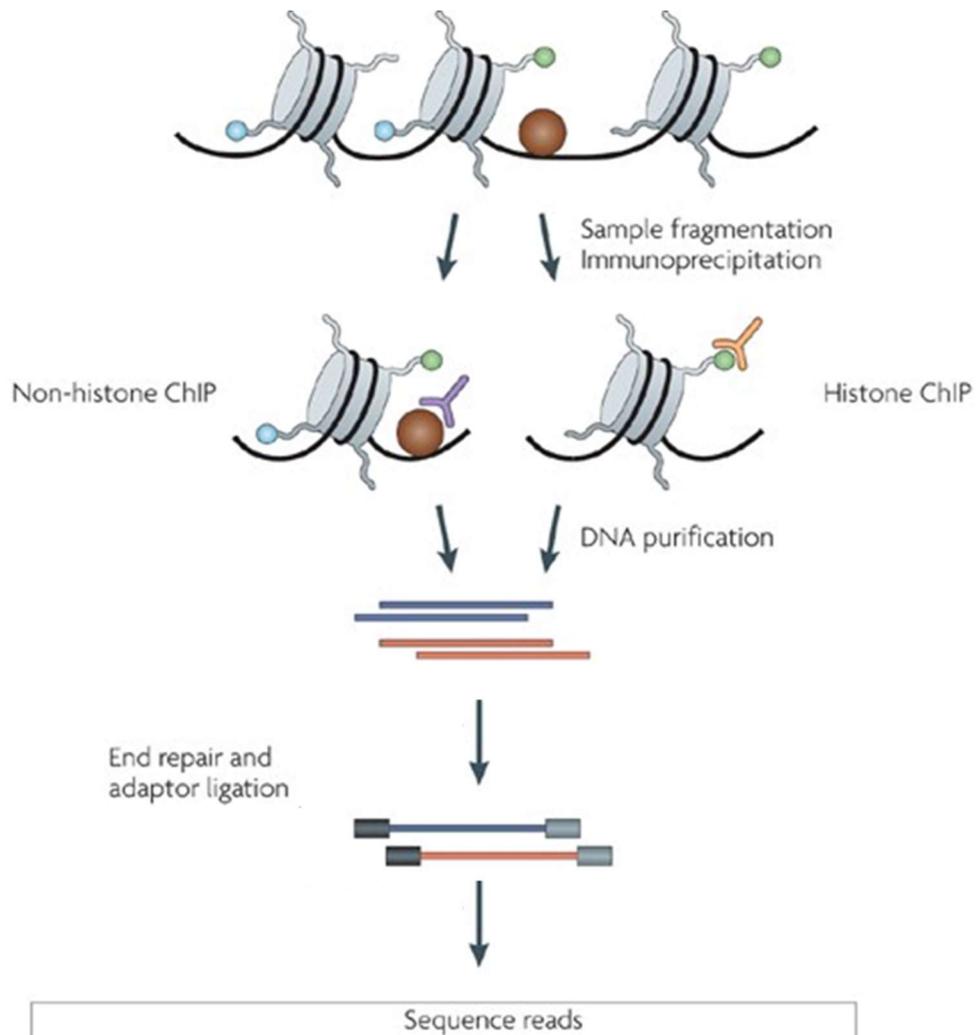


Figure 1.4. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) (Park, 2009). To investigate protein-DNA interactions, proteins are cross-linked to the DNA, followed by DNA fragmentation and immunoprecipitation using antibodies specific to the protein of interest (histone, or non-histone protein). The cross-linking is reversed, followed by purification of the DNA fragments, adaptor ligation and sequencing of the DNA fragments.

Cleavage under targets and tagmentation (CUT&Tag) is a newer ChIP-based technique which allows higher resolution mapping of protein-DNA interactions under native conditions. It involves incubating the non-crosslinked sample with an antibody specific to the protein of interest. Following antibody-protein binding, protein A-conjugated Tn5 transposase is added, which binds the antibody-protein complex and tags DNA surrounding the binding site (Kaya-Okur *et al.*, 2019). The tagmentation process involves DNA cleavage and sequencing adaptor loading, meaning the tagged DNA can then be isolated, indexed and sequenced directly (figure 1.5) (Kaya-Okur *et al.*, 2019). CUT&Tag is more efficient than traditional ChIP-seq and can therefore be used with lower cell numbers. It also allows higher resolution analysis, and the fact crosslinking is not necessary removes the possibility of crosslinking artefacts in the resultant data; however, this does also mean that CUT&Tag is mainly suited to assessing histone modifications and/or proteins which bind directly to DNA with high stability.

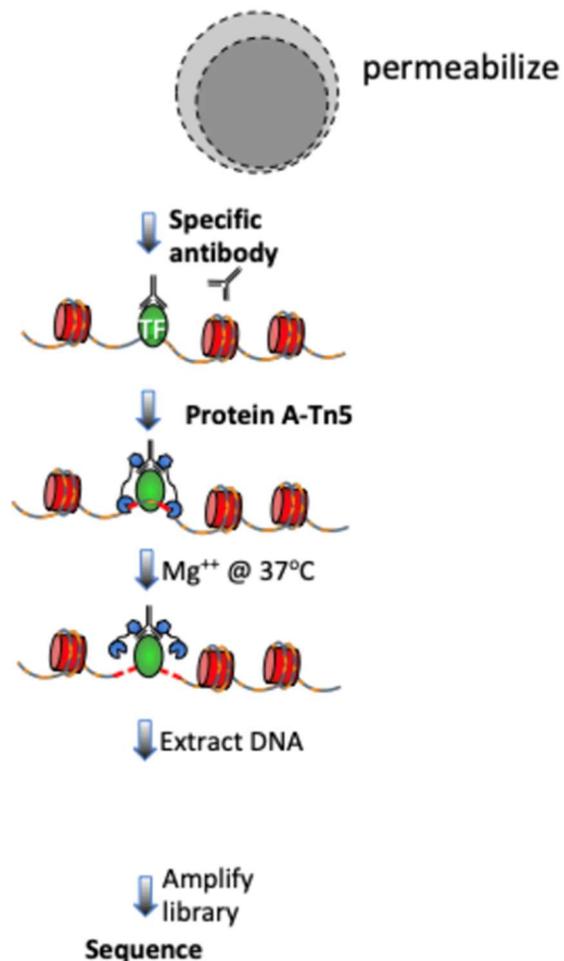


Figure 1.5. Cleavage under targets and tagmentation (CUT&Tag) followed by sequencing (Kaya-Okur *et al.*, 2019). Cells are permeabilised, and an antibody specific to the protein of interest is added. The transposase Tn5 is fused to protein A, which directs Tn5 to the site of the protein of interest, by binding the antibody. Tn5 is activated by addition of Mg²⁺, and simultaneously cleaves the DNA and inserts adapter sequences. The DNA fragments can then be extracted and sequenced.

1.2 DNA damage

1.2.1 Causes of DNA breaks

DNA breaks can either occur on a single DNA strand (single-strand breaks (SSB)), or across both DNA strands (double-strand breaks (DSB)). DNA breaks can be caused by both external factors, such as ultraviolet or ionising radiation, and cellular

processes, including reactive oxygen species generated by metabolism, or by the collapse of a DNA replication fork (Rothkamm *et al.*, 2015; Lindahl, 1993). DNA breaks are deliberately induced by cells, during V(D)J recombination allowing T- and B-cells to form unique antigen receptors, and during meiotic recombination in germ cells, increasing genetic diversity (Rothkamm *et al.*, 2015). DNA breaks are also induced by topoisomerases during DNA replication, to relieve torsion caused by overwound/supercoiled DNA ahead of DNA polymerase (Morimoto *et al.*, 2019).

1.2.2 DNA damage repair

Our DNA is exposed to DNA-damaging agents daily, causing an estimated 10^4 - 10^5 DNA lesions in each cell; however, these are, in most cases, resolved by DNA damage repair pathways. DSBs are highly genotoxic, and if left unrepaired or mis-repaired, can lead to apoptosis or DNA mutations, deletions, chromosome translocations and carcinogenesis. Most cancers present genomic instability and DNA damage repair genes are commonly mutated (Negrini, Gorgoulis and Halazonetis, 2010).

1.2.2.1 DNA repair pathways

The two main DSB repair pathways are homologous recombination (HR) and non-homologous end-joining (NHEJ) (Hustedt and Durocher, 2016). These two pathways differ in whether they require a template to repair DNA breaks, and when in the cell cycle they are used. NHEJ is used throughout the cell cycle apart from mitosis, and the DNA ends are either processed, or not, before being ligated back together (Hustedt and Durocher, 2016). During NHEJ, the DSB is bound by a Ku70/80 heterodimer, which recruit NHEJ polymerases, nucleases and ligases (figure 1.6) (Chang *et al.*, 2017). The DNA strands are processed and resected by Artemis, DNA-

dependent protein kinase catalytic subunit (DNA-PKcs) and the MRE11/Rad50/NBS1 (MRN) complex, before the DNA strands are ligated back together by the XLF/XRCC4/Ligase IV complex (Lieber, 2010). Due to a lack of template, NHEJ is error-prone. In addition, the resection can lead to nucleotide loss, and the repaired DNA strand may contain insertions and/or deletions (Chang *et al.*, 2017). Contrarily, HR uses the sister chromatid as a template to re-synthesise the sequence where the DNA break was found, and is thus restricted to the S and G2 cell cycle phases (Hustedt and Durocher, 2016). First, the DSB is bound by the MRN complex, which resect the 5' ends of DNA around the DSB together with CtIP and exonuclease 1 (EXO1), forming single stranded 3' overhangs (figure 1.6). The overhangs are bound by replication protein A (RPA), which is exchanged for Rad51 by BRCA2. With the aid of Rad51, the single-stranded DNA invades the sister chromatin, where it aligns with its complementary DNA strand (San Filippo, Sung and Klein, 2008). A DNA polymerase extends the invading strand, which then serves as a template for synthesis of the resected DNA strand, followed by ligation (Li and Heyer, 2008).

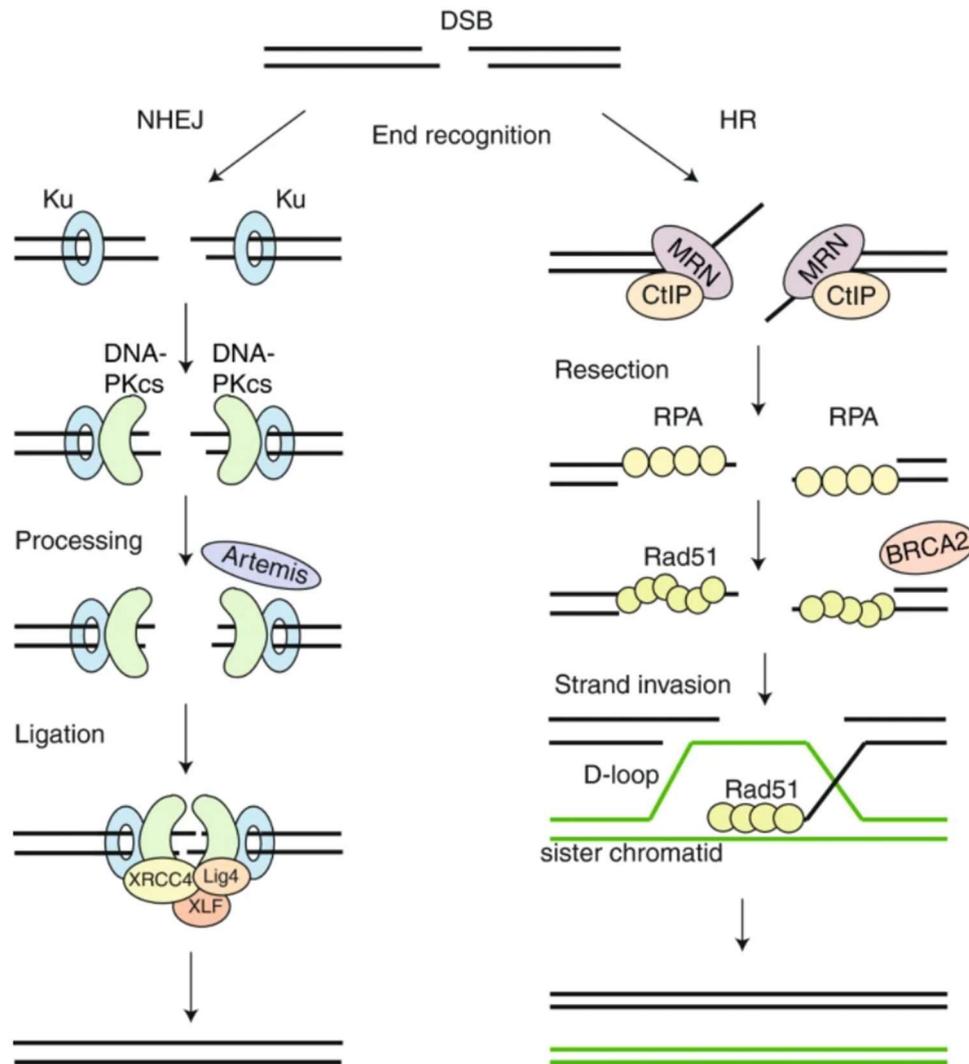
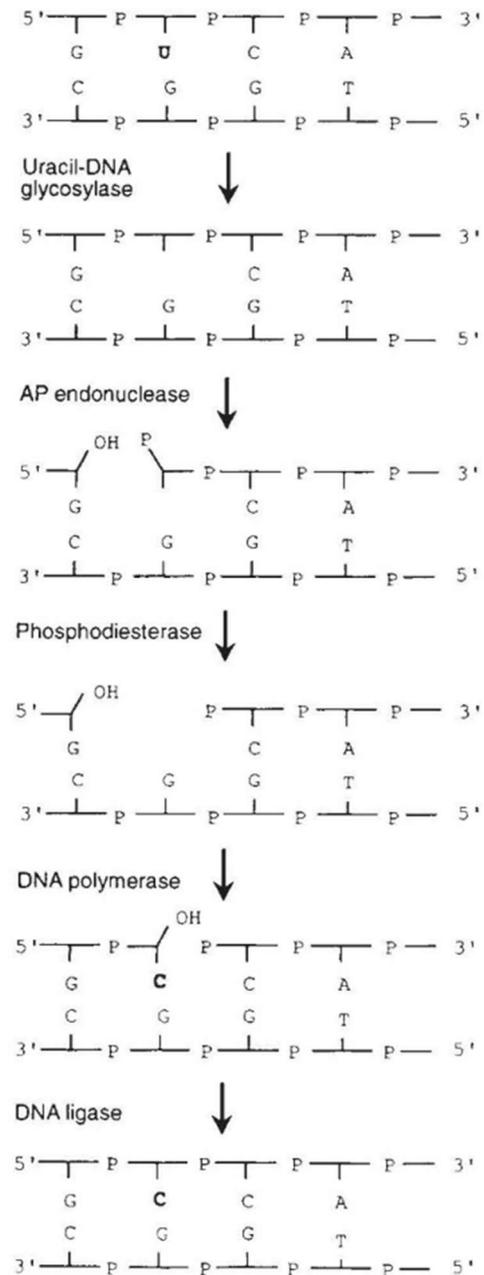


Figure 1.6. Non-homologous end joining (NHEJ) and homologous recombination (HR) DSB repair pathways (Brandsma and Gent, 2012). In NHEJ (left), DSBs are recognised by Ku70/80 heterodimers, serving as a scaffold to recruit downstream NHEJ factors. This is followed by recruitment of DNA-PKcs kinase, which then recruits and activates the nuclease Artemis, and DSB ends are processed. With the aid of XRCC4 and XLF, Ligase 4 interacts with the broken DNA ends to ligate them back together. In HR (right), DSBs are bound by the MRN complex which together with CtIP resect the DNA, forming 3' ssDNA overhangs, which can be hundreds of basepairs long. The ssDNA overhangs are bound by RPA, which BRCA2 then replaces with Rad51. The Rad51-bound ssDNA invades the sister chromatin, where it aligns with the homologous template strand. Extension of the invading strand (through extension of the D-loop) is followed by synthesis of the resected strand, and ligation of the DNA strands.

Other DNA repair pathways include nucleotide excision repair (NER), mismatch repair and base excision repair (BER). BER removes non-helix distorting bases, caused by processes such as oxidation or deamination of nucleotides. The lesion is recognised by DNA glycosylases, which cleave the N-glycosidic bond between the nitrogenous base and the sugar molecule of the damaged nucleotide, creating an abasic (AP) site (figure 1.7) (Chatterjee and Walker, 2017). The AP site is recognised by AP endonucleases, which excise the base. DNA polymerase inserts a new base, and the DNA strand is ligated (Chatterjee and Walker, 2017).

Figure 1.7. Base excision repair pathway

(Lindahl, 1993). Processes including spontaneous cytidine deamination leads to presence of uracil in the DNA strand. Uracil DNA glycosylase cleaves the bond attaching the uracil base to the deoxyribose in the phosphodiester (P) backbone, leading to formation of an apyrimidinic (AP) site. The AP site is recognised by AP endonuclease, which cleave the phosphodiester backbone, and phosphodiesterase remove the deoxyribose-phosphate residue remaining from the excised base. DNA polymerase then synthesises a new base filling the gap in the DNA strand, and DNA ligases joins the strand together.



1.2.2.2 Chromatin remodelling in DNA repair

To allow access of the DNA repair machinery, chromatin remodelling occurs in the vicinity of the DSB. This includes nucleosome sliding, nucleosome ejection or histone exchange. This takes place concurrently with DNA damage repair, and involves multiple ATP-dependent protein complexes, including SWI/SNF, WSTF, ISWI, INO80

and p400, which uses the energy generated by ATP hydrolysis to rearrange nucleosomes (Jeggo and Downs, 2014). For example, chromatin remodelling by ISWI and WSTF near DSBs is required for recruitment of repair factors such as BRCA1, Rad51, Ku70/80 and XRCC4 (Aydin, Vermeulen and Lans, 2014). p400 exchanges histone H2AZ onto nucleosomes, which leads to more open and accessible chromatin surrounding the DSB (Xu *et al.*, 2012).

1.2.2.3 Chromatin state affecting choice between HR and NHEJ

HR and NHEJ repair pathways are both present during S and G2 phases, and some evidence suggests that the location of the DNA damage in chromatin can influence which repair method is used. DNA breaks in or around active genes are thought to mainly be repaired by HR, as seen by Rad51, which is associated with HR repair, being recruited to DNA breaks in an RNA polymerase II-dependent manner (Aymard *et al.*, 2014). In addition, Rad51 presence coincides with histone marks, such as Histone 3 lysine 36 trimethylation (H3K36me), which are mostly found in transcriptionally active regions (Clouaire *et al.*, 2018). Chromatin marks appearing at the DNA break site may also be linked to specific repair pathways (Clouaire *et al.*, 2018). Whereas some histone marks are found equally at sites prone to repair by either HR or NHEJ, some modifications are specific to HR but not NHEJ and vice versa (Clouaire *et al.*, 2018). For example, H4 serine 1 phosphorylation and deposition of macroH2A appears irrespective of whether the site uses HR or NHEJ to repair DNA breaks, whereas H2AX serine 139 phosphorylation is mainly found at sites prone to HR, together with H1 depletion and ubiquitination (Clouaire *et al.*, 2018).

1.3 H2A histone variant H2AX and its involvement in DNA damage repair

1.3.1 H2A variant H2AX is a minor constituent of chromatin

The H2A histone variant H2AX make up around 25% of total H2A in mammals, depending on cell type, and is found in around 10% of nucleosomes (Cowell *et al.*, 2007; Iacovoni *et al.*, 2010). Interestingly, some evidence suggests H2AX is not incorporated evenly throughout the genome (Iacovoni *et al.*, 2010; Bewersdorf, Bennett and Knight, 2006; Savic *et al.*, 2009). Instead, it is found to be particularly enriched in gene-rich regions (Iacovoni *et al.*, 2010). H2AX differs from the other histone H2A variants by its conserved C-terminal tail motif S-Q-E-Y (residues 139-142) (Mannironi, Bonner and Hatch, 1989). The tyr142 residue is conserved in metazoans including mammals, *Xenopus laevis* and *Drosophila melanogaster* (S-Q-A-Y), but not in unicellular *Saccharomyces cerevisiae* (S-Q-E-L) (Xiao *et al.*, 2009). Several of the H2AX residues undergo post-translational modifications, including phosphorylation of serine 139 and tyrosine 142, ubiquitination of lysine 119 and acetylation of lysine 5, all of which are important for the role of H2AX in the DNA damage response (Corujo and Buschbeck, 2018).

Dysregulation of H2AX has been associated with various cancers, including breast, lung, and colorectal cancer and has also been linked to cancer stem cell maintenance and therapeutic resistance (Celeste *et al.*, 2002; Bonner *et al.*, 2008). Altered H2AX expression or function can contribute to genomic instability, impaired DNA repair, and oncogenic transformation (Bonner *et al.*, 2008; Corujo and Buschbeck, 2018). Human H2AX is encoded by the *H2AFX* gene on chromosome 11, located in an area frequently mutated in cancers, and around 1/3 of breast cancer patients have been found to present *H2AFX* copy number alterations (Bassing *et al.*, 2002; Bassing *et al.*, 2003; Srivastava *et al.*, 2008). H2AX post-translational modifications play a crucial role in initiating the DNA damage response (see section 1.3.2), and many cancer therapies act by introducing high levels of DNA damage to induce cell death. Tripe-

negative breast cancers displaying constitutive H2AX serine 139 phosphorylation have significantly poorer prognosis (Nagelkerke *et al.*, 2011), whereas chemotherapy-induced degradation of H2AX, caused by chronic oxidative stress, is associated with better response to treatment (Gruosso *et al.*, 2016).

1.3.2 γ H2AX: H2AX is phosphorylated at ser139 in response to DNA damage

In mammalian cells, detection of DNA double-strand breaks is followed by a cascade of events. This is initiated by phosphorylation of H2AX at its serine 139 residue (forming γ H2AX), which spreads to megabase chromosomal regions surrounding the DSB (seen as γ H2AX 'foci' by microscopy) (Celeste *et al.*, 2003; Kim *et al.*, 2007; Rogakou *et al.*, 1998; Rogakou *et al.*, 1999). Phosphorylation of H2AX upon ionising radiation is also conserved in other species, including *S. cerevisiae* (phosphorylation of H2A), *D. melanogaster* and *X. laevis* (Rogakou *et al.*, 1999). γ H2AX formation in response to DSBs was first described over twenty years ago (Rogakou *et al.*, 1998). The role of serine 139-phosphorylated H2AX in the DNA damage response has since been extensively described. γ H2AX formation is one of the earliest steps in the DNA damage response pathway, but γ H2AX is also formed during apoptosis, in response to DNA fragmentation (Rogakou *et al.*, 2000)

H2AX is phosphorylated by phosphoinositide 3-kinase-related protein kinases (PIKKs): ATR (ATM and Rad3-related), ATM (ataxia teleangiectasia mutated), or DNA-PKcs, which become activated upon DNA damage (Rothkamm *et al.*, 2015). In DNA damaged nuclei, the γ H2AX signal can be visualised as foci around the break site (Rothkamm *et al.*, 2015). The γ H2AX signal appears seconds after DNA damage induction in mammalian cells and around half of the H2AX is phosphorylated to γ H2AX within 10 minutes. The signal then decreases over the next few hours, as the break sites undergo repair (Rogakou *et al.*, 1998). γ H2AX marks double-strand DNA

breaks, as seen by its formation in response to double-strand break-inducing ionising radiation, ultraviolet A radiation, and treatment with the radiomimetic agent bleomycin (Rogakou *et al.*, 1998). However, H₂O₂ treatment or ultraviolet C light, which mainly induce single-strand DNA breaks, do not lead to detectable levels of γ H2AX (Rogakou *et al.*, 1998).

γ H2AX forms megabase domains surrounding the DNA break site; and the spread and pattern of these γ H2AX domains is regulated by the underlying chromatin confirmation (Arnould *et al.*, 2021). The chromosome confirmation capture technique Hi-C revealed that TAD boundaries coincide with the boundaries of γ H2AX domains (Arnould *et al.*, 2021). Cohesin accumulates at the DSB (irrespective of HR or NHEJ repair pathways), and induces one-sided loop extrusion on both sides of the DSB (Arnould *et al.*, 2021). ATM, which is responsible for the formation of the majority of γ H2AX upon DNA damage, does not spread across the TAD, but is situated near the site of the DSB (Arnould *et al.*, 2021). A model was therefore suggested, whereby ATM phosphorylates H2AX as it passes through during formation of these loops (figure 1.8) (Arnould *et al.*, 2021).

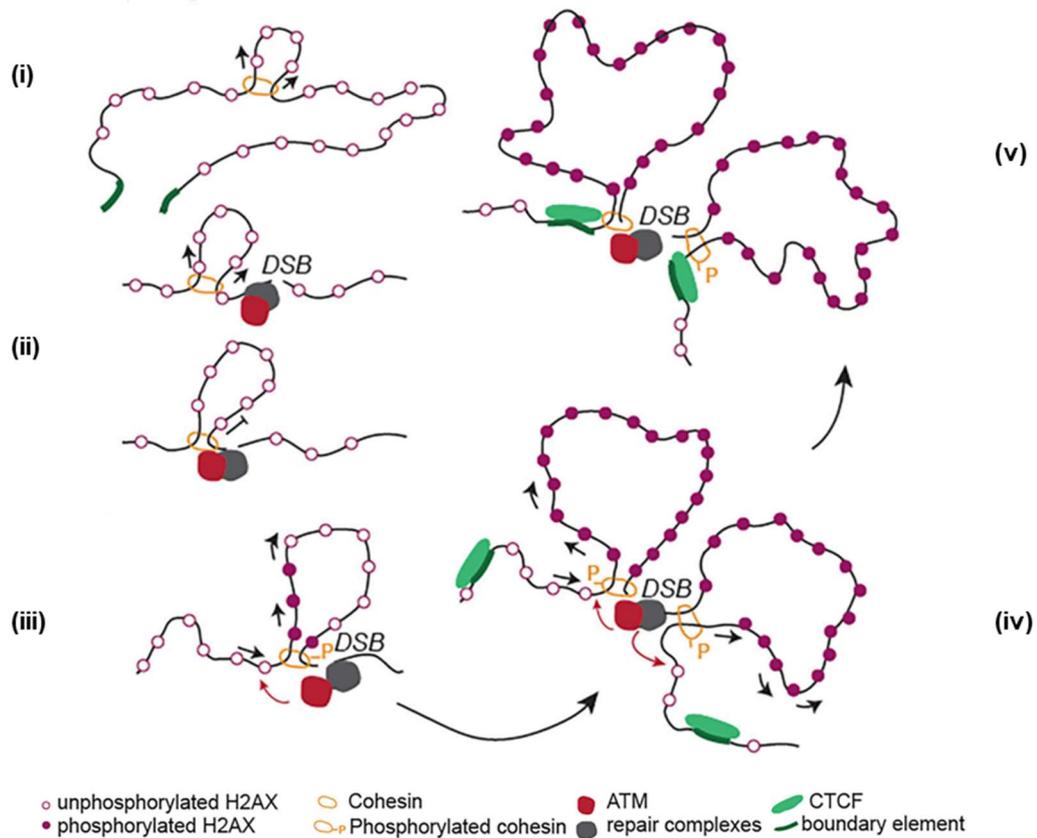


Figure 1.8. Suggested model for spread of the γ H2AX signal through loop extrusion mechanisms (Arnould *et al.*, 2021). (i) Bi-directional loop extrusion is a cohesin-mediated process in which cohesin binds and translocates along chromatin, extruding both strands behind it as it goes. (ii) Upon DSB formation, loop extrusion is blocked on the side of the break by break and/or repair complexes. (iii) As loop extrusion progresses unidirectionally, H2AX is phosphorylated by ATM kinase as it passes through the base of the loop, forming γ H2AX (ATM also phosphorylates cohesin). (iv) The same is occurring on the other side of the DSB, and unidirectional loop extrusion on both sides of the DSB leads to spread of the γ H2AX signal in both directions. (v) The loop extrusion halts when boundary elements (marking TAD borders) are reached, such as CTCF-bound loci.

1.3.3 γ H2AX recruits DNA damage repair proteins

γ H2AX formation is followed by recruitment of DNA damage repair factors, including Mediator of DNA damage checkpoint protein 1 (MDC1), p53-binding protein 1

(53BP1), Rad51 and ATM ser1981ph (activated) kinase, which initiate DNA damage repair (Rothkamm *et al.*, 2015; Stucki *et al.*, 2005; Paull *et al.*, 2000). MDC1 interacts with γ H2AX via its tandem C-terminal BRCT domains (Stucki *et al.*, 2005). Interestingly, MDC1 is unable to interact with γ H2AX in cells expressing tyr142ala mutation. Although γ H2AX foci still form in these cells, MDC1 and downstream DNA damage response factors 53BP1, NSB1 or phospho-S1981 ATM are not recruited to these foci, highlighting the importance of γ H2AX-led recruitment of DNA repair factors (Stucki *et al.*, 2005).

Not every DSB displays γ H2AX, and this histone mark is found less frequently at DNA breaks in heterochromatin, unless the heterochromatin regions are undergoing replication (Cowell *et al.*, 2007). However, the importance of γ H2AX in the DNA damage repair pathway is seen in mice lacking H2AX, which display genomic instability and increased sensitivity to irradiation (Celeste *et al.*, 2002).

1.3.4 H2AX is constitutively phosphorylated at tyr142

More recently, it was discovered that H2AX can also be phosphorylated at its tyrosine 142 residue (tyr142) (Xiao *et al.*, 2009). In fact, H2AX is constitutively phosphorylated at tyr142, under normal conditions, by WSTF–ISWI ATP-dependent chromatin-remodelling (WICH) complex member Williams–Beuren syndrome transcription factor (WSTF) (Xiao *et al.*, 2009). WSTF was not previously known to have kinase activity, and does not contain domains with sequence homology to known kinase domains. Its N-terminal WSTF/Acf1/cbp146 (WAC) domain alone is capable of phosphorylating H2AX tyr142; however, presence of both WTSF N- and C terminal domains are required for high levels of H2AX tyr142 phosphorylation (Xiao *et al.*, 2009).

1.3.5 H2AX tyr142 phosphorylation is lost concurrently with ser139 phosphorylation

Tyr142 phosphorylation is gradually lost upon DNA damage induction (Xiao *et al.*, 2009; Cook *et al.*, 2009). The dephosphorylation of H2AX tyr142 is mediated by the phosphatase enzymes EYA (1/3), mammalian homologues of *Drosophila melanogaster eyes absent (eya)*. EYA1 and 3 are thought to combine to form a tyrosine phosphatase complex, as they are both required to fully dephosphorylate H2AX tyr142 (Cook *et al.*, 2009). In response to DNA damage, EYA3 is phosphorylated at its ser129 residue, through ATM/ATR kinase activity, which is required for the interaction between EYA (1/3) with γ H2AX (Cook *et al.*, 2009).

1.3.6 Blocking H2AX tyr142 phosphorylation perturbs the DNA damage response

H2AX tyr142ph appears to play an important role in the DNA damage response. WSTF knockdown or kinase silencing leads to a rapid loss of γ H2AX in DNA damaged cells (Xiao *et al.*, 2009), and similarly, mutating tyrosine 142 to phenylalanine, which cannot be phosphorylated, leads to a strongly reduced γ H2AX signal (Cook *et al.*, 2009). However, this may be caused by MDC1 being unable to interact with γ H2AX with mutated C-terminal tyrosine residue, affecting the downstream DNA damage response and recruitment of factors including ATM (Stucki *et al.*, 2005). A loss of H2AX tyr142 phosphorylation also affects other factors of the DNA damage response. WSTF knockdown or silencing leads to reduced MDC1 foci formation and ATM serine 1981 phosphorylation, in DNA damaged cells (Xiao *et al.*, 2009). Therefore, H2AX tyr142 phosphorylation appears to be required for a proper initiation and/or maintenance of the DNA damage response.

Interestingly, while a lack of H2AX tyr142 *phosphorylation* impairs the DNA damage response, a lack of H2AX tyr142 *dephosphorylation* also affects it. Knocking down EYA3 leads to a loss of MDC1-H2AX interactions in DNA damaged cells, underlining the importance of H2AX tyr142 dephosphorylation for this interaction (Cook *et al.*, 2009). ZNF506, another factor required for H2AX tyr142 dephosphorylation, is frequently mutated in various cancers, including T-cell prolymphocytic leukaemia, which commonly exhibits altered DNA damage response pathways (Nowsheen *et al.*, 2018). Upon DNA damage, ZNF506 is recruited by MDC1 to γ H2AX foci (Nowsheen *et al.*, 2018). This happens through ZNF506 thr140 phosphorylation by ATM, allowing direct interaction between ZNF506 and MDC1's forkhead-associated domain (which binds phosphorylated serine/threonine domains) (Nowsheen *et al.*, 2018). ZNF506's important role in the DNA damage response is evidenced by the fact that ZNF506 knockdown leads to reduced MDC1 foci formation, and reduced recruitment of DNA damage response factors downstream of MDC1: RNF8, BRCA1 and 53BP1 (Nowsheen *et al.*, 2018). Similar to EYA 1/3 knockdown, ZNF506 knockdown leads to loss of H2AX tyr142 dephosphorylation in DNA damaged cells (Nowsheen *et al.*, 2018). EYA1 and ZNF506 interact directly upon DNA damage, through ZNF506 zinc finger domains, and EYA1 is unable to bind at repair foci in absence of ZNF506 (Nowsheen *et al.*, 2018). The interaction is dependent on the EYA3 ser129 residue, which becomes phosphorylated upon DNA damage (Nowsheen *et al.*, 2018). Interestingly, EYA1 and MDC1 are unable to interact in DNA damaged cells in absence of ZNF506 (and its thr140 residue) (Nowsheen *et al.*, 2018). This suggests a role for thr140-phosphorylated ZNF506 in recruiting EYA1 to the site of damage, causing/allowing downstream dephosphorylation of H2AX tyr142 and continued DNA damage response (Nowsheen *et al.*, 2018). A suggested model for H2AX tyr142 dephosphorylation upon DNA damage is shown in figure 1.9.

Cells lacking ZNF506/harboured ZNF506 thr140 mutations are hypersensitive to radiation, underlining its importance in DNA damage response (Nowsheen *et al.*, 2018). Investigating several cancer-associated ZNF506 mutations from patient samples, including from T-PLL, demonstrated that although γ H2AX formation was unaffected, MDC1, BRCA1 and 53BP1 (repair proteins downstream of MDC1) failed to localise to γ H2AX foci in response to DNA damage (Nowsheen *et al.*, 2018). This was rescued by expressing WT, but not the thr140 phospho-mutant ZNF506 (Nowsheen *et al.*, 2018).

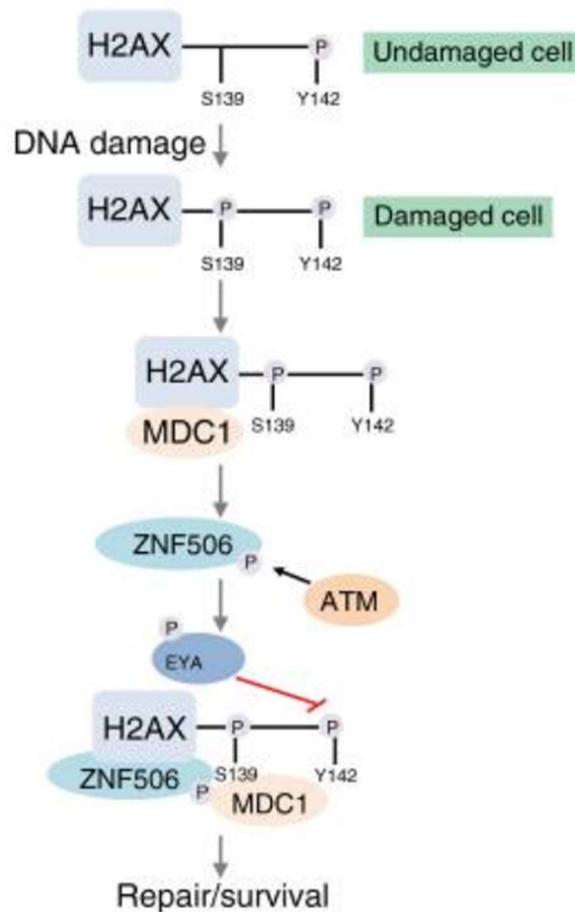


Figure 1.9. Suggested model for H2AX tyr142 dephosphorylation during the DNA damage response (Nowsheen *et al.*, 2018). H2AX is phosphorylated at tyr142 in undamaged cells. Upon DNA damage, H2AX rapidly becomes phosphorylated at ser139, leading to formation of doubly phosphorylated H2AX. MDC1 is recruited to ser139-phosphorylated H2AX (although evidence suggests MDC1 does not bind di-phosphorylated H2AX), which recruits ATM-phosphorylated (activated) ZNF506. ZNF506 recruits EYA phosphatases, which dephosphorylates H2AX at the tyr142ph residue, allowing recruitment of other DNA repair factors to the site of the DSB.

1.3.7 Di-γH2AX interacts with pro-apoptotic factors

H2AX can be mono-phosphorylated at ser139 and tyr142, and can also be doubly phosphorylated at ser139 and tyr142 (di-γH2AX). H2AX ser139 phosphorylation is important for facilitating interactions with DNA repair factors, such as MDC1, Mre11

and Rad50 (Cook *et al.*, 2009). Di- γ H2AX is unable to interact with MDC1, Mre11 and Rad50, suggesting that the repair machinery is unable to bind di- γ H2AX (Cook *et al.*, 2009; Campbell, Edwards and Glover, 2010). Instead, di- γ H2AX interacts with pro-apoptotic factor JNK1 (Cook *et al.*, 2009). This is mediated by the suggested pro-apoptotic factor Fe65, which interacts both with JNK1 and H2AX in DNA damaged cells (Cook *et al.*, 2009). Fe65 binds H2AX through its phosphotyrosine-binding domain 2 (PTB2), in a H2AX tyr142ph-dependent manner (Cook *et al.*, 2009). Contrary to this, the DNA damage response factor MCPH1, which mediates DNA repair factor recruitment, is able to interact with di- γ H2AX *in vitro* (Singh *et al.*, 2012). However, studies using Terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL) assay, have demonstrated reduced induction of apoptosis in cells expressing H2AX tyr142phe (un-phosphorable), further supporting a proposed role for H2AX tyr142 phosphorylation in apoptosis (Cook *et al.*, 2009).

Cook *et al* hypothesised that di- γ H2AX is a pro-apoptotic signal, whereas γ H2AX induces DNA break repair. However, more research is needed to fully explain the role of di- γ H2AX and H2AX tyr142ph in the DNA damage response. γ H2AX is formed rapidly upon DNA damage, whereas active removal of the H2AX tyr142 phosphorylation mark by EYA (1/3) is inherently a slow process. This could lead to a prolonged apoptosis-inducing di- γ H2AX signal in every cell upon DNA damage. However, the proportion of H2AX which are tyr142 phosphorylated in steady-state conditions, is unknown. Di- γ H2AX might be a minor signal, or the signal might also have other functions.

1.3.8 The di-phosphorylated form of H2AX is proposed to be present during early phases of the DNA damage response

Presence of H2AX tyr142ph in undamaged cells and di- γ H2AX during the DNA damage response has been investigated by multiple research groups. Some of these have used anti-phosphotyrosine antibodies to indirectly detect tyrosine-phosphorylated H2AX (Cook *et al.*, 2009; Nowsheen *et al.*, 2018), although H2AX has 4 tyrosine residues (position 39, 50, 57 and 142). Singh *et al* developed a novel antibody against di- γ H2AX; however, this antibody showed cross-reactivity with both ser139 and tyr142 phosphorylated H2AX C-terminal peptides (Singh *et al.*, 2012). This suggests that the di- γ H2AX signal detected early upon DNA damage may instead represent the highly abundant γ H2AX signal.

Cook *et al* described a role for di- γ H2AX in cell fate decisions, where doubly phosphorylated H2AX interacts with pro-apoptotic factors. JNK1 was found to interact with doubly phosphorylated H2AX peptides, but not peptides phosphorylated at ser139 only (Cook *et al.*, 2009). Co-immunoprecipitation showed interactions between JNK1 and Fe65 with WT H2AX in DNA damaged cells, but not with H2AX tyr142phe (non-phosphorylatable), although the phosphorylation state of the WT H2AX investigated is unknown (Cook *et al.*, 2009).

The current view of H2AX phosphorylation in undamaged/damaged cells (figure 1.9) suggests that di- γ H2AX only forms on H2AX with tyr142 phosphorylation; however, H2AX tyr142ph are below detection level with mass spectrometry, suggesting that it is a low-abundance histone mark (Hatimy *et al.*, 2015). Further investigations into di- γ H2AX and its hypothesised role in the DNA damage response thus requires new tools to allow specific and systematic detection of it.

1.4 Induction of DNA breaks

1.4.1 DNA break induction methods

To study histone modifications and protein recruitment at the DNA break site, DNA breaks can be induced in cells in a controlled manner. Irradiation and genotoxic drugs are commonly used, but induce DSBs randomly throughout the genome in every cell, creating a heterogenous cell population and thus making it unsuitable for e.g. ChIP analysis of protein-DNA interactions (Iacovoni *et al.*, 2010). Also, as irradiation creates random breaks, many of them will be located in heterochromatin (Clouaire *et al.*, 2018). Other techniques include using the modified yeast I-SceI nuclease, which creates a single double-strand break per cell (Rouet, Smih and Jasin, 1994); however, if the aim is to investigate DNA breaks in various different genomic regions, this may not generate desirable data. Expressing the endonuclease I-PpoI generates several DNA breaks per cell, but these are mainly found in ribosomal regions, and would therefore not allow studying DNA breaks in a range of different genomic regions, in addition to being unsuitable for ChIP analysis (Berkovich, Monnat and Kastan, 2007; Iacovoni *et al.*, 2010).

1.4.2 AsiSI-ER DNA double-strand break induction

The DSB Inducible via AsiSI (DIVA) cell line is a DSB-inducible U2OS cell line expressing the restriction enzyme AsiSI fused to the estrogen receptor (ER) ligand binding domain. The AsiSI cleavage site is GCGAT⁺CGC. In presence of 4-Hydroxytamoxifen (OHT), the fusion protein translocates to the nucleus and generates 200 breaks per cell in G1, equal to irradiation with 5-10 Gy, but does not induce apoptosis (Iacovoni *et al.*, 2010; Massip *et al.*, 2010). DSBs generated can also be detected as early as 15 minutes after addition of OHT, which allows investigation of temporal recruitment of DNA repair proteins (Massip *et al.*, 2010).

Compared to the irregular nature of DNA breaks formed by irradiation, AsiSI-ER creates 'clean' DSBs. After DNA breaks have been induced by AsiSI, histone modifications or protein recruitment can be investigated, both by ChIP-seq or ChIP polymerase chain reaction (ChIP-qPCR), by using primers specific for known AsiSI cutting sites (Iacovoni *et al.*, 2010).

Although AsiSI creates several DSBs across the genome, the enzyme is sensitive to methylated CpG islands, which are commonly found in inactive genes, and creates few breaks in heterochromatin (Clouaire *et al.*, 2018; Iacovoni *et al.*, 2010). Another potential problem with investigating DNA breaks and DNA repair mechanisms in cells expressing AsiSI-ER is that once OHT has been added, the enzyme will create breaks continuously, and may cut the same site several times after the break has been repaired (Aymard *et al.*, 2014). To prevent this, Aymard *et al.* created another fusion protein, with AsiSI-ER fused to an auxin-inducible degron, where addition of auxin leads to degradation of the fusion protein, stopping further break induction (Aymard *et al.*, 2014; Nishimura *et al.*, 2009).

1.5 The nuclear hormone receptor estrogen receptor

1.5.1 The role of estrogen receptor in regulation of gene expression and in cancer

ER is a TF which regulates expression of genes involved in cell division (Carroll, 2016). ER has two isoforms, ER α (*ESR1*) and ER β (*ESR2*). Estrogen receptor has two key domains: the ligand-binding domain (LBD), which binds estrogen, and DNA-binding domain (DBD, figure 1.10 A, B). Upon estrogen binding at the LBD, ER translocates to the nucleus, where it hetero/homo-dimerises and binds to specific ER motifs (known as estrogen response elements (EREs), commonly found within enhancers and promoters) via its DBD and activates transcription of its target genes.

The ERE consists of two inverted half sites, spaced by 3 basepairs, and each half site is bound by an ER, together making up the dimer (figure 1.10 C) (Mason *et al.*, 2010). ER acts as a TF, regulating the expression of thousands of genes, which are involved in processes including cell proliferation and reproduction (figure 1.10 D) (Yang *et al.*, 2017). The role ER plays in regulating cell division means that it is a prime candidate for mutation in cancer, and mutations affecting the regulation of ER leads to uncontrolled cell division and tumorigenesis. In fact, three quarters of breast cancers are characterized by presence of ER, and are termed ER+. ER α is the main isoform of ER (hereby referred to as ER).

Forkhead box A1 (FOXA1), another TF, displays pioneer activity; this means that it can bind compacted DNA and make it more accessible for the recruitment of additional factors (figure 1.10 D) (Cirillo *et al.*, 2002). FOXA1 motifs are highly enriched at/around ER binding sites (as categorised by ChIP-seq), and all ER-binding sites associated with ER+ cancer are dependent on FOXA1 expression (Hurtado *et al.*, 2011; Carroll *et al.*, 2005).

As well as ER and FOXA1, GATA binding protein 3 (GATA3) is also commonly mutated in ER+ cancer (Perou *et al.*, 2000). There is clear conservation in the spacing between ER, FOXA1 and GATA3 motifs within ER-binding sites (Serandour *et al.*, 2013), and it seems that these three factors may associate with adjacent regions of DNA in order to cooperate, form the estrogen response complex, and activate gene expression. Interestingly, mutating GATA3 at different sites affects the EREs bound by ER and FOXA1, suggesting that GATA3 may function to regulate ER binding patterns and ER-FOXA1 interactions (Theodorou *et al.*, 2013). It is worth noting that differential binding of ER to chromatin is associated with clinical outcome in breast cancer, which indicates that GATA3 mutations may be a key determinant of prognosis (Ross-Innes *et al.*, 2012).

In order to activate target gene expression, ER recruits chromatin remodellers, as well as transcriptional coactivators and machinery (Métivier, Reid and Gannon, 2006). Interestingly, there is some evidence that the recruitment of proteins involved in recognising and repairing DNA damage is important for establishing high levels of expression of ER target genes (Fong, Cattoglio and Tjian, 2013). It is thought that these proteins aid in chromatin modification and remodelling, facilitating the recruitment of transcriptional machinery, and downstream gene activation.

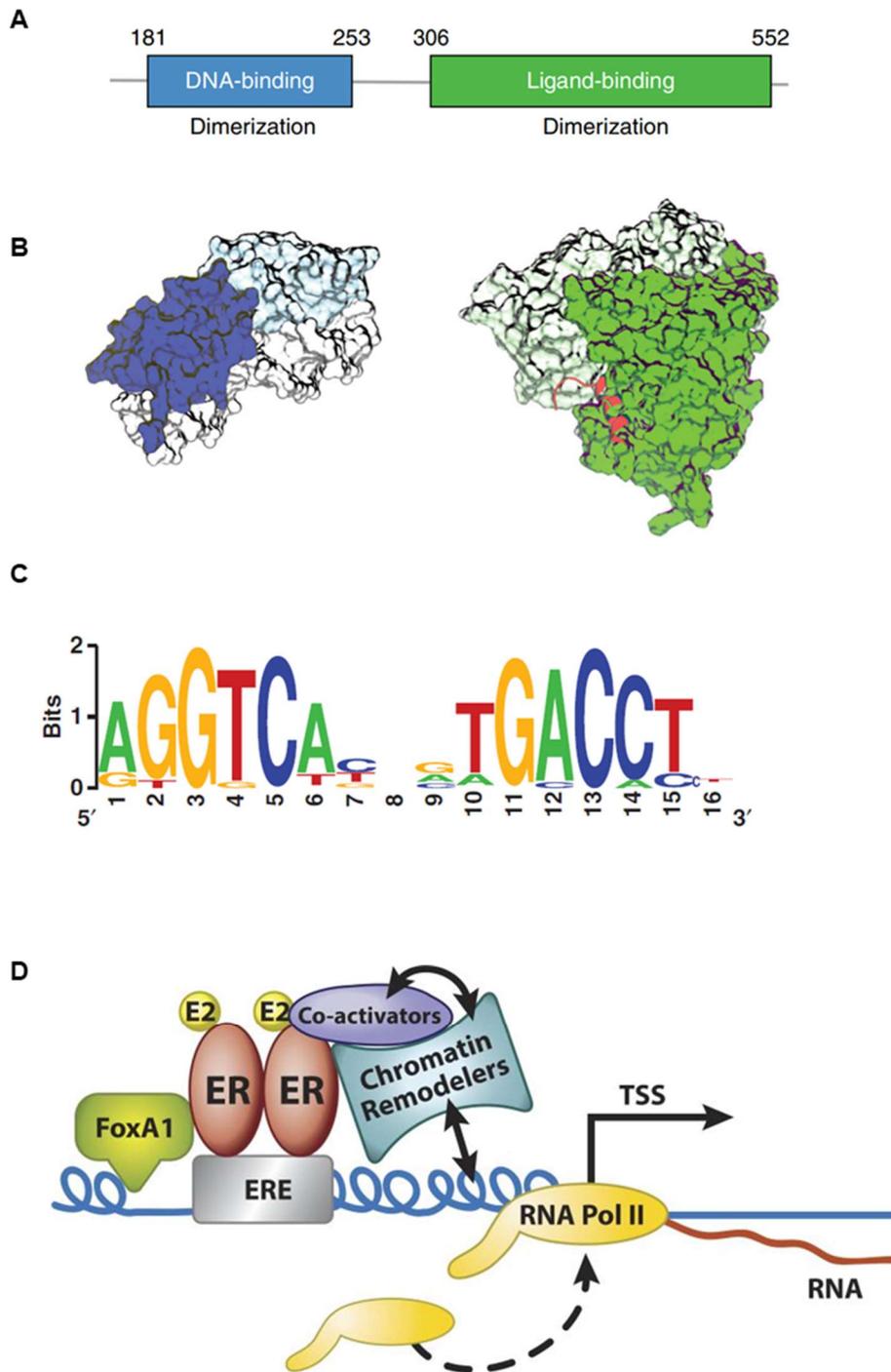


Figure 1.10. The estrogen receptor (ER) and estrogen response element (ERE).

- (A) ER contains a highly conserved DNA-binding domain (blue) and ligand binding domain (green). ER α forms homodimers upon ligand binding (Huang *et al.*, 2018).
- (B) Crystal structure of ER DNA-binding domain homodimer (light and dark blue) bound to its DNA motif, and ER ligand-binding domain homodimer (light and dark green)

bound to estradiol and a coactivator peptide (TIF2). The C-terminal helix H12, which holds the ligand in place, is shown in red (Huang *et al.*, 2018).

(C) The position weight matrix of the 15 nucleotide long ERE (Welboren *et al.*, 2009).

(D) The pioneer factor FOXA1 modifies chromatin making it more accessible to ER. Ligand-bound ER dimers bind EREs, and interacts with transcriptional co-activators and chromatin remodellers to improve accessibility of RNA polymerase II to transcriptional start sites (TSS) and initiate transcription of target genes (Hewitt, Winuthayanon and Korach, 2016).

1.5.2 Estrogen and ER lead to formation of DNA damage

ER+ breast cancers display high levels of genomic instability, caused by copy number aberrations, rearrangements, insertions/deletions and substitutions (Nik-Zainal *et al.*, 2016). Several single/double nucleotide substitution signatures have been identified, in addition to kataegis; localised hypermutations where a small chromatin region display high levels of mutations (Nik-Zainal *et al.*, 2012). Interestingly, somatic mutations seen in breast tumours (duplications, deletions, inversions, translocations) are significantly enriched at estrogen-regulated genes, suggesting that these genomic regions are susceptible to DNA damage (Stork *et al.*, 2016; Nik-Zainal *et al.*, 2016).

The presence of estrogen and activation of ER leads to genome-wide formation of DSBs, many of which are found near estrogen-target genes (Periyasamy *et al.*, 2015; Stork *et al.*, 2016). ER activates transcription of its target genes, which is associated with formation of R-loops (Santos-Pereira and Aguilera, 2015). In fact, estrogen induces formation of R-loops in a transcription-dependent manner on many ER target genes (Stork *et al.*, 2016). These R-loops lead to induction of DSBs, as seen by γ H2AX formation in regions surrounding the R-loops (Stork *et al.*, 2016). The DSBs are thought to be generated through clearance of the co-transcriptionally formed R-

loops by the DNA damage response (Stork *et al.*, 2016). However, another possibility is that R-loop formation leads to exposure of ssDNA on the opposite strand, and this ssDNA could be the target for DNA damage induction via APOBEC cytidine deamination (see following section).

1.6 APOBEC3B

1.6.1 APOBEC3B is a cytidine deaminase

Apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3 (APOBEC3B, A3B) is a member of the APOBEC3 (A3) enzyme family, seven closely related cytidine deaminases: A3A/B/C/D/F/G/H (Swanton *et al.*, 2015). The APOBEC3 family of enzymes are part of our innate immune defence against DNA, RNA and retroviruses including hepatitis B virus and HIV, and retrotransposons, inducing mutations by deaminating cytosine to uracil on ssDNA or RNA (Swanton *et al.*, 2015; Henderson and Fenton, 2015). A3B can also bind throughout the rest of the genome, and its substrates include ssDNA structures such as DNA synthesis lagging strand and R-loops (Stork *et al.*, 2016; Hoopes *et al.*, 2016). A3B contains a tandem zinc-coordinating C-terminal cytidine deaminase domain, which catalyses the cytosine to uracil deamination (Shi *et al.*, 2015).

APOBEC3s drive mutations in many cancers, including breast, by C-U deaminations leading to C-T/C-G transitions/transversions (APOBEC3 signature mutations) preferably at cytosines preceded by a thymine and followed by a thymine or adenine (TCW) (Zou *et al.*, 2017; Roberts *et al.*, 2013). A3B deamination can also cause kataegis in cancers, including breast cancer (Taylor *et al.*, 2013; Maciejowski *et al.*, 2020), and over 50% of primary breast cancers present increased A3B levels (Burns *et al.*, 2013). Overexpressed A3B leads to formation of DSBs, as seen by formation

of γ H2AX, suggesting that A3B activity is a source of DNA damage (Burns *et al.*, 2013).

1.6.2 APOBEC3B is involved in estrogen-target gene expression

A3B is expressed in both ER negative and positive breast cancers; however, high A3B expression levels are only linked with poor survival in ER-positive patients (Periyasamy *et al.*, 2015). This led to an investigation of the relationship between A3B and ER, and a novel role for A3B in ER target gene expression was discovered (Periyasamy *et al.*, 2015). In ER-positive breast cancer cell lines, A3B is recruited to ER binding sites, in an estrogen-dependent manner (Periyasamy *et al.*, 2015). This includes binding at the ERE's of estrogen-regulated genes TFF1, GREB1, FOS and CTSD, whose expression is reduced in absence of A3B, or A3B's cytidine deaminase domain (Periyasamy *et al.*, 2015). Although A3B is recruited by ER to ER-binding sites, ER binds these sites independently of A3B (Periyasamy *et al.*, 2015). Interestingly, upon recruitment to estrogen-target genes, A3B deaminates cytosines in the vicinity of the A3B/ER binding site (Periyasamy *et al.*, 2015). This is followed by recruitment of uracil-DNA glycosylase (UNG) by A3B, which excises uracil as part of the BER pathway (Periyasamy *et al.*, 2015). A3B cytidine deaminase activity and excision of the generated uracil is required for optimal gene expression, as seen by loss of UNG leading to reduced expression of the estrogen target genes (Periyasamy *et al.*, 2015; Zhang *et al.*, 2022). In fact, activation of the DNA damage response, following cytidine deamination, appears to be crucial for full expression of estrogen target genes.

1.6.3 APOBEC3B activity leads to formation of DSBs and γ H2AX

In the ER positive breast cancer cell line MCF7, DNA breaks lead to γ H2AX formation in an estrogen/ER/A3B-dependent manner. Performing ChIP-seq for γ H2AX, identified that 50% of γ H2AX binding sites co-localised with binding of A3B and/or ER, and 55% of the binding sites occupied by both A3B and ER were also occupied by γ H2AX (Periyasamy et al., 2015). The formation of γ H2AX is likely a result of excision of uracil by UNG leading to formation of abasic sites and DNA strand breaks, as the γ H2AX formation is downstream of A3B and UNG recruitment (Periyasamy et al., 2015). The positive regulatory function of the induced DNA damage is seen by formation of active histone marks (H3K9ac and H3K4me3), recruitment of RNA pol II, and recruitment of SWI-SNF chromatin remodelling factor BRG1 near estrogen target genes, in an estrogen and A3B-dependent manner (Periyasamy et al., 2015; Zhang et al., 2022). Together, this highlights a necessary role for DNA damage-induced by A3B in expression of estrogen-target genes (figure 1.11).

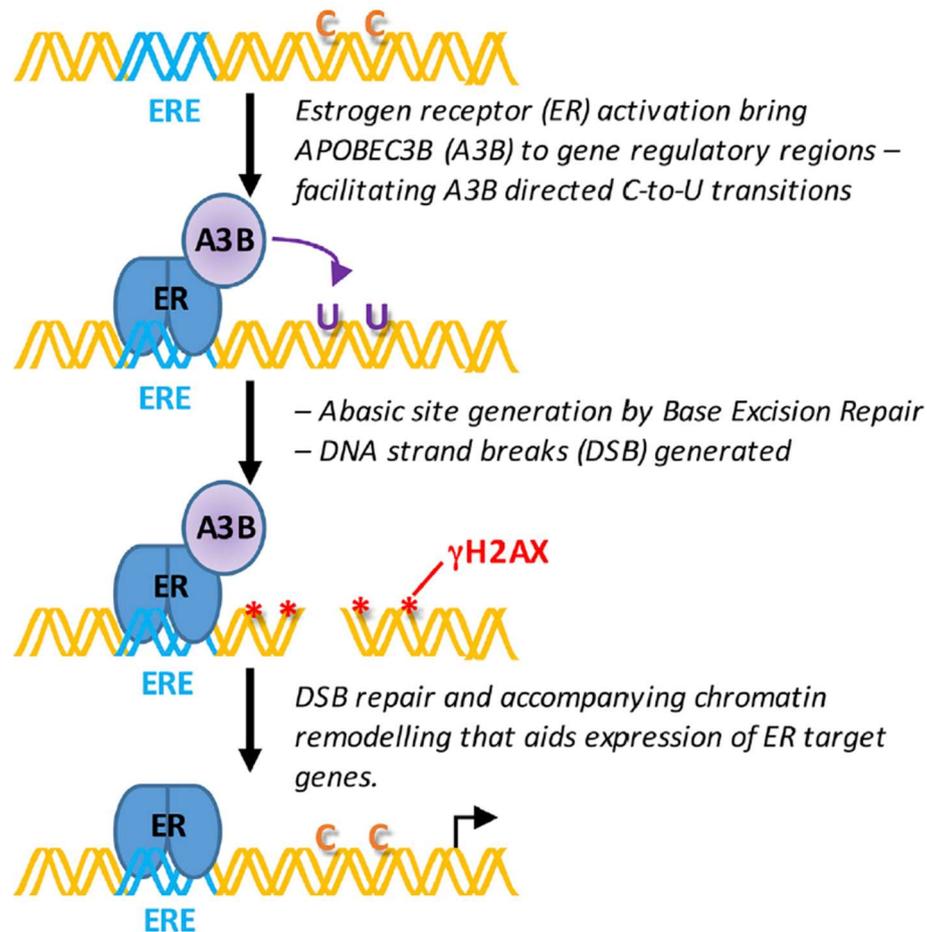


Figure 1.11. APOBEC3B cytidine deamination induces expression of ER target genes (Periyasamy *et al.*, 2015). ER binds ERE and recruits A3B, which deaminates nearby cytosines to uracil. Uracil is excised by BER, and DSBs are generated as seen by formation of γ H2AX. This leads to recruitment of DNA repair factors including chromatin remodellers, which aid in the expression of ER target genes.

1.6.4 APOBEC3B binds R-loops to regulate ER activity

Presence of estrogen leads to increased transcription, and increased R-loop formation near estrogen-responsive genes (Stork *et al.*, 2016). This is followed by formation of R-loop-dependent DSBs, suggesting another source of estrogen-induced DSBs (Stork *et al.*, 2016). However, a recent pre-print proposes a role for R-loops in A3B-mediated regulation of ER-target genes, where ssDNA in the R-loop

serves as a substrate for A3B. In agreement with findings in MCF7-cells (Stork et al., 2016; Periyasamy et al., 2015), the ER-positive T-47D breast cancer cell line display increased A3B binding (and R-loop formation) in presence of estrogen (Zhang et al., 2022). By expressing a UNG inhibitor in T-47D cells, investigation of estrogen-induced A3B deamination, without the removal of deaminated cytosines, was possible (Zhang et al., 2022). Interestingly, over half of estrogen-induced A3B binding sites are found in close proximity to R-loops; both estrogen-induced and pre-existing R-loops (Zhang et al., 2022). A3B binding sites near estrogen-induced R-loops are more enriched for TSS and transcribed enhancers, than A3B binding sites distal to R-loops, suggesting that A3B's involvement in transcriptional regulation is associated with its binding to R-loops (Zhang et al., 2022). In addition, A3B-signature mutations TC-TT are found on the displaced ssDNA strand of R-loops (Zhang et al., 2022). Of the estrogen-induced DSBs in T-47D cells, around a third are A3B-dependent and found near R-loops (Zhang et al., 2022). These DSBs are not formed upon loss of the transcription-coupled NER component Cockayne syndrome group B protein, which is required for converting R-loops into DNA breaks (Sollier et al., 2014). This suggests a role for R-loops in estrogen-induced A3B-dependent DSB formation, where BER excision of A3B deaminated bases forms SSB on the displaced strand, together with R-loop induced SSBs, form double-strand breaks (figure 1.12) (Zhang et al., 2022).

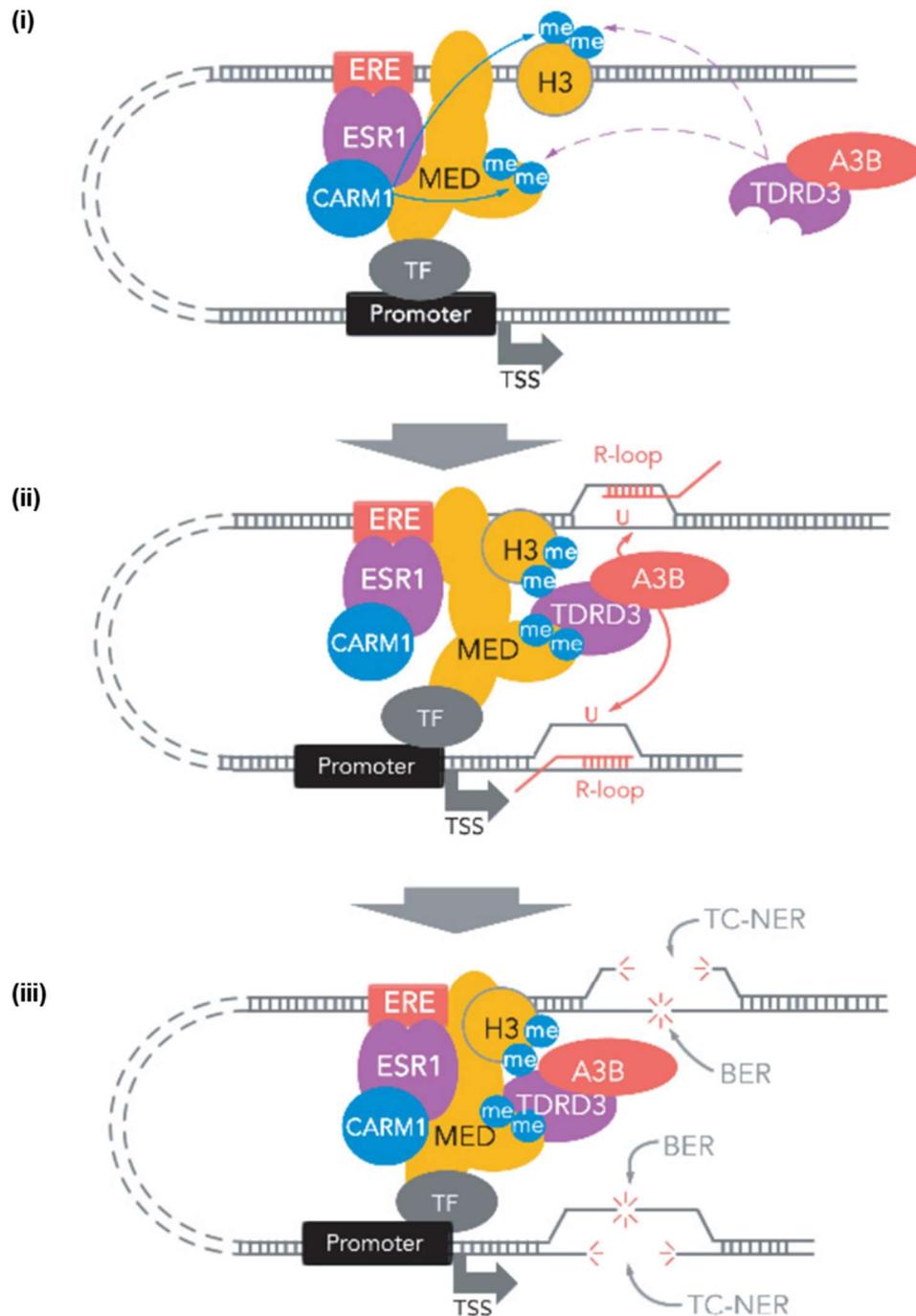


Figure 1.12. Suggested model of DNA double-strand break formation in response to estrogen receptor (ER)-mediated transcription and APOBEC3B activity (Zhang *et al.*, 2022). (i) ER (ESR1) homodimerises and binds ER-motifs (EREs) near promoters/transcriptional start sites (TSS) in presence of estrogen. TDRD3, which is involved in resolving R-loop structures, is recruited by asymmetric dimethylarginine (ADMA)-modified proteins (MED, H3) by CARM1. APOBEC3B (A3B) is recruited via TDRD3. (ii) Transcriptional activation by ER/ESR1 together with other transcription factors (TF) leads to formation of co-

transcriptionally formed R-loops. The displaced ssDNA strand serves as a substrate for A3B cytidine deamination (forming uracil). (iii) Uracil is excised by the base excision repair (BER) pathway and the DNA-RNA hybrid is removed by transcription-coupled nucleotide excision repair (TC-NER) factors lead to formation of single-strand breaks (which if in close proximity form double-strand breaks).

Together, this suggests a mechanism where estrogen-induced gene expression leads to transcription-associated R-loop formation at estrogen-regulated genes and enhancers. The R-loop provides a ssDNA-substrate for A3B, which deaminates cytosines in its vicinity, leading to formation of DSBs, and induction of estrogen-target gene expression (Zhang et al., 2022). For example, this is seen at the estrogen-regulated gene *RARA*, where the A3B peak overlaps with an estrogen-induced R-loop and A3B-dependent DSB (Zhang et al., 2022).

1.7 Key questions addressed in this thesis

This thesis is comprised of three related sub-projects. The first addresses γ H2AX detection at DSBs using the novel technique CUT&Tag; the second addresses the investigation of the hypothesised histone mark di- γ H2AX in the DNA damage response; and the third addresses the role of A3B in up-regulating expression of ER target genes. **Results chapter one:** The recently developed protein-chromatin interaction assay CUT&Tag was used to investigate whether γ H2AX can be detected at DSBs in a DSB-inducible cell system. This allowed investigation of the γ H2AX signal at and surrounding DSBs at higher resolution than that which is possible using traditional ChIP-seq methods (Chapter 3). **Results chapter two:** The di- γ H2AX histone mark is proposed to play a role in cell fate decisions, by leading the cell towards programmed cell death, instead of repair and survival. Using a novel di- γ H2AX antibody with high specificity developed by the Ellis group, the presence of di- γ H2AX in DNA damaged cells was probed by western blot, immunofluorescence microscopy and ChIP-qPCR (Chapter 4). **Results chapter three:** A3B-induced DSB formation is important for ER-regulated gene expression. However, A3B binds at many chromatin regions independently of ER. Bioinformatic analysis of publicly available, previously published data was used to investigate A3B's binding sites with or without ER, to further study the role of A3B in regulation of gene expression (Chapter 5).

2.0 Chapter 2: Methods

2.1 Cell culture

All cell lines were cultured at 37°C under humidified conditions, with 5% CO₂.

HEK293T (human embryonic kidney 293T), mouse embryonic fibroblast (MEF) and MEF H2AX knock-out (KO) cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Gibco, 41966052) supplemented with 10% Fetal Bovine Serum (FBS, Gibco, 11570516) and 1% L-Glutamine–Penicillin–Streptomycin (Sigma-Aldrich, G6784).

MCF10A (human mammary gland epithelial) cells were cultured in Minimum Essential Medium α , GlutaMAX Supplement (Gibco, 32561029), supplemented with 5% Horse Serum (Sigma-Aldrich, H1138), 1% L-Glutamine–Penicillin–Streptomycin (Sigma-Aldrich, G6784), 0.5 μ g/mL Hydrocortisone (Sigma-Aldrich, H6909), 10 μ g/mL Insulin (Sigma-Aldrich, I9278), 20ng/mL epidermal growth factor (Sigma-Aldrich, E9644) and 100ng/mL cholera toxin (Sigma-Aldrich, C8052).

DivA (DSB Inducible via AsiSI, U2OS human bone osteosarcoma epithelial) cells were cultured in DMEM, High Glucose, GlutaMAX Supplement (Gibco, 61965026) supplemented with 10% FBS (Gibco, 11570516), 1% Penicillin-Streptomycin (Gibco, 15140148), 1mM Sodium Pyruvate (Gibco 11360070) and 1 μ g/ml Puromycin (InvivoGen, ant-pr-1).

Cells were maintained in the logarithmic growth phase by passaging every 2-3 days, and were detached from the cell culture vessel by trypsinisation. Cells were washed twice in phosphate-buffered saline (PBS) before 0.02mL trypsin (Gibco™ Trypsin-EDTA (0.25%), 11560626) per cm² surface area of the cell culture vessel was added, until all cells were detached. Cells were plated at 20% confluency for each cell culture vessel.

2.2 The novel γ H2AX and di- γ H2AX antibodies used in this study

Generation of the novel di- γ H2AX antibody used in this study (table 2.1) was performed under contract by PTM Biolabs. Briefly, three rabbits were immunised with a peptide matching the terminal 12 residues of human H2AX, doubly phosphorylated at ser139 and tyr142. The immunogen sequence: CGGKKATQAS^PQEY^P and the peptide was conjugated to keyhole limpet hemocyanin (KLH) as an adjuvant to increase immunogenicity. Immunised rabbits were boosted twice (i.e. three injections in total) to boost antibody production. Following test bleeds to check specificity of the resulting antibodies (performed by PTM Biolabs, data not shown), sera from rabbits #2 and #3 were purified by one round of immunoselection using the original immunogen peptide, and one round of immunodepletion to remove antibodies cross-reacting with other phospho-forms (peptides used for immunodepletion: CGGKKATQAS^PQEY, CGGKKATQASQEY^P and CGGKKATQASQEY). After further specificity testing, the serum from rabbit #2 was subjected to a second round of immunodepletion to remove residual cross-reactive antibodies. Antibodies AP13 and AP14 (table 2.1) are the first- and second-round purified antibodies from rabbit #2 of this project.

In a parallel project, a novel mouse monoclonal γ H2AX antibody was made. For this, four 6-week mice were immunised with a peptide corresponding to the C-terminal tail of ser139-phosphorylated mouse H2AX, conjugated to KLH. The immunogen sequence was CKKASQAS^PQEY and mice were immunised four times at two-weekly intervals. Spleen cells harvested from the mice postmortem were fused with sp2/0 myeloma cells, and individual candidate antibody-producing clones were screened for reactivity against the various different phospho-forms of γ H2AX. The final clone was shown to produce an antibody specific for γ H2AX with no cross-reaction to other phospho-forms.

2.3 DNA damage induction

DNA-damaging agents were diluted in the respective media.

HEK293T and MCF10A cells were treated with 5 μ M bleomycin (Thermo Scientific, 15483439) for 2 hours, or for 2 hours followed by 2 hours of recovery in untreated media (where indicated).

DlvA cells were treated with 300nM 4-hydroxytamoxifen (OHT, Sigma-Aldrich, H7904) for 4 hours (where indicated).

2.4 Immunofluorescence (IF)

Cells were grown on glass coverslips, pre-treated with 1mg/mL Poly-D-Lysine (Sigma-Aldrich, A-003-E) to aid cell adhesion. Prior to fixation, all buffers were used at 4°C and incubation steps were also performed at 4°C, to reduce protein degradation.

Immunofluorescence was performed as previously described (Sawasdichai *et al.*, 2010). After DNA damage induction, media was removed by washing cells twice with PBS, then incubated for 1 minute in cytoskeletal (CSK) buffer (10mM PIPES, 300mM sucrose, 100mM NaCl, 3mM MgCl₂, 1mM EDTA, 1mM DTT, Protease and Phosphatase Inhibitor (Thermo Scientific, 15662249), pH 6.8) with 0.1% Triton X-100, to permeabilise cells and extract cytoplasmic and loosely bound nuclear proteins. Cells were washed twice with PBS, and incubated for 20 minutes in CSK with 0.5% Triton X-100, to extract tightly bound proteins. Cells were washed three times with PBS, followed by fixation in 4% paraformaldehyde (PFA). To reduce autofluorescence, cells were incubated in 50mM NH₄Cl (in PBS) for 15 minutes. Cells were washed twice with PBS, then washed twice in tris-buffered saline with 0.025% Triton X-100 (TBS-T, 20mM Tris, 150mM NaCl, pH 7.6, preferred when studying phosphorylated proteins) for 5 minutes, to permeabilise cells. Cells were blocked for

2 hours in TBS with 10% FBS and 1% bovine serum albumin (BSA, Sigma-Aldrich, A2153), to reduce non-specific binding. Cells were then incubated overnight with primary antibodies, as indicated (table 2.1). All antibodies were diluted in TBS with 1% BSA, and incubation was performed in a humidity chamber at 4°C. The next day, unbound primary antibody was washed off by rinsing cells three times for 5 minutes with TBS 0.025% Triton X-100. Secondary antibodies (table 2.2), diluted in TBS with 1% BSA, were then added to cells, followed by incubation in a humidity chamber for 2 hours at room temperature. Unbound secondary antibody was then removed by washing cells three times, for 5 minutes, in TBS. Coverslips were then mounted onto Superfrost microscope slides (Fisher Scientific, Fisher Scientific) using Antifade Mounting Medium with DAPI (Vector Labs, H-1200), where the minor groove-binding dye DAPI stains DNA/nuclei.

2.5 Microscopy and imaging

Microscopy was performed on an Olympus BX-61 microscope with a CCD camera and DAPI, FITC and Texas Red filters, using SmartCapture imaging software (version 4).

Images were edited using FIJI and GIMP image processing software. Colour levels in each channel were similarly enhanced across images from each cell line/experiment.

2.6 Cell lysis and protein extraction

After DNA damage induction, cells were trypsinised (as described in section 2.1 Cell culture), collected in centrifuge tubes, and pelleted (200 x g for 5 minutes). The next steps were performed at 4°C to prevent protein degradation. Cells were lysed in modified radioimmunoprecipitation assay (RIPA) buffer containing the nuclease

benzonase (25mM Tris-HCl, 150mM NaCl, 1% NP-40, 1% sodium deoxycholate, 1% SDS, 1mM EDTA, Protease and Phosphatase Inhibitor (Thermo Scientific, 15662249), 0.2U/uL benzonase, pH 7.6). RIPA buffer was added 1:1 to the cell pellet volume, and mixed by pipetting and gentle vortexing. Lysed cells were incubated on ice for 30 minutes, then centrifuged for 30 minutes (13,000 x g, 4°C). The supernatant, containing solubilised protein, was transferred to a new tube.

2.7 Protein quantitation

Protein levels in histone extracts were quantified using the BCA protein assay kit (Thermo Scientific, 23235), following the manufacturer's instructions. Proteins present in the lysate reduce Cu^{2+} to Cu^{1+} , in an alkaline environment. Bicinchoninic acid (BCA) forms complexes with Cu^{1+} , whose absorbance at 562nm is linear with increasing protein concentrations. Protein levels were detected by imaging on Multiscan FC Photometer (Thermo Scientific, 51119000). Standards containing BSA at known concentrations were included to allow calculation of exact protein concentrations in the histone extracts.

2.8 SDS-PAGE and western blot (WB)

Lysates were mixed with Laemmli sample buffer (10% glycerin, 60mM Tris-HCl, 2% SDS, 0.1M DTT, 0.01% bromophenol blue, pH 6.8) and heated to 95°C for 5 minutes to denature protein and add a net negative charge, ensuring separation towards the anode is dependent solely on the protein's molecular weight. 25µg of protein was loaded per sample. Samples were run on 4-12% Bis-Tris Mini Protein Gels (Invitrogen, NP0321BOX) at 150V for 1 hour in MOPS SDS Running Buffer (Invitrogen, NP0001). Before transfer, PVDF Western Blotting Membranes (Roche, 03010040001) were activated/hydrated in 100% methanol, before membranes and

gels were incubated in Bjerrum and Schafer-Nielsen transfer buffer (48mM Tris, 39mM Glycine, pH 9.0-9.4) for 15 minutes, to equilibrate, and remove salts and detergents from the gels. Proteins were transferred from gel to membrane by semi-dry transfer using Trans-Blot SD Semi-Dry Electrophoretic Transfer Cell (Bio-Rad, 170-3940), at 15V for 20 minutes.

To prevent non-specific binding of antibodies to the membrane, membranes were blocked (5% BSA, TBS-T) for 1 hour, before being incubated with primary antibody (table 2.1, diluted in blocking buffer) overnight at 4°C. Unbound or loosely bound primary antibody was removed by washing five times for 5 minutes in TBS-T. Membranes were then incubated in secondary antibodies (table 2.2) diluted in blocking buffer, for 2 hours at room temperature. Unbound secondary antibody was then removed by washing five times for 5 minutes in TBS-T. Membranes were then incubated for 5 minutes in ECL Western Blotting Substrate (Thermo Scientific, 10590624), before imaging on Syngene G:BOX Chemi XX6 using GeneSys image capture software, for chemiluminescent detection of proteins.

2.9 Dot blot

Peptides were diluted in DEPC-treated H₂O (Thermo Fisher, AM9906) to final concentrations of 256, 64, 16, 4 and 1 ng/μL (table 2.3).

1.5μL of peptide solution was pipetted onto nitrocellulose membranes, which were then left to dry. Membranes were blocked and incubated with primary and secondary antibodies (table 2.1, 2.2), as above.

2.10 Chromatin Immunoprecipitation (ChIP)

The day before fixation, 5×10^6 cells were seeded in 140mm dishes, which yield approximately 300 μ g chromatin each, sufficient for 3 IPs. All buffers were made up with molecular biology grade H₂O (Corning, 46-000-CM) and filter sterilised before use.

Per sample, 100 μ L protein A/G beads (Sigma P-7786 and P-3296, ratio 1:1) for pre-clearing and 100 μ L protein A/G beads for the IP were prepared. A mix of protein A and G beads were used, to allow use of both rabbit- and mouse-derived antibodies for the IPs. Beads were washed three times in IP buffer (1:9 nuclear lysis buffer to dilution buffer, see below), resuspended in 100 μ L per sample IP buffer with 0.5mg/mL BSA and incubated for four hours at 4°C, to reduce non-stringent binding to beads. Beads were then washed twice in IP buffer, and resuspended in 100 μ L per sample IP buffer.

ChIP was performed as previously described (Tyteca *et al.*, 2006). After DNA damage-induction, DlvA cells were incubated in 1% formaldehyde in cell culture media for 15 minutes, crosslinking proteins and DNA. Formaldehyde was then quenched by incubating in 0.125M glycine for 5 minutes. Cells were washed twice in PBS and collected by scraping in 1mL cold PBS per dish, followed by centrifugation at 2000rpm for 5 minutes. Cell pellets were kept at -80°C overnight. Pellets were resuspended in 200 μ L per million cells cellular lysis buffer (5mM PIPES pH 8, 85mM KCl, 0.5% NP-40, protease inhibitor cocktail (Roche, 11873580001), incubated for 10 minutes on ice, and homogenised on ice using a Dounce homogeniser and B pestle (Kimble Chase Tissue Grinder, 885303-0015 and Tissue Grind Pestle Sc15, 885302-0015, 20 passes, 2 minute rest, 20 passes), lysing the cells by shear stress. Nuclei were pelleted by centrifuging at 4000rpm for 5 minutes at 4°C, resuspended in 100 μ L per million cells nuclear lysis buffer (50mM Tris pH 8.1, 10mM EDTA, 1% SDS, protease inhibitor cocktail), and incubated for 10 minutes on ice. To fragment the

DNA, nuclei were sonicated ten times for 10 seconds (Branson sonifier 450 with microtip, power setting 5, 50% duty cycle), creating fragments with average length of 500bp. DNA content was then estimated using NanoDrop 1000 Spectrophotometer (Nucleic Acids application module, DNA-50). Samples were diluted ten times in dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2mM EDTA, 16.7mM Tris pH 8.1, 167mM NaCl, protease inhibitor cocktail), with 200µg DNA per sample. To pre-clear the samples, to reduce non-specific binding to beads, 100µL of the previously prepared beads were added per sample, and incubated for 1-2 hours at 4°C, rotating. Samples were then centrifuged at 4000rpm for 10 minutes, supernatants transferred to new tubes, and 100µL input removed and stored at -20°C. Antibodies were added (table 2.1) and incubated overnight at 4°C, rotating. Each round of ChIP included beads-only controls, with no antibodies added.

Proteins of interest, and associated chromatin, were immunoprecipitated by adding 100µL of previously prepared A/G beads per IP and incubated for 2 hours at 4°C, rotating. Samples were then centrifuged at 2000rpm for 1 minute; beads were washed once in 1mL per sample dialysis buffer (2mM EDTA, 50mM Tris pH 8.1, 0.2% sarkosyl, protease inhibitor cocktail) and transferred to DNA LoBind tubes (Eppendorf, 0030108051). Beads were washed five times in wash buffer (100mM Tris pH 8.8, 500mM LiCl, 1% NP-40, 1% sodium deoxycholate, protease inhibitor cocktail), incubated for 10 minutes at 4°C, rotating per wash. Beads were then washed twice in TE (10mM Tris-HCl, 1mM EDTA, pH 8) buffer and resuspended in 200µL TE per IP. Input samples were included from this step (resuspended in 100µL TE to same final volume as ChIP samples). To degrade RNA which can interfere with the DNA purification, RNase A (50µg/mL) was added, and samples were incubated for 30 minutes at 37°C. To reverse formaldehyde crosslinks, SDS (0.2%) was added overnight with shaking (1200rpm) at 70°C. To digest proteins (including nucleases),

and further reverse the cross links, proteinase K (0.2µg/mL, Roche, 03115828001) was added for 1.5 hours at 45°C, shaking (1200rpm).

To extract the immunoprecipitated and input DNA, 1:1 volume of phenol (Invitrogen, 15593-031) was added, samples were vortexed and centrifuged for 5 minutes 1400rpm. The aqueous, top layer containing DNA was transferred to new DNA LoBind tubes. 200µL per IP TE was added to the lower, organic/phenol fraction, vortexing and centrifuging was repeated, and this aqueous fraction was added to the first (400µL final volume per IP). Phenol extractions leave small amounts of phenol in the aqueous phase. Chloroform extracts phenol from the aqueous phase, and two successive chloroform (400µL per IP, Sigma, 25666) extractions were performed, with vortexing and centrifuging as above. Final volume of 350µL per IP. DNA was then precipitated in 70% ethanol, 83mM sodium acetate and 4ng/µL glycogen (Invitrogen 10814-010) overnight at -20°C. The next day, samples were centrifuged at 1400rpm 4°C for 30 minutes; pellets were washed in 70% ethanol, centrifuged as before for 10 minutes, air dried and resuspended in 100µL nuclease-free H₂O per pellet (Promega, P119C).

2.11 Cleavage Under Targets and Tagmentation (CUT&Tag)

CUT&Tag was performed using the Hyperactive In-Situ ChIP Library Prep Kit for Illumina (Vazyme, TD902), following the manufacturer's instructions. The following volumes are per sample.

Concanavalin A (ConA, lectin binding extracellular glycoproteins)-coated magnetic beads were mixed gently with binding buffer (20mM HEPES pH 7.9, 10mM KCl, 1mM CaCl₂ and 1mM MnCl₂, 10µL beads in 1.5mL buffer). Tubes were placed on a magnetic stand to clear, and supernatant was removed. This was repeated twice, resuspending beads in 100µL, then 10µL of binding buffer. Throughout the protocol,

tubes were placed on a magnetic stand, to remove supernatant without transferring beads.

Before permeabilisation of cells, all steps were performed at room temperature, to reduce stress on the live cells. All volumes are per sample, unless otherwise stated.

2.11.1 Preparing testis samples

The testis samples were prepared by Frances Burden, Ellis group. Both testes were harvested from a mouse, and the tunica vaginalis and tunica albuginea were removed. The seminiferous tubules were cut in small pieces, and placed in a centrifuge tube with 18mL DMEM, no glucose, no glutamine, no phenol red (Gibco, A1443001). 2mL of 2.5% Trypsin, no phenol red (Gibco, 15090046) was added, and the tubules were incubated for 30 minutes at 31°C with gentle mixing, to dissociate cells from the tissue. Trypsin was quenched by adding 1.8mL FBS and the cell suspension was passed through a 40uM cell strainer to form a single-cell suspension and remove Sertoli cells. The cells were pelleted (500 x g, 15 minutes, room temperature), and the pellet resuspended in 5mL DMEM, no glucose, no glutamine, no phenol red. Cells count and viability was assessed by diluting the cell suspension with trypan blue (1:10) and counting cells excluding the dye on a haemocytometer. Cell viability >90% is recommended for CUT&Tag to reduce background noise. 100,00 cells were used per sample.

2.11.2 Preparing DivA cells

DivA cells (treated/untreated) were harvested at room temperature. Cell viability was assessed by staining cells with 0.2% trypan blue solution. 80,000 cells were used per sample.

2.11.3 Antibody recognition of targets and tagmentation

Cells were centrifuged at 600g for 3 minutes and supernatant was discarded. Cells were resuspended in 500 μ L wash buffer (20mM HEPES pH 7.5, 150mM NaCl, 0.5mM spermidine, Complete Protease Inhibitor EDTA-Free tablet (Roche, 5056489001)); centrifugation was repeated, and supernatant discarded. Cell pellets were resuspended in 100 μ L wash buffer, transferred to new tubes, and 10 μ L prepared bead slurry was added while vortexing gently, ensuring thorough mixing, binding cells to the beads. Samples were incubated on a rotator for 5-10 minutes, then gently centrifuged and supernatant containing unbound cells was removed. To permeabilise cell membranes, 50 μ L of ice-cold antibody buffer (dig-wash buffer with 2mM EDTA and 0.1% BSA) was added.

Primary antibodies (table 2.1) were added while gently vortexing, and samples were incubated, rocking overnight at 4°C. Supernatants were discarded, removing any unbound primary antibody, and 50 μ L of secondary antibody (table 2.2) diluted in dig-wash buffer (wash buffer with 0.05% digitonin) was added while vortexing. Samples were incubated at room temperature for 60 minutes, rocking. Adding secondary antibodies increases protein G/A binding to the antibody complex. Supernatants were discarded, 800 μ L dig-wash buffer was added, and tubes were inverted 10 times to wash the beads/cells, removing un-bound secondary antibodies. This was performed 3 times.

Hyperactive protein G/protein A-Tn5 transposon was mixed with dig-300 buffer (20M HEPES pH 7.5, 300mM NaCl, 0.5mM spermidine, 0.01% digitonin, Complete Protease Inhibitor EDTA-Free tablet), final concentration 0.04 μ M. The increased NaCl concentration reduces pA-Tn5 binding to accessible regions of chromatin (i.e. non-targeted binding). Supernatants were removed from samples/beads. To bind hyperactive pA/pG-Tn5 transposon fusion enzyme to antibodies binding target proteins, 100 μ L of hyperactive pA/pG-Tn5 transposon/dig-300 buffer mix was added

while vortexing, and incubated for 1 hour at room temperature, rocking. Supernatants were discarded, removing un-bound pA/pG-Tn5, and beads were washed three times by adding 800µL dig-300 buffer, inverting tubes 10 times to mix and discarding supernatants. In presence of Mg²⁺, Tn5 transposon tagments DNA, by cutting the DNA and adding adapter sequences to both ends of the fragments. 300µL of tagmentation buffer (5 mL Dig-300 buffer with 0.01M MgCl₂) was added, samples were vortexed to mix gently and incubated at 37°C for 1 hour. The tagmentation reaction was terminated by adding 16.7µM EDTA, 0.1% SDS and 166.7 mg/mL Proteinase K at room temperature. This digests proteins (including nucleases) within the samples and releases the DNA fragments for further processing. Samples were vortexed gently, and incubated at 37°C overnight.

To extract the DNA fragments, 300µL phenol:chloroform:isoamyl alcohol (25:24:1) was added, and mixed by vortexing at high speed. Samples were centrifuged at 16,000 x g for 5 minutes at room temperature. The upper, aqueous layer containing DNA was transferred to a new tube, leaving behind proteins in the organic, lower phase. To extract phenol residues from the aqueous phase, 300µL chloroform was added, samples were mixed by inverting 10 times and centrifugation repeated. The aqueous layer was transferred to a new tube containing 750µL 100% ethanol, and mixed by pipetting. Samples were chilled on ice and centrifuged at 16,000 x g for 15 minutes at 4°C. Supernatant was then discarded, and the pelleted DNA was washed in 1mL of 100% ethanol, followed by centrifugation at 16,000 x g for 1 minute at 4°C. Supernatants were removed and DNA pellets air-dried. Pellets were then resuspended in 25/30µL TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA), and samples stored at -20°C.

2.11.4 Library amplification

For each sample, the following was prepared in a sterile PCR tube: 24µL purified DNA product, 5µL ddH₂O, 10µL 5x TAB (Trueprep Amplify Buffer), 5µL i5 indexed primer, 5µL i7 indexed primer, 1µL TAE (TruePrep Amplify Enzyme). Samples were mixed gently by pipetting and PCR was performed using a Multigene OptiMax Thermal cycler (Labnet, TC9610), as follows: hot lid of 105°C, 72°C 3 mins; 98°C 30s then 98°C 15s, 60°C 30s, 72°C 30s for 15 cycles; followed by 72°C 5 mins, and 4°C hold. A different barcode was used per sample (table 2.4), from the TruePrep Index Kit V2 for Illumina (Vazyme, TD202).

DNA concentration was measured by Qubit 4 Fluorometric quantification assay (Thermo Fisher, Q33239), using the Qubit dsDNA Quantitation HS kit (Invitrogen, Q32851), following the manufacturer's protocol.

Samples with DNA concentration below 0.5ng/µL were re-amplified by PCR. Per reaction, 10µL 5X Q5 Reaction buffer (NEB, B9027S), 1µL 10mM dNTPs (Invitrogen, 18427013), 5µL forward and reverse primers (as above), 1-3ng DNA, 0.5µL Q5 High-Fidelity DNA polymerase (NEB, M0491), nuclease-free H₂O (up to 50µL). PCR was performed as above, with 6 additional cycles.

2.11.5 PCR product purification

VAHTS DNA clean beads (Stratech, N411-01-VAZ) were equilibrated to room temperature. 60µL/1.2x volume of beads (giving DNA fragments over 300bp) were pipetted into each PCR product, mixed, and the PCR product/bead mix was incubated for 5 minutes at room temperature, binding DNA fragments to the beads. Tubes were placed on a magnetic stand, and supernatant was removed. Beads were then washed twice with 200µL 80% ethanol; each time, ethanol was added, beads incubated in ethanol for 30 seconds, and ethanol removed, without removing tubes from the

magnetic stand. Beads were then air-dried for 3-5 minutes, to remove all ethanol, and resuspended in 22 μ L nuclease-free H₂O. Beads were incubated in water for 5 minutes, to elute DNA fragments; tubes were then placed on a magnetic stand, and 20 μ L of the supernatant (containing purified DNA library) was transferred to a new tube, and stored at -20°C.

2.12 Quantitative real-time Polymerase Chain Reaction (qPCR)

ChIP-qPCR: Samples were heated at 65°C for 5 minutes. Each reaction was performed in duplicate, including input. Per reaction: 2 μ L input/sample mixed with 1 μ L each of forward and reverse primers (table 2.5), 6 μ L TB Green Premix Ex Taq (Takara, RR420L), 2 μ L nuclease-free H₂O, 12 μ L total volume per well.

CUT&Tag-qPCR: Each reaction was performed in duplicate. Per reaction: 5 μ L SYBR Green Master Mix (Thermo Fisher, A25742), 800nM forward and reverse primer (table 2.5), 1 μ L DNA, nuclease-free H₂O to 10 μ L final volume.

qPCR was run on Quantstudio 3 Real-time PCR system (Thermo Fisher, A28567), following manufacturer's protocol, as follows: 50°C for 2 minutes, 95°C for 2 minutes; then 95°C for 15 seconds; 59°C for 15 seconds; 72°C for 1 minute, for a total of 40 cycles, followed by 95°C for 15 seconds, 60°C for 1 minute and 95°C for 15 seconds.

2.13 Bioanalyzer

To assess size of DNA fragments produced by CUT&Tag, samples were run on a bioanalyzer (Agilent, G2938C). This was performed at the Earlham Institute by Ashleigh Lister.

2.14 Next-generation sequencing

CUT&Tag libraries were sequenced by paired-end Illumina MiSeq v2 (50 cycles). This was performed at the Earlham Institute by the Geomics Pipelines Group.

2.15 ChIP-seq and BLESS analysis

GEO accession numbers for files analysed: GSE56979 (A3B and input (non-IP'ed) ChIP-seq), GSE57426 (γ H2AX ChIP-seq), GSE14664 (ER α ChIP-seq), GSE31755 (H3K27ac ChIP-seq), GSE86714 (H3K4me1 ChIP-seq), GSE96506 (H3K4me3 ChIP-seq). ArrayExpress accession numbers for DivA files analysed: E-MTAB-5817 (DivA γ H2AX ChIP-seq and BLESS).

A summary of the ChIP-seq and BLESS processing and analysis is shown in figure 2.1. Files were downloaded and quality of the sequencing data was assessed using the FastQC tool (version 0.11.9), analysing the following metrics: per base sequence quality, per tile sequence quality, per sequence quality score, per base sequence content, per sequence GC content, per base N content, sequence length distribution, sequence duplication levels, overrepresented sequencing and adapter content (Andrews, 2010).

Reads were then aligned to the human genome (hg19) using bowtie2 (version 2.3.5, including input file). Hg19 was used for analysis for two reasons: firstly, to be compatible and comparable with datasets presented in Clouaire *et al* and Periyasamy *et al*; secondly, due to its well-established annotation and abundance of compatible analysis tools. BLESS paired-end reads were aligned to the human genome using bowtie2 (--local --very-sensitive-local (trimming of reads to maximise alignment), -l 10 -X 500 (minimum and maximum fragment length for paired-end alignments = 10 and 500, respectively)). Samtools (version 1.10) was used to convert SAM files to binary BAM files (samtools view), which were sorted by chromosomal position (samtools

sort) and indexed (samtools index), and PCR duplicates removed (samtools rmdup). Number of reads mapping to any given genomic position was calculated using bedtools (version 2.29.2) genomecov. bedGraphToBigWig (version 4) was used to convert bedGraph files to bigWig files. For the DivA files, bamCoverage (bin size = 20, normalise read coverage by counts per million (CPM, reads x 1 million / total number of reads sequenced)) was used to convert BAM files to bigWig coverage tracks. bigWig files were viewed on the University of California Santa Cruz (UCSC) genome browser.

To identify genomic areas where significant numbers of reads align – where proteins of interest are interacting with DNA – peak calling was performed, using MACS2 (version 2.2.7, control = input file, q = 0.5 for A3B files and 0.01 for ER α / γ H2AX files, broad peak and q = 0.05 for DivA γ H2AX file, control = untreated file and q = 0.01 for DivA BLESS files) or LanceOtron (Find and Score Peaks) (Zhang *et al.*, 2008; Hentges *et al.*, 2022). Overlapping peaks from different datasets were identified using bedtools intersect, and blacklisted regions, containing repeat elements or other regions leading to high background signal, were removed (Amemiya, Kundaje and Boyle, 2019).

To identify motifs underlying the peaks, *de novo* motif analysis was performed using the meme-chip tool within the MEME suite (MEME database: HOCOMOCO human) (Bailey *et al.*, 2015). Peak regions were annotated, and gene and genome ontology analysis performed using HOMER (version 4.11) annotatePeaks (Heinz *et al.*, 2010). A list of predicted MCF7 enhancers was downloaded from EnhancerAtlas 2.0 (Gao and Qian, 2020). Heatmaps were made using deepTools (version 3.5.0) computeMatrix (reference point = center) and plotHeatmap (interpolation method = nearest), where bigwig tracks were plotted across regions from a bed-file (as stated). Instances of full/half ERE motifs in the genome were identified using motif sequence from HOCOMOCO (ESR1_HUMAN.H11MO.0.A and ESR1_HUMAN.H11MO.1.A for

full and half ERE, respectively) and PWMTools (PWMscan) (Ambrosini, Groux and Bucher, 2018). A list of human transcriptional start sites (TSS) was generated by downloading RefSeq whole gene region list, and creating a TSS region file containing start of genes +1 bp. Association between genomic regions from different datasets was analysed by performing permutations tests using regioneR (version 1.26.1, permTest, filter chromosomes = hg19 autosomes and sex chromosomes, number of permutation tests = 1000, randomise function = randomizeRegions, count.once = TRUE, set.seed = 123, mc.set.seed = FALSE, evaluate function = numOverlaps). Functional enrichment analysis of genes regulated by promoters/enhancers of interest was performed using Metascape (Express analysis, multiple gene lists) (Zhou *et al.*, 2019).

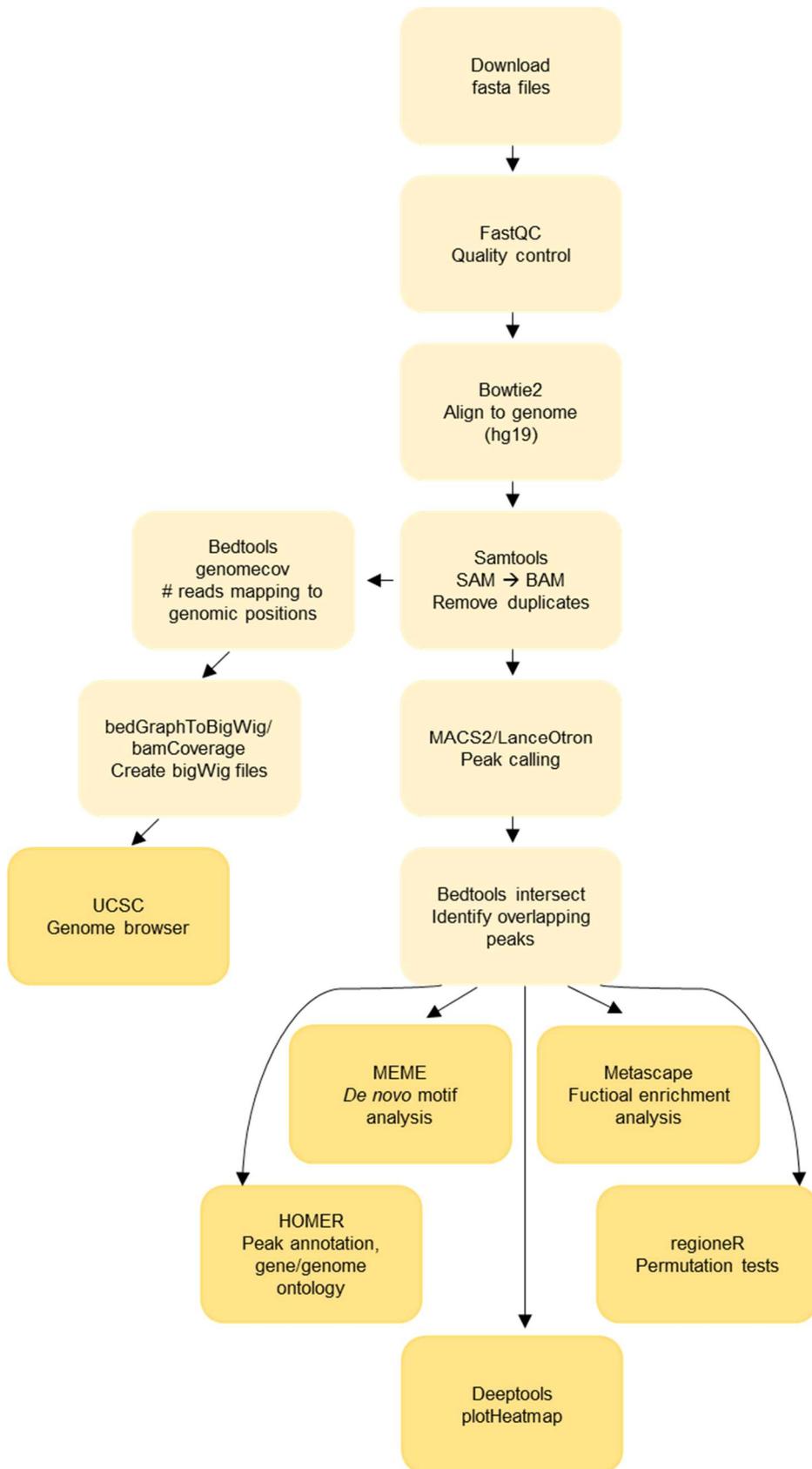


Figure 2.1: Summary of CHIP-seq and BLESS processing and analysis.

2.16 CUT&Tag-seq analysis

CUT&Tag-seq analysis was performed as previously described (Kaya-Okur *et al.*, 2019), following a CUT&Tag analysis protocol (Zheng, Ahmad and Henikoff, 2020). A summary of the processing and analysis is shown in figure 2.2.

File quality was assessed using FastQC, as described above. Reads were aligned to the genome (hg19 for DlvA files (see section 2.15), mm10 for mouse testis files) using bowtie2 (--end-to-end (for perfect/untrimmed alignment) --very-sensitive (very sensitive/accurate alignment) --no-mixed (only allows paired-end alignment for pairs) --no-discordant (only allows concordant alignment) --phred33 (Phred quality score plus 33) -l 10 -X 70 (minimum and maximum fragment length for valid paired end alignment 10 and 70, respectively)). Duplication rate was assessed by sorting the sam files using Picard SortSam, and marking the duplicates using Picard MarkDuplicates (version 2.25.5). Samtools (version 1.10) was then used to convert SAM files to binary BAM files (view -bS (input SAM/output BAM) -F 0x04 (exclude unmapped reads)). Coverage tracks were created by calculating number of reads per bin using deeptools bamCoverage (bin size = 20, normalisation method = CPM), and bigWig files were viewed on UCSC genome browser.

Samtools idxstats was used to get number of reads per chromosome for the testis samples. To adjust for ploidy, the number of reads per autosome were halved. DlvA cell analysis was focused around the top 80 AsiSI sites (Clouaire *et al.*, 2018). Heatmaps were created as above, but the top 80 AsiSI region file was organised into descending BLESS +OHT bigwig track-signal for all heatmaps (bedtools plotHeatmap --outFileSortedRegions). Clustering of different datasets was investigating using deeptools multiBigwigSummary to compute average scores for each bigWig file in every genomic region, followed by principle component analysis (PCA) using deeptools plotPCA.

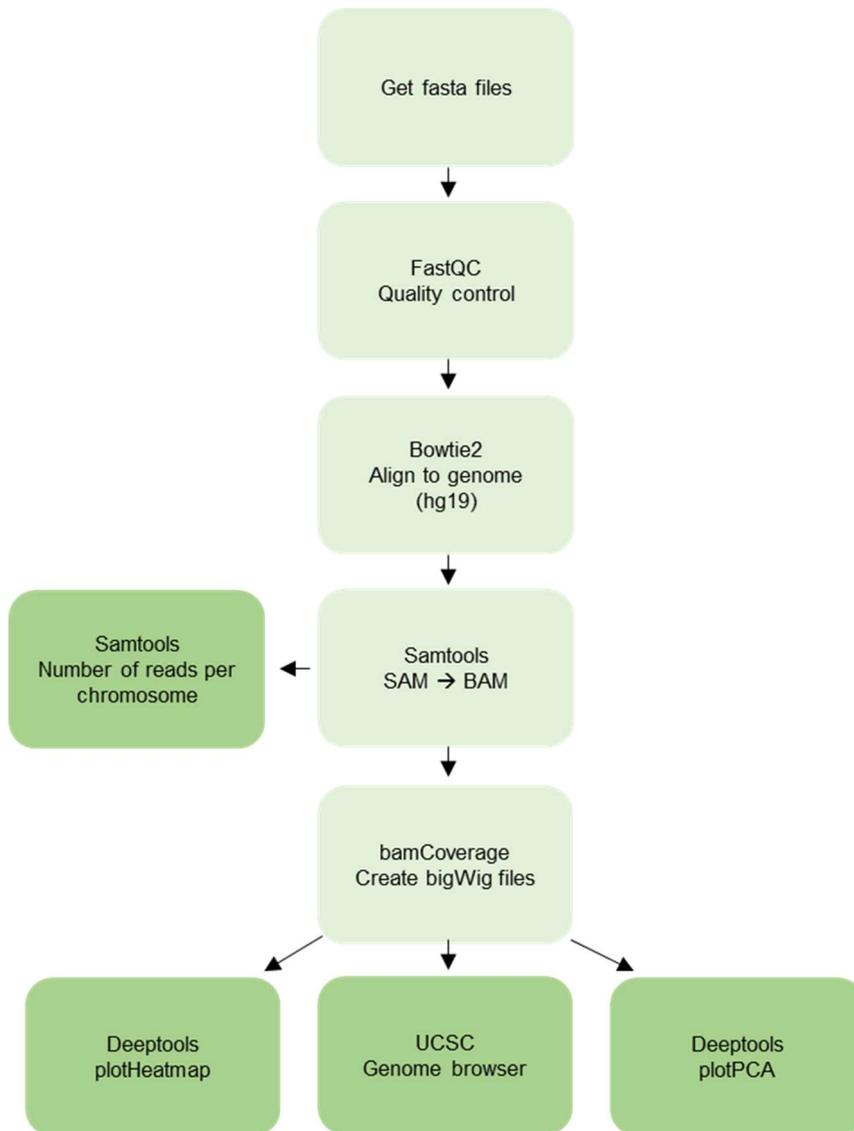


Figure 2.2: Summary of CUT&Tag processing and analysis.

Table 2.1: Primary antibodies

	Company	Clonality	Species	Target	Concentration	Dilution/ usage
AP2	Abcam ab94602	Polyclonal	Rabbit	Anti-Histone H2AX (phospho Y142)	0.6 mg/mL	WB/dot blot 1:5000 Per ChIP 3µL
AP3	Millipore 05-636	Monoclonal	Mouse	Anti-phospho- Histone H2AX (Ser139)	1 mg/mL	IF 1:400 WB/dot blot 1:5000 CUT&Tag 1:100
AP4	Millipore ABE1457	Polyclonal	Rabbit	Anti-phospho- H2AX (Tyr142)	0.15 mg/mL	WB/dot blot 1:1000
AP12	PTM Biolabs 128938FA16	Monoclonal	Mouse	Anti-H2AX (pS139)	1 mg/mL	IF 1:250, 1:500 WB/dot blot 1:5000
AP13	PTM Biolabs Z388F629P2	Polyclonal	Rabbit	Anti-H2AX (pS139/Y142) 1 st purification	1.4 mg/mL	WB/dot blot 1:5000
AP14	PTM Biolabs Z388F921P2	Polyclonal	Rabbit	Anti-H2AX (pS139/Y142) 2 nd purification	1.44 mg/mL	IF 1:50, 1:100 WB/dot blot 1:1000 1:5000 Per ChIP 2µL

AP20	Abcam ab140498	Polyclonal	Goat	Anti-H2AX	1 mg/mL	WB 1:5000
AP21	Abcam ab20669	Polyclonal	Rabbit	Anti-H2AX	0.8 mg/mL	WB 1:5000
AP26	Abcam ab18255	Polyclonal	Rabbit	Anti-H2A	1 mg/mL	WB 1:1000
-	Abcam ab81299	Monoclonal	Rabbit	Anti-H2AX (phosphor S139)	0.733 mg/mL	IF 1:250 Per ChIP 1µL
-	Millipore 07-627	Polyclonal	Rabbit	Anti-H2AX	-	Per ChIP 1µL
-	Abcam ab6002	Monoclonal	Mouse	Anti-H3K27me3	1 mg/mL	CUT&Tag 1:100

Table 2.2: Secondary antibodies

	Company	Clonality	Species	Target	Concentration	Dilution/ usage
AS1	Abcam ab6785	Polyclonal	Goat	Anti-Mouse IgG H&L (FITC)	2 mg/mL	IF 1:500
AS2	Abcam ab6719	Polyclonal	Goat	Anti-Rabbit IgG H&L (Texas Red)	2 mg/mL	IF 1:500
AS3	Bio-Rad 170-6516	Polyclonal	Goat	Goat Anti-Mouse IgG (H + L)-HRP Conjugate	NA	WB/dot blot 1:10,000
AS4	Bio-Rad 170-6515	Polyclonal	Goat	Goat Anti-Rabbit IgG (H + L)-HRP Conjugate	NA	WB/dot blot 1:10,000
AS5	Abcam ab97110	Polyclonal	Donkey	Donkey Anti-Goat IgG (H + L)-HRP Conjugate	1 mg/mL	WB/dot blot 1:10,000
AS7	Abcam ab6787	Polyclonal	Goat	Anti-Mouse IgG H&L (Texas Red)	2 mg/mL	IF 1:500
AS8	Abcam ab6717	Polyclonal	Goat	Anti-Rabbit IgG H&L (FITC)	2 mg/mL	IF 1:500
-	Abcam ab46540	Polyclonal	Rabbit	Anti-Mouse IgG H&L	1 mg/mL	CUT&Tag 1:100

Table 2.3: H2AX C-terminal tail peptides

Company	#	Phosphorylation	C-terminal tail
PTM Biolabs	PCTM-350	H2AX (no phosphorylation)	SQEY
PTM Biolabs	CPTM-037	Phospho H2AX (S139/Y142)	pSQEpY
PTM Biolabs	CPTM-038	Phospho H2AX (S139)	pSQEY
PTM Biolabs	PPTM-484	Phospho H2AX (Y142)	SQEpY

Table 2.4: CUT&Tag indexed primers for PCR library amplification

Sample	Antibody (IP)	i5 primer	Index sequence	i7 primer	Index sequence
DlvA control	γH2AX	N501	TAGATCGC	N701	TAAGGCGA
DlvA treated	γH2AX	N502	CTCTCTAT	N702	CGTACTAG
DlvA positive control	H3K27me3	N503	TATCCTCT	N703	AGGCAGAA
DlvA negative control	Secondary only	N504	AGAGTAGA	N704	TCCTGAGC
Testis sample	γH2AX	N505	GTAAGGAG	N705	GGACTCCT

Table 2.5: ChIP/CUT&Tag-qPCR oligonucleotide primers

Company	Name	Sequence	Target
Eurogentec	FW884	CCCATCTCAACCTCCACACT	No DSB
	REV885	CTTGTCCAGATTCGCTGTGA	
Eurofins	FW2166	AGCACATGGGATTTTGCAGG	No DSB
	REV2167	TTCCCTCCTTTGTGTCACCA	
Eurofins	FW1007	GATTGGCTATGGGTGTGGAC	DSB proximal (80bp)
	REV1008	CATCCTTGCAAACCAGTCCT	
Eurofins	FW1497	ACATGGGTCTTCCAGGTGAC	DSB distal (3kb)
	REV1498	GGAACTTACAACCCACACTT	

3.0 Chapter 3: CUT&Tag can be used to study γ H2AX formation upon DNA damage

3.1 Introduction

The development of Chromatin ImmunoPrecipitation (ChIP) has allowed thorough investigation of protein-DNA interactions, including transcription factors and epigenetic histone marks, increasing our understanding of regulation of gene expression, and regulation of chromatin (Furey, 2012). ChIP can be coupled with quantitative PCR (qPCR) to investigate protein interactions with specific loci of interest. For genome-wide, unbiased investigation of protein binding, ChIP can be coupled with DNA microarray hybridisation (chip) or next generation sequencing (seq). The latter has become less expensive in recent years, and provides many advantages, including higher resolution and lower signal-to-noise ratio (Ho *et al.*, 2011; Park, 2009).

ChIP has allowed “dissection” of the formation of γ H2AX in response to DNA damage. γ H2AX forms at large domains spanning the site of the DSB (~0.5-2 megabases), but is depleted in the immediate vicinity of the DSB in mammals (Massip *et al.*, 2010; Iacovoni *et al.*, 2010; Berkovich, Monnat and Kastan, 2007). In yeast, the H2AX (γ H2AX) ortholog H2A (γ H2A) follows a similar pattern (Shroff *et al.*, 2004; Kim *et al.*, 2007).

The DivA cell line has been used to extensively study γ H2AX domains. DivA cells express an estrogen-receptor-AsiSI restriction endonuclease fusion protein, which translocates to the nucleus upon OHT treatment, where AsiSI induces DNA breaks at around 200 known sites (Iacovoni *et al.*, 2010; Massip *et al.*, 2010). This allows investigation of γ H2AX formation at pre-determined DSB sites. γ H2AX domains are asymmetrical around the site of the DSB, and discontinuous within the domains themselves (Iacovoni *et al.*, 2010). The γ H2AX domain borders are defined by

topologically associated domain (TAD) boundaries (Arnould *et al.*, 2021). Areas within the γ H2AX domains depleted of γ H2AX are typically enriched for active gene promoters, allowing continued transcription uninterrupted by DNA damage repair (Iacovoni *et al.*, 2010).

γ H2AX formation is commonly used as a marker of DSBs, however, this is an indirect measurement of DNA damage formation. Thus, when studying γ H2AX formation, another method for detecting DSBs is needed. Breaks labelling, enrichment on streptavidin, and next-generation sequencing (BLESS) was developed to measure DSBs directly (Crosetto *et al.*, 2013). Here, DSBs are captured *in situ* by biotin-labelled linker DNA and extracted on streptavidin beads. The fragments are amplified by PCR with primers specific for the linkers, and fragments are then sequenced, revealing the exact locations of the DSBs.

Although ChIP is widely used to investigate protein-DNA interactions, it does have certain limitations. ChIP demands high cell numbers (~10 million cells), and the formaldehyde cross-linking of cells can mask target epitopes, causing the low signal-to-noise-ratios frequently observed. This has led to recent development of other methodologies for investigating protein-DNA interactions, including Cleavage Under Targets and Tagmentation (CUT&Tag) (Kaya-Okur *et al.*, 2019). Similar to ChIP, the protein of interest is targeted by an antibody, but this then recruits the hyperactive Tn5 transposome fused to protein A. Addition of Mg^{2+} activates Tn5, leading to cutting of nearby DNA and insertion of specific adapters (tagmentation). DNA fragments with adapters on both ends are amplified by PCR, and downstream analyses, such as qPCR or sequencing can be performed. CUT&Tag requires fewer cells than ChIP (60,000-100,000, or even single cells), produces low levels of background noise, and requires ~10X fewer sequencing reads than ChIP-seq, reducing costs (Kaya-Okur *et al.*, 2019; Wu *et al.*, 2021). CUT&Tag is fixation-free, thus allowing investigation of interactions under native conditions, and integrates adapters directly during

tagmentation, meaning separate library preparations are not required. However, there is a risk of proteins loosely bound to DNA going undetected by CUT&Tag, and it is most suited for histones and histone modifications, or proteins that are known to have a long half-life on DNA.

Aims

To test whether CUT&Tag can be used to identify γ H2AX phosphorylation marks formed after DNA damage, CUT&Tag-qPCR and -seq was performed in DivA cells, using an antibody against γ H2AX. The AsiSI-ER fusion protein expressed by DivA cells creates ~200 DSBs at known sites (Massip *et al.*, 2010). ChIP-qPCR and -seq for γ H2AX in DivA cells is widely published, therefore serving as a good positive control to benchmark the CUT&Tag data against. In addition, to investigate whether CUT&Tag-seq can also identify histone marks spanning entire chromosomes, testis samples were included, and probed with antibodies targeting H3K27me3 and γ H2AX.

3.2 Results

3.2.1 CUT&Tag-seq identifies chromosome-wide histone marks

In male germ cells, the X and Y chromosomes are transcriptionally silenced through meiotic sex chromosome inactivation, which is an important step in spermatogenesis (Turner, 2007). This leads to ATM-dependent γ H2AX formation across the length of the sex chromosomes (Abe *et al.*, 2020). The autosomal chromosomes of male germ cells are marked by large H3K27me3 domains, formed by a polycomb repressive complex protein (Maezawa *et al.*, 2018). To investigate whether CUT&Tag can be used to detect histone marks spanning entire chromosomes, CUT&Tag was performed on mouse testis samples probing for γ H2AX and H3K27me3. The

sequencing files were aligned to the mouse genome (mm10), with high alignment rates, and low duplication levels were detected in the sequencing reads (table 3.1).

CUT&Tag Tn5 tagmentation typically occurs near the entry point to the nucleosome, creating DNA fragments that are the length of a nucleosome (expected size of 180 bp for one complete linker region between adjacent nucleosomes), and multiples of this. Analysing the fragment size distribution revealed that most fragments were the size of one nucleosome (~180 bp) (figure 3.1 A, B). The H3K27me3-file had more fragments the length of two nucleosomes than the γ H2AX file, and the γ H2AX file had more fragments the length of half a nucleosome than H3K27me3 (figure 3.1 A, B).

To investigate whether the expected increased presence of H3K27me3 on autosomes and γ H2AX on the sex chromosomes in mouse testis could be detected by CUT&Tag, the number of aligned reads was plotted for each chromosome (figure 3.1 C, D). There was significantly more H3K27me3 on the autosomes compared to sex chromosomes (Wilcoxon rank test, $p < 0.01$), and significantly more γ H2AX on the sex chromosomes versus autosomes (Wilcoxon rank test, $p < 0.01$) (figure 3.1 C, D).

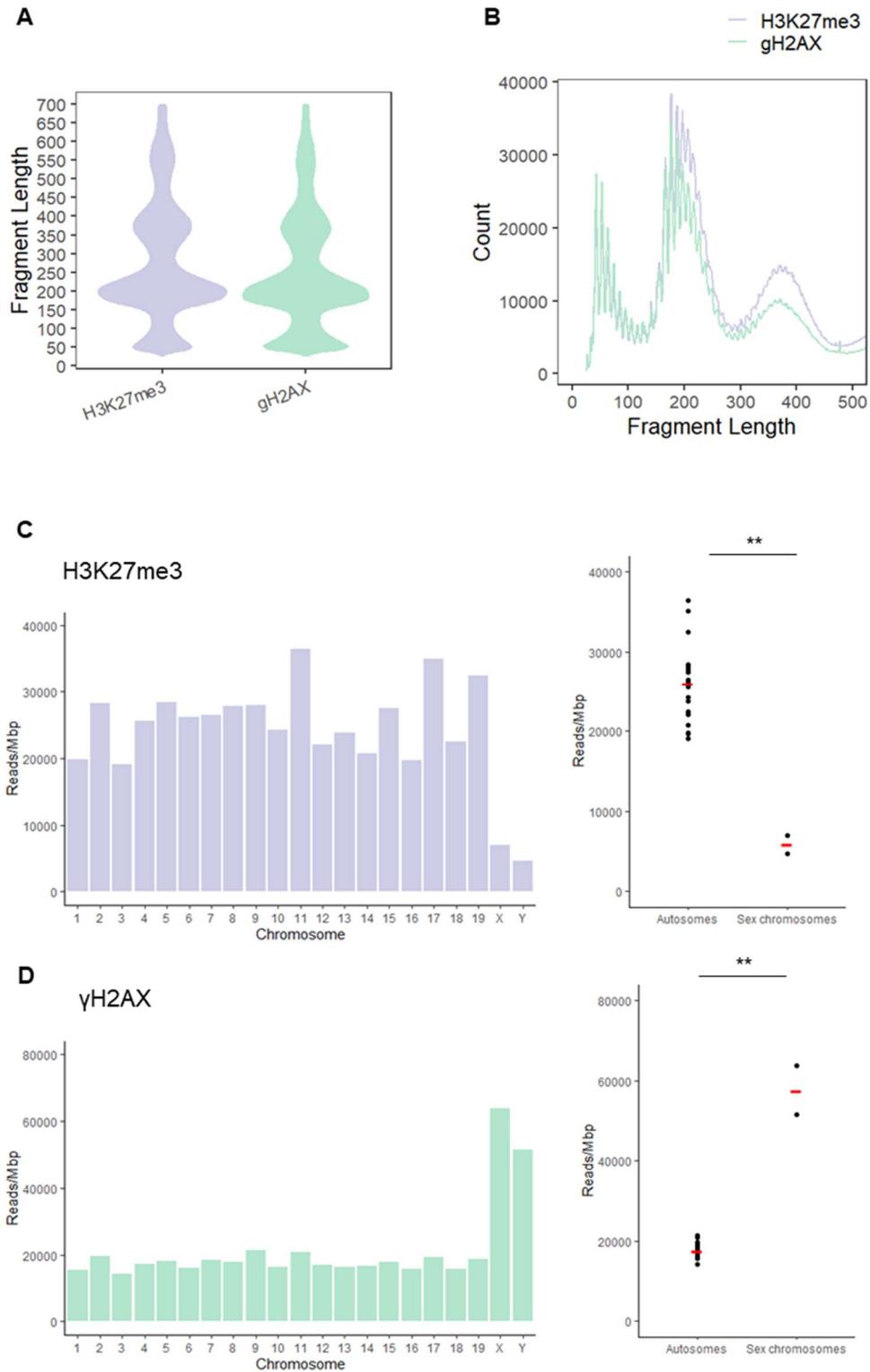


Figure 3.1. CUT&Tag can be used to detect chromosome-wide histone modifications in mouse testis.

(A) Violin plot of fragment length distribution of aligned fragments from CUT&Tag-seq.

(B) Number of aligned fragments, and their fragment length, from CUT&Tag-seq.

(C), (D) Number of reads per mega base pair across all chromosomes, adjusted for ploidy (left) and Wilcoxon rank-sum test comparing reads per mega base pair at autosomes versus sex-chromosomes (right, ** denotes $p < 0.01$).

3.2.2 OHT-induced γ H2AX formation in DivA cells

To verify that DNA damage was induced in DivA cells upon OHT treatment, DivA cells were incubated with OHT for 4 hours followed by immunofluorescence staining and fluorescence microscopy. OHT treatment led to a strong induction of DNA damage, as seen by a striking increase in number of γ H2AX foci in the OHT-treated DivA cells (figure 3.2).

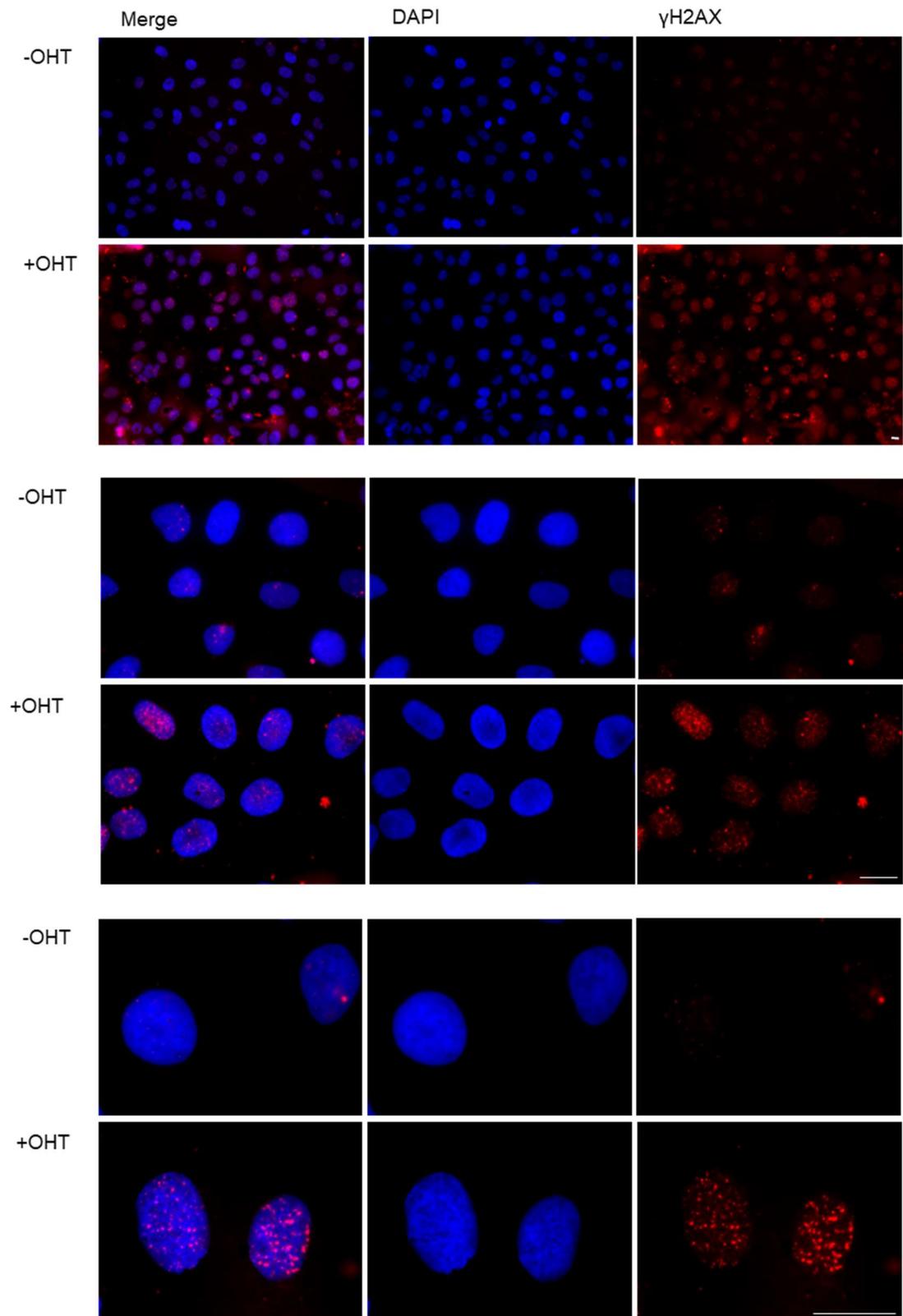


Figure 3.2. OHT induces γ H2AX formation in DlvA cells. DlvA cells were treated with OHT (or not) for 4 hours, and immunostained for γ H2AX, visualising nuclei with the DAPI DNA stain.

Top two panels: 20x objective, middle two panels: 60x objective, bottom two panels: 100x objective. 20 μ m scale bar. Images have been similarly enhanced.

3.2.3 qPCR validation of CUT&Tag efficiency

DSBs are expected to form at ~200 known AsiSI target sites in DlvA cells (Massip *et al.*, 2010). Successful immunoprecipitation of γ H2AX was confirmed by qPCR targeting a control (no DSB) site, a site near a known DSB (DSB proximal), and a site further away from the DSB (DSB distal) (figure 3.3 A, B). When assessed by ChIP-qPCR, OHT treatment led to increased γ H2AX signal at the DSB proximal site, and particularly strong enrichment of γ H2AX at the DSB distal site (figure 3.3 A). CUT&Tag-qPCR indicated high levels of γ H2AX at both the DSB proximal and distal site, with little difference between the two sites (figure 3.3 B). H3K27me3 was included as a positive control, and expected to be found at similar levels across all three sites. Enrichment levels of the repressive histone mark H3K27me3 was similar at the DSB distal and no DSB site, however it displayed a surprising enrichment at the DSB proximal site (figure 3.3 B).

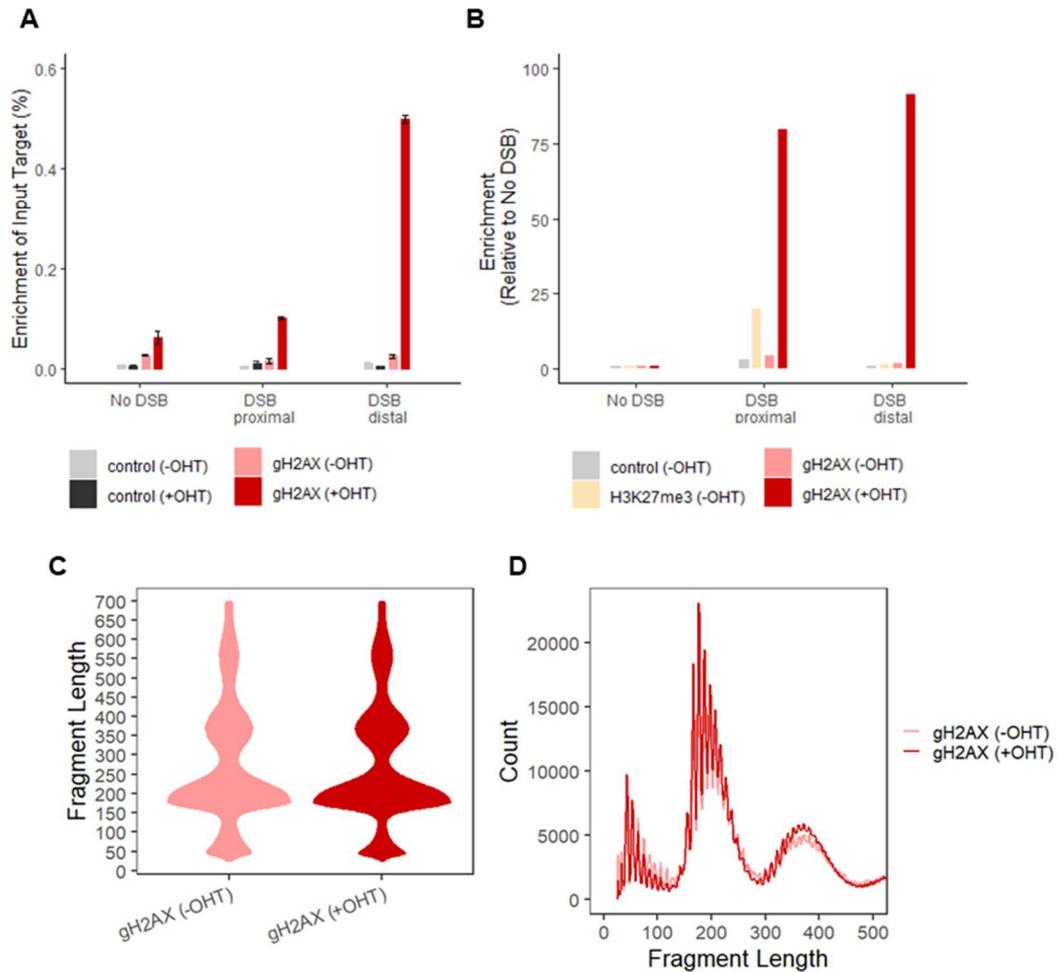


Figure 3.3. CUT&Tag detects γ H2AX formation near DSBs induced by OHT in DivA cells.

- (A) γ H2AX and control (no antibody) ChIP-qPCR at a no DSB control site (primer pair 2166/2167), a DSB proximal site (1007/1008) and a DSB distal site (1497/1498). Enrichment is calculated as a percentage of the input signal. Results from two repeats, errorbars +/- SE
- (B) γ H2AX, H3K27me3 and control (no antibody) CUT&Tag-qPCR at a no DSB control site, a DSB proximal site and a DSB distal site. Enrichment is relative to the signal at the no DSB site. Results from one repeat.
- (C) Violin plot of fragment length distribution of aligned fragments from CUT&Tag-seq.
- (D) Number of aligned fragments, and their fragment length, from CUT&Tag-seq.

3.2.4 Quality control and genome alignment of γ H2AX CUT&Tag-seq files

The DivA γ H2AX -OHT and +OHT CUT&Tag fragment libraries were sequenced. Quality control checks of the raw sequencing files was performed using the FastQC tool (Andrews, 2010). All files (R1 and R2 γ H2AX -OHT files, and R1 and R2 γ H2AX +OHT files) failed “Per Base Sequence Content”, which tests for proportion of each DNA base in the sequence. This indicates that the difference between A/T/G/C was greater than 20% at one/several positions along these sequences. However, this could be due to Tn5 preference or adapter sequences not being trimmed off the genomic sequences. In addition, the γ H2AX -OHT files failed “Sequence Duplication Levels”, suggesting that over 50% of the total sequences were non-unique. CUT&Tag produces very low background signal, and in the untreated cells, the γ H2AX antibody is expected to immunoprecipitate very low levels of γ H2AX, leading to few DNA fragments being present. The quality control for “Overrepresented sequences” also failed for the γ H2AX -OHT files. This was due to 4 sequences of which 2 made up more than 1% of total γ H2AX -OHT sequences. It also gave a warning for the γ H2AX +OHT files, due to 2 sequences making up more than 0.1% of total sequences. These sequences were shared between the 4 files, and are likely to represent adapter sequences. Tagmentation does not fragment DNA completely randomly, due to inherit preference of Tn5, which may lead to high presence of overrepresented sequences.

The paired-end sequencing files for γ H2AX -OHT and +OHT were aligned to the human genome (version hg19), with alignment rates >93%, indicating the sequencing data is of high-quality (table 3.1). The insertion of adapters by Tn5 in CUT&Tag-seq is affected by DNA accessibility in the vicinity of where the Tn5 is tethered. This can lead to formation of DNA fragments that are identical, increasing the duplication rate. Unlike duplicates formed by PCR, these duplicates may represent true fragments.

Here, the duplication rate of mapped fragments was low for both samples ($\leq 0.75\%$), and duplicates were not removed (table 3.1).

The discrepancy between sequence duplication rates predicted by FastQC versus Bowtie2 is due to the criteria with which each tool defines a duplicate. FastQC defines duplicates based on sequence content, regardless of the position of that sequence within a given read; because most sequencing-based techniques have inherent sequence biases, this often leads to unexpectedly high duplication rates. On the other hand, the bowtie2 peak caller defines peaks as read pairs who have the same start and end positions as one another. In such cases, the duplicate sequences are likely the result of PCR duplication rather than a biological phenomenon. Therefore, if there are a lot of reads clustered in a similar area with some degree of overlap in sequence, FastQC will likely report a high duplication rate, and fewer duplicates will be identified using an aligner such as bowtie2.

Table 3.1: Number of CUT&Tag sequencing reads and their alignment

	Histone mark	Sequencing depth	Genome build	Mapped fragment number	Alignment rate	Duplication rate
DivA	γ H2AX	2,312,344	hg19	2,177,190	94.16%	0.75%
	-OHT					
DivA	γ H2AX	2,428,024	hg19	2,282,123	93.99%	0.16%
	+OHT					
Testis	γ H2AX	5,299,757	mm10	5,127,918	96.76%	0.10%
Testis	H3K27me3	6,884,132	mm10	6,425,216	93.33%	0.44%

Fragment length was plotted, and the majority of DNA fragments for both γ H2AX - OHT and +OHT samples were ~ 180 bp long (as expected for CUT&Tag fragments,

but fragments representing half, 2 and 3 nucleosomes were also present (figure 3.3 C, D). Tn5 can insert adapters within nucleosomes too, likely causing formation of fragments with length of half a nucleosome. In addition, a 10bp sawtooth pattern could be seen, due to the pitch of the DNA helix winding around nucleosomes, exposing some residues more than others to Tn5 (figure 3.3 D). The 10bp periodicity is typical of a successful CUT&Tag experiment, and is seen in ATAC-seq experiments which also rely on Tn5 insertion into chromatin (Buenrostro *et al.*, 2013; Zhang *et al.*, 2021)

3.2.5 CUT&Tag-seq can be used to identify γ H2AX near DSBs

Bigwig peak files were generated using the bamCoverage tool, which creates coverage tracks from the files containing reads aligned to the genome. This is done by counting the number of reads per bin, along the genome, normalised by counts per million (number of reads per bin/number of mapped reads). These coverage tracks were then viewed on the UCSC genome browser. CUT&Tag-seq γ H2AX files were compared to previously published ChIP-seq γ H2AX (+OHT) and BLESS-seq (+/-OHT) files, and the coverage tracks were compared around the top 80 AsiSI sites in DivA cells previously identified by BLESS (Clouaire *et al.*, 2018). Three of these AsiSI/DivA sites are shown in figure 3.4. BLESS gave a stacked line at the site of each DSB (figure 3.4). The γ H2AX signal from ChIP-seq and CUT&Tag-seq both formed wide domains around the DSBs in DNA damaged cells, spanning around 1 million bases. The γ H2AX peaks from CUT&Tag-seq followed a similar pattern to those of ChIP-seq, forming discontinuous, asymmetrical peaks around DSBs, displaying near identical peaks and valleys across the γ H2AX domains (figure 3.4) (Massip *et al.*, 2010; Iacovoni *et al.*, 2010).

The γ H2AX (+OHT) coverage track from CUT&Tag-seq displayed less background signal than the ChIP-seq coverage track (figure 3.4). This is likely due to the

differences between the two techniques, where the absence of formaldehyde fixation and of DNA shearing by sonication in CUT&Tag is likely to reduce background signal formed. Notably, this was achieved with sequencing reads of 2.3-2.4 million for the CUT&Tag-seq files, whereas the γ H2AX CHIP-seq file had more than 22 million sequencing reads. Thus, using CUT&Tag, it was possible to achieve improved results investigating γ H2AX formation near DSBs, with reduced sequencing cost.

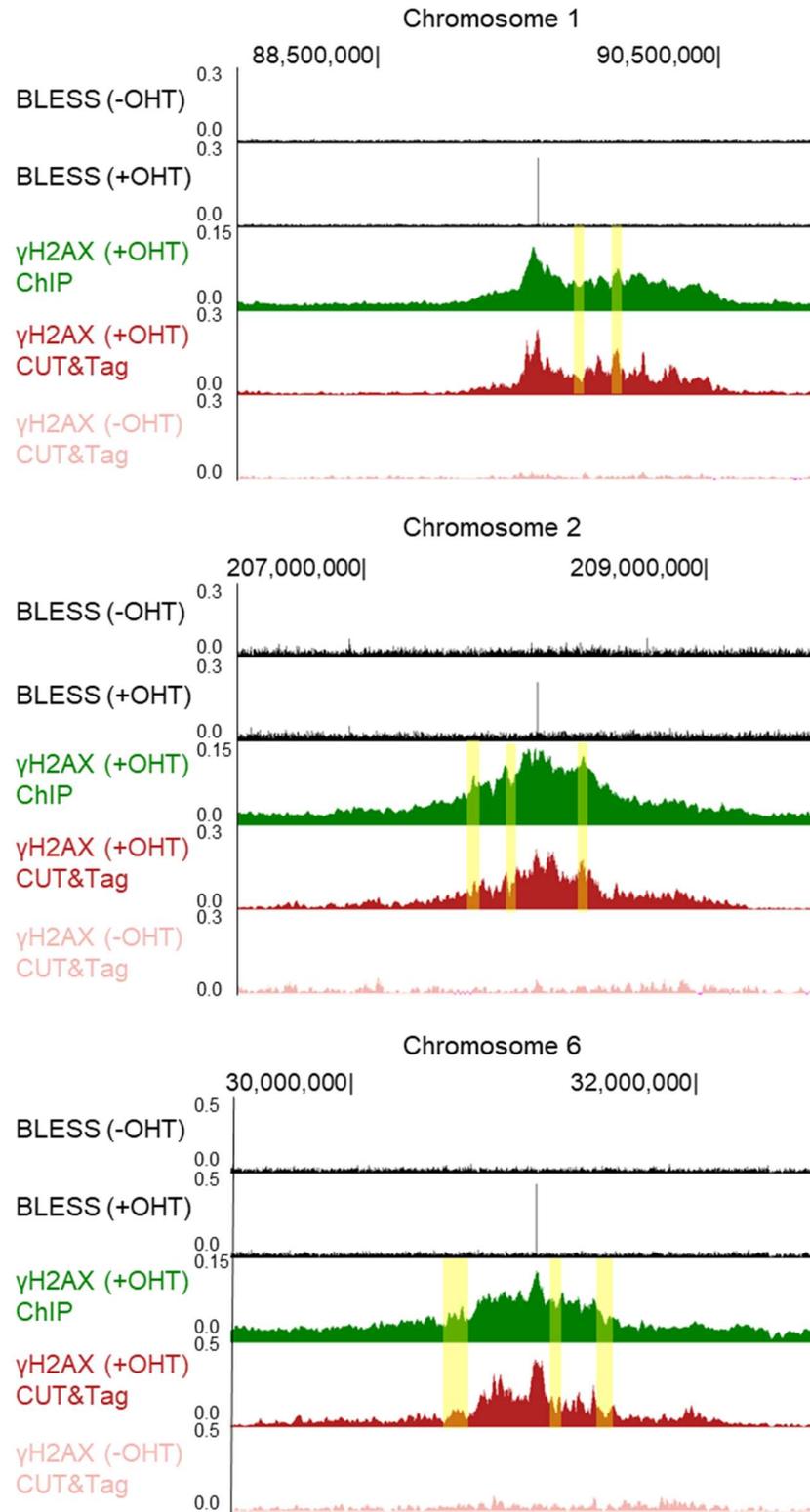


Figure 3.4. γ H2AX peaks from CUT&Tag are indistinguishable from γ H2AX ChIP-seq peaks. UCSC genome browser snapshots of γ H2AX CUT&Tag and ChIP-seq, and BLESS at three DSB sites, from DivA cells treated with OHT (or not) to induce DSBs. Examples of

overlapping peaks/dips in γ H2AX signal between ChIP and CUT&Tag (+OHT) are shown in yellow.

When plotting the γ H2AX signal from CUT&Tag and ChIP, and BLESS, across the top 80 AsiSI DSB sites in DlvA cells with a \pm 0.5 Mb window, BLESS gave a stacked peak at the centre of the DSB, whereas γ H2AX formed a peak over the centre of the DSB with both techniques (figure 3.5 A, B, C). However, these summary plot displayed a few stacked lines, which are likely a result of high signal at problematic/blacklisted regions which give high signal in next generation sequencing experiments, but were not removed for this analysis (Amemiya, Kundaje and Boyle, 2019). Interestingly, when plotting the γ H2AX signal with a narrower \pm 5kb window, the shape of the peak at the very centre of the DSB differ between the two techniques (figure 3.6 A, B). The γ H2AX signal from ChIP-seq formed a dip at the centre of the DSB, as previously described (figure 3.6 A) (Iacovoni *et al.*, 2010). This could also be seen when plotted across the \pm 0.5 Mb window (figure 3.5 B). In contrast, γ H2AX from CUT&Tag displayed a peak at the centre of the DSB (figure 3.6 B). This clear difference in γ H2AX signal shape was only apparent when viewing the files in a \pm 5kb window.

To further compare the γ H2AX signal across DSBs between ChIP and CUT&Tag, principal component analysis (PCA) was performed on the γ H2AX files from CUT&Tag and ChIP, and the BLESS files. Here, the PCA was performed on an array produced by computing the average score (fragments aligned) in every genomic region, for each file. γ H2AX +OHT from ChIP and CUT&Tag clustered closely together, and also with BLESS +OHT (figure 3.6 C). This indicates that γ H2AX is identified similarly by ChIP and CUT&Tag, across DSBs.

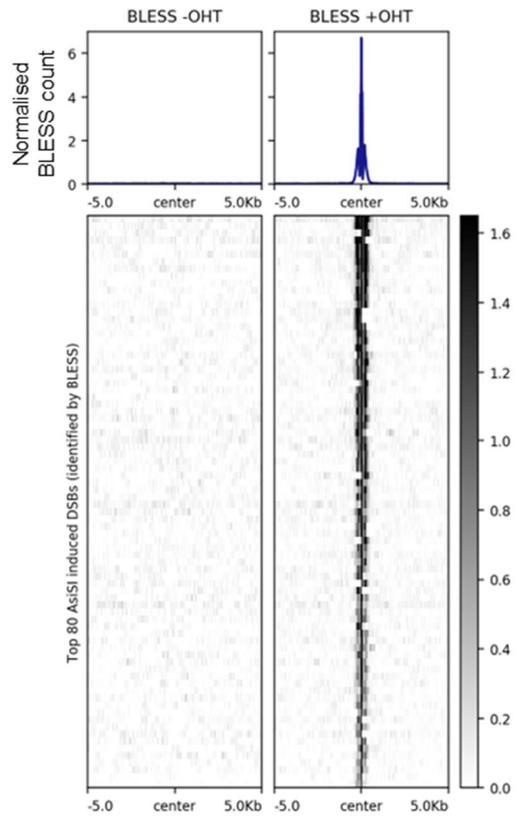
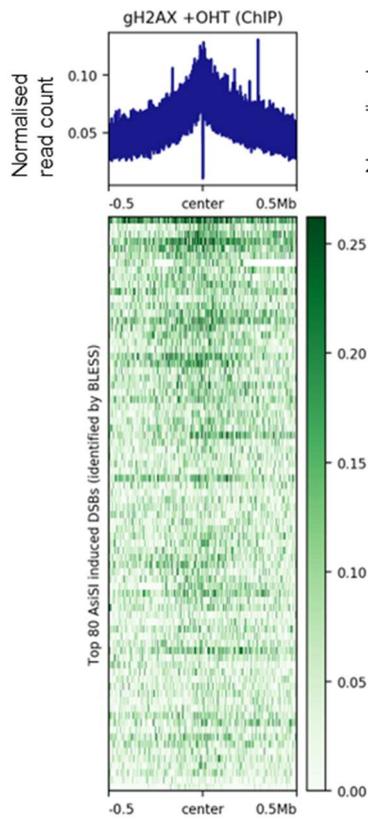
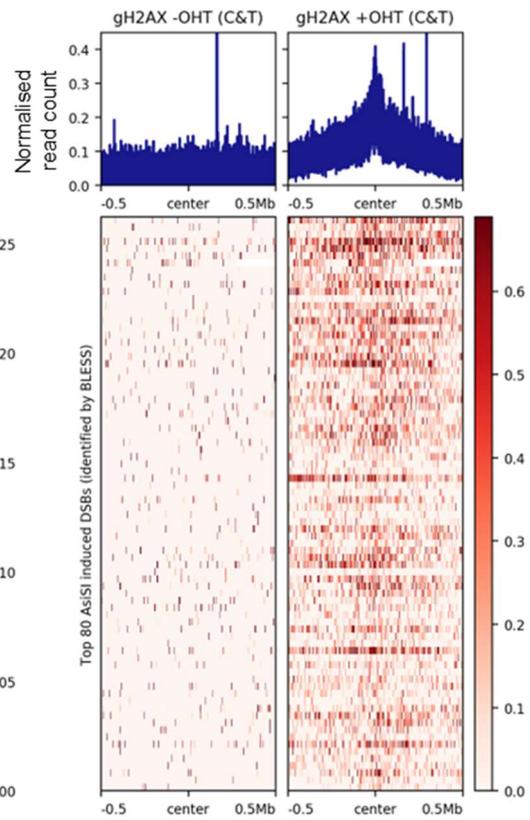
A**B****C**

Figure 3.5. γ H2AX enrichment at the top 80 AsiSI-induced DSBs in DivA cells can be detected by CUT&Tag.

(A) Heatmaps and summary plots show BLESS signal at previously identified top 80 DSB sites in induced DivA cells (Clouaire *et al.*, 2018).

(B) Heatmap and summary plot show γ H2AX ChIP-seq signal at top 80 DSB sites.

(C) Heatmaps and summary plot show γ H2AX CUT&Tag signal at the top 80 DSB sites.

DivA cells were treated with OHT (or not) to activate AsiSI. The order of the DSB regions are the same for all heatmaps, based on the BLESS signal (BLESS +OHT).

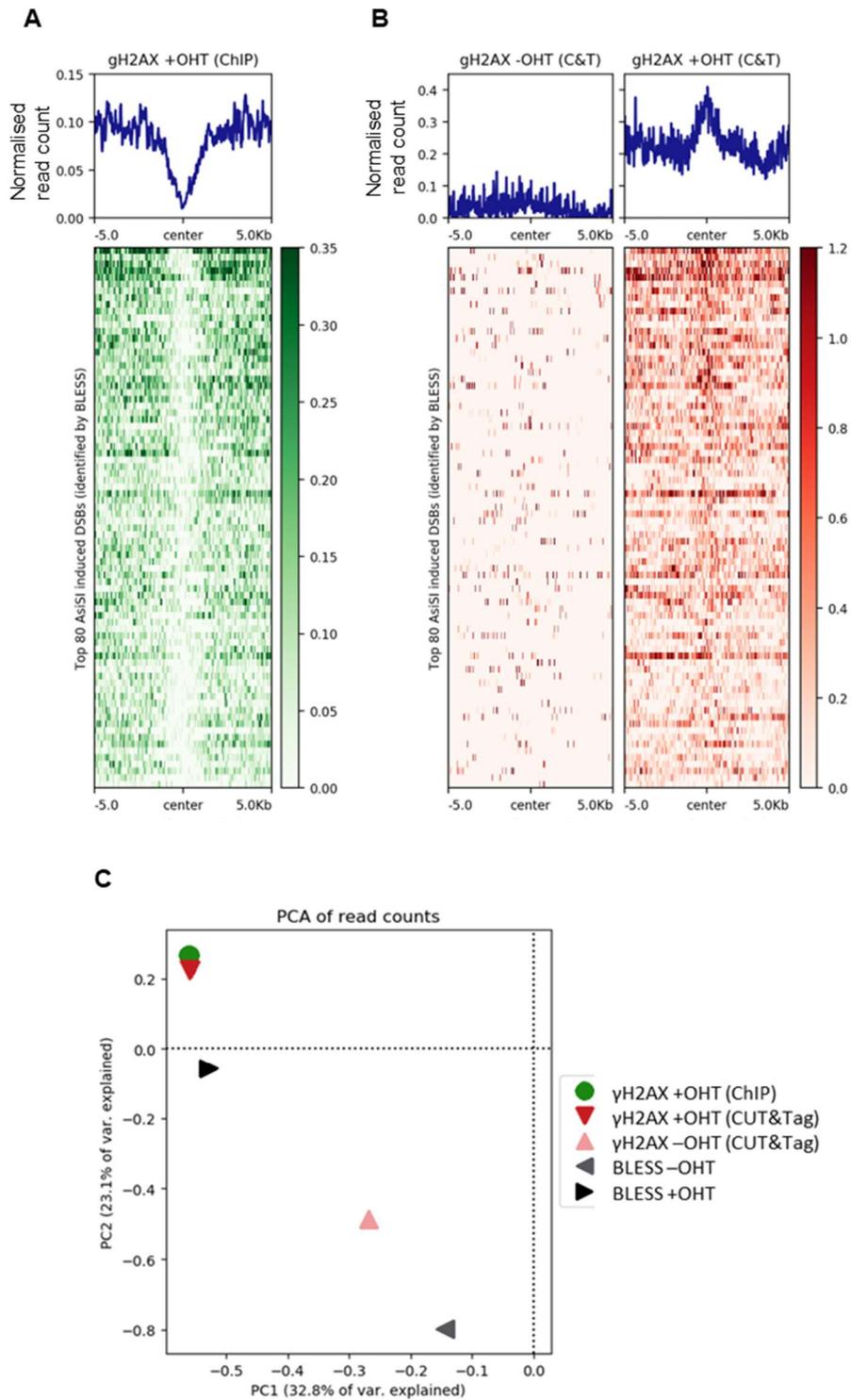


Figure 3.6. γ H2AX signal at the centre of DSBs forms a “dip” when generated by ChIP-seq and a “peak” when generated by CUT&Tag.

(A) Heatmap and summary plot show γ H2AX ChIP-seq signal at previously identified top 80 DSB sites in induced DivA cells (Clouaire *et al.*, 2018).

(B) Heatmap and summary plot show γ H2AX CUT&Tag signal at top 80 DSB sites.

(C) PCA plot showing clustering of γ H2AX ChIP and CUT&Tag, and BLESS files at the top 80 DSB sites.

DIvA cells were treated with OHT (or not) to activate AsiSI. The order of the DSB regions are the same for all heatmaps, based on the BLESS signal (BLESS +OHT, figure 3.5).

3.3 Discussion

The results presented in this chapter indicate that the recently developed CUT&Tag technique, followed by next generation sequencing, can be used to detect and investigate DNA damage-induced histone modifications ranging from the scale of kilo/megabase domains to chromosome-wide. CUT&Tag is recommended for investigation of histones and histone modifications, and was utilised successfully to assay histone marks under the conditions investigated here (Kaya-Okur *et al.*, 2019).

Mouse testes are expected to display γ H2AX solely on the sex chromosomes, and H3K27me3 solely on the autosomes. Therefore, it was possible to use these samples as a positive control to investigate whether CUT&Tag could be utilised for chromosome-wide detection of histone modifications. CUT&Tag-seq of a testis sample gave a significant enrichment of γ H2AX on the sex chromosomes, and H3K27me3 on the autosomes. The DNA fragments generated by CUT&Tag-seq aligned to the genome with very high alignment rates (>96% for γ H2AX and >93% for H3K27me3), indicating a successful CUT&Tag experiment.

γ H2AX formation around DSBs has been extensively studied by ChIP-qPCR/seq, including in the DSB-inducible DIvA cell line (Iacovoni *et al.*, 2010; Massip *et al.*, 2010; Arnould *et al.*, 2021). Performing γ H2AX CUT&Tag in the DIvA cell line therefore allowed direct comparison of the γ H2AX signal between the two techniques, at induced DSBs at known locations. Similar to γ H2AX ChIP-qPCR, CUT&Tag-qPCR

using the same γ H2AX antibody, detected increased γ H2AX signal near a DSB site in DNA damaged cells. However, whereas γ H2AX signal as assayed by ChIP-qPCR showed a high enrichment only at the DSB distal site, γ H2AX assayed by CUT&Tag-seq displayed high enrichment both at the DSB proximal and distal site, and H3K27me3 (control) was also enriched at the DSB proximal site (see Chapter 6, section 6.2). CUT&Tag-qPCR requires that the tagmented immunoprecipitated DNA fragments contain the sequences which the CUT&Tag-qPCR (site-specific) primers anneal to. Tn5 exhibits some bias for particular DNA sequences and states (e.g. is more likely to tagment regions of open chromatin or those with high GC content) (Yan *et al.*, 2020). Tn5 biases can influence whether sites of tagmentation encompass or exclude qPCR primer-complementary sequences, thereby influencing whether these sites will be amplified and detected by the downstream qPCR steps. Although γ H2AX could be detected at both DSB proximal and distal sites in a DSB-dependent manner, further repeats will determine whether CUT&Tag-qPCR can be used robustly to measure enrichment of γ H2AX in response to DSB.

The CUT&Tag fastq sequencing files presented some errors when analysed using the FastQC sequencing quality control tool; however, these were likely a result of the mechanism via which CUT&Tag generates fragments; the Tn5 bias and presence of adapter sequences. Nevertheless, the fragments were aligned to the genome with high alignment rate (>93% for both files) and low duplication rates. In addition, the majority of the fragments produced were approximately one nucleosome long, as expected. Together, these results indicated a successful CUT&Tag experiment.

Viewing the γ H2AX signal from CUT&Tag and ChIP-seq on the UCSC genome browser revealed very high similarity in the γ H2AX signal generated by the two techniques. γ H2AX domains from CUT&Tag displayed a similar asymmetric distribution around the site of the DSB, and peaks/dips within the domains, as is characteristic of γ H2AX ChIP-seq. However, the background signal was much lower

in the CUT&Tag-files, as expected, which means that CUT&Tag requires fewer sequencing reads to distinguish real signal from background noise (Kaya-Okur *et al.*, 2019). γ H2AX from CUT&Tag and ChIP-seq was found to be enriched at the top 80 DSB sites in DivA cells, although the signal in the near vicinity of the DSB differs between the two techniques (see Chapter 6, section 6.2). PCA further indicated the similarity in γ H2AX signal from CUT&Tag and ChIP-seq at the top 80 DSB sites, in DNA damaged-induced cells.

CUT&Tag followed by qPCR or sequencing was able to detect ser139-phosphorylated H2AX and H3K27me3 as peaks/chromosome-wide signals in a human cell line and in primary mouse testis cells. CUT&Tag required 5-10-fold fewer sequencing reads than conventional ChIP, and 10-fold fewer cells; moreover, it presented very low background levels. These improved results using fewer cells and reduced sequencing reads by 10-fold will revolutionise research in this area, allowing groups on fixed budgets to conduct protein-DNA interaction assays which have traditionally been high-cost. CUT&Tag's ability to detect signal using fewer cells is predominantly due to its efficient method of library preparation. Whereas traditional ChIP-seq sequencing libraries are prepared via multiple inefficient enzymatic steps (end repair, A-tailing, adaptor ligation, PCR), CUT&Tag uses a one-step library preparation method, in which Tn5 adds adaptors directly to antibody-bound chromatin in a single, very efficient reaction (Kaya-Okur *et al.*, 2019). This is followed immediately by PCR amplification of the adaptor-ligated library (Kaya-Okur *et al.*, 2019). Not only does CUT&Tag produce better γ H2AX signal:noise compared to ChIP, it also offers significant advantages in terms of cell number, sequencing and cost when bench-marked against traditional ChIP-seq. CUT&Tag therefore represents an efficient and sensitive technique for assaying epigenetic marks or histone modifications involved in the DNA damage response.

4.0 Chapter 4: Investigating the role of doubly phosphorylated H2AX (serine 139 and tyrosine 142) in the DNA damage response

4.1 Introduction

Our genomes are targeted daily by DNA-damaging agents, leading to formation of DNA breaks, most of which are resolved by DNA damage repair pathways. Unrepaired DSBs are highly genotoxic, and DSBs left unrepaired can lead to chromosome translocations, deletions, and carcinogenesis.

Upon DNA damage, histone H2AX is phosphorylated at ser193, forming γ H2AX (Rogakou *et al.*, 1999; Rogakou *et al.*, 1998). This signal spreads to large domains around the DNA break site, and can be seen as distinct foci by IF staining (Rogakou *et al.*, 1999). These foci serve as the base for recruitment of other DNA damage repair factors, such as MDC1 (Paull *et al.*, 2000; Caron *et al.*, 2015). H2AX is phosphorylated by PI3K kinase ATM, ATR and DNA-PKcs, with ATM being responsible for the majority of γ H2AX formation (Caron *et al.*, 2015). Within minutes of DNA damage induction, γ H2AX spreads over large chromatin domains, spanning up to 1-2 megabases surrounding the site of the double-strand break (Iacovoni *et al.*, 2010). Within these domains, the γ H2AX distribution is not uniform, and some regions contain little γ H2AX, predominantly active promoters/active gene regions (Iacovoni *et al.*, 2010). The borders of the γ H2AX domains coincide with borders of topologically associated domains, suggesting that formation of γ H2AX domains is regulated by underlying chromatin organisation (Arnould *et al.*, 2021). Together, this demonstrates that pre-established chromatin organisation and transcriptional activity can influence the spread and formation of γ H2AX signal.

H2AX can also be phosphorylated at another residue, at tyrosine 142 (H2AX tyr142ph). H2AX is phosphorylated at tyr142 under steady state conditions, by WT5F (Xiao *et al.*, 2009). Upon DNA damage, the H2AX tyr142 phosphorylation mark is lost,

as tyr142 is dephosphorylated by EYA phosphatases, recruited to H2AX by ZNF506 (Cook *et al.*, 2009; Nowsheen *et al.*, 2018). However, H2AX tyr142ph appears to play a role in the DNA damage response. Cells lacking WSTF are unable to initiate a strong γ H2AX signal, and recruitment of DNA repair factors to DSBs is reduced (Cook *et al.*, 2009; Xiao *et al.*, 2009). Interestingly, a loss of H2AX tyr142 dephosphorylation has a similar effect. Cells lacking EYA or ZNF506 display reduced recruitment of DNA repair factors to DSBs (Nowsheen *et al.*, 2018; Cook *et al.*, 2009).

Interestingly, H2AX can also be doubly phosphorylated at ser139 and tyr142 (di- γ H2AX). Unlike γ H2AX, di- γ H2AX is unable to interact with DNA repair factors including MDC1, and instead, interacts with the pro-apoptotic factor JNK1 via the pro-apoptotic factor Fe65 (Cook *et al.*, 2009; Campbell, Edwards and Glover, 2010). A role for di- γ H2AX in inducing apoptosis was therefore suggested, where γ H2AX promotes repair and survival, and a sustained di- γ H2AX signal promotes cell death (figure 4.1).

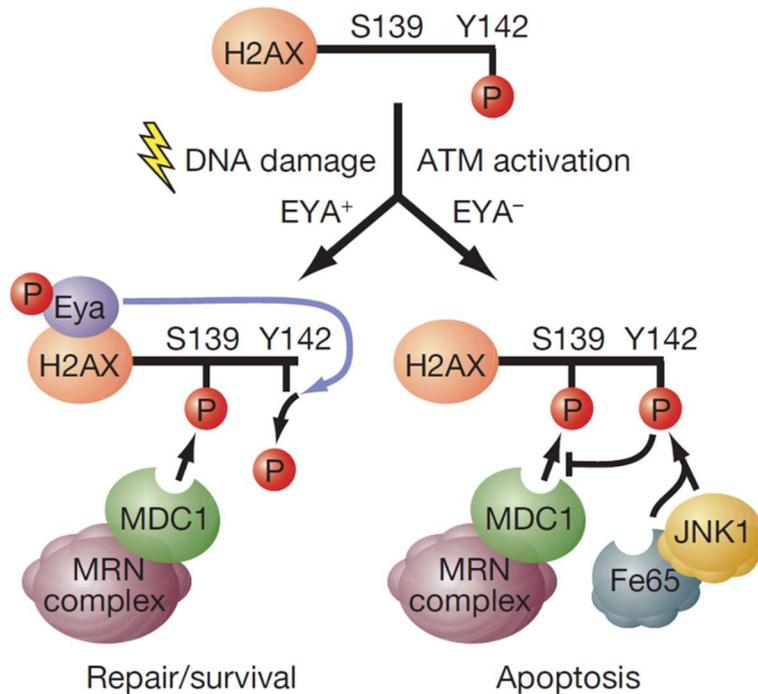


Figure 4.1. The H2AX ser139 and tyr142 phosphorylation marks (Cook *et al.*, 2009).

In undamaged cells, some H2AX is phosphorylated at tyr142. Upon DNA damage, H2AX is phosphorylated at ser139 by ATM/ATR/DNA-PKcs (not shown). EYA phosphatases are activated, and remove the H2AX tyr142 phosphorylation mark, leaving H2AX monophosphorylated at ser139ph (γ H2AX). γ H2AX recruits MDC1 and other DNA repair factors to the site of the DSB, initiating DNA repair and survival. Alternatively, if the H2AX tyr142ph mark is not lost, H2AX phosphorylated at both sites (di- γ H2AX) is unable to interact with repair factors, and instead bind the pro-apoptotic factors JNK1 and Fe65.

H2AX tyr142ph is gradually lost upon DNA damage, and di- γ H2AX may therefore (also) serve as an intermediate step between H2AX tyr142ph in undamaged cells, and γ H2AX in damaged cells. Di- γ H2AX has also been reported to form foci upon DNA damage, like γ H2AX (Singh *et al.*, 2012). If di- γ H2AX is an intermediate step between the two phosphorylation states, di- γ H2AX would be expected to be found within γ H2AX foci, in DNA damaged cells. H2AX tyr142ph is below detection level by mass spectrometry, suggesting that the histone mark is present at very low levels

(Hatimy *et al.*, 2015). If di- γ H2AX is only formed at tyr142 phosphorylated H2AX, it might be expected to also be found at low levels. Singh *et al* developed a di- γ H2AX antibody to investigate its role in the DNA damage response; however, this antibody showed cross-reactivity with γ H2AX. Therefore, there was a need for development of a new antibody specifically targeting the di- γ H2AX phosphorylation mark.

Aims

Using a di- γ H2AX antibody developed by the Ellis group, the formation of di- γ H2AX upon DNA damage was investigated by western blot and IF, with an aim to increase our understanding of the role of di- γ H2AX in the DNA damage response. Di- γ H2AX formation was also probed by CHIP-qPCR, using the DSB-inducible DlvA cell line, which generates DSBs at known locations. In addition, a commercial H2AX tyr142ph-antibody was used to investigate the H2AX tyr142ph in undamaged and DNA damaged cells.

4.2 Results

4.2.1 Testing specificity of H2AX-antibodies against differentially phosphorylated H2AX C-terminal peptides

To test the specificity of the novel di- γ H2AX antibody, a novel γ H2AX antibody and commercial H2AX antibodies used in later experiments, dot blots were performed. High antibody specificity is crucial, as these H2AX-forms differ only by a single phosphorylation mark. H2AX C-terminal peptides (table 2.3, Chapter 2) either unphosphorylated, serine-139 phosphorylated, tyrosine-142 phosphorylated or serine-139 & tyrosine-142 phosphorylated were spotted onto nitrocellulose

membranes in a four-fold dilution series (256 ng/μL to 1 ng/μL), and probed with the different H2AX antibodies (figure 4.2).

The novel and commercially available γH2AX antibodies (AP12 and AP3) displayed strong signal at the serine 139-phosphorylated H2AX peptide, with a sensitivity down to 4ng/μL (figure 4.2). However, AP3 also gave a weak signal at the SQEpY (256 and 64ng/μL) and pSQEpY (256ng/μL) peptides, indicating some cross-reactivity with these phospho-H2AX forms (figure 4.2). The AP3 antibody was raised against a peptide with amino acids corresponding to the H2AX C-terminal tail, specifically amino acids in position 134-142.

The two H2AX tyr142ph antibodies (AP2 and AP4) reacted predominantly with the SQEpY peptide, with the AP2 antibody being four-fold more sensitive than the AP4 antibody (giving signal at 4 ng/μL vs 64 ng/μL, figure 4.2). Both antibodies displayed cross-reactivity with the pSQEpY peptide, at the highest concentration (256 ng/μL).

Two novel antibodies against di-γH2AX were tested, from the first and second purification of the antibodies (AP13 and AP14, respectively). AP13 gave strongest signal with the tyrosine-phosphorylated peptides (SQEpY and pSQEpY), but reacted with all peptides at concentrations 256-64 ng/μL (figure 4.2), indicating low specificity for the doubly-phosphorylated H2AX C-terminal tails. AP14 reacted with pSQEpY at concentrations of 256-64 ng/μL, when used at dilutions of 1:1000 and 1:5000 (figure 4.2). At both dilutions, AP14 displayed a weak signal with SQEY peptide at 256-64 ng/μL, but did not react with SQEY peptides singly phosphorylated at the serine or tyrosine residues (figure 4.2). The AP14 antibody was therefore used for further experiments, due to its specificity for doubly phosphorylated H2AX C-terminal tails.

Two commercially panH2AX antibodies were also tested (AP20 and AP21). AP20 reacted strongly with the unphosphorylated SQEY peptide, with a sensitivity down to 4 ng/μL peptide, but also reacted weakly with the SQEpY peptide at 256 and 64 ng/μL,

indicating low levels of reactivity with H2AX peptides phosphorylated at the terminal tyrosine (figure 4.2). AP21 reacted with the SQEY peptide at lower concentrations than the AP20 antibody (1 ng/ μ L), but also reacted with the SQEpY peptide (down to 4 ng/ μ L, figure 4.2). The AP20 antibody was chosen for further experiments, as it reacts little with H2AX-peptides with different C-terminal phosphorylation marks. The AP26 H2A antibody did not react with any of the H2AX peptides (figure 4.2), as expected.

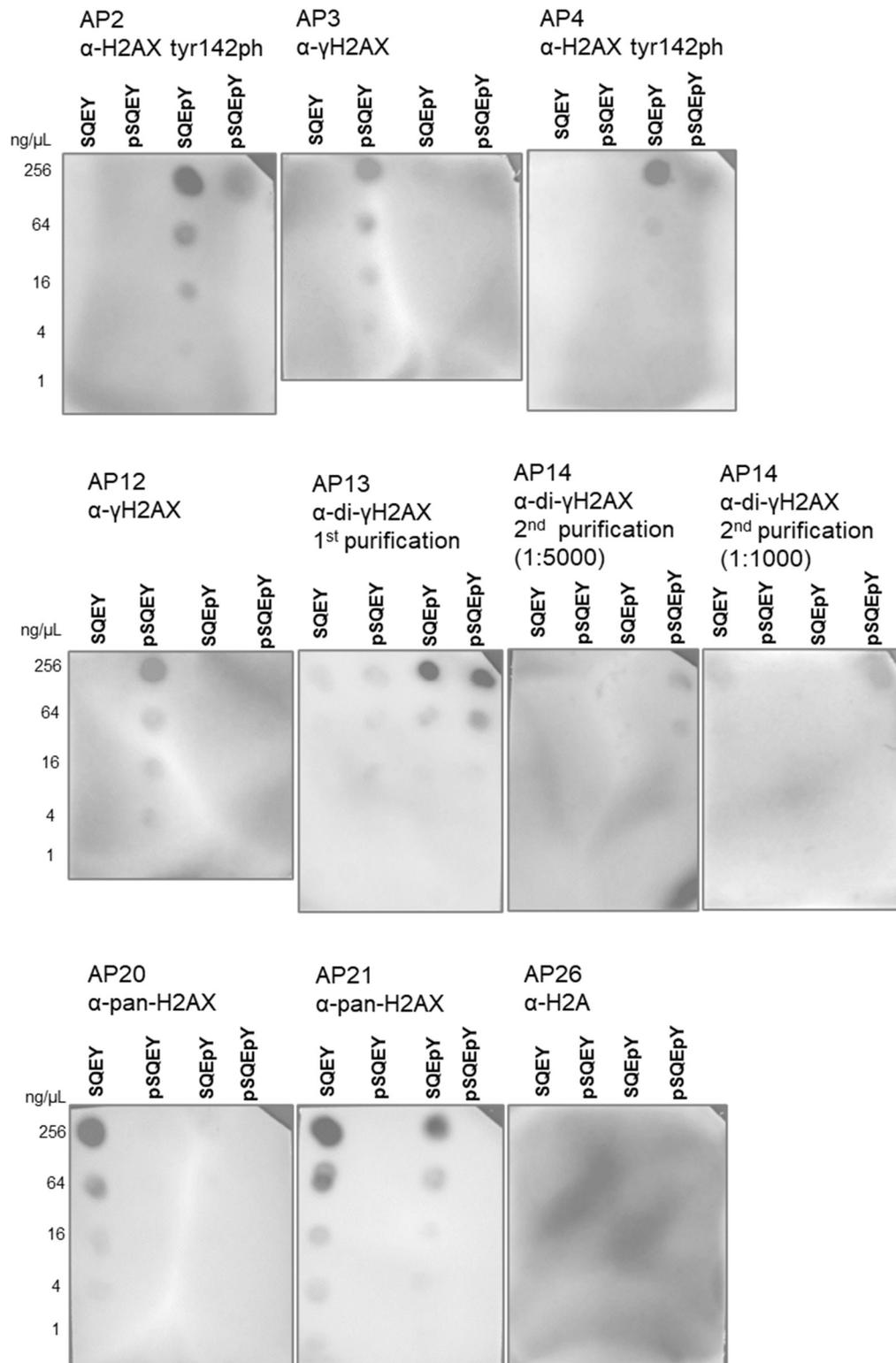


Figure 4.2. Testing specificity of H2AX-antibodies used for immunostaining and western blot. Decreasing concentrations of H2AX C-terminal peptides were pipetted onto nitrocellulose membranes, followed by probing with H2AX-antibodies.

4.2.2 Investigating di- γ H2AX and γ H2AX in DNA damaged cells by immunofluorescence staining

To further investigate di- γ H2AX formation upon DNA damage, IF experiments were performed in two immortalised human cell lines. The cells were treated with the radiomimetic drug bleomycin (5 μ M) for two hours, to induce DNA damage. Cells were then fixed and stained with γ H2AX and di- γ H2AX antibodies in combination, to investigate the location of di- γ H2AX, comparing it to the well characterised γ H2AX staining pattern (figure 4.3).

In the human embryonic kidney cell line HEK293T, γ H2AX appeared as small distinct foci upon DNA damage (figure 4.3) (Rogakou *et al.*, 1998; Iacovoni *et al.*, 2010). γ H2AX foci are already present in untreated cells, as immortalised cell lines constitutively display low levels of DNA damage. Treating cells with bleomycin for 2 hours led to an increase in γ H2AX, as expected (figure 4.3). Unlike the distinct pattern of γ H2AX staining, di- γ H2AX was seen as small nucleus-wide speckles, with no obvious foci or clusters formed (figure 4.3). The signal patterning was similar in both untreated and bleomycin-treated cells, with an increased intensity upon bleomycin treatment (figure 4.3).

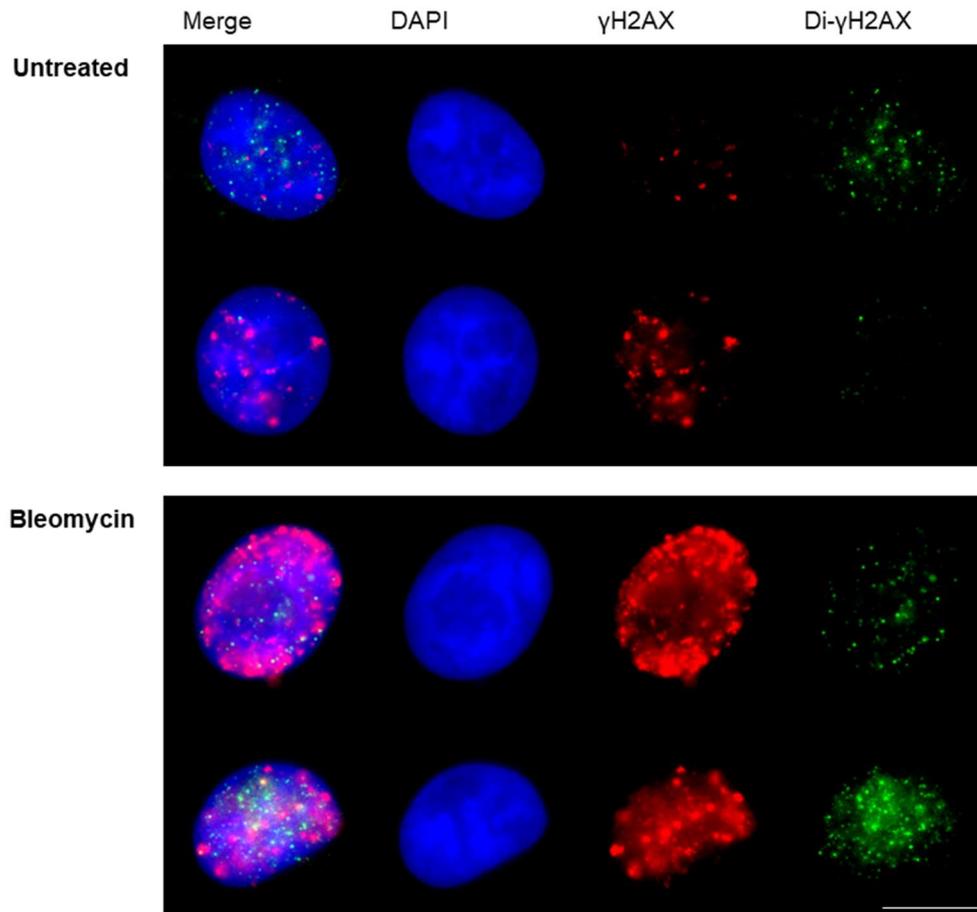


Figure 4.3. Di- γ H2AX forms nucleus-wide speckles in DNA damaged HEK293T cells.

DNA damage was induced by bleomycin treatment (5 μ M for 2 hours), followed by immunostaining for γ H2AX (AP3) and di- γ H2AX (AP14, 1:100), visualising nuclei with the DAPI DNA stain. Images are similarly enhanced. Scalebar 10 μ m.

The HEK293T cell line is an immortalised cell line with an unstable genome. To investigate H2AX phosphorylation in a more stable cell line, the non-tumourigenic human mammary epithelial cell line MCF10A was used. HEK293T and MCF10A were treated with 5 μ M bleomycin for 2 hours and then left to recover for 2 hours to investigate H2AX phosphorylations in cells undergoing DNA repair. The MCF10A cells displayed less DNA damage in untreated cells compared to HEK293T cells, as seen by fewer γ H2AX foci in untreated cells (figure 4.4). Similar to 2 hours of bleomycin treatment without recovery, HEK293T cells displayed an increase in

γ H2AX foci upon bleomycin treatment (figure 4.4). For both cell lines, the di- γ H2AX signal was seen as nucleus-wide speckles in both treated and untreated cells (figure 4.4). Interestingly, the speckled di- γ H2AX signal appeared to be independent of DNA damage-status of the cell, as seen by it forming in a HEK293T cell with very little γ H2AX (figure 4.4).

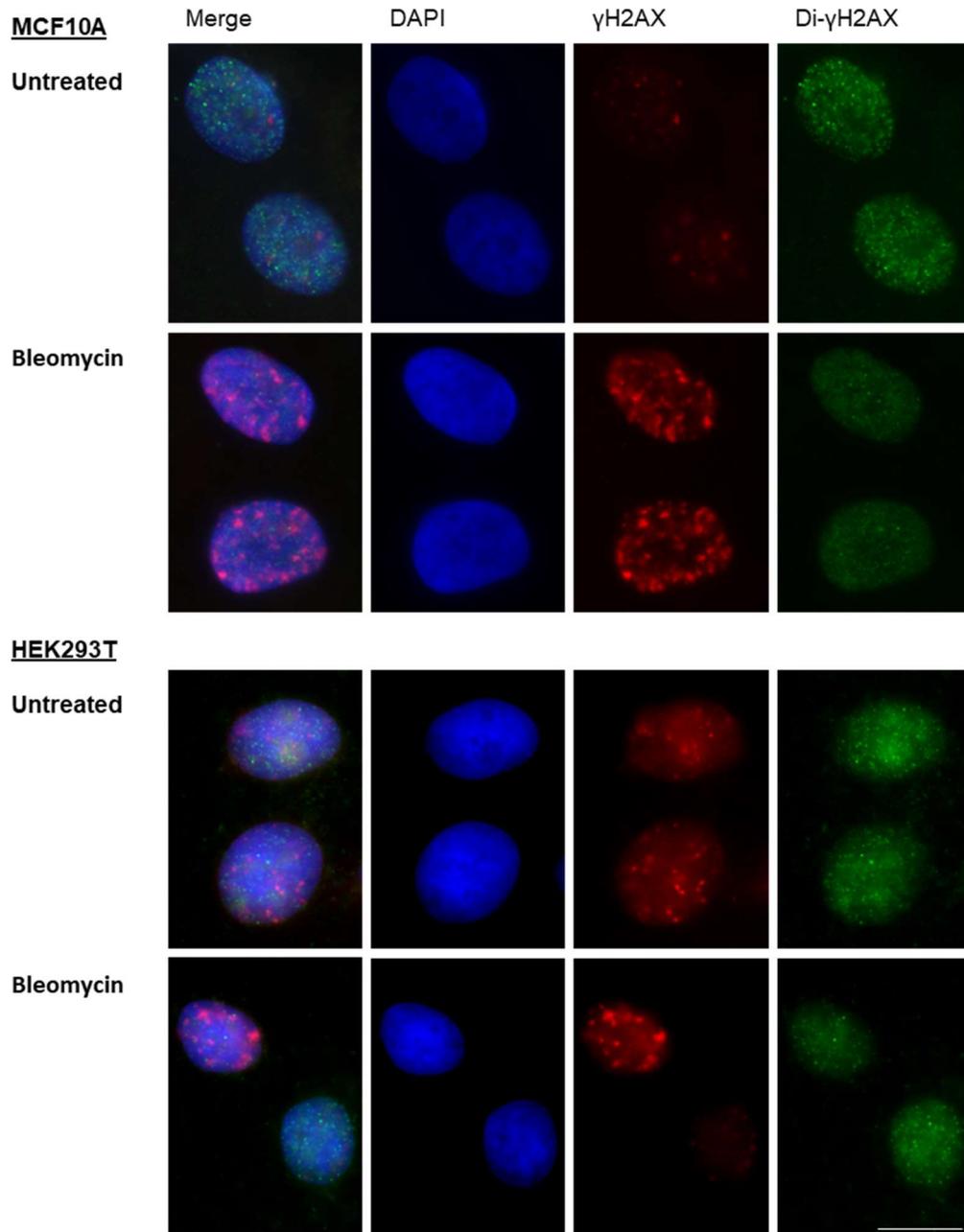


Figure 4.4. Di- γ H2AX forms nucleus-wide speckles in MCF10A and HEK293T cells recovering from DNA damage. DNA damage was induced by bleomycin treatment (5 μ M for 2 hours) of MCF10A (top) and HEK293T (bottom), followed by recovery (no bleomycin, 2 hours). Cells were immunostained for γ H2AX (AP3) and di- γ H2AX (AP141, 1:100), visualising nuclei with the DAPI DNA stain. Images are similarly enhanced (within cell line). Scalebar 10 μ m

In both cell lines investigated, DNA damage induction led to formation of distinct γ H2AX foci (repair foci), as expected. In contrast, the di- γ H2AX signal did not localise to the repair foci specifically, and its pattern appeared random. To exclude the possibility that the lack of di- γ H2AX localisation to the repair foci was an artefact of the two antibodies binding histones in a mutually exclusive fashion, a novel γ H2AX-antibody (AP12, developed by the Ellis group) was tested. HEK293T cells were treated with bleomycin for 2 hours and stained with the di- γ H2AX and AP12 γ H2AX antibodies in combination. Again, bleomycin-treated cells displayed distinct γ H2AX foci, whereas the di- γ H2AX signal was seen as diffuse nucleus-wide speckles (figure 4.5 B).

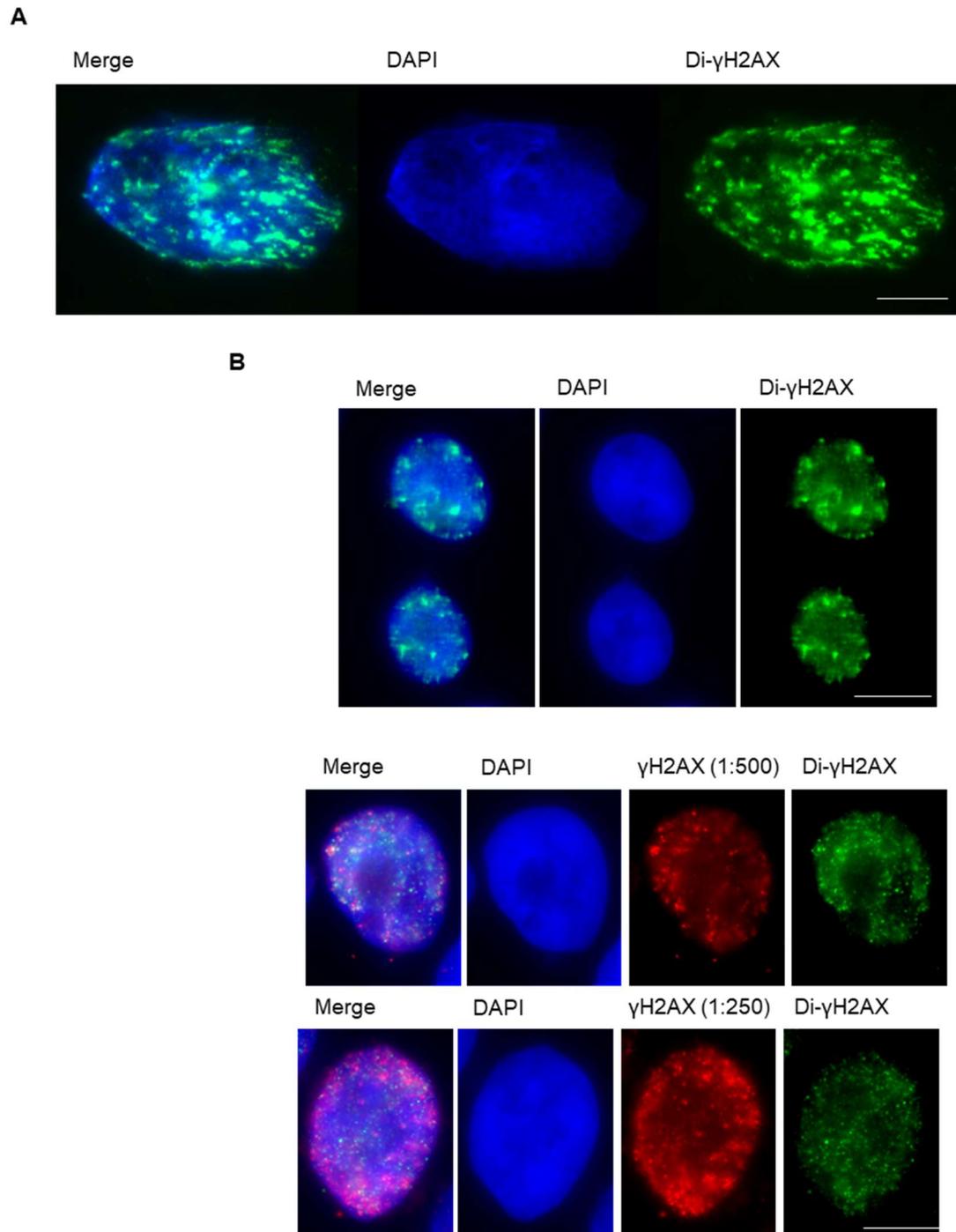


Figure 4.5. Di- γ H2AX forms foci in DNA damage only in absence of γ H2AX antibody.

(A) DNA damage was induced in HEK293T cells by bleomycin treatment (5 μ M for 2 hours), followed by immunostaining for di- γ H2AX (AP14, 1:50). Image represents 5-10% of cells displaying di- γ H2AX foci.

(B) DNA damaged cells were immunostained for di- γ H2AX (AP14, 1:50) only, or together with γ H2AX (AP12) at the normal and reduced concentrations (1:250 and 1:500).

Nuclei are visualised with the DAPI DNA stain. Images are similarly enhanced. Scalebar 10µm.

This is in conflict with the findings of Singh *et al* (2012), who reported that di-γH2AX staining formed foci similar to γH2AX foci. This is most likely due to residual cross-reactivity of their di-γH2AX antibody with γH2AX. While the results shown in figures 4.3, 4.4 and 4.5 above were obtained by simultaneous dual staining with AP14 and AP3/12 (di-γH2AX and γH2AX respectively), when bleomycin-damaged HEK293T cells were stained with AP14 antibody alone, 5-10% of cells displayed distinct di-γH2AX foci resembling γH2AX foci (figure 4.5 A, B), as reported by Singh *et al* (2012). Subsequently, a co-staining experiment was performed using AP14 and a lower concentration of AP3 antibody (figure 4.5). This excluded the di-γH2AX signal from the repair foci, leaving only the nucleus-wide speckles as seen in figures 4.3, 4.4 and 4.5. Although the dot blots (figure 4.2) did not indicate cross-reactivity between the di-γH2AX antibody and serine 139-only phosphorylated H2AX peptides, the AP14 di-γH2AX antibody appears to be able to interact with γH2AX under the conditions used for immunostaining, which is prevented when the binding sites are occupied by a γH2AX antibody. Another potential but less likely explanation is that di-γH2AX is present at the γH2AX foci, but the di-γH2AX antibody is unable to interact with these sites in presence of γH2AX antibody. This is however supported by the AP3 antibody displaying cross reactivity with doubly phosphorylated H2AX C-terminal tail peptides (pSQEpY) when tested on a dot blot (figure 4.2).

Although γH2AX foci are formed in response to bleomycin treatment in both HEK293T and MCF10A cell lines, di-γH2AX is seen as nucleus-wide speckles when cells are co-stained with antibodies against both γH2AX and di-γH2AX. Di-γH2AX only appear as foci in absence of γH2AX antibody, but these foci disappear in presence of γH2AX antibody, suggesting cross-reactivity of the di-γH2AX antibody with γH2AX.

4.2.3 Chromatin Immunoprecipitation to investigate di- γ H2AX at DNA double-strand breaks in DivA cells

The initial IF experiments revealed two potential patterns of di- γ H2AX staining. The first was focal staining that mirrored the typical pattern expected from γ H2AX in damaged cells. This was only seen in single staining experiments, and never in co-staining experiments. Thus, it seems likely that the apparent di- γ H2AX foci seen in some stains are an artefact of residual cross-reactivity with γ H2AX, since this signal is blocked by a γ H2AX antibody. The second pattern was diffuse nucleus-wide speckles that could be background, or it could be coming from DSBs at a different point in the repair process, either very early foci (where ser139ph has been added but tyr142ph has not yet been removed) or potentially late foci (if tyr142ph gets replaced before ser139ph is removed).

To distinguish between these, the DSB inducible DivA system, where DSBs can be induced at known loci was used, to look for γ H2AX and/or di- γ H2AX accumulation. DivA cells express the restriction enzyme AsiSI fused to the ER ligand binding domain (Iacovoni *et al.*, 2010). Addition of OHT to the growth medium leads to activation of the estrogen receptor ligand binding domain, and subsequent relocalisation of the AsiSI-ER complex to the nucleus (Iacovoni *et al.*, 2010). Once inside the nucleus, AsiSI induces approximately 200 DSBs, at known sites, which allows investigation of DNA repair factor recruitment by techniques such as ChIP-qPCR (Iacovoni *et al.*, 2010; Massip *et al.*, 2010). To gain insight into presence of the di- γ H2AX and H2AX tyr142ph signals near double-strand breaks, ChIP experiments were performed. This was followed by qPCR, allowing investigation of interactions between di- γ H2AX and chromatin surrounding double-strand breaks. A lack of di- γ H2AX enrichment at the known break sites would suggest that the speckled di- γ H2AX signal seen by IF is background, and not associated with the DSB response.

To verify that DSBs were induced in DivA cells upon OHT treatment, DivA cells were incubated with OHT for four hours, before fixing and staining with a γ H2AX antibody. Although the untreated DivA cells displayed some DNA damage (as seen by γ H2AX foci), OHT treatment led to a large increase in γ H2AX foci, indicating that DSBs had been induced by AsiSI in these cells (figure 4.6).

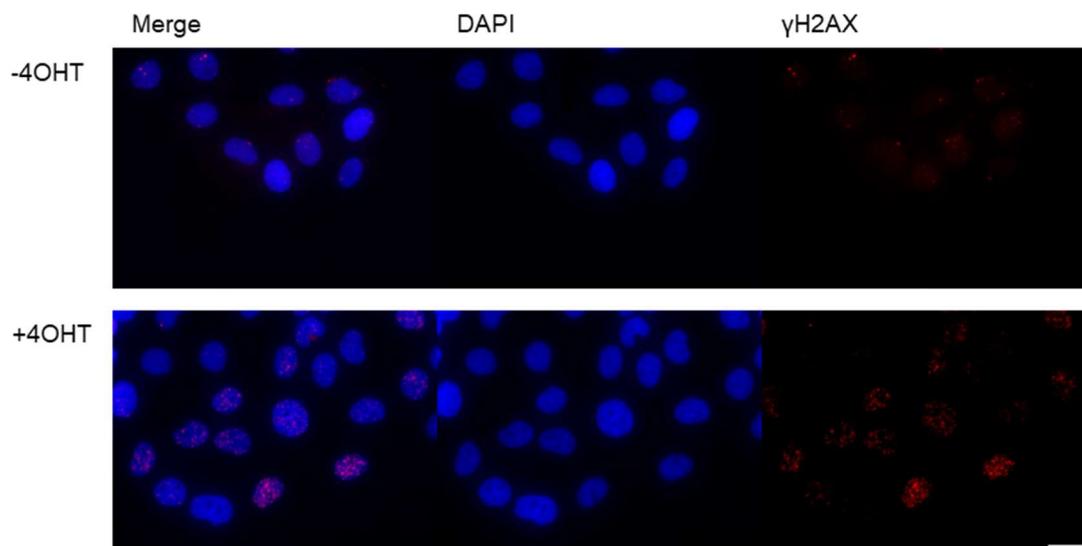


Figure 4.6. OHT treatment of DivA cells induces γ H2AX formation. DivA cells were treated with OHT (or not) for 4 hours, and immunostained for γ H2AX (Abcam ab81299), visualising nuclei with the DAPI DNA stain. Images are similarly enhanced. Scalebar 20 μ m.

H2AX tyr142ph and di- γ H2AX occupancy at three different regions were then investigated: 1) an undamaged site, 2) a DSB proximal site (80bp away), and 3) a DSB distal site (3kb from a DSB). DivA cells were treated with OHT for four hours, to induce DNA damage, followed by ChIP-qPCR (four repeats in total), including γ H2AX for comparison (in two of the repeats). H2AX has been found to be constitutively phosphorylated at the tyr142 residue, and is thought to become dephosphorylated by EYA phosphatase upon DNA damage (Xiao *et al.*, 2009; Krishnan *et al.*, 2009). H2AX tyr142ph is hypothesised to be lost from sites around DSBs, upon DNA damage

induction. Di- γ H2AX occupancy near DSBs, which is thought to be an intermediate step between constitutive H2AX tyr142ph and DNA damage-induced γ H2AX, has not previously been studied. CHIP-qPCR was therefore performed to investigate whether di- γ H2AX is present in untreated cells, and whether it can be found close to DSBs, and/or at DSB distal sites, and how its occupancy compares to that of γ H2AX.

Upon DSB induction (e.g. by AsiSI), little γ H2AX is found in the regions immediately surrounding the DSB, but a stronger induction can be seen across larger domains surrounding the break site, ranging from 0.5-1.7 mega-basepairs (Iacovoni et al. 2010). After OHT treatment, γ H2AX enrichment at the DSB distal site was over 5-fold higher, than at the non-DSB site (figure 4.7). Only a slight increase in enrichment of γ H2AX at the DSB proximal site was observed, which is in agreement with low γ H2AX level in regions immediately surrounding DSBs (Iacovoni *et al.*, 2010). This is due to resection of DNA at the DSB site as part of the DNA repair process, which involves ejection of nearby nucleosomes (Jeggo and Downs, 2014).

H2AX tyr142ph displayed a loss of enrichment at the DSB distal site upon DSB induction (figure 4.7), which is in line with the signal being removed as H2AX becomes serine phosphorylated (Xiao et al. 2009, Krishnan et al. 2009). H2AX tyr142ph levels were also slightly increased at the control site upon DSB induction, and little change was seen at the DSB proximal site (figure 4.7). However, CHIP-qPCR was not repeated for this H2AX phosphorylation mark.

Interestingly, in repeats 1 and 2, di- γ H2AX followed a similar pattern to that of γ H2AX, with an increased enrichment at the DSB distal site upon DNA damage induction (repeat 1: ~2 fold increase of di- γ H2AX and ~16 fold increase of γ H2AX, repeat 2: ~5 fold increase of di- γ H2AX and ~25 fold increase of γ H2AX, versus untreated control) (figure 4.7). However, this was not seen in repeats 3 and 4 (figure 4.7). Although di- γ H2AX displayed a ~1.5 fold enrichment at the DSB distal site upon DNA damage induction in repeat 3, this level was close to enrichment levels before/after DNA

damage induction at the other sites (figure 4.7). In repeat 3, control enrichment levels (no antibody) were higher than in other repeats, and similar to those of di- γ H2AX, which could be due to non-specific binding to beads (figure 4.7). In repeat 4, di- γ H2AX enrichment was higher in the untreated cells, at all three sites (figure 4.7). γ H2AX was not included in repeat 3 and 4, and it is unclear whether the di- γ H2AX results seen are a cause of technical errors or representing biological events, although the high background in negative control suggests the former. Although di- γ H2AX, like γ H2AX, appeared to be formed distally to a DSB site in response to DNA damage, this finding was not reproducible.

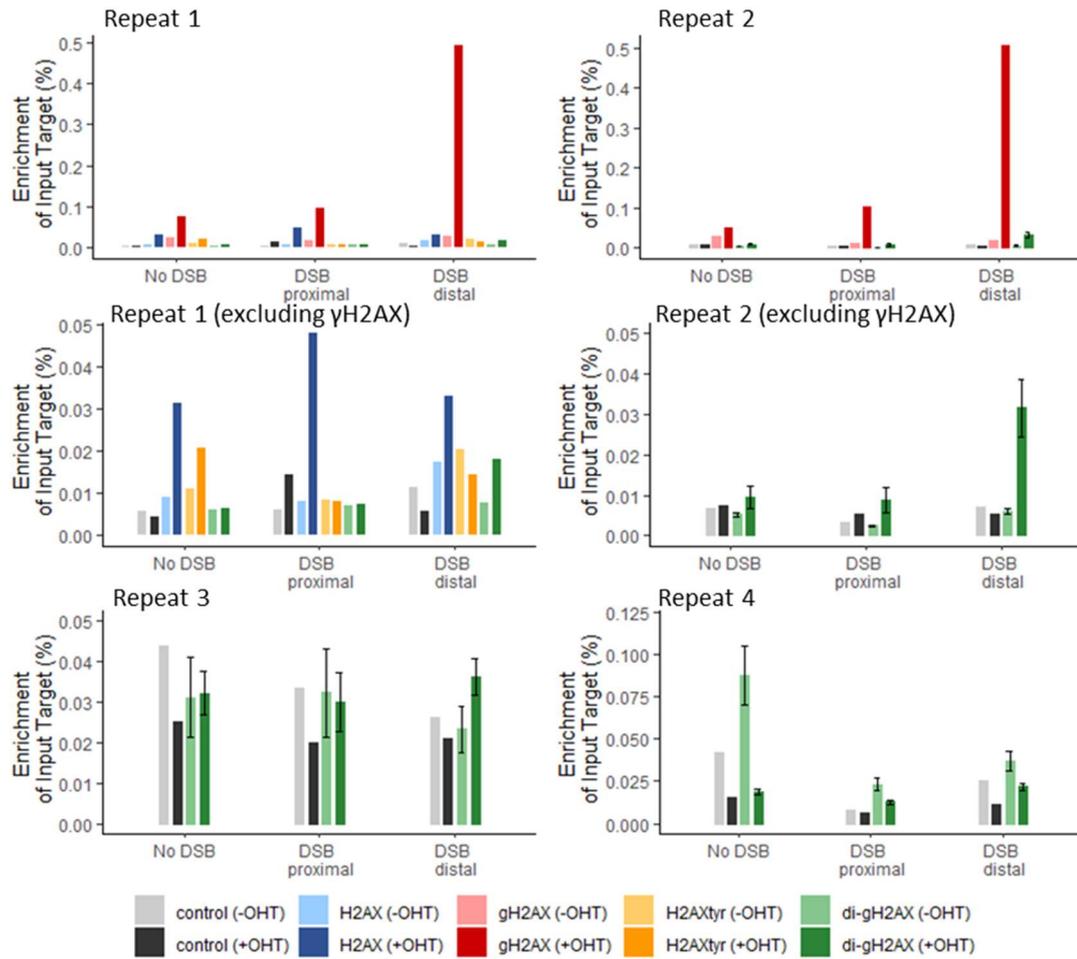


Figure 4.7. Di- γ H2AX is not reproducibly detected near DSBs by ChIP-qPCR. DivA cells were treated with OHT (or not) for 4 hours to induce DSBs, followed by ChIP-qPCR with control (no antibody), H2AX (Millipore 07-627), γ H2AX (Abcam ab81299), H2AX tyr142ph (H2AXtyr, AP2) and di- γ H2AX (AP14) antibodies, at three sites; a control site (no DSB, primer pair 884/885 for repeat 1 and 2, and 2166/2167 for repeat 3 and 4), a DSB proximal site (1007/1008) and a DSB distal site (1497/1498). Enrichment is calculated as a percentage of input signal. Errorbars +/- SE, where technical repeats are included for di- γ H2AX (repeat 2: n = 2 for -OHT & n = 3 for +OHT, repeat 3: n = 2, repeat 4: n = 6).

4.2.4 The di- γ H2AX and H2AX tyr142ph antibodies show signal via western blot in an H2AX knock-out cell line

H2AX has previously been reported to be phosphorylated at its tyrosine 142 residue under normal conditions, but this residue is dephosphorylated upon DNA damage (Xiao *et al.*, 2009; Cook *et al.*, 2009; Nowsheen *et al.*, 2018). After DNA damage induction, di- γ H2AX has been found to appear rapidly, but is gradually lost and only present at very low levels after 4 hours (Singh *et al.*, 2012). However, in these previous studies, H2AX tyrosine 142 phosphorylation was detected using antibodies against tyrosine-phosphorylated H2AX (although not specific to the 142 residue) (Cook *et al.*, 2009; Nowsheen *et al.*, 2018). Singh *et al.* developed an antibody against the doubly phosphorylated di- γ H2AX, but this antibody displayed some cross-reactivity against H2AX singly phosphorylated at the same residues (Singh *et al.*, 2012).

Using an antibody against di- γ H2AX with no cross-reactivity against singly phosphorylated H2AX (figure 4.2), di- γ H2AX levels in whole cell lysates from HEK293T, MEF and H2AX KO cell lines was investigated by western blot (figure 4.8). DNA damage was induced in these cells by treatment with bleomycin for 2 hours (figure 4.8). Antibodies against γ H2AX and H2AX tyr142ph were included for comparison, and H2A was included as a loading control, indicating approximately even loading of protein in each sample (figure 4.8).

H2AX was not present in the H2AX KO cell line (figure 4.8), as expected (Celeste *et al.*, 2002). The 2-hour bleomycin treatment led to an increase in γ H2AX, in the HEK293T and MEF cell lines (figure 4.8). However, both di- γ H2AX and H2AX tyr142ph bands were found in all cell lines, including the H2AX KO cell line (figure 4.8). These were both low level signals, and long exposure times were required for their detection. This strongly suggests that these antibodies were capable of binding non-H2AX proteins. The dot blots indicated that these antibodies were specific

against the epitopes they were raised against, with little cross-reactivity with other H2AX phospho-forms (figure 4.2), and it is unclear what other proteins these antibodies were binding in the cell lysates. Together, these results strongly suggest that this di- γ H2AX antibody cannot be used to study H2AX doubly phosphorylated at serine 139 and tyrosine 142.

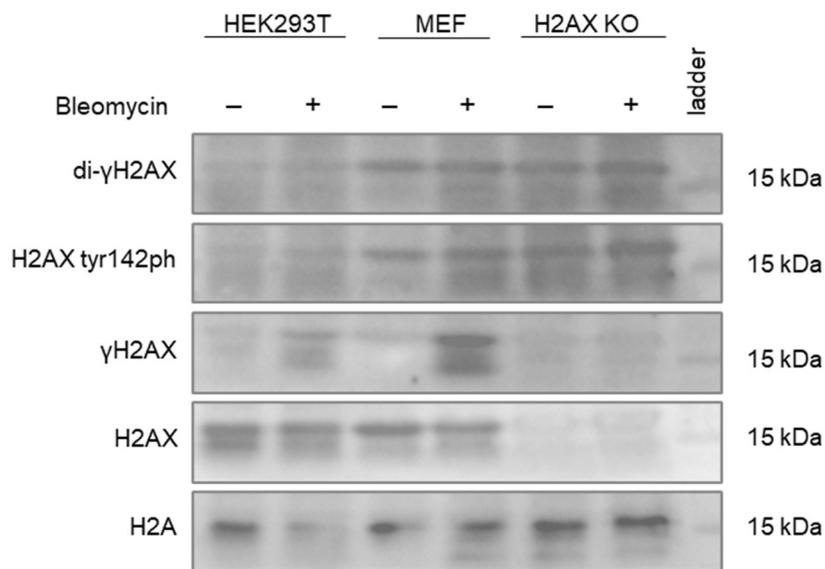


Figure 4.8. The di- γ H2AX antibody AP14 is nonspecific. DNA damage was induced by bleomycin treatment (5 μ M for 2 hours) in HEK293T cells, WT MEFs and H2AX KO MEF, followed by immunblotting for di- γ H2AX (AP14), H2AX tyr142ph (AP2), γ H2AX (AP3), H2AX (AP20) and H2A (AP26), loading 25 μ g protein per sample.

4.3 Discussion

H2AX doubly phosphorylated at ser139 and tyr142, di- γ H2AX, has previously been detected in DNA damaged cells. Di- γ H2AX is thought to be an intermediate step between two phosphorylation states; H2AX tyr142ph which is present in undamaged cells but lost upon DNA damage, and H2AX ser139ph (γ H2AX), which is formed in response to DNA damage (Nowsheen *et al.*, 2018). In addition, di- γ H2AX has been

found to interact with pro-apoptotic factors, and it is hypothesised that a persistent di- γ H2AX signal directs the cell towards apoptosis, instead of repair and survival (Cook *et al.*, 2009). However, many of these results have been generated using phosphotyrosine antibodies agnostic to the site of the phosphorylation, use of synthetic di- γ H2AX peptides, or use of a di- γ H2AX antibody with cross-reactivity to γ H2AX (Krishnan *et al.*, 2009; Cook *et al.*, 2009; Nowsheen *et al.*, 2018; Singh *et al.*, 2012). Therefore, a new di- γ H2AX antibody was developed by the Ellis group, to allow investigation of di- γ H2AX and its role in the DNA damage response *in situ*.

When tested against H2AX C-terminal peptides, the novel di- γ H2AX antibody showed specificity towards di- γ H2AX, and no other phospho-forms of H2AX, however it did display some cross-reactivity with unphosphorylated H2AX. The di- γ H2AX antibody was therefore used to investigate di- γ H2AX in DNA damaged cells. Like γ H2AX foci, di- γ H2AX foci have previously been found to form upon DNA damage (Singh *et al.*, 2012). Here, upon DNA damage induction, di- γ H2AX formed nucleus-wide speckles; this signal was unchanged after two hours of recovery post-DNA damage induction, and was observed in two different cell lines (HEK293T and MCF10A). Di- γ H2AX only formed foci when cells were not co-stained for γ H2AX, but this was only seen in a small number of cells (5-10%). Singh *et al.* reported high levels of di- γ H2AX foci formation in cells, although this antibody cross-reacted with γ H2AX, and these results must be interpreted with caution.

Previous results indicate that upon DNA damage, H2AX tyr142ph is gradually depleted, and di- γ H2AX is formed (Singh *et al.*, 2012; Nowsheen *et al.*, 2018; Krishnan *et al.*, 2009). Preliminary results in the DNA damage-inducible DivA cell line suggested di- γ H2AX recruitment to a site 3kb away from a DSB, following a similar pattern to γ H2AX. However, this result was not reproducible. H2AX tyr142ph did display depletion around a DSB, but this result is based on only one repeat.

Importantly, both the di- γ H2AX and H2AX tyr142ph antibodies gave signal by western blot in H2AX KO cells. Although dot blots indicated specificity for these antibodies, it is unclear what they bind *in vivo*, but this signal is most likely explained by non-specific aggregation of proteins via weak hydrophobic interactions. New tools for the investigation of the many unanswered questions regarding the role of di- γ H2AX in the DNA damage response are therefore still required.

5.0 Chapter 5: APOBEC3B is recruited to promoters and enhancers in presence of estrogen, and binds estrogen response pathway genes

5.1 Introduction

5.1.1 APOBEC3B cytidine deamination regulates expression of estrogen target genes

APOBEC3B (A3B) is a cytidine deaminase which, in humans, primarily functions to protect the cell from DNA, RNA and retroviruses, by deaminating cytosine to uracil in ssDNA and RNA (Swanton *et al.*, 2015; Henderson and Fenton, 2015). A3B targets cytosines which are preceded by a thymine and followed by a thymine or adenine (TCA or TCT) (Roberts *et al.*, 2013). A3B can also deaminate cytosine residues throughout the human genome, which makes it a key driver of mutation in a number of cancers. A3B is implicated in both ER positive and negative breast cancers; however, A3B activity is only correlated with poor prognosis in ER positive cancers (Periyasamy *et al.*, 2015).

Recently, a role for A3B in regulation of estrogen-target gene expression was discovered. A3B binds estrogen response elements (EREs) in an estrogen-dependent fashion; this includes recruitment at the EREs of a number of well-known estrogen-regulated genes, such as TFF1, GREB1, FOS and CTSD (Periyasamy *et al.*, 2015). In the absence of A3B (or its catalytic domain), expression of these genes is reduced, exposing A3B's role in activating the expression of ER-regulated genes. A3B's role in activating gene expression has been linked directly to its deaminase activity and the downstream activation of the DNA damage base excision repair (BER) pathway, particularly the process of uracil excision mediated through uracil-DNA glycosylase (UNG) (Periyasamy *et al.*, 2015). The majority of loci co-bound by A3B and ER are marked with γ H2AX, signalling DSB formation. This is likely caused by UNG-dependent uracil excision and subsequent abasic site and DSB formation,

and their repair appear to facilitate expression of estrogen target genes (Periyasamy *et al.*, 2015). Interestingly, A3B appears to bind ssDNA exposed within R-loops (Zhang *et al.*, 2022). Estrogen treatment leads to an up-regulation of ER-regulated genes, and this increased transcription provokes an increase in R-loop formation (Zhang *et al.*, 2022; Stork *et al.*, 2016). In fact, the majority of estrogen-dependent A3B binding sites are found in close proximity to R-loop hotspots, and these regions in turn are enriched for TSSs and enhancers (Zhang *et al.*, 2022).

Together, this suggests a positive feedback loop wherein estrogen treatment causes an increase in transcription of ER-regulated genes. This increased transcription leads to R-loop formation, and A3B recruitment to the exposed ssDNA. A3B can then deaminate cytosine residues, provoking DNA repair via the BER pathway and further induction of ER target gene expression.

5.1.2 APOBEC3B binds chromatin together with/without ER

A3B ChIP-seq in MCF7 cells stimulated with estrogen identified 24,486 A3B binding sites across the genome, the majority of which were found in introns and gene-distal regions (likely gene regulatory regions, e.g. enhancers), which is also where most ER binding occurs (Periyasamy *et al.*, 2015; Welboren *et al.*, 2009). However, only around half of these estrogen-dependent A3B binding events overlapped with ER binding sites, and whether A3B has any ER-independent roles in regulating transcription is unknown (Periyasamy *et al.*, 2015).

Aims

With an aim to further investigate A3B binding at chromatin both in presence and absence of ER, bioinformatic tools developed or updated since the discovery of A3B's

involvement in ER target gene expression was used to analyse published A3B and ER ChIP-seq data. A3B binding regions and motifs targeted, enrichment at promoters and enhancers, and genes bound by A3B in presence or absence of ER was assessed, to investigate the potential difference in role between A3B binding chromatin independently, or together with ER.

5.2 Results

5.2.1 Quality control check of sequencing data by FastQC

ChIP-seq was performed by Periyasamy *et al* for A3B and γ H2AX, and Welboren *et al* for ER (Periyasamy *et al.*, 2015; Welboren *et al.*, 2009). The quality of the sequencing files was assessed by FastQC (Anderson, 2010), and the analysis modules that gave warnings/failed are shown in table 5.1. The input, A3B and γ H2AX files, all from Periyasami *et al*, gave warnings of the “Per tile sequence quality”. This was due to errors with a few of the flowcell tiles, which could be caused by bubbles or other debris in the flowcell used for sequencing of these samples. All files gave warnings or failed “Overrepresented sequences”, which is caused by presence of sequence(s) making up more than 0.1% or 1% of total sequences, respectively. This could indicate a high abundance of PCR duplicates or low library complexity. The input, A3B and γ H2AX files were all contaminated with adapter sequences. The ER files were contaminated by a 32 bp long adenosine oligonucleotide. The input and A3B files did not pass the test for “Per base sequence content”, as the difference between A, T, C and G was greater than 10% (warning) or 20% (fail) along the sequences. This appeared to be due to high levels of A but low levels of T in the first position of the sequencing reads, likely because adapter sequences had not been removed. All files apart from A3B (estrogen) and ER (estrogen) gave warnings/failed “Per sequence GC content”. It is expected that the GC content of the libraries

represents the GC content of the genome. Deviations from this could be a result of adapter dimer contamination, which is in agreement with these files containing adapter sequences. The GC content of A3B (estrogen) and ER (estrogen) contained expected GC content, which may be due to these libraries containing the highest levels of “true” fragments bound by the proteins of interest, and the least background noise. The adapter sequences were not removed, as it is not essential for alignment of the sequencing reads to the genome (see Chapter 2, section 2.14).

Table 5.1: FastQC analysis modules that gave warnings and errors for ChIP-seq files

	Per tile sequence quality	Per base sequence content	Per sequence GC content	Overrepresented sequences
Input (vehicle)	Warning	Warning	Fail	Warning
A3B (vehicle)	Warning	Fail	Warning	Warning
A3B (estrogen)	Warning	Warning		Warning
ER (vehicle)			Warning	Fail
ER (estrogen)				Warning
γH2AX (vehicle)	Warning		Warning	Warning
γH2AX (estrogen)	Warning		Warning	Fail

5.2.2 Sequence alignment to the human genome

The sequencing files were aligned to the human genome assembly hg19, using bowtie2 (table 5.2). The alignment rates for input and A3B were a few percentage points higher than previously found (Periyasamy *et al.*, 2015). Most files were aligned at high rates, apart from the γ H2AX files, where just over half of the reads were aligned to the genome. Low alignment rates could be caused by presence of adapter sequences, which is in agreement with the results from FastQC. Or it could be caused by contamination by another species (e.g. bacterial infection of the cells). Periyasamy *et al.* reported different number of total reads for the γ H2AX files; 5 million and 13 million reads in absence/presence of estrogen, respectively. The alignment rate was over 97% for both files (Periyasamy *et al.*, 2015). These files were also aligned to hg19, however using the older version of bowtie, bowtie1.

Table 5.2: Number of sequencing reads and their alignment rate to the human genome (genome build hg19)

	Number of reads	Alignment rate
Input (vehicle)	18,032,282	89.23%
A3B (vehicle)	25,996,991	98.67%
A3B (estrogen)	32,913,006	98.55%
ER (vehicle)	3,334,448	68.72%
ER (estrogen)	12,652,745	92.43%
γH2AX (vehicle)	12,265,523	62.08%
γH2AX (estrogen)	11,528,336	55.40%

5.2.3 Calling ChIP-seq peaks using MACS2

The genome alignment creates files containing information about where the DNA fragments are enriched throughout the genome. To determine the significance of this enrichment, peak calling was performed using Model-based Analysis of ChIP-seq 2 (MACS2). The input file was included in the peak calling, providing expected background noise for these files. The peak calling creates an output with coordinates corresponding to genomic regions that have significant fragment enrichment, which represent the genomic binding sites of the proteins of interest.

The peak calling identified 1,124 A3B binding sites in absence of estrogen, and 10,592 A3B binding sites in presence of estrogen (table 5.3). 1062 of these binding sites were present in both absence/presence of estrogen, making up ~10% of the total A3B binding sites in presence of estrogen (figure 5.1 A). This is lower than the number of A3B binding sites identified by Periyasamy *et al.* Here, A3B was found to bind at 3052 sites in absence of estrogen, and 24,486 sites in presence of estrogen (Periyasamy *et al.*, 2015). 2291 A3B binding sites were independent of presence of estrogen, again making up ~10% of total A3B binding sites in presence of estrogen.

Peak calling of the ER ChIP-seq file identified 27,006 binding sites in presence of estrogen (table 5.3), which is over 2.5 fold more than the 10,205 binding events identified by Welboren *et al.*, but less than the 35,663 ER binding events reported by Periyasamy *et al.* ER was found to bind 3812 sites in absence of estrogen, with 535 binding sites bound in both presence and absence of estrogen (table 5.3, figure 5.1 A).

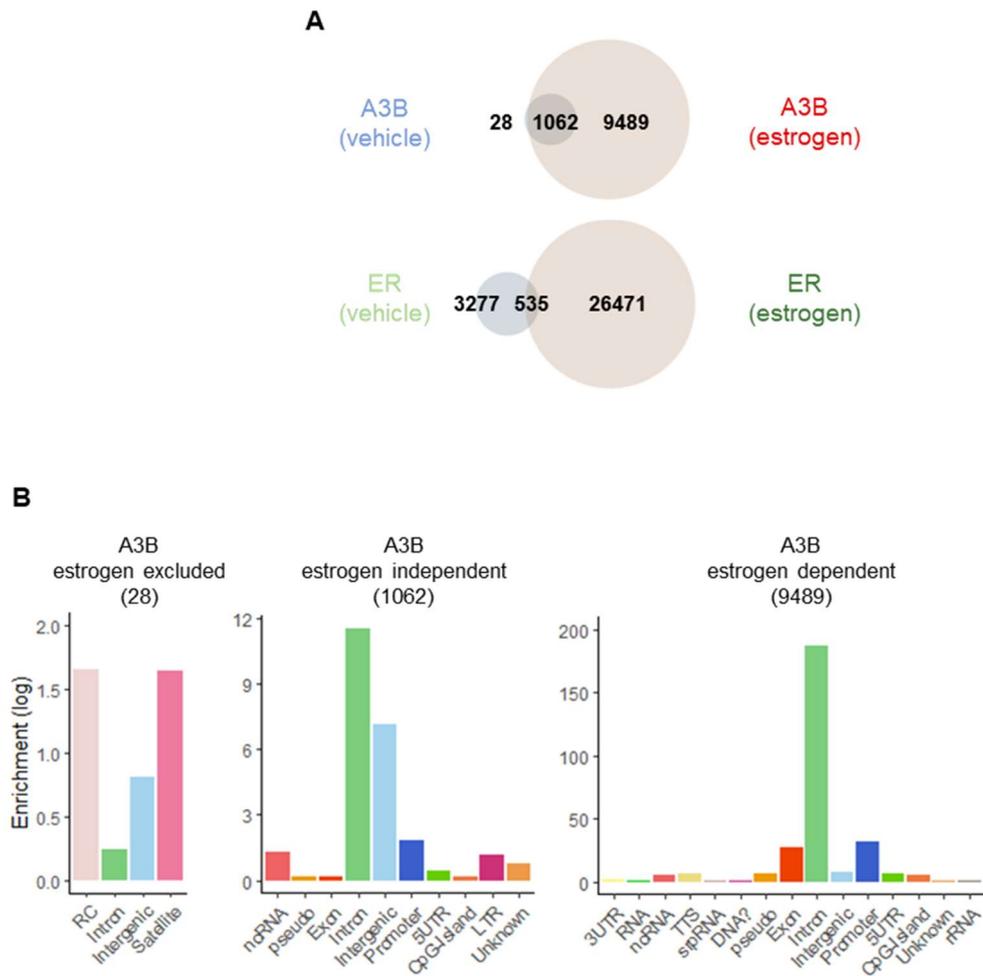


Figure 5.1. APOBEC3B binds introns, intergenic regions and promoters in presence and absence of estrogen in MCF7 cells.

(A) Overlap between A3B (top) and ER (bottom) ChIP-seq binding regions in MCF7 cells treated with estrogen or not (vehicle).

(B) HOMER peak annotation of A3B binding regions.

γ H2AX was enriched at 4499 sites in absence of estrogen, and 5730 sites in presence of estrogen, with γ H2AX found at 3161 of these binding sites under both conditions (table 5.3, figure 5.2 A). This indicates that most of the γ H2AX is formed independently of estrogen. γ H2AX was previously reported at over three-fold more locations; at 17,892 sites across the genome in presence of estrogen (Periyasamy *et*

al., 2015). Zhang *et al* reported that in T-47D cells, estrogen induced over 15,000 DSBs, which represented 30% of detectable DSBs in the genome. Of these, only a third were found to be A3B-induced. The A3B-independent estrogen-induced DSBs were not linked to presence/formation of R-loops, suggesting that these DSBs are made through another mechanism in response to estrogen stimulation (Zhang *et al.*, 2022).

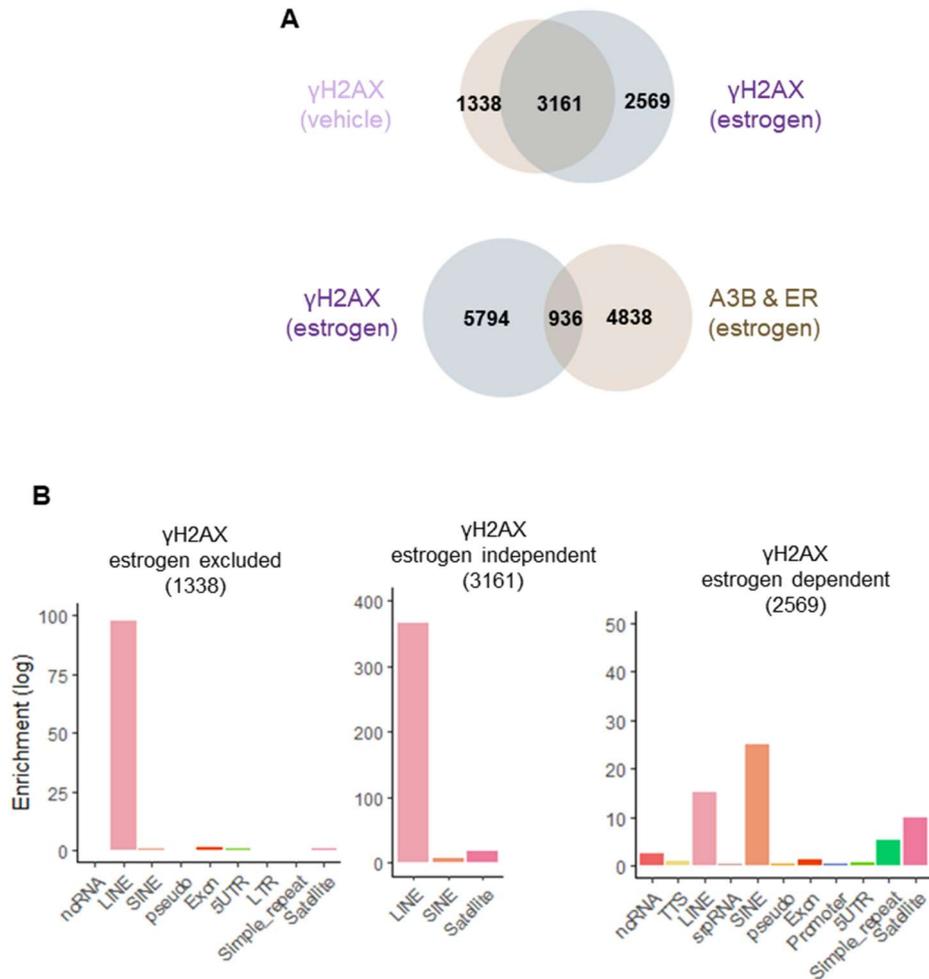


Figure 5.2. Only a small number of γ H2AX binding regions overlap with APOBEC3B and ER binding regions in presence of estrogen in MCF7 cells.

(A) Overlap between γ H2AX binding sites in absence/presence of estrogen (top). Overlap between the sites bound by γ H2AX in presence of estrogen, and the sites co-occupied by A3B & ER in presence of estrogen (bottom). Binding sites identified by ChIP-seq in MCF7 cells.

(B) HOMER peak annotation of γ H2AX binding regions.

The difference in number of binding sites identified is likely due to usage of different tools or versions of the tools. Welboren *et al* used the peak caller FindPeaks, which is no longer updated. Periyasamy *et al* used MACS, an older version of the MACS

peak caller. MACS2 is the updated version, and offers more accurate peak calling. MACS2 also handles peaks with varying signal types better (e.g. broad or punctate peaks), whereas MACS works best for punctate peaks. The higher number of peaks reported by Periyasamy *et al* may be a result of MACS splitting broader peaks into several, punctate peaks. However, the peak regions generated by Periyasmy *et al* are not provided, preventing a direct comparison between the peak regions generated here using MACS2, and by Periyasamy *et al* using MACS. Furthermore, MACS2 gives more parameters which can be changed to suit the particular experiment. The generation of peak files is influenced by the parameters used by the peak caller to calculate the significance of the DNA fragment enrichment, which will be affected by the p-value or false discovery rate (FDR) used (Landt *et al.*, 2012). Whereas MACS2 permits the user to specify the p-value or FDR to be used for the peak calling, MACS only allows the user to specify a p-value (Zhang *et al.*, 2008). During peak calling, a window is slid along the genome and the statistical significance of potential enrichment observed in that region is calculated (Zhang *et al.*, 2008). This results in numerous statistical tests being performed per peak calling, and FDR might be a preferred parameter, to reduce the number of false positives.

Table 5.3: Total peaks identified from the peak callers MACS, MACS2 and LanceOtron

	Total peaks		
	MACS (Periyasamy <i>et al.</i> 2015)	MACS2	LanceOtron (-noise)
A3B (vehicle)	3052	1090	2925
A3B (estrogen)	24,098	10,551	6531
ER (vehicle)	NA	3812	26
ER (estrogen)	35,663	27,006	3114
γ H2AX (vehicle)	NA	4499	441
γ H2AX (estrogen)	17,892	5730	607

5.2.4 Peak calling using the novel tool LanceOtron

More recently, a new bioinformatics tool for peak calling was developed. Many peak calling tools, such as MACS, identify peaks based comparing enrichment of DNA fragments to the (assumed random) background signal. LanceOtron was developed to eliminate dependence of background signal, as this can be impacted by factors such as chromatin state or underlying sequence, and thus it is not random across the genome (Hentges *et al.*, 2022). LanceOtron uses a convolutional neural network, a type of artificial neural network, which has been trained on ENCODE datasets to recognise ChIP-seq peaks, based on their characteristic shape (Hentges *et al.*, 2022). This is used alongside analysis of the ChIP-seq fragment enrichment to identify peaks (Hentges *et al.*, 2022). When tested on ChIP-seq files of transcription factor or histone modification datasets with known peak regions, LanceOtron outperformed MACS2, as seen by the higher sensitivity, selectivity and accuracy (Hentges *et al.*, 2022). Presence of an input file to provide background noise only improves LanceOtron peak

calling slightly, whereas it has a larger effect on MACS2 (Hentges *et al.*, 2022). However, MACS2 performs better for broad peaks (>1kb) than LanceOtron (Hentges *et al.*, 2022).

Peak calling of the A3B, ER and γ H2AX datasets by LanceOtron was performed, and compared to peaks identified by MACS2. Total number of peaks identified for A3B in absence of estrogen was nearly 3 times higher than identified by MACS2 (2925 vs 1090 peaks, table 5.3). However, LanceOtron identified fewer peaks for A3B in presence of estrogen than MACS2 (6531 vs 10,551 peaks, table 5.3). The most prevalent peak type was super spread and punctate in absence/presence of estrogen, respectively (figure 5.3). A3B was previously found to form punctate peaks spanning regions of around 1kb on average (Periyasamy *et al.*, 2015). Peaks of A3B binding in the absence of estrogen were mostly jagged peaks spanning wide regions, which is different than the sharp, distinctive A3B peaks previously reported (in presence of estrogen) (Periyasamy *et al.*, 2015). Around 15% of the A3B binding sites in absence of estrogen identified by LanceOtron were also identified by MACS2. In comparison, 60% of the A3B binding sites in presence of estrogen identified by LanceOtron overlapped with the binding regions identified by MACS2 (table 5.4). This indicates that there is a stronger agreement between LanceOtron and MACS2 in identifying binding sites occupied by A3B in presence of estrogen. This likely reflects MACS2's superiority in identifying broader peaks, but inferiority in identifying punctate peaks.

The number of peaks discovered by LanceOtron for ER (in the presence of estrogen) and γ H2AX (in the presence/absence of estrogen) was around 10% of the number of peaks identified by MACS2 (table 5.3). This was even lower for number of ER peaks in absence of estrogen, where LanceOtron identified less than 0.7% of the total number of peaks discovered by MACS2 (table 5.3). However, unlike A3B, most of the ER/ γ H2AX binding regions discovered by LanceOtron overlapped with binding regions discovered by MACS2 (table 5.4).

The majority of the ER peaks (in presence of estrogen) identified by LanceOtron were punctate peaks (figure 5.3), which is similar to the ER peak morphology previously seen (Periyasamy *et al.*, 2015; Welboren *et al.*, 2009). The most prevalent peaks for ER (in absence of estrogen) and γ H2AX (presence/absence of estrogen) were classified as noise by LanceOtron, and therefore excluded. However, punctate was the second most prevalent peak type for all three files. This is in agreement with the appearance of the ER/ γ H2AX peaks previously reported for these files (Periyasamy *et al.*, 2015).

Table 5.4: Number of binding regions shared between the peak region lists identified by LanceOtron and MACS2

	LanceOtron only (-noise)	Shared	MACS2 only
A3B (vehicle)	2473	452	655
A3B (estrogen)	2410	4121	6246
ER (vehicle)	2	24	3970
ER (estrogen)	12	3102	24,100
γH2AX (vehicle)	77	364	4874
γH2AX (estrogen)	46	561	6092

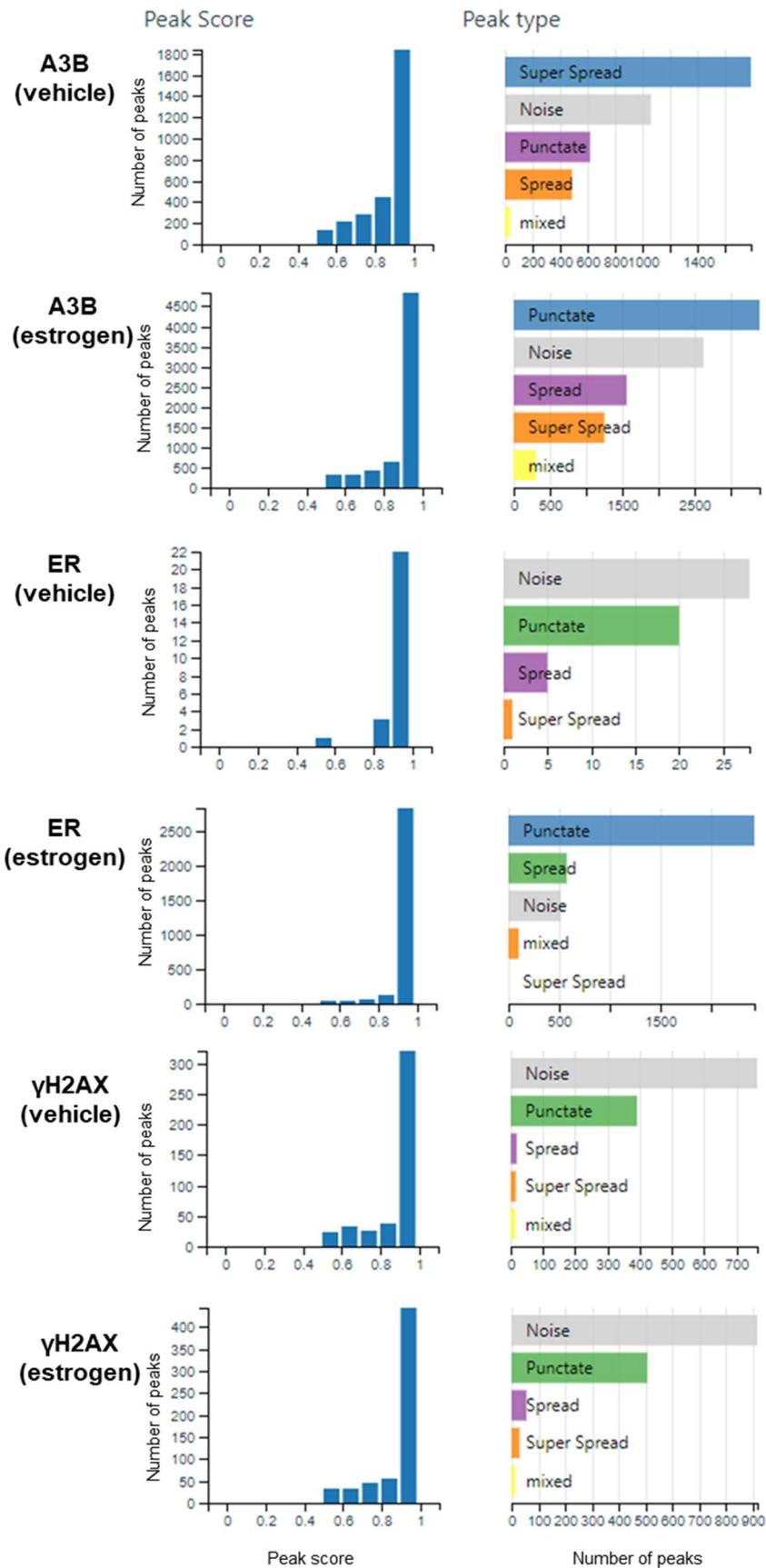


Figure 5.3. ChIP-seq peak calling using the novel tool LanceOtron. Binding regions for ChIP-seq files from vehicle/estrogen-treated MCF7 cells were identified using LanceOtron (find and score peaks). Peak scores (left) and peak types (right) are shown, excluding peaks identified as noise. Note that the y-axes are on different scales, reflecting the number of peaks identified per group.

LanceOtron offers a highly user-friendly interface, where files are uploaded directly and analysis performed, and where the peaks identified can be viewed on a genome browser, and detailed information about each peak (including peak score, peak type, peak width) is provided. The peak regions identified by LanceOtron for ER and γ H2AX overlapped highly with the peak regions identified by MACS2; however, MACS2 also identified thousands more peaks for each file, meaning some information could potentially be lost by using LanceOtron for peak calling for this particular experiment. Nevertheless, the LanceOtron peak files did include peaks at expected regions, near estrogen-target genes TFF1, GREB1, FOS, CTSD, RARA and EGR3 (figure 5.4, 5.5, 5.6, 5.7, 5.8, 5.9). Peak calling using LanceOtron on A3B ChIP-seq data (in the presence/absence of estrogen) identified more peaks which did not overlap with peak regions identified by MACS2. However, as MACS was used in the original publication of the A3B ChIP-seq files, MACS2 was chosen as peak caller for this analysis.

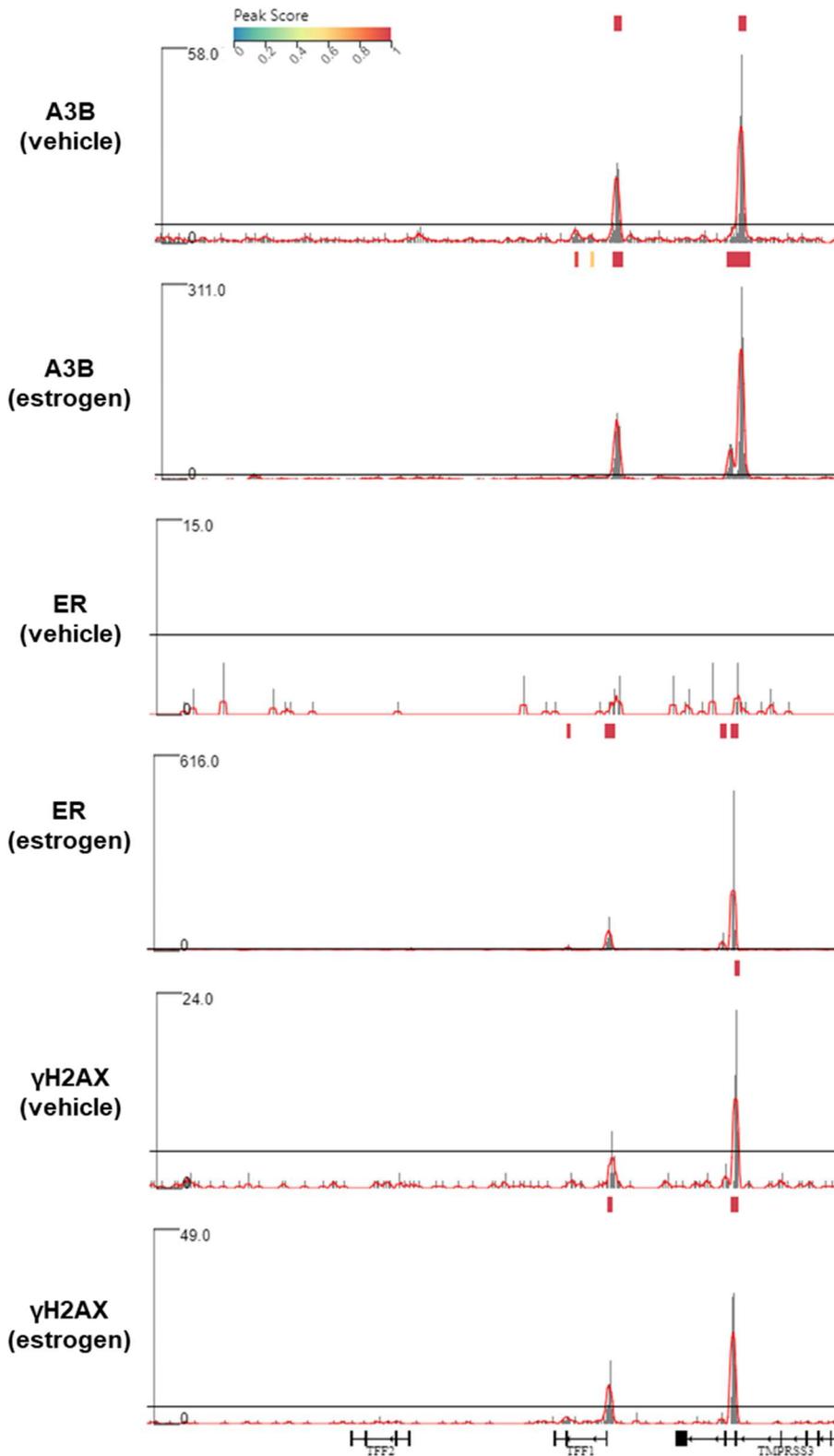


Figure 5.4. LanceOtron identifies APOBEC3B, ER and γ H2AX binding near estrogen target gene TFF1, in estrogen treated MCF7 cells. LanceOtron genome browser snapshot of ChIP-seq peaks, and significant binding regions identified by LanceOtron (blocks, coloured by peak score). Y-axes not standardised.

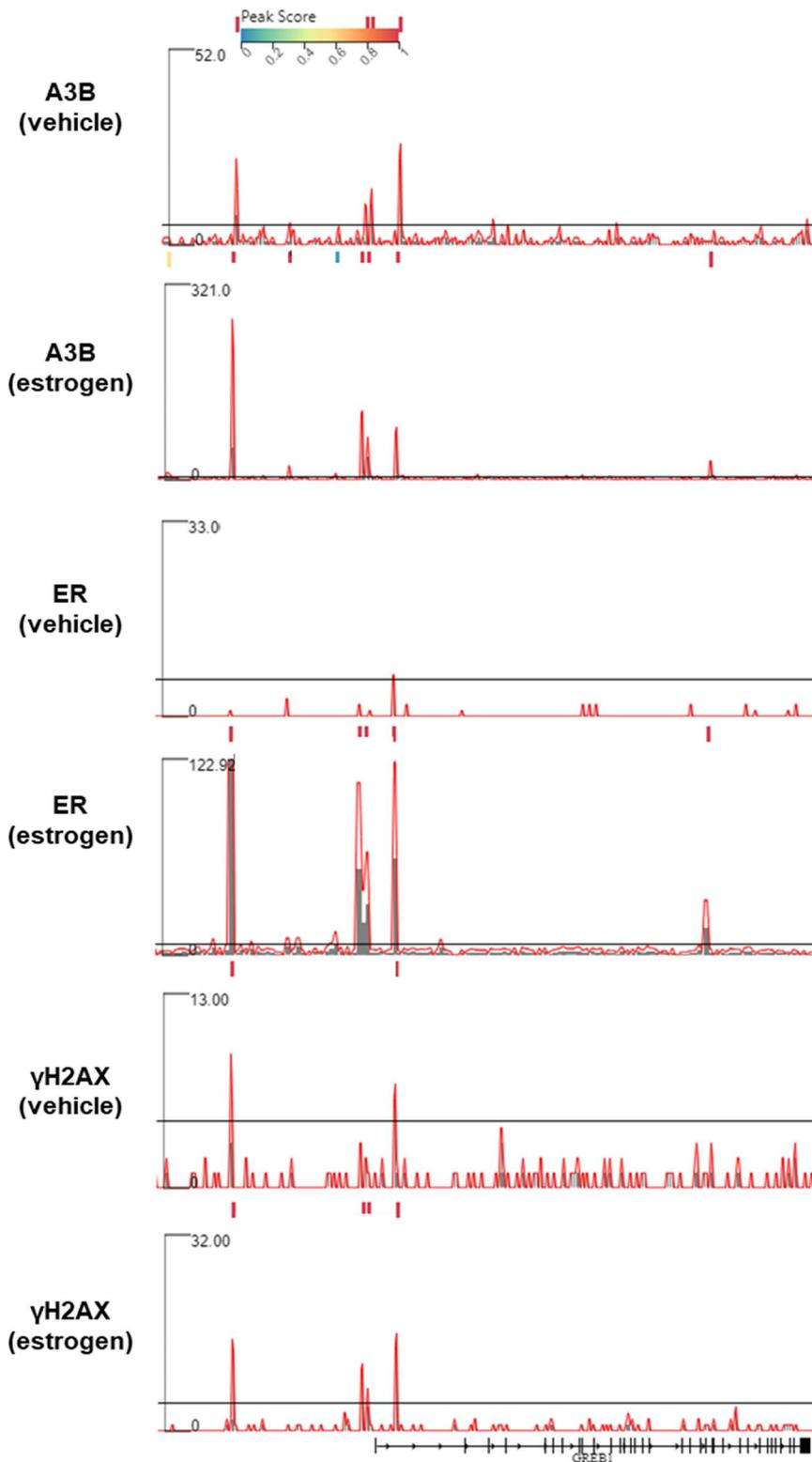


Figure 5.5. LanceOtron identifies APOBEC3B, ER and γ H2AX binding near estrogen target gene GREB1, in estrogen treated MCF7 cells. LanceOtron genome browser snapshot of ChIP-seq peaks, and significant binding regions identified by LanceOtron (blocks, coloured by peak score). Y-axes not standardised.

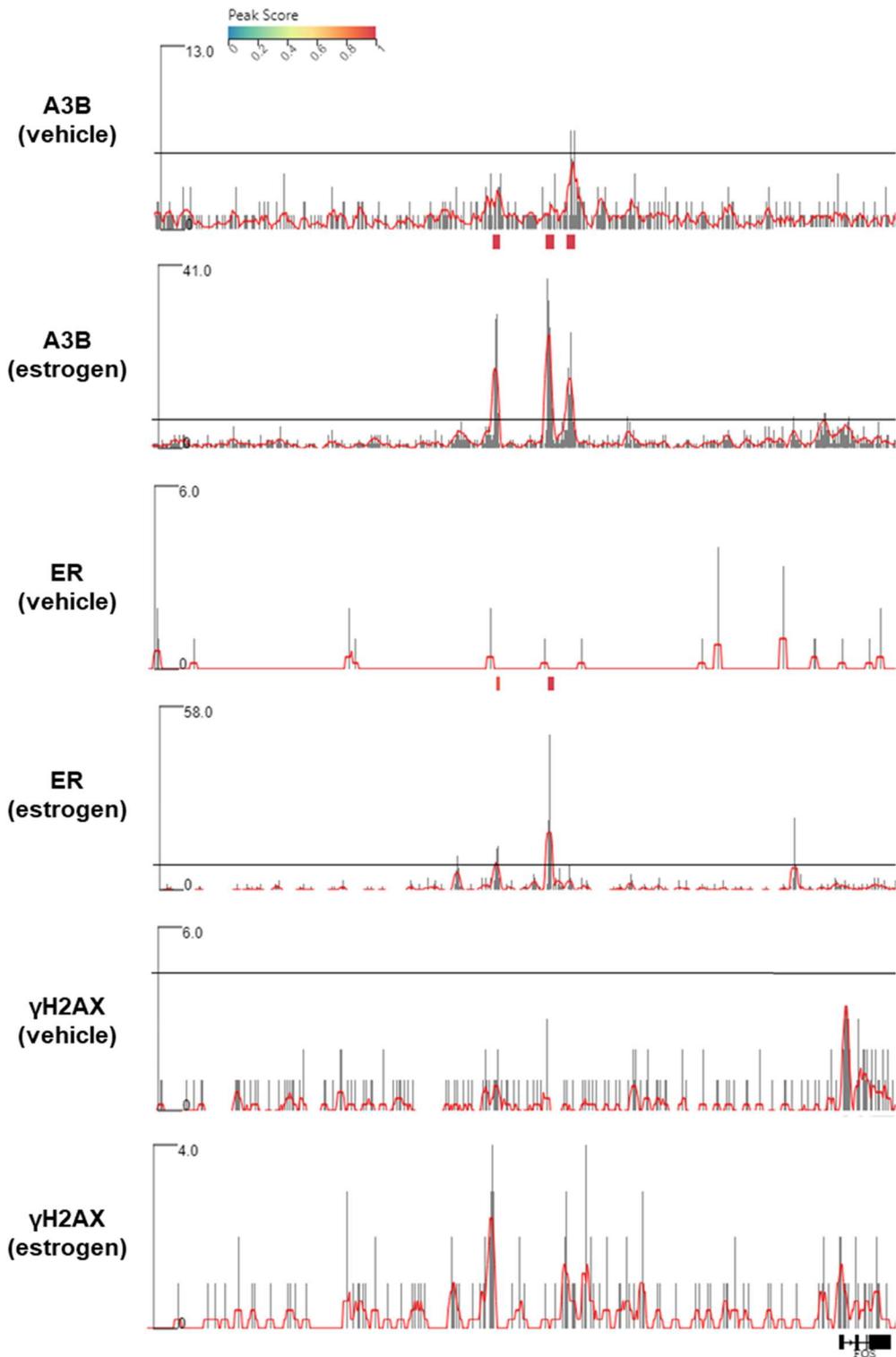


Figure 5.6. LanceOtron identifies APOBEC3B and ER binding near estrogen target gene FOS, in estrogen treated MCF7 cells. LanceOtron genome browser snapshot of ChIP-seq peaks, and significant binding regions identified by LanceOtron (blocks, coloured by peak score). Y-axes not standardised.

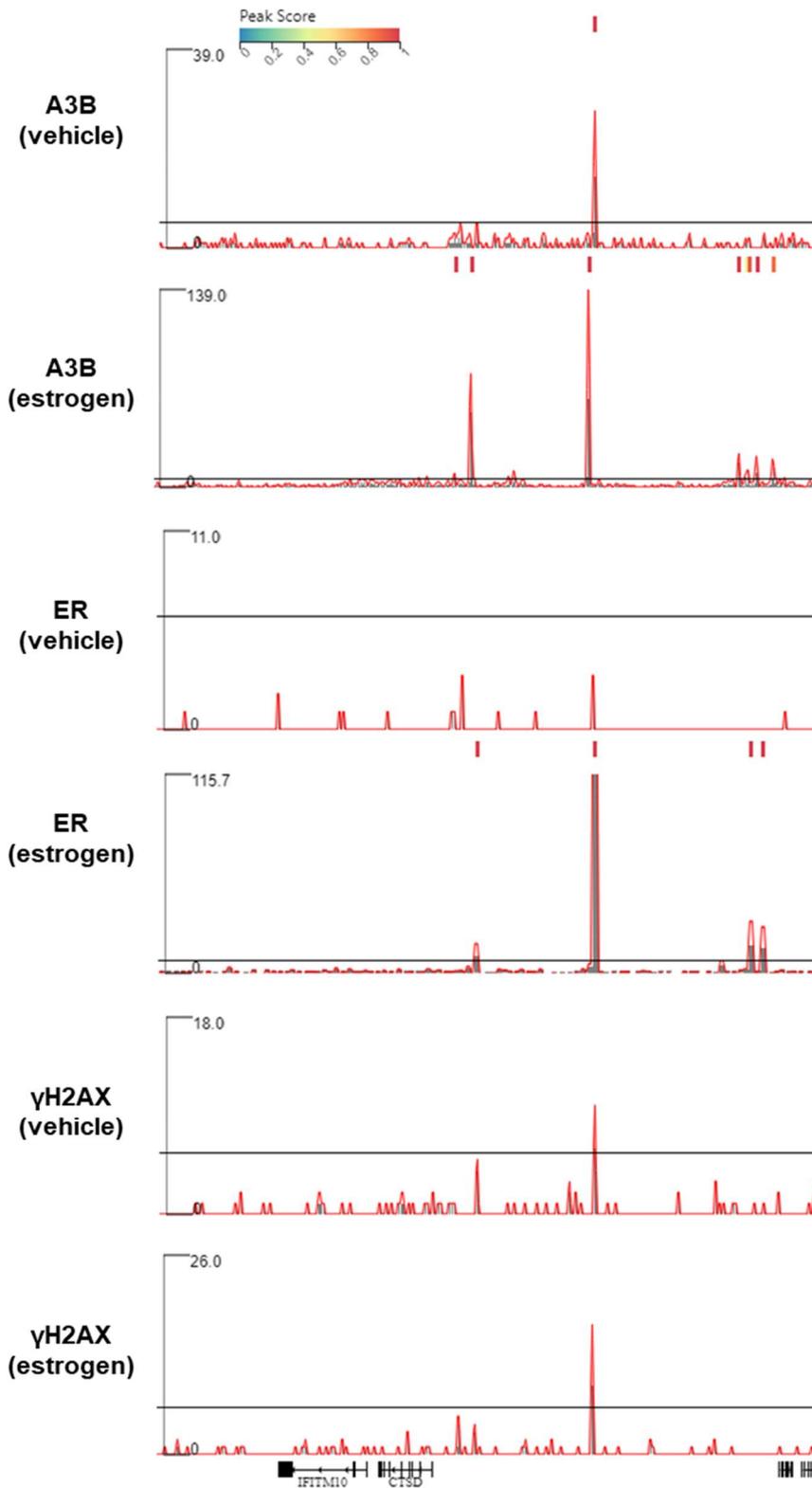


Figure 5.7. LanceOtron identifies APOBEC3B and ER binding near estrogen target gene CTSD, in estrogen treated MCF7 cells. LanceOtron genome browser snapshot of ChIP-seq peaks, and significant binding regions identified by LanceOtron (blocks, coloured by peak score). Y-axes not standardised.

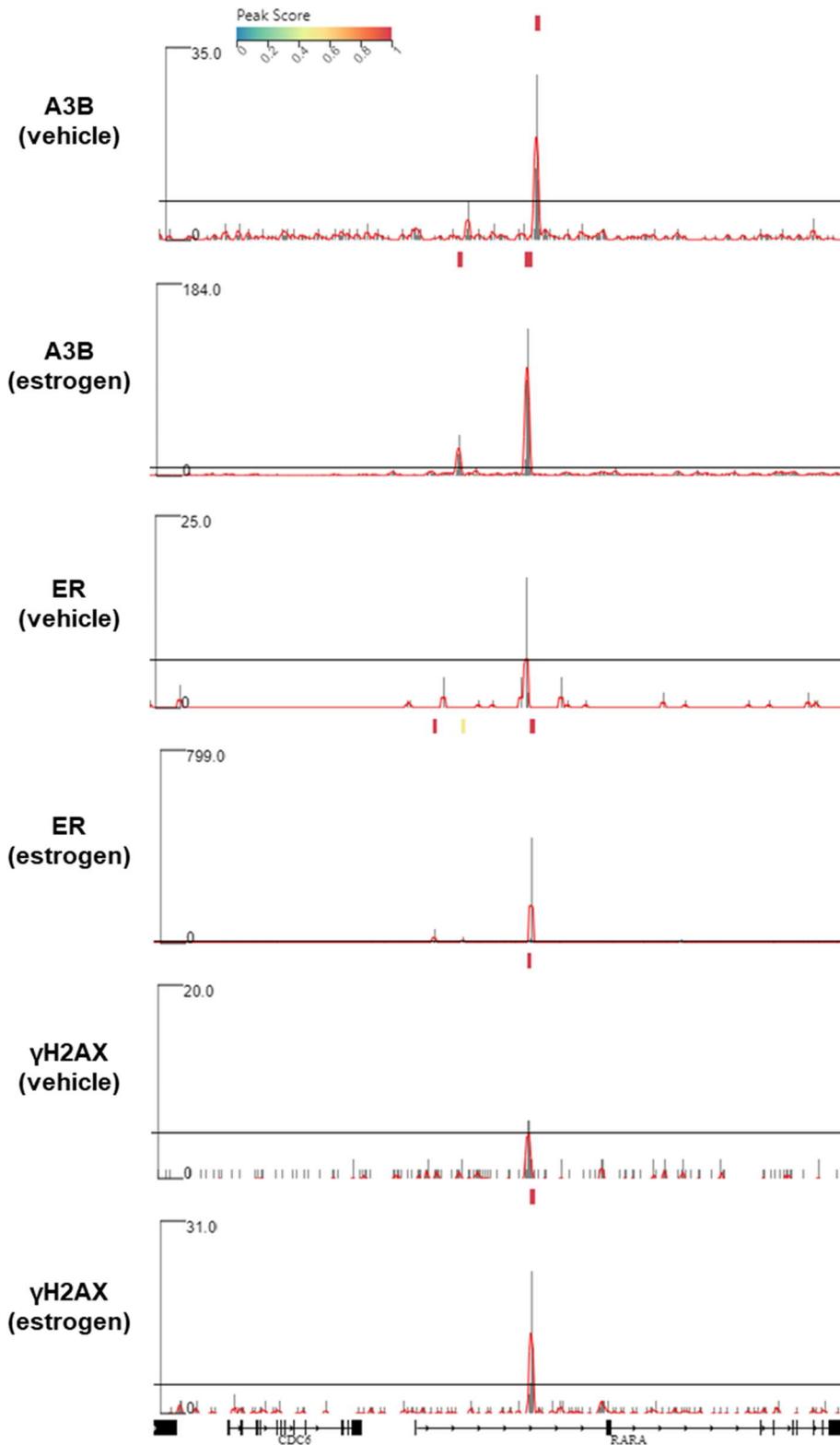


Figure 5.8. LanceOtron identifies APOBEC3B, ER and γ H2AX binding near estrogen target gene RARA, in estrogen treated MCF7 cells. LanceOtron genome browser snapshot of ChIP-seq peaks, and significant binding regions identified by LanceOtron (blocks, coloured by peak score). Y-axes not standardised.

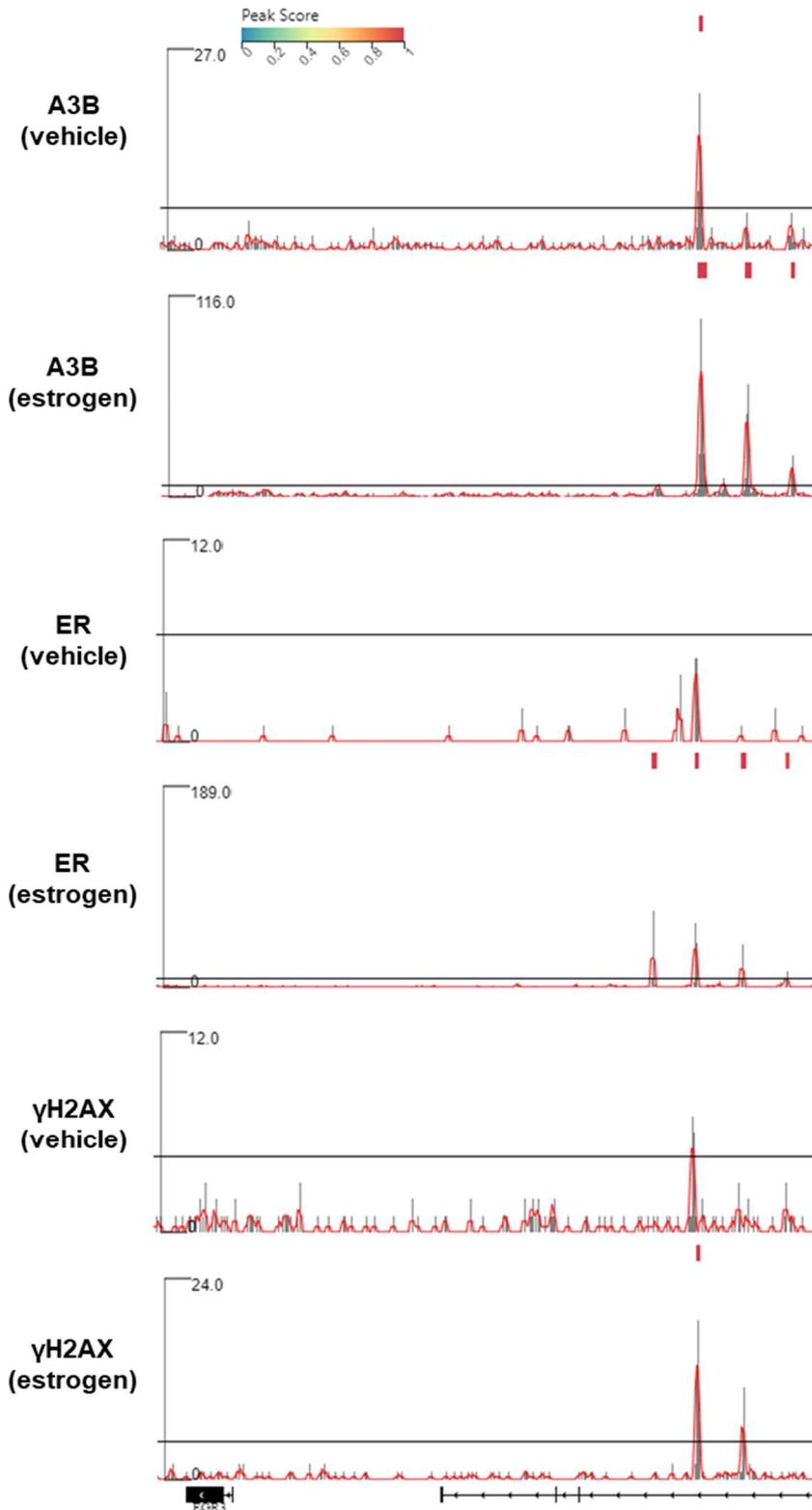


Figure 5.9. LanceOtron identifies APOBEC3B, ER and γ H2AX binding near estrogen target gene EGR3, in estrogen treated MCF7 cells. LanceOtron genome browser snapshot of ChIP-seq peaks, and significant binding regions identified by LanceOtron (blocks, coloured by peak score). Y-axes not standardised.

5.2.5 A3B and ER binds estrogen-target genes in presence of estrogen

A3B was previously found to bind 24,098 sites in presence of estrogen, of which 46% were regions also bound by ER (Periyasamy *et al.*, 2015). Here, 54% of the 10,551 A3B binding sites were found to overlap with ER binding sites (figure 5.10 A).

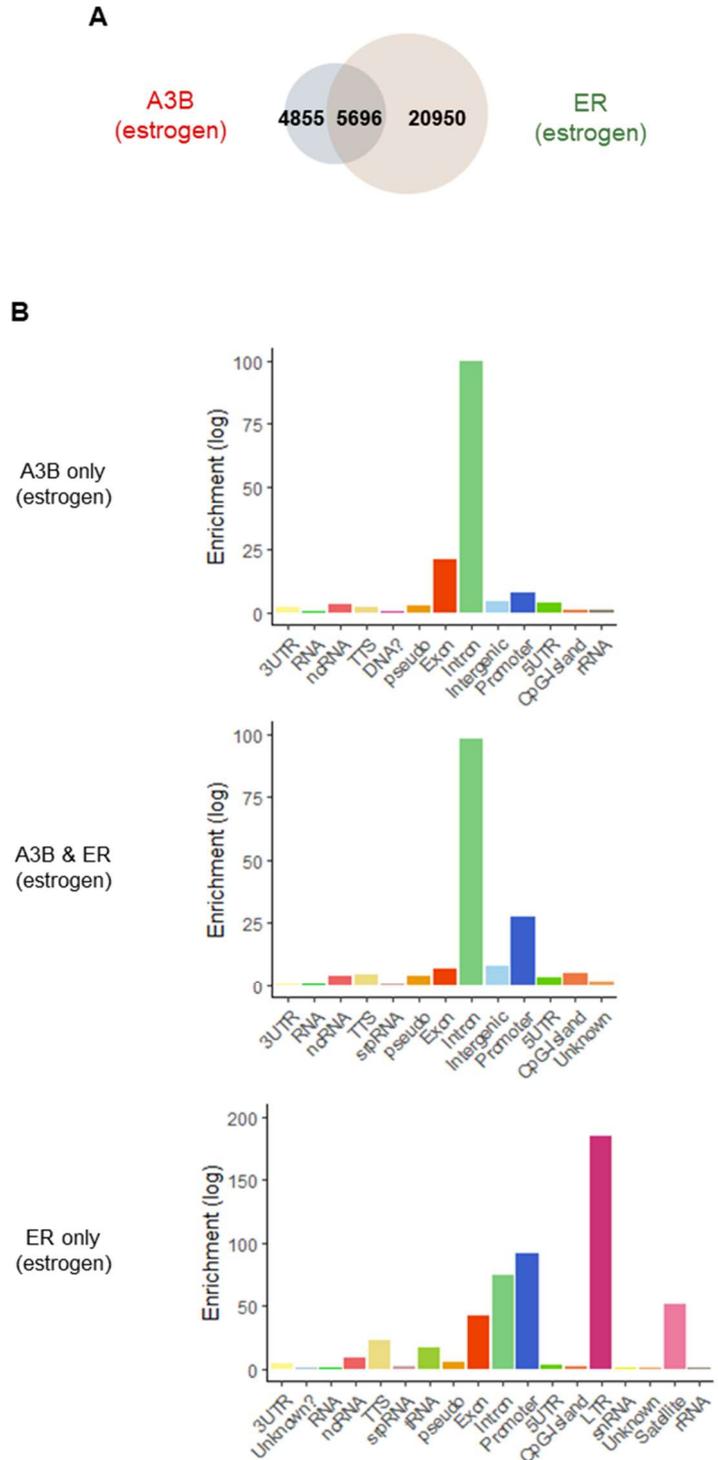


Figure 5.10. Half of APOBEC3B binding sites overlap with ER binding sites in presence of estrogen in MCF7 cells.

- (A) Overlap between A3B and ER ChIP-seq binding regions in presence of estrogen.
- (B) HOMER peak annotation of A3B and ER overlapping and nonoverlapping binding regions.

A3B was previously found at ER binding sites near estrogen-target genes, in an ER- and estrogen-dependent manner (Periyasamy *et al.*, 2015). Despite differences in total number of A3B/ER peaks detected here and by Periyasamy *et al.*, A3B and ER were found at overlapping regions near TFF1, PDZK1, RARA, EGR3, GREB1, FOS and CTSD in an estrogen-dependent manner (figure 5.11, 5.12). In addition, A3B and ER were found at overlapping binding sites near three genes previously found to be bound by ER in MCF7-cells, but where A3B binding has not previously been investigated; DICER1, CHPT1 and MAX (figure 5.12) (Welboren *et al.*, 2007).

Higher levels of γ H2AX formation were previously detected at TFF1, PDZK1, RARA and EGR3 in an ER/A3B-dependent manner (Periyasamy *et al.*, 2015). Here, γ H2AX was found at increased levels in regions overlapping with A3B and ER, at TFF1, PDZK1, RARA, EGR3, GREB1, CTSD and DICER1 (figure 5.11, 5.12). However, γ H2AX was found at these regions also in absence of estrogen (although at lower levels), suggesting that some DNA damage is occurring at these sites without estrogen stimulation. This could be a result of estrogen not being completely depleted from the vehicle treated cells. FOS, CHPT1 and MAX displayed low levels of γ H2AX with little difference in absence/presence of estrogen (figure 5.12). This suggests that A3B/ER are not inducing DSBs at all sites they bind in MCF7 cells.

Interestingly, γ H2AX typically spreads to megabase wide-domains around the site of the DSB (Iacovoni *et al.*, 2010; Massip *et al.*, 2010). However, the γ H2AX peaks

overlapping with A3B/ER were narrow, around 1kb wide (figure 5.11, 5.12). This was also true for γ H2AX peaks not overlapping with neither A3B nor ER (not shown).

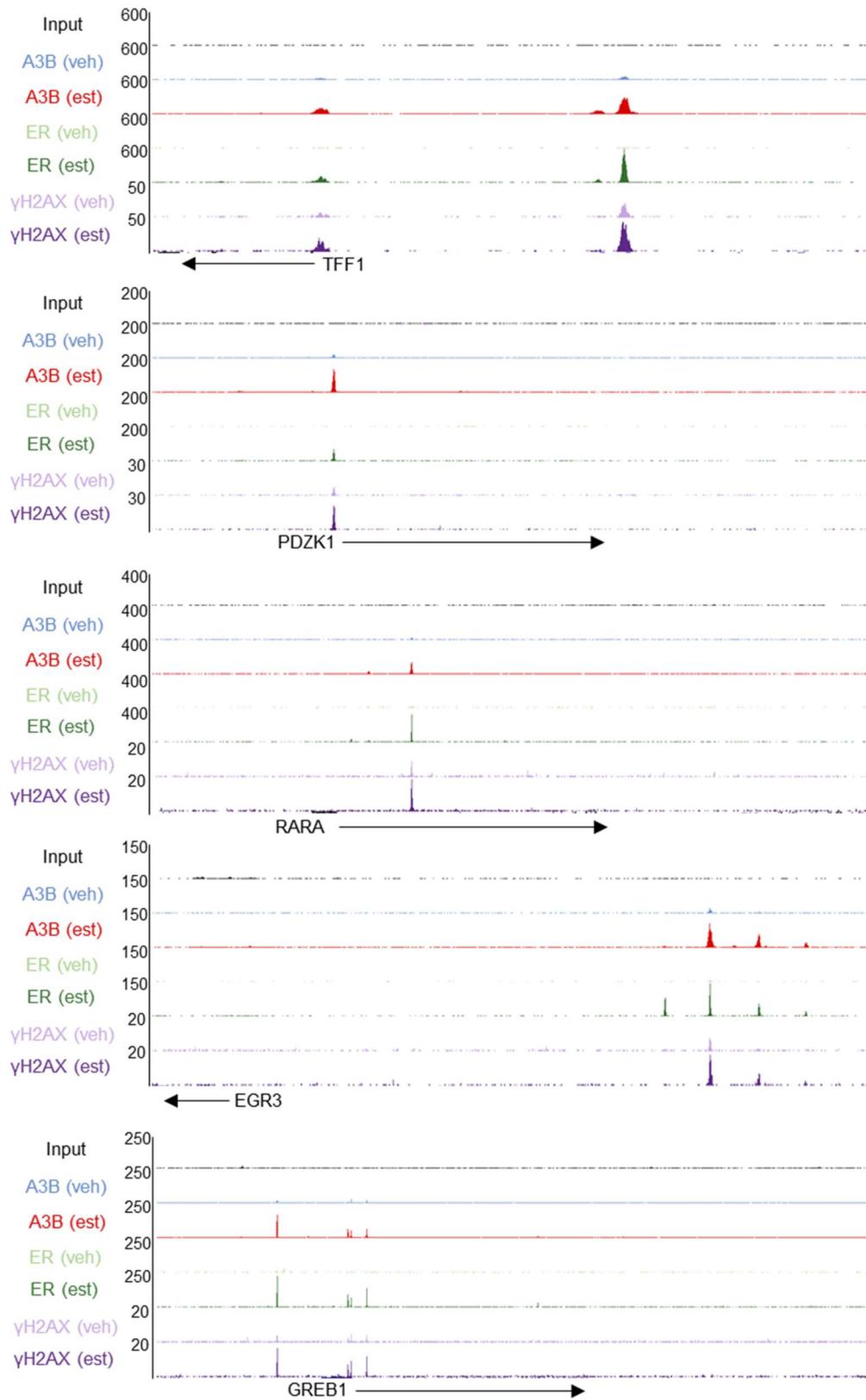


Figure 5.11. APOBEC3B, ER and γ H2AX bind near estrogen target genes TFF1, PDZK1, RARA, EGR3 and GREB1, in estrogen treated MCF7 cells. UCSC genome browser snapshot of A3B, ER and γ H2AX ChIP-seq peaks and input (control) from untreated (veh) and estrogen (est) treated MCF7 cells.

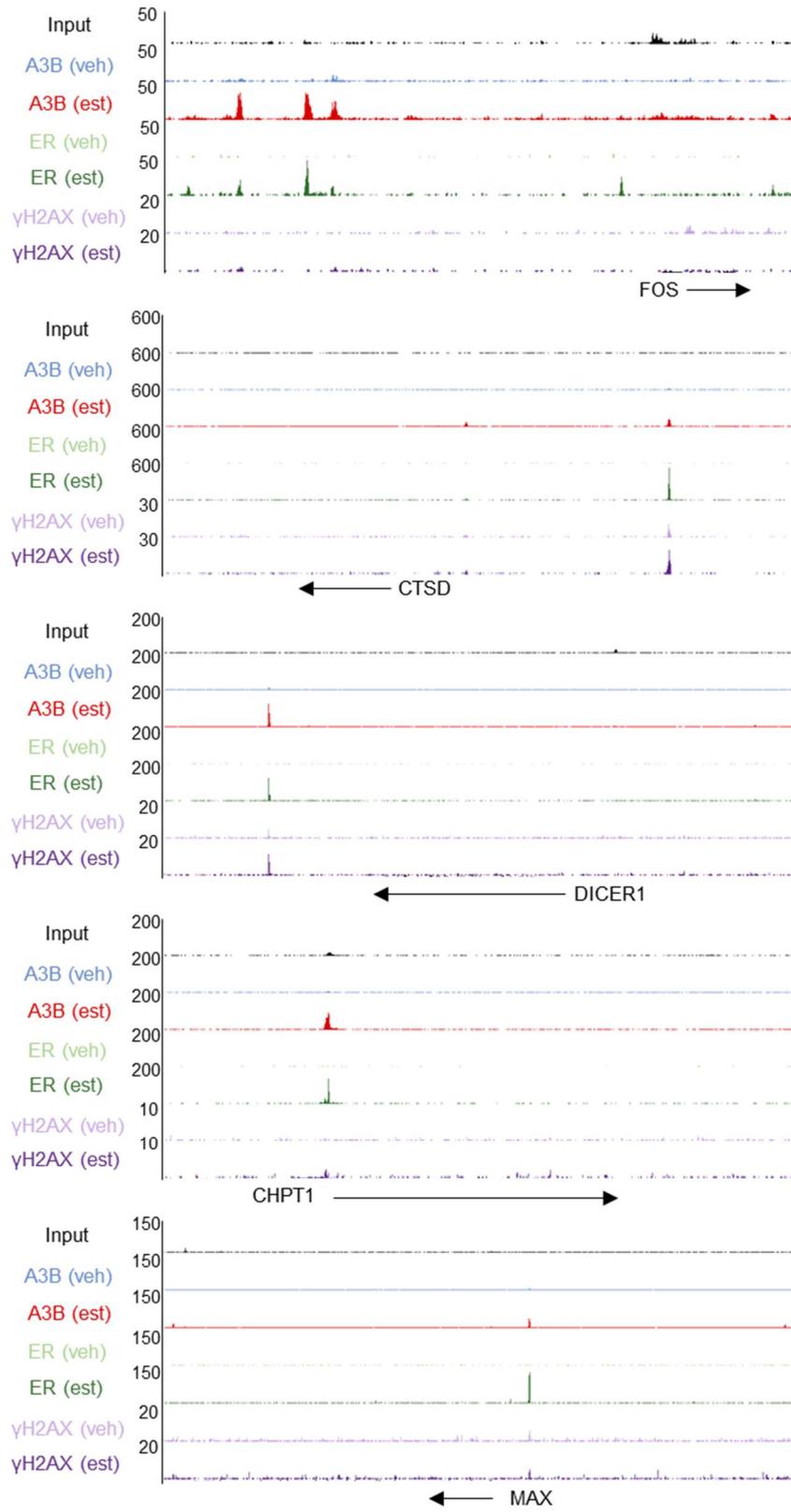


Figure 5.12. APOBEC3B, ER and γ H2AX bind near estrogen target genes FOS, CTSD, DICER1, CHPT1 and MAX, in estrogen treated MCF7 cells. UCSC genome browser snapshot of A3B, ER and γ H2AX ChIP-seq peaks and input (control) from untreated (veh) and estrogen (est) treated MCF7 cells.

5.2.6 Peak region annotation of A3B/ER binding sites

To further investigate the underlying features of the chromatin regions bound by A3B and ER, gene and genome ontology analysis, and peak region annotation, was performed using the next-generation sequencing tool Hypergeometric Optimization of Motif EnRichment (HOMER). The 28 peaks found specifically in the absence of estrogen (estrogen excluded) were mostly found at satellite regions and relaxed circular (RC) DNA (figure 5.1 B). Satellite DNA is made up of repeating sequences, which are therefore challenging to map. This group of peaks might therefore mostly consist of noise resulting from inaccurate alignment. The 1062 A3B peaks found both in absence and presence of estrogen (estrogen independent) were mostly found at introns and intergenic regions, which could represent enhancer regions (figure 5.1 B). Although A3B is found at these binding sites independently of estrogen, 89% of these sites overlapped with ER binding sites (not shown). The 9489 regions exclusively bound by A3B in presence of estrogen were mostly found at introns, followed by exons and promoters (figure 5.1 B), in agreement with the previous report (Periyasamy *et al.*, 2015).

ER was previously found to bind intronic and gene-distal regions (Welboren *et al.*, 2009). Here, in presence of estrogen, near half of the A3B binding regions overlapped with ER binding regions (5696 peaks in total). These were mostly found at introns and promoter regions (figure 5.10 B), which is in line with A3B and ER binding near estrogen-target genes to regulate their expression (Periyasamy *et al.*, 2015). The 4855 sites bound by A3B independently of ER were enriched for introns and exons,

with less enrichment for promoter regions (figure 5.10 B). ER bound 20,950 sites independently of A3B; these regions were mostly found at long terminal repeats (LTR), followed by promoters, introns, satellite regions and exons (figure 5.10 B). As mentioned previously, mapping to repeat regions, such as LTRs or satellite regions should be interpreted with caution. However, this does suggest that there is a set of genes which ER can regulate independently of A3B, as ER-specific binding regions were enriched for promoters.

Periyasamy *et al* reported that the majority of the γ H2AX formed in presence of estrogen overlapped with ER and A3B binding sites. Here, only 16% (940 peaks) of γ H2AX binding sites were found to overlap with sites occupied by A3B and ER (figure 5.2 A). The majority of γ H2AX was found at repetitive elements both in absence and presence of estrogen, such as Long Interspersed Nuclear Elements (LINEs), Short Interspersed Nuclear Elements (SINEs), simple repeats and satellite regions (figure 5.2 B). However, some estrogen-dependent γ H2AX peaks were found at exons and promoters, which likely represents the A3B/ER-bound regions where DNA damage and γ H2AX formation is induced.

5.2.7 A3B overlaps significantly with transcriptional start sites with/without ER

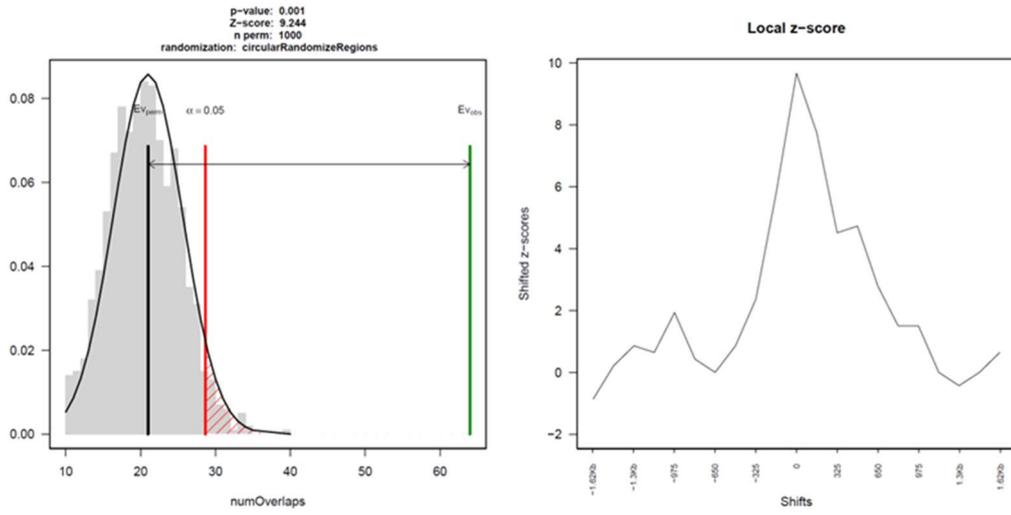
The majority of A3B binding regions with/without ER, are found at gene-distal regions. To investigate whether A3B, also in absence of ER, is significantly enriched at transcriptional start sites (TSS, hg19 Refseq whole gene), regioneR was used to assess whether these peak regions overlap more with TSSs than expected (Gel *et al.*, 2016). regioneR creates a list of randomised regions, which is equivalent to the inputted list of ChIP-seq peak regions of interest (same number of regions, and same width of regions). Permutation tests are then performed, where number of overlaps between the ChIP-seq peak and TSS regions of interest are computed, along with

number of overlaps with the randomised regions. This process is repeated 1000 times with different randomised coordinates, and the overlap between these sites and TSSs computed, to calculate the probability of peaks overlapping with the TSS regions of interest by chance. For these permutation tests, the p-value is restricted by the number of permutations performed (lowest possible p-value 0.001 with 1000 permutation tests). The Z-score is also computed, which represents the distance between the number of overlaps for the region list of interest, and the mean overlap for the random regions, divided by the standard deviation of the overlap for the random regions (Gel *et al.*, 2016). To investigate whether the association between the ChIP-seq peaks and the TSSs is linked to the exact position of the ChIP-seq peaks, these regions are shifted 5' and 3' from the original position, and Z-scores from these altered positions are calculated (Gel *et al.*, 2016).

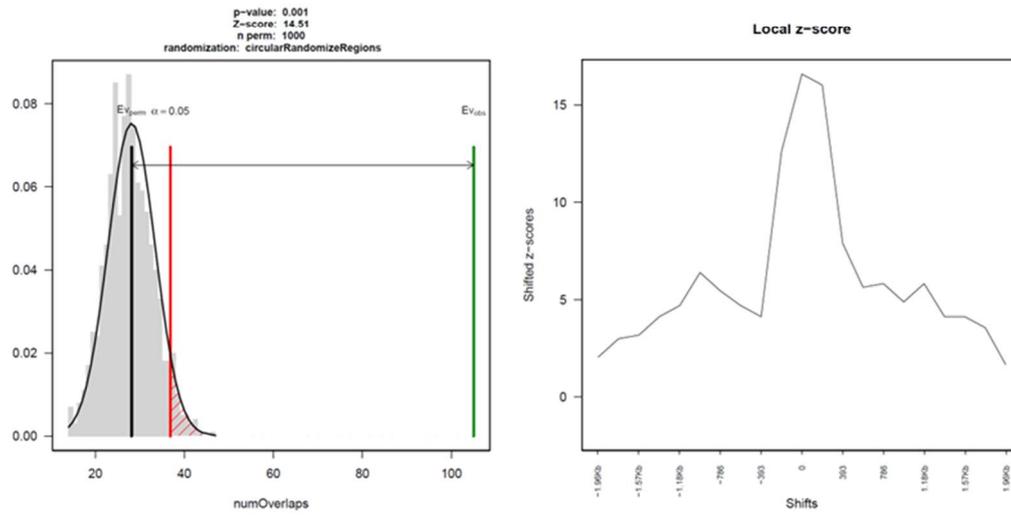
Regions occupied by A3B and ER together significantly overlapped with TSSs (figure 5.13, $p < 0.001$), which is in line with their role in transcriptional regulation, and binding near estrogen target genes (Periyasamy *et al.*, 2015; Zhang *et al.*, 2022). Interestingly, regions occupied by A3B alone, also overlapped significantly with TSSs (figure 5.13, $p > 0.001$). A3B can be recruited to potential promoter regions independently of ER, although the role of A3B at these sites needs further investigation. Similarly, regions occupied by ER alone also overlapped significantly with TSSs (figure 5.13, $p > 0.001$). The association between TSSs and ChIP-seq peak regions were dependent on the exact position of the ChIP-seq peaks, as seen by peaks formed at the centre of the shifted Z-score plots (figure 5.13).

Transcriptional start sites

A3B only (estrogen)



A3B & ER (estrogen)



ER only (estrogen)

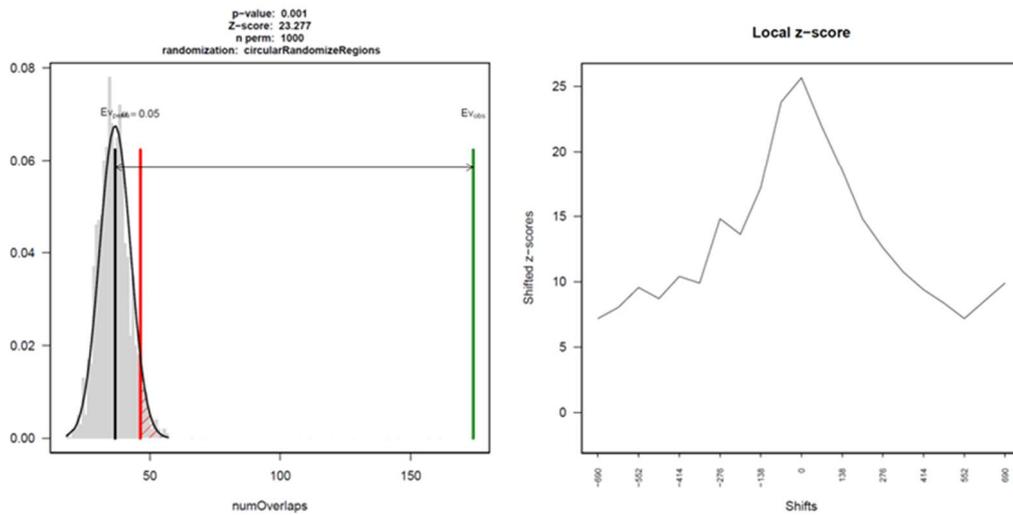


Figure 5.13. APOBEC3B and ER binding sites overlap significantly with transcriptional start sites, in estrogen-treated MCF7 cells. Permutations tests performed using regioneR to test significance of overlap between the binding regions and TSS. Observed number of overlaps (green), significance limit (red), observed number of overlaps for randomised regions (black) and Z-score (arrow) are shown (left). Local Z-score (right) is calculated by shifting the regions 5' and 3', calculating the Z-score to assess whether the association is linked to the exact position of the binding regions.

5.2.8 Gene ontology enrichment of peak regions

A3B binds near various genes, including estrogen-regulated genes. To investigate the biological function of the genes near A3B and/or ER, gene ontology enrichment was performed using HOMER (top five categories, table 5.5). The genes near regions bound by A3B exclusively (A3B-only), were enriched for categories relating to estrogen/ER targets (table 5.5). This is surprising, as ER was not detected at these regions. The genes occupied by both A3B & ER were all enriched for estrogen/ER-related processes (table 5.5). This is in agreement with A3B regulating estrogen response genes together with ER (Periyasamy *et al.*, 2015). Interestingly, this was not the case for the highest enriched genes bound exclusively by ER (ER-only) (table 5.5). These genes were related to neoplasm/tumour of haematopoietic/lymphoid tissue, pancreas and the central nervous system (table 5.5).

Table 5.5: Enriched gene ontology categories (top 5)

	Term	P-value
	GOZGIT_ESR1_TARGETS_DN	2.375e-33
	system development	6.054e-23
A3B only	HALLMARK_ESTROGEN_RESPONSE_EARLY	7.464e-22
(estrogen)	multicellular organism development	1.357e-21
	MEISSNER_BRAIN_HCP_WITH_H3K4ME3_AND_H3K27ME3	1.655e-21
	BHAT_ESR1_TARGETS_NOT_VIA_AKT1_UP	1.911e-82
	HALLMARK_ESTROGEN_RESPONSE_EARLY	4.033e-80
A3B & ER	BHAT_ESR1_TARGETS_VIA_AKT1_UP	1.293e-76
(estrogen)	HALLMARK_ESTROGEN_RESPONSE_LATE	1.425e-47
	DUTERTRE ESTRADIOL_RESPONSE_6HR_UP	1.162e-38
	haematopoietic_and_lymphoid_tissue-lymphoid_neoplasm	1.339e-69
	pancreas-carcinoid-endocrine_tumour	9.908e-63
ER only	central_nervous_system-brain-	4.215e-62
(estrogen)	primitive_neuroectodermal_tumour-medulloblastoma	
	carcinoid-endocrine_tumour	1.625e-61
	haematopoietic_and_lymphoid_tissue-lymphoid_neoplasm-Burkitt_lymphoma	1.220e-55

5.2.9 De novo motif analysis: A3B with/without ER binds the ERE

The ERE (ESR1-motif for ER α) is a palindromic sequence, made up of two halves, separated by 3 nucleotides: AGGTCAnnnTGACCT (figure 5.14 A) (Gruber *et al.*,

2004). ER can bind both full- and half-length EREs. A3B preferably deaminates cytosines preceded by a thymidine (McDaniel *et al.*, 2020), which is present within the ERE, and A3B binding sites were previously found to be enriched for the ERE (Periyasamy *et al.*, 2015). To investigate which motifs A3B binds in absence of ER, *de novo* motif analysis was performed using the meme-chip tool within the MEME suite. Here, short, un-gapped sequences that were significantly enriched in the sequences underlying the regions bound by A3B/ER were identified. This was then compared to a motif database, and known or similar motifs were reported.

In line with previous findings, the top 2 most significantly enriched motifs for regions co-occupied by A3B and ER were the full- and half-length ERE (figure 5.14 B). 70% of these regions overlapped with (at least) the half-length ERE, and 50% of these regions overlapped with the full-length ERE (figure 5.14 B). Interestingly, 29% of the A3B-only binding sites were enriched for the half-length ERE, but the full-length ERE was not enriched, suggesting that A3B only binds full-length ERE in presence of ER (figure 5.14 B). The second most enriched motif for sites bound by A3B-only, was a 15 base long sequence made up of cytosine triplets spaced by 2-3 nucleotides (figure 5.14 B). This is similar to the motif for MYC-associated zinc finger protein (MAZ), POZ-, AT hook-, and zinc finger-containing protein 1 (PATZ1) and Sp2 motif, factors which are all involved in regulation of gene expression. Regions bound exclusively by ER were enriched for the ERE half and full motifs, as expected (figure 5.14 B). However, the most enriched version of the full-length motif was 30 nucleotides long, with 3 and 12 nucleotides preceding/succeeding the typical ERE motif. The thymidine residue at position 10 of the ERE motif was replaced by adenine (position 13, figure 5.14 B). However, only 3% of ER-only binding regions overlapped with this motif. A higher percentage of ER-only binding regions overlapped with the 15 nucleotide ER (30%), but this motif was less significantly enriched. The E-value, used by MEME to report the significance of a given motif, is an estimate of the likelihood of getting the

same motif sequence in a randomised sequence pool (Bailey *et al.*, 2015). The significance of a motif is affected by both the frequency of a motif and the confidence of it.

γ H2AX is formed in response to DNA double-strand breaks, including when A3B recruited by ER creates DNA breaks to induce expression of estrogen target genes (Periyasamy *et al.*, 2015; Bonner *et al.*, 2008). The DNA break is formed when the uracil (deaminated cytidine) is excised as part of BER, creating DNA nicks (Periyasamy *et al.*, 2015). To investigate A3B/ER/ γ H2AX binding at the half- and full length ERE motifs, a list of these motifs in the hg19 genome was generated. With position weight matrices (PMW) from the HOCOMOCO database for the ERE half- and full-length motifs (figure 5.15 A), PWMtools was used to identify genomic regions at which these motifs occur. 119,580 and 167,227 full- and half-length EREs were identified, respectively (figure 5.15 B). Regions occupied by γ H2AX & A3B & ER showed a similar enrichment for ERE full-length motif to regions occupied by A3B & ER with no γ H2AX formation (41% vs 36% of peaks with the motif, figure 5.15 B). Consistently, regions occupied by γ H2AX & A3B & ER had similarly low levels of ERE half-length motif per peak, when compared to regions occupied by A3B & ER with no γ H2AX formation (4% vs 10% of peaks with the motif, figure 5.15 B). This suggests that there is little difference in γ H2AX formation and presence of the ERE full- and half-length motifs, but that A3B and ER preferentially bind the ERE full-length motif.

A

ERE (ESR1)



B

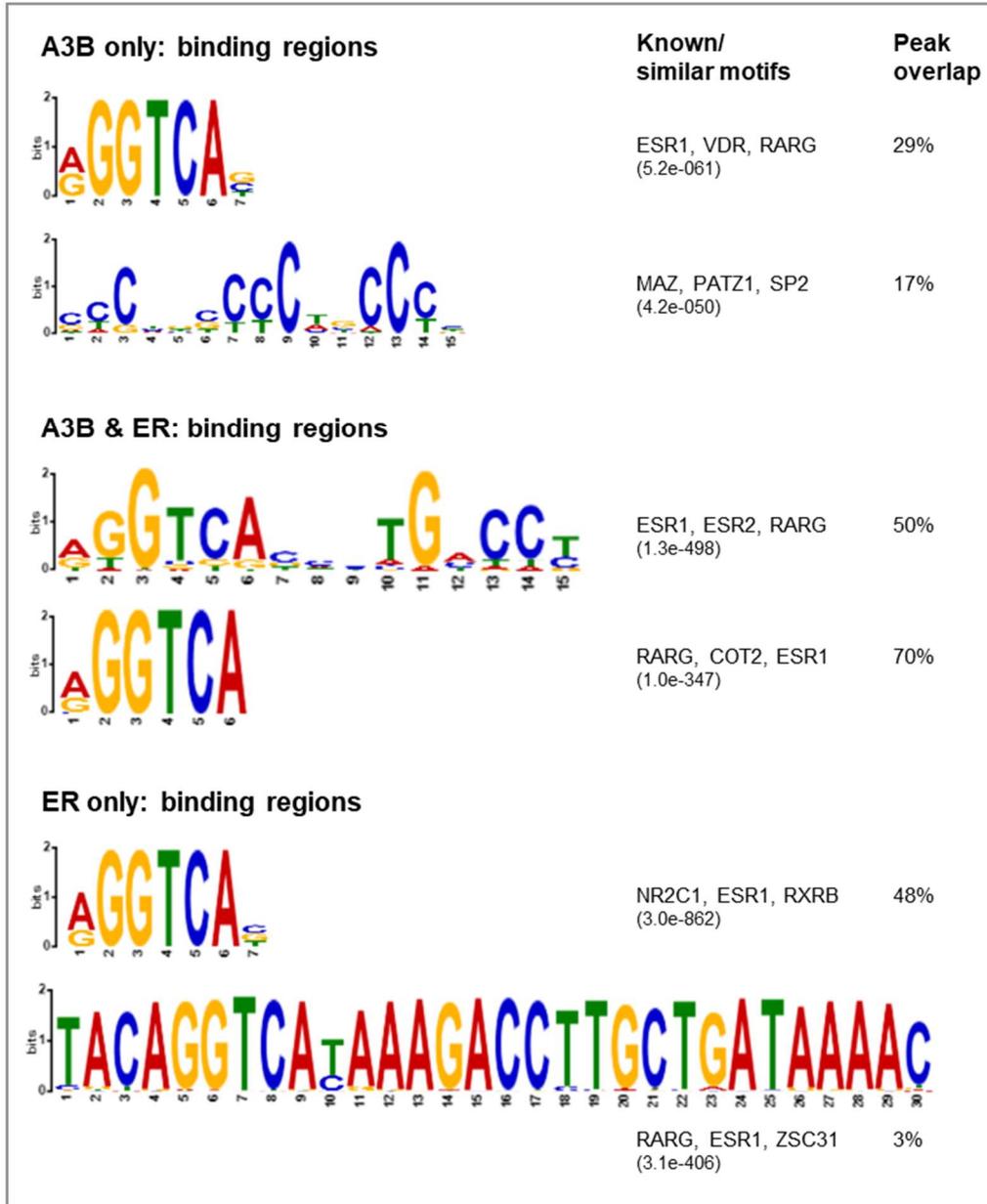


Figure 5.14. Top motifs underlying regions bound by APOBEC3B alone, APOBEC3B and ER together, and ER alone.

(A) Full-length ERE (HOCOMOCO database).

(B) Top two motifs discovered by MEME *de novo* motif analysis, and the percentage of peaks overlapping with the motifs.

A

Model	LOGO	Transcription factor	Model length
			>= 20
ESR1_HUMAN.H11MO.0.A Full motif		ESR1 (GeneCards)	15
ESR1_HUMAN.H11MO.1.A Half motif		ESR1 (GeneCards)	10

B

ERE full motif (PWMTools): 119,580

	γ H2AX	γ H2AX & A3B & ER	A3B & ER
Overlapping peaks	42	386	1756
% Peaks with motif	0.7%	41%	36%

ERE half motif (PWMTools, excluding full motifs): 167,227

	γ H2AX	γ H2AX & A3B & ER	A3B & ER
Overlapping peaks	39	41	482
% Peaks with motif	0.7%	4%	10%

Figure 5.15. Regions occupied by γ H2AX & A3B & ER, or A3B & ER, are enriched for the full-length ERE.

(A) ERE (ESR1) full and half-length motifs from the HOCOMOCO database.

(B) Number of ERE full and-half length motifs genome-wide (identified by PWMTools), and their overlap with γ H2AX, A3B and ER binding regions, in estrogen-treated MCF7 cells.

5.2.10 A3B and ER are enriched at promoters and enhancers in presence of estrogen

A3B and ER bind directly to promoters of some genes, but are also thought to regulate gene expression through interactions with enhancers. A3B was previously found to bind 7% of active promoters and 93% of active enhancers in MCF7 cells (Periyasamy *et al.*, 2015). The promoters were defined as ER binding sites overlapping with regions within TSS and 3000bp upstream, and the enhancers were defined as all other regions (Periyasamy *et al.*, 2015). In T-47D cells, A3B binding sites that overlap with R-loops, are highly enriched for TSS and transcribed enhancers (Zhang *et al.*, 2022).

Here, the overlap between A3B and ER in absence/presence of estrogen in MCF7 cells, in ChIP-seq files processed using more recent tools, was analysed. Promoters were defined by their epigenetic marks; H3K27ac and H3K4me3 for promoters, and H3K27ac and H3K4me1 for enhancers (Consortium, 2012). These epigenetic marks were specific for the MCF7 cell line, and A3B and ER signal was plotted across these regions. As expected, A3B and ER were found to be enriched near promoters and enhancers in presence of estrogen (figure 5.16). To validate this result, a list of MCF7 enhancer regions was obtained from EnhancerAtlas 2.0 (Gao and Qian, 2020). A3B and ER were also found to be enriched in presence of estrogen at these enhancer regions (figure 5.17).

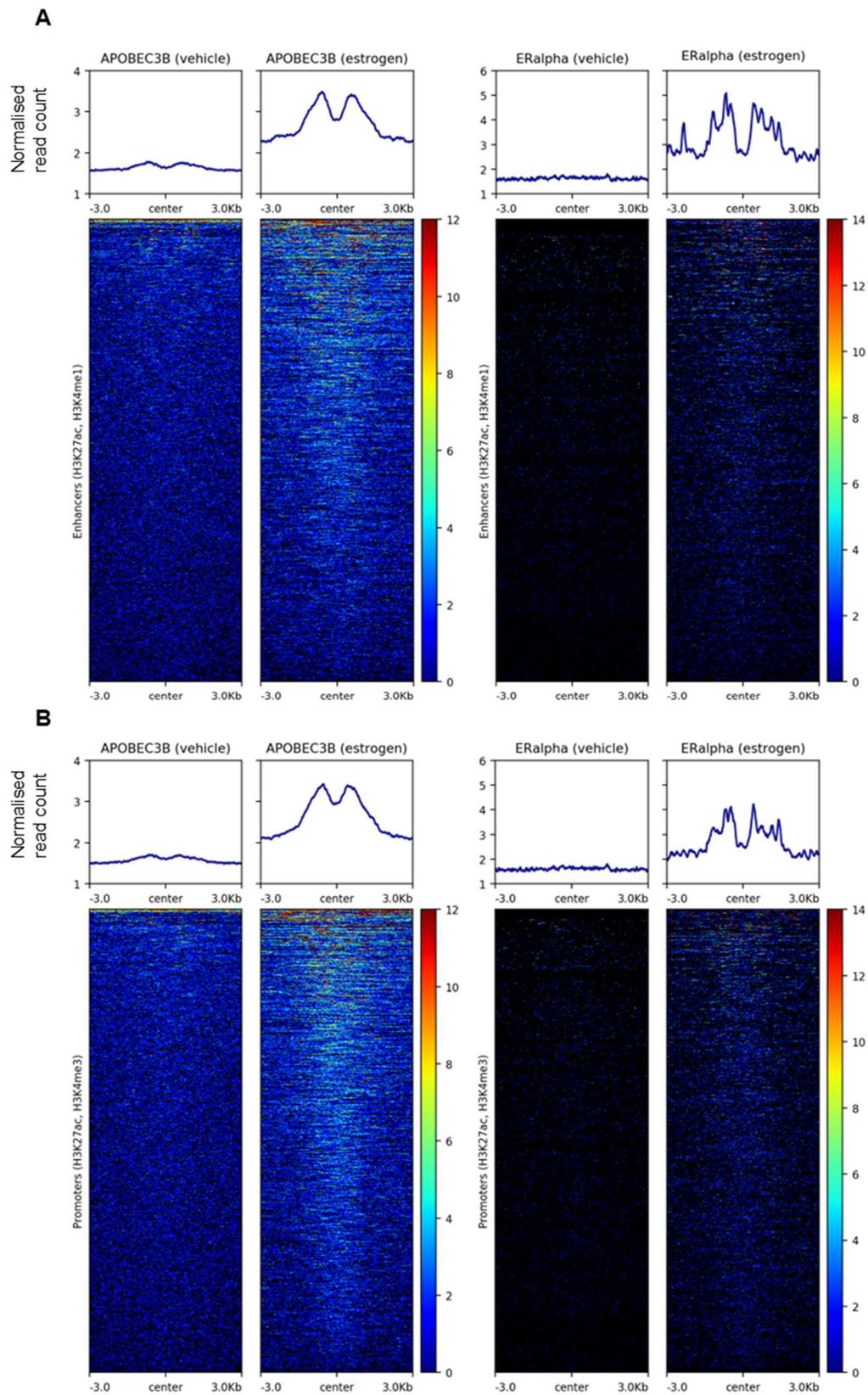


Figure 5.16. APOBEC3B and ER are enriched at enhancers and promoters in presence of estrogen in MCF7 cells.

- (A) Heatmap and summary plot show APOBEC3B and ER CHIP-seq signal at enhancers identified by presence of H3K27ac and H3K4me1.
- (B) Heatmap and summary plot show APOBEC3B and ER CHIP-seq signal at promoters identified by presence of H3K27ac and H3K4me3.

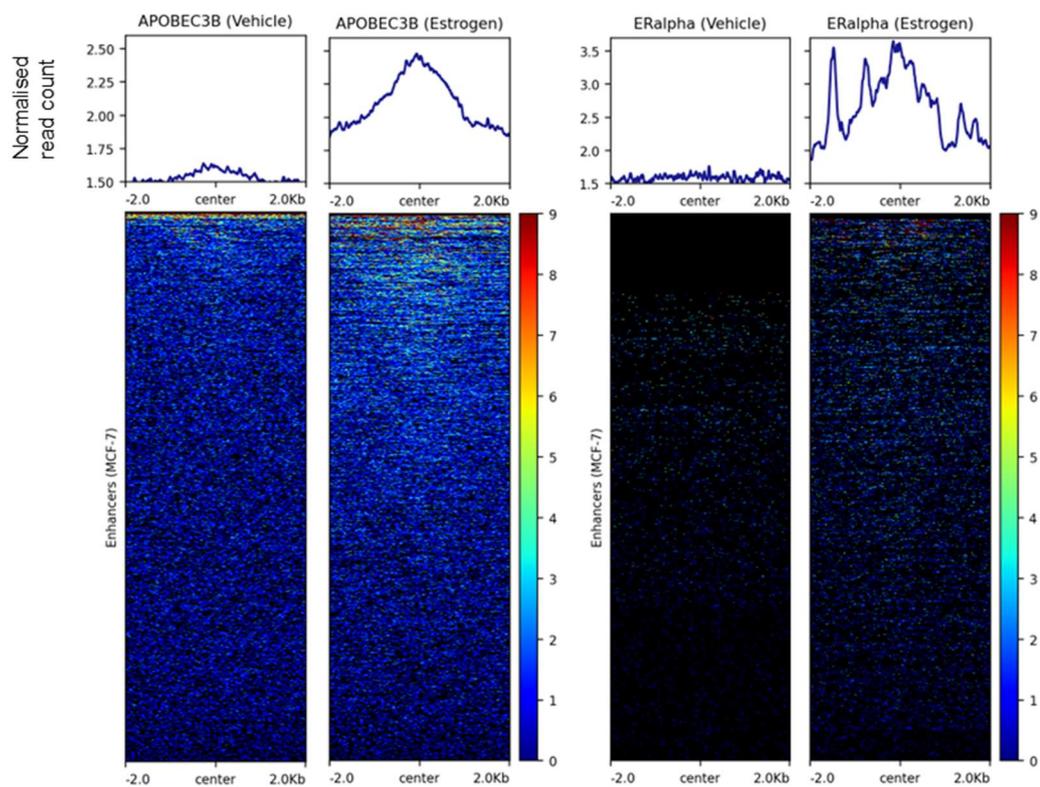


Figure 5.17. APOBEC3B and ER are enriched at enhancers in presence of estrogen in MCF7 cells. Heatmap and summary plot show APOBEC3B and ER CHIP-seq signal at MCF7 enhancer regions (EnhancerAtlas 2.0).

5.2.11 A3B-only/A3B & ER/ER-only bind genes belonging to similar biological pathways

A3B binds near promoters and is significantly enriched at TSSs, both with/without ER (figure 5.1 B, 5.10 B, 5.13). To investigate genes A3B and/or ER bind near, and the biological processes these genes regulate, functional enrichment analysis was performed using Metascape. Metascape was developed to provide analysis of “omics” datasets, integrating over 40 databases, for up-to-date gene annotation, interactome and functional enrichment analysis (Zhou *et al.*, 2019).

Lists of genes bound by A3B/ER were generated from HOMER peak annotation, where TSS regions were defined as -1kb to +100bp to the start of the gene. ER by itself bound the highest number of genes, followed by A3B & ER, with A3B by itself binding the fewest genes (figure 5.18 A). Very few genes were simultaneously bound by A3B-only, A3B & ER, and ER-only, suggesting that the groups regulate different gene sets, rather than simply binding at different regions within the same genes (figure 5.18 A). However, many of the genes bound by each group belonged to the same ontology terms (figure 5.18 A). This suggests that A3B (without ER), A3B and ER together, and ER (without A3B) are found at different genes, but some of these genes are involved in similar biological processes.

To further investigate the enriched ontology terms, Metascape clusters significantly enriched terms based on similarities between the genes these terms represent. The term with the lowest p-value within each cluster is selected as the representative term for that cluster of enriched terms, to reduce redundancy (Zhou *et al.*, 2019). A3B & ER bound promoters of genes significantly enriched for terms such as “Estrogen metabolism”, “PID ERA GENOMIC PATHWAY” (nuclear estrogen receptor alpha network) and “Estrogen receptor pathway” (figure 5.18 B). ER-only also binds promoters of “Estrogen receptor pathway”-genes, and “Estrogen dependent gene expression” (figure 5.18 B). It also has high enrichment for genes involved

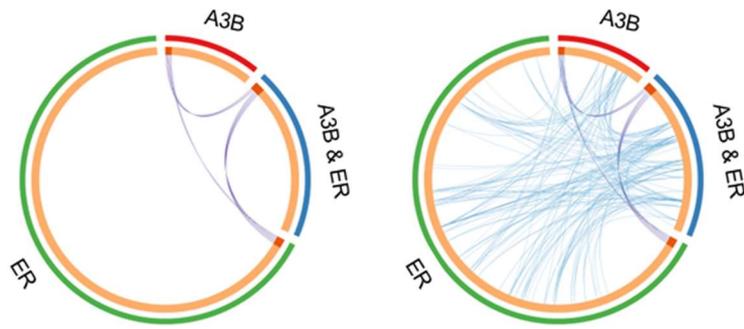
transcription/translation, such as “ribonucleoprotein complex biogenesis”, “translation”, “mRNA metabolic process” and “ncRNA metabolic process” (figure 5.18 B). Only two terms were significantly enriched for the genes bound by A3B-only; “carbohydrate synthetic process” and “response to hormone”, the latter shared with ER-only (figure 5.18 B). Only a subset of A3B binding sites were found at promoters, although these genes might be involved in response to estrogen (sub-group of “response to hormone”). A3B & ER, and ER-only both bound genes involved in estrogen response pathways, although the most enriched genes for ER-only were genes relating to transcription and translation.

To investigate which other transcription factors (TFs) may regulate these genes, Metascape performs Transcription Regulatory Relationships Unraveled by Sentence-based Text Mining (TRRUST). No other TFs were found to bind the promoters bound by A3B-only (figure 5.18 C). Promoters bound by A3B & ER, were recognised as regulated by ESR1 (ER) (figure 5.18 C). Interestingly, promoters bound by ER-only were not identified to be bound by ESR1/ER, further suggesting that ER binds (and regulates) its “typical” targets together with A3B. In addition, the pioneer factor FOXA1, which is important for ER regulation of target genes, was found to regulate the genes bound by A3B & ER, but not ER-only (figure 5.18 C). Promoters bound by A3B & ER were also regulated by the oncogenic TF JUN, and tumour suppressor BRCA1 (figure 5.18 C). Expression of AP-1 family member JUN is regulated by estrogen, and JUN subfamily member c-JUN is thought to recruit ER to DNA (indirectly, not through ERE), to regulate gene expression (Chiappetta *et al.*, 1992; He *et al.*, 2018). BRCA1 is an ER antagonist, inhibiting ER-mediated gene expression (Gorski *et al.*, 2009).

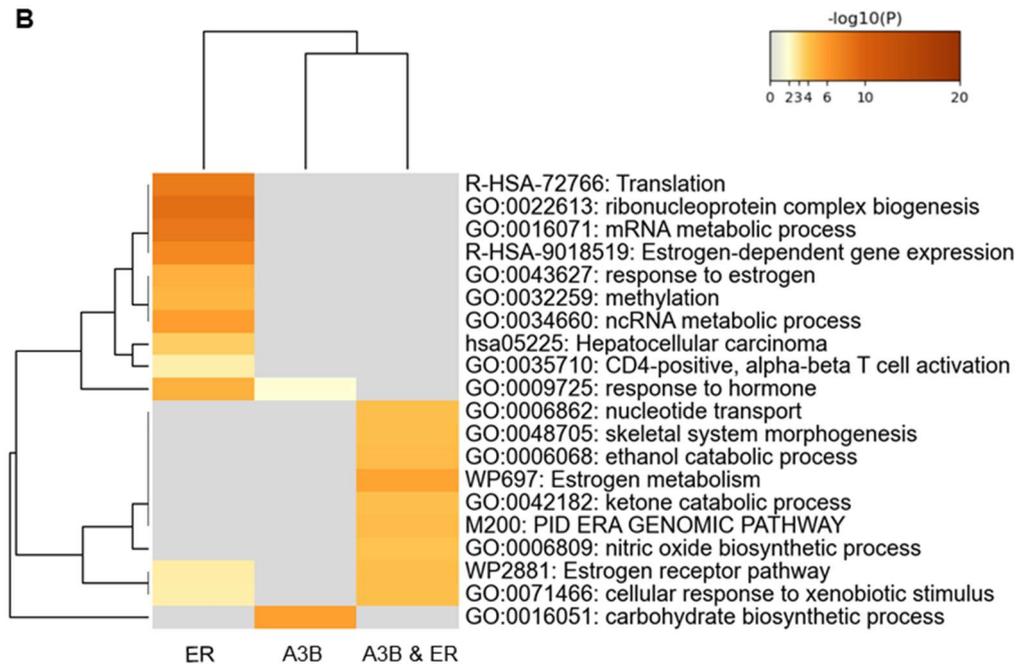
The genes occupied by ER-only were also regulated by BRCA1 (only shared TF), but the TFs with highest enrichment was tumour suppressor RB1, and transcriptional repressor ZNF382 (figure 5.18 C). RB1 is a negative regulator of the G1/S-phase of

the cell cycle, through mechanisms including (direct) transcriptional inhibition of genes required for the transition from G1 to S-phase (Giacinti and Giordano, 2006). ZNF382 encodes a KRAB domain zinc finger transcription factor (KZNF), is a pro-apoptotic suppressor, and represses AP-1 signalling pathway in the cell line HEK293 (Cheng *et al.*, 2010). Whereas ER (and A3B) regulate genes related to cell proliferation, some of the TFs identified by TRRUST for A3B & ER and ER-only binding genes may in fact be repressors their expression (BRCA1, RB1, ZNF382).

A



B



C

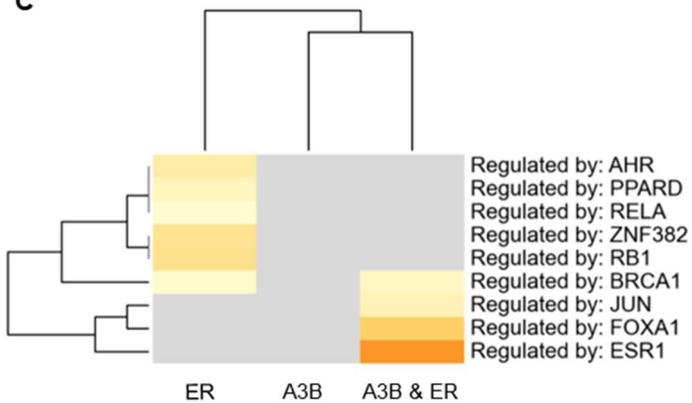


Figure 5.18. Genes bound by APOBEC3B & ER, and ER alone, in estrogen-treated MCF7 cells, belong to estrogen-related pathways.

- (A) List of genes (orange) bound by A3B alone, A3B & ER, and ER alone and their overlap (purple lines, left), and ontology terms shared between gene lists (blue lines, right).
- (B) Significantly enriched ontology terms for genes bound by A3B alone, A3B & ER, and ER alone, with dendograms displaying clustering of ontology terms, and of gene lists.
- (C) TRRUST analysis identifying transcription factors binding the genes bound by A3B alone, A3B & ER, and ER alone. Dendograms display clustering of transcription factors, and of gene lists.

Plots were generated using Metascape.

5.2.12 Functional enrichment analysis on genes whose enhancers are bound by A3B and ER

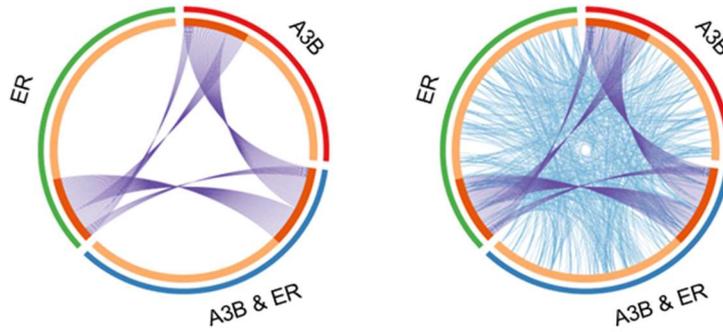
The majority of A3B and ER binding regions are found at gene distal regions, and A3B and ER are enriched at enhancers in presence of estrogen (figure 5.16, 5.17) (Periyasamy *et al.*, 2015; Welboren *et al.*, 2009). To investigate the enhancers bound by A3B and ER, and what genes they regulate, functional enrichment analysis was performed using Metascape. The list of enhancers was generated by downloading the enhancer list for MCF7 cells from EnhancerAtlas 2.0, and identifying overlapping regions with A3B-only, A3B & ER, and ER-only (Gao and Qian, 2020). The enhancer regions of interest were then annotated using EnhancerAtlas 2.0, to identify the genes regulated by these enhancers.

There was some overlap between the genes regulated by the enhancers bound by A3B-only, A3B & ER and ER-only (figure 5.19 A); however, many of these genes belong to the same ontology group (figure 5.19 A). Most of the enriched pathways of the genes whose enhancers were bound by A3B & ER, were also shared with ER-

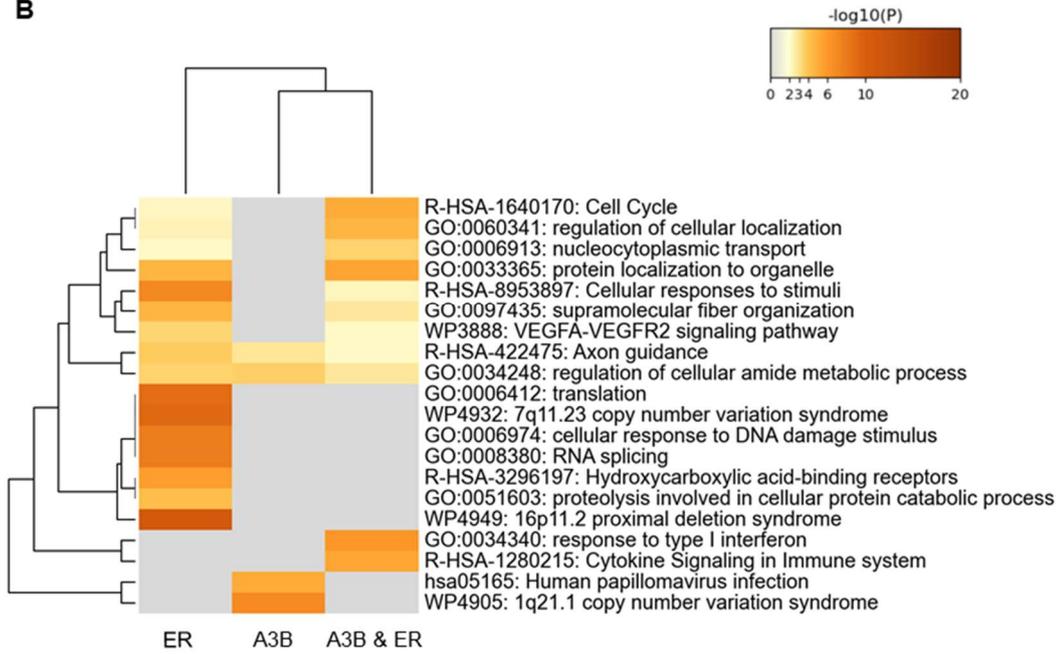
only (figure 5.19 B). These belonged to pathways such as “Cell cycle”, “protein localisation to organelle” and “regulation of cellular localization” (figure 5.19 B). The most significantly enriched ontology terms for genes whose enhancers were bound by ER-only, were “16p11.2 proximal deletion syndrome”, “translation” and “copy number variation syndrome” (figure 5.19 B). The most enriched ontology terms for the genes whose enhancers A3B-only bound, were “1q21.1 copy number variation syndrome” and “human papillomavirus infection” (figure 5.19 B).

The genes whose enhancers were bound by A3B & ER, were also regulated by BRCA1 and FOXA1, as were the genes directly bound by A3B & ER (figure 5.18 C, 5.19 C). These genes were also regulated by tumour suppressor p53, transcription factor SP1 and SP1-related SP3, metastasis-associated protein 1 (MTP1) and RELA/p65, a component of the transcription factor complex NF- κ B (figure 5.19 C).

A



B



C

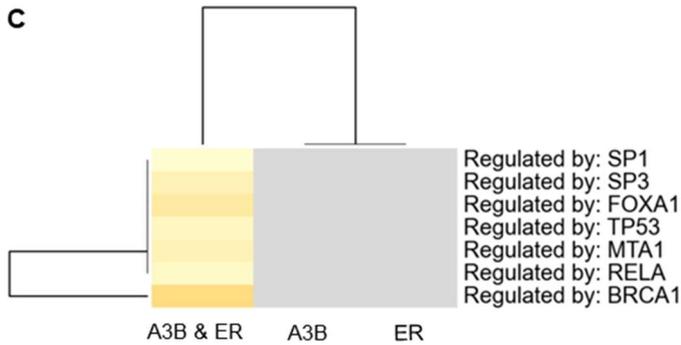


Figure 5.19. Genes regulated by enhancers bound by APOBEC3B & ER, and ER alone, in estrogen-treated MCF7 cells, belong to the same ontology terms.

- (A) List of genes (orange) regulated by enhancers bound by A3B alone, A3B & ER, and ER alone and their overlap (purple lines, left), and ontology terms shared between gene lists (blue lines, right).
- (B) Significantly enriched ontology terms for genes regulated by enhancers bound by A3B alone, A3B & ER, and ER alone, with dendograms displaying clustering of ontology terms, and of gene lists.
- (C) TRRUST analysis identifying transcription factors binding the genes regulated by enhancers bound by A3B alone, A3B & ER, and ER alone. Dendograms display clustering of transcription factors, and of gene lists.

Plots were generated using Metascape.

5.3 Discussion

A3B is required for full expression of estrogen-dependent ER target genes (Periyasamy *et al.*, 2015). In this chapter, ChIP-seq analysis was performed on previously published datasets, to verify their findings, and to further investigate which genes A3B may be regulating where it binds together with, or independently of, ER.

5.3.1 Re-analysis of A3B and ER ChIP-seq datasets previously analysed by Periyasamy *et al*

Bioinformatic tools which have been developed or updated since the publication of Periyasamy *et al* were used to re-analyse their ChIP-seq datasets and to compare findings from the original publication to those generated using these more sophisticated tools (Periyasamy *et al.*, 2015). There was an overall agreement in findings; however, analysis with new tools offered several novel insights. Periyasamy

et al identified between 1-3 fold more A3B and ER binding regions using the peak caller MACS, than what was identified here using MACS2. This is likely caused by changes in the updated tool, including the ability to specify an FDR value, rather than just a p-value for MACS. Whereas the p-value gives the probability of a false positive on a *single* test, the peak caller will perform thousands of statistical tests as the statistical significance of enrichment is calculated across the entire genome (Zhang *et al.*, 2008). Because the number of false positives will add up for each statistical test that is performed, it will overall lead to discovery of peaks which are false-positives. MACS2 allows the user to specify the FDR instead, defining the number of significant results which will result in false positives. This should reduce the number of false positive peaks discovered.

Another difference between the results generated by Periyasamy *et al*, and results generated here, is the proportion of regions co-occupied by A3B, ER and γ H2AX. Periyasamy *et al* reported that over half of the sites bound by A3B and ER also displayed γ H2AX; the analysis presented here found that only around 15% of the regions occupied by A3B and ER overlap with γ H2AX. Although an increase in γ H2AX was seen in presence of estrogen, in agreement with previous reports (Stork *et al.*, 2016; Zhang *et al.*, 2022), it puts into question whether the generation of DSBs by A3B, and subsequent γ H2AX formation, is an essential part of A3B's role in regulating the estrogen response. A dependence on DSB induction for the expression of estrogen-target genes also comes with risks, as erroneous repair of these DSBs could lead to introduction of mutations. ER activation led to A3B recruitment to EREs, but these results suggest that γ H2AX formation at these sites is not essential.

5.3.2 γ H2AX forms narrow peaks in response to A3B/estrogen-induced DNA damage

Surprisingly, the estrogen-induced γ H2AX formed narrow peaks, both at A3B/ER binding regions, and regions independent of A3B/ER. This is different to the megabase-wide γ H2AX domains typically seen forming around DSBs (Rogakou *et al.*, 1999; Iacovoni *et al.*, 2010; Massip *et al.*, 2010). In the MCF7 and T-47D ER+ breast cancer cell lines, γ H2AX is formed by A3B deamination and R-loop resolving (Periyasamy *et al.*, 2015; Stork *et al.*, 2016; Zhang *et al.*, 2022). Zhang *et al.* reported that around 30% of DSBs detected in T-47D cells were estrogen-induced. Of these, around a third were A3B-dependent, although these were only formed following R-loop formation (Zhang *et al.*, 2022). Periyasamy *et al.* reported that A3B cytidine deamination leads to activation of the BER repair pathway. The H2AX glu141 residue becomes ADP-ribosylated upon DNA damage, leading to recruitment of BER factors (Chen *et al.*, 2021). Loss of H2AX glu141 ADP-ribosylation leads to formation of γ H2AX instead, and downstream recruitment of DSB response factors, rather than BER activation (Chen *et al.*, 2021). However, estrogen induction initiates formation of γ H2AX (Stork *et al.*, 2016; Periyasamy *et al.*, 2015; Zhang *et al.*, 2022). The repair method used to repair these DSBs may influence the γ H2AX spread. Whereas HR can take several hours to repair a DSB, NHEJ is much faster and repairs breaks in approximately 30 minutes (Mao *et al.*, 2008). The narrow γ H2AX peaks may be a result of the DSBs being repaired quickly by NHEJ, before γ H2AX has time to spread to megabase domains.

5.3.3 Disagreement in A3B binding regions discovered by the peak callers MACS2 and LanceOtron

A3B, ER and γ H2AX binding regions were defined using the peak caller MACS2, and also using the novel peak caller LanceOtron. There was large agreement between the binding regions identified by LanceOtron and MACS2 for ER and γ H2AX, however this was not the case for A3B. In general, MACS2 performs better when identifying broad binding regions than LanceOtron (Hentges *et al.*, 2022); however, MACS2 was used with the settings for discovery of narrow peaks. The peaks identified by LanceOtron-only may be broad peaks, and therefore not discovered by MACS2, although the A3B peaks identified by MACS2 and LanceOtron located near estrogen-response genes and displayed narrow peaks, as expected (Periyasamy *et al.*, 2015). Hentges *et al* found that peaks called by LanceOtron, but missed by MACS2, were often located in areas of open chromatin, with high background signal; because of this, even strongly enriched peaks are frequently excluded by MACS2 (Hentges *et al.*, 2022; Auerbach *et al.*, 2009). It would be interesting to investigate the A3B binding regions identified by LanceOtron exclusively, and whether these are also found near ER binding sites and estrogen target genes, or near genes not regulated by estrogen. Whereas over half of the A3B binding sites in estrogen-induced cells identified by LanceOtron were also discovered by MACS2, only a small proportion of A3B binding sites in absence of estrogen were shared between both peak callers. Because few of these peaks could be robustly identified by both peak callers, it raises the question of whether the majority of A3B binding sites in absence of estrogen are in fact noise. This is supported by these binding regions displaying lower peak scores compared to the other datasets. Further replicates investigating A3B binding in absence of estrogen in MCF7 cells would determine whether A3B binds these regions consistently, suggesting this binding may serve a function, or whether the regions are random and likely to be noise.

5.3.4 A3B binds genes related to the estrogen pathway independent of ER

Around half of the estrogen-induced A3B binding sites are independent of ER. However, these binding sites are enriched for genes related to the estrogen response pathway, and are enriched for the half-length ERE motif, which suggests that ER might also be involved in the recruitment of A3B to these regions. One possibility is that ER binds transiently at the half-length EREs, which may be sufficient to recruit A3B, but below detection-level for ER by ChIP-seq. Zhang *et al* reported that the oncogene and transcriptional activator TDRD3 is involved in estrogen-induced A3B recruitment to R-loops. A3B may be recruited to these ER-independent binding sites by TDRD3, and it would be interesting to investigate whether these binding sites overlap with estrogen-induced R-loops.

When investigating the genes regulated by enhancers bound by A3B and ER, no estrogen-related genes were found to be regulated by these enhancers. However, these results rely on the enhancer file used to detect overlaps with A3B and ER binding sites being up to date, and EnhancerAtlas being able to recognise and annotate the gene targets of these enhancers. There may therefore be enhancers bound by A3B and ER, and gene targets, which were not identified by this analysis. Interestingly, enhancers bound by A3B-only regulate genes involved in human papillomavirus (HPV) infection, during which A3B is strongly upregulated (Warren *et al.*, 2017). It would be interesting to investigate these HPV-related genes whose enhancers are bound by A3B further, and whether A3B plays a role in regulating their expression during HPV infection.

A3B appears to play an important role in regulating cell proliferation in ER+ breast cancers. siRNA-mediated knockdown of A3B reduces growth of both MCF7 and T-47D tumours (Periyasamy *et al.*, 2015). This is likely linked to A3B's role in expression of ER-target genes, of which many are involved in cell proliferation (Yang *et al.*, 2017). Interestingly, CRISPR-Cas9-mediated A3B KO human keratinocytes (which are not

regulated by estrogen) are nonviable, suggesting that A3B may also regulate cell proliferation through processes independent of estrogen (Dr Tim Fenton, personal correspondence). Further investigations into A3B binding at gene regulatory regions, including cells not regulated by ER, would increase the understanding of A3B's role in regulating gene expression.

5.3.5 Further steps

There are several directions in which this work could be extended. The first would be to investigate the expression of genes which are found to be bound by A3B, particularly those where ER is not also detected. This could be done using RNA-seq in WT and siRNA-mediated A3B knockdown MCF7 cells; global transcriptomic investigation in this context and manner has not been performed before, and it would give a detailed picture of A3B's role in gene regulation both in the presence and absence ER. Zhang *et al* found that, in T-47D cells, 87% of estrogen-regulated genes require the BER processes, downstream of A3B recruitment, for complete and appropriate regulation upon estrogen treatment (Zhang *et al.*, 2022). Investigating global transcriptomic regulation in siRNA-mediated A3B knockdown cells would reveal whether these effects are indeed dependent on A3B.

All experiments in Periyasamy *et al* and Zhang *et al* were conducted in bulk populations of cells; this means that it is impossible to detect whether A3B and ER are recruited simultaneously in individual cells, or whether they each bind the same loci but in separate cells. The second direction would be to repeat analysis of A3B and ER recruitment in single cells, for example, using ChIP-re-ChIP to investigate whether ER and A3B bind simultaneously. ChIP-re-ChIP is a technique used to investigate protein co-occupancy on DNA at specific genomic regions (Beischlag, Prefontaine and Hankinson, 2018). It involves performing an initial ChIP experiment

to isolate the target protein-DNA complexes, followed by a second round of CHIP using antibodies against a different protein of interest (Beischlag, Prefontaine and Hankinson, 2018). This enables the study of protein-protein interactions and their co-localisation on the same genomic regions. This is particularly interesting for distinguishing binding sites at which proteins bind individually versus in pairs, for example those bound by A3B and ER together/separately.

6.0 Chapter 6: Discussion

6.1 Thesis outline

This thesis is composed of three results chapters. The first explores the use of a recently developed technique for assaying protein-chromatin interactions (CUT&Tag) in investigating DNA damage, and in particular the γ H2AX histone modification. This chapter shows that CUT&Tag has significant advantages in terms of cost and cell number requirements when compared to more commonly used assays, and furthermore, demonstrates that CUT&Tag offers greater resolution despite using fewer resources.

The second chapter explores the deposition of the di- γ H2AX histone modification following DNA damage. More specifically, it addresses the H2AX phosphorylation events which occur in cells upon DNA damage, which are hypothesised to play a role in determining cell survival versus apoptosis. However, the antibody used in experiments throughout this chapter was found to be non-specific, making the results obtained difficult to interpret.

The final results chapter investigates the role of A3B in regulating gene expression. In particular, it probes A3B's contribution to gene regulation following estrogen treatment, and shows that A3B is predominantly recruited to activating elements such as enhancers and promoters, that there is significant overlap in A3B and ER binding sites, and that A3B and ER bind near genes associated with common biological processes.

A more thorough discussion of each chapter is included below.

6.2 CUT&Tag detects γ H2AX formation in response to DNA damage using fewer cells and with less background signal than ChIP-seq

- CUT&Tag can be used to detect γ H2AX formation around DSBs, providing near-identical results to γ H2AX ChIP-seq, but with fewer sequencing reads and cell input requirements
- The γ H2AX signal at the site of the DSB differs between CUT&Tag and ChIP, where a dip in signal is seen with ChIP, and a peak in signal is detected by CUT&Tag

CUT&Tag is a technique for investigation of protein-chromatin interactions (Kaya-Okur *et al.*, 2019). ChIP-seq has been the “gold standard” for many years, but CUT&Tag offers many advantages, by requiring fewer cells, and providing better signal-to-noise ratios. Whereas ChIP-seq requires 5-20 million cells and 10-50 million paired-end sequencing reads to detect enrichment, CUT&Tag requires at least 10-fold fewer cells and fewer sequencing reads. 2 million reads from a CUT&Tag experiment is equivalent to 20 million ChIP-seq reads (Kaya-Okur *et al.*, 2019). Here, CUT&Tag sequencing reads here were between 2-7 million.

To allow investigation of H2AX phosphorylation near DSBs, the D1vA cell line was utilised, where DSBs are induced at around 200 known sites by the restriction endonuclease AsiSI. Immunostaining of γ H2AX indicated an increase in γ H2AX foci, as expected, which verified induction of DNA damage. CUT&Tag was used to assess formation of γ H2AX around DSBs, and compared to previously published γ H2AX ChIP-seq. As expected, γ H2AX could be detected in the vicinity of DSBs by CUT&Tag, as seen by both qPCR and analysis of paired-end sequencing data. However, the γ H2AX signal directly at the site of the DSB differ between CUT&Tag and ChIP-seq. The γ H2AX signal generated by ChIP-seq is depleted in the immediate area surrounding the DSB, likely due to chromatin remodelling during DSB end processing evicting histones in the vicinity of the DSB (Jeggo and Downs, 2014).

Surprisingly, CUT&Tag results indicated that γ H2AX is enriched directly over DSBs. An unexpected enrichment in the vicinity of a DSB was also seen for H3K27me3 by qPCR. During DSB repair by the HR pathway, DNA resection creates 3' single-stranded overhangs (San Filippo, Sung and Klein, 2008). Recently, Liu *et al* discovered that the MRN complex recruits RNA polymerase III (RNA pol III) to DSBs. Here, RNA pol III synthesises RNA strands, which form DNA-RNA hybrids with the 3' overhangs (figure 6.1) (Liu *et al.*, 2021). This is an early event in the HR pathway, and is thought to have a protective effect on the single strand 3' overhangs, reducing the risk of genetic loss (Liu *et al.*, 2021). Interestingly, the Tn5 transposome used in CUT&Tag is capable of binding and tagmenting DNA-RNA hybrids, in addition to dsDNA (Lu *et al.*, 2020). The peak in γ H2AX at the site of DSBs could be a result of Tn5 binding and tagmenting RNA-DNA hybrids at DSBs repaired by HR (figure 6.1). However, histones are not likely to be present at these sites, as DSB end processing involves histone ejection. The recruitment of protein A-Tn5 fusion protein to these sites during CUT&Tag is therefore unclear. Repeating the CUT&Tag under similar conditions including RNase H, which degrades RNA of RNA-DNA hybrids, will allow investigation of whether the increased CUT&Tag signals at DSBs are caused by Tn5 tagmentation of RNA-DNA hybrids formed at DSBs.

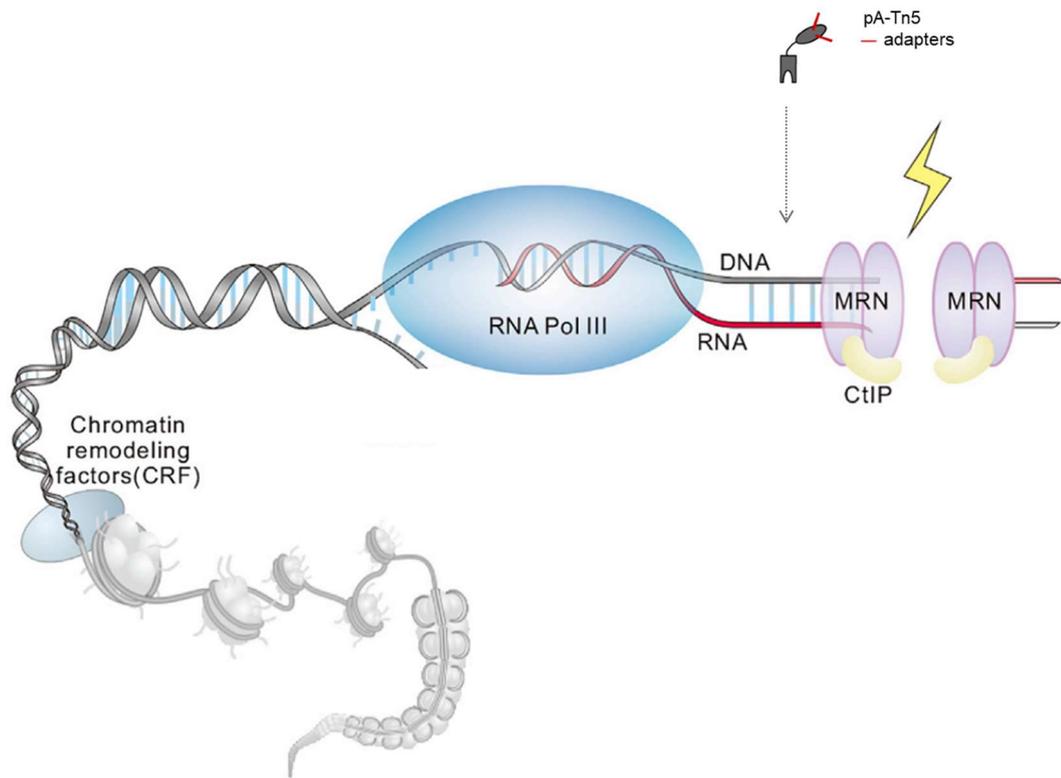


Figure 6.1. In homologous repair-mediated DSB repair, RNA polymerase III synthesises an RNA strand forming an RNA-DNA hybrid structure, which the Tn5 transposome may target (adapted from Liu *et al.*, 2021 and Kaya-Okur *et al.*, 2019). In response to DSBs (lightning bolt), nucleosomes are displaced by chromatin remodelling factors, to increase access of repair factors to DNA. The MRN complex and CtIP are recruited, leading to end-resection of the DNA strands, forming 3' ssDNA overhangs. RNA polymerase III is recruited by MRN, and synthesise an RNA strand, which forms an RNA-DNA hybrid, protecting the 3' overhang. Tn5 transposome (which in CUT&Tag is tethered to protein A) is capable of binding and tagmenting RNA-DNA hybrids, and may target the RNA-DNA hybrids formed near DSBs during homologous recombination.

Overall, there was a remarkable similarity in the megabase γ H2AX domains formed at DSBs. These domains form a unique shape at each DSB; they are asynchronous (with regard to the site of the DSB) as the γ H2AX boundaries are defined by TAD boundaries, and the γ H2AX signal is discontinuous within each domain, as γ H2AX is

depleted at sites occupied by RNA pol II (Arnould *et al.*, 2021; Iacovoni *et al.*, 2010). Similar γ H2AX domain span size, and γ H2AX peaks/valleys within domains, was detected using both techniques. This highlights the fact that the underlying chromatin state influences the γ H2AX signal, as it can be replicated at a different time, using a different technique. In addition, CUT&Tag detected chromosome-wide significant enrichment of γ H2AX on sex chromosomes and H3K27me3 on autosomes in mouse testis, as expected.

6.3 The novel di- γ H2AX antibody is not specific, and binds other targets than H2AX

- Dot blots suggested that the novel di- γ H2AX antibody is specific to H2AX C-terminal peptides doubly phosphorylated at the ser139 and tyr142 residues, with no cross-reactivity to mono-phosphorylated H2AX
- The di- γ H2AX antibody binds targets in cells lacking H2AX, indicating that the antibody is unspecific, and capable of binding non-H2AX proteins

H2AX is rapidly phosphorylated at ser139 in response to DNA damage, which instigates recruitment of DNA repair factors, leading to repair of the break site (Rogakou *et al.*, 1998; Rogakou *et al.*, 1999). H2AX is also phosphorylated at another residue, tyr142, under steady-state conditions, and this phosphorylation mark is removed upon DNA damage (Krishnan *et al.*, 2009; Cook *et al.*, 2009; Xiao *et al.*, 2009). H2AX can also be doubly phosphorylated at both ser139 and tyr142 residues – forming di- γ H2AX – which may be an intermediate step in the DNA damage response, on H2AX which has not yet lost its tyr142 phosphorylation mark, but has been ser139-phosphorylated (Nowsheen *et al.*, 2018; Cook *et al.*, 2009).

Di- γ H2AX foci have previously been detected in DNA damaged cells, resembling DNA-damage induced γ H2AX foci (Singh *et al.*, 2012). However, the antibody used

showed cross-reactivity with γ H2AX, and it is therefore unclear whether these foci represent γ H2AX or di- γ H2AX. Here, a novel di- γ H2AX antibody was used, which showed no cross-reactivity with other H2AX phosphorylation forms when tested by dot blot using H2AX C-terminal tail peptides with different phosphorylation marks. This antibody also detected di- γ H2AX foci in DNA damaged cells, although this was only seen when staining for di- γ H2AX alone, and not when co-staining for γ H2AX. However, both the novel di- γ H2AX antibody and a commercial H2AX tyr142ph antibody detected proteins in an H2AX KO cell line by western blot, indicating that both antibodies are capable of binding non-H2AX proteins. The H2AX tyr142ph-antibody was raised against a peptide corresponding to H2AX residue 100-142. H2A is 12 residues shorter than H2AX, but the first 123 residues share 95% similarity with H2AX, and the H2AX tyr142ph antibody may be capable of interacting with H2A.

Because the di- γ H2AX antibody used here is unspecific, the results generated using this antibody must be disregarded. Similarly, the results generated by Singh *et al* using another di- γ H2AX antibody with cross-reactivity to γ H2AX must be interpreted with caution. Thus, the role of di- γ H2AX in the DNA damage response remains unclear. Both H2AX tyr142 phosphorylation, and tyr142 dephosphorylation upon DNA damage, are important for the initiation of γ H2AX formation and the downstream DNA damage response. Interestingly, H2AX ser139 and tyr142 phosphorylation also play a role in erythropoiesis, specifically in the terminal erythroid maturation step, where condensation of chromatin is followed by enucleation (nucleus extrusion). During this process, organelles are lost, through Caspase-3 activity, which resembles similar events during apoptosis (Carlile, Smith and Wiedmann, 2004). In mice lacking H2AX, the enucleation process is perturbed, although these mice are still viable (Zhao *et al.*, 2016; Celeste *et al.*, 2002). During the terminal erythroid maturation step, γ H2AX levels increase, although it does not appear to correlate with high levels of DNA damage, but might be linked to the suggested role for γ H2AX marking active

replication forks (Jeffery *et al.*, 2021; Dhuppar, Roy and Mazumder, 2020). Interestingly, erythroid cells express high levels of WSTF, and the high γ H2AX levels coincide with H2AX tyr142ph (Jeffery *et al.*, 2021). Whether this leads to formation of di- γ H2AX, or whether H2AX is solely monophosphorylated on ser139 and tyr142 is not known (Jeffery *et al.*, 2021). However, a loss of WSTF led to inhibition of terminal erythroid maturation and reduced Caspase-3 activation of downstream factors in this apoptosis-like pathway (Jeffery *et al.*, 2021). A possible explanation for this is that di- γ H2AX is present, and able to interact with apoptosis-related factors also found in erythrocytes, to initiate terminal maturation (Jeffery *et al.*, 2021).

The hypothesised role for di- γ H2AX in initiating apoptosis raises many questions. The early DNA damage response factor MDC1 is unable to bind di- γ H2AX, and would not be able to recruit downstream DNA damage factors until the tyr142 phosphorylation mark is removed. Also, di- γ H2AX is hypothesised to act as a pro-apoptotic mark, but the tyr142 phosphorylation mark is removed rapidly upon DNA damage, leaving the pro-repair/survival mark γ H2AX (Cook *et al.*, 2009). Thus, how the di- γ H2AX is sustained in DNA damaged cells destined for apoptosis is unclear (Lukas and Bartek, 2009).

There have been many attempts at investigating di- γ H2AX and its role in DNA damaged cells; however, antibodies with cross-reactivity to other phosphorylation forms/histones has been a recurring problem. Detection of H2AX tyr142 phosphorylation by mass spectrometry was unsuccessful, likely due to the histone mark being very low abundance, and detection of di- γ H2AX by this technique might be challenging for the same reason (Hatimy *et al.*, 2015). In addition, phosphopeptides are challenging to detect by mass spectrometry due to the intrinsic negative charge of the phosphate meaning they do not ionise well (which is required for their detection by the mass spectrometer). It would therefore be expected that a

diphospho-peptide would be even harder to detect by mass spectrometry, than a mono-phosphorylated peptide.

6.4 Estrogen recruits APOBEC3B to enhancers and promoters, where it binds genes involved in the estrogen response pathway

- In estrogen-induced cells, A3B recruitment is mostly detected at introns and intergenic regions, and is enriched at promoters and enhancers. A3B binding regions overlap significantly with TSSs
- Around 50% of A3B binding sites are co-occupied by ER, and A3B binds near genes involved in the estrogen response pathway both with/without ER
- A3B and ER bind genes associated with similar biological pathways, but the genes bound by A3B & ER or ER-only are co-occupied by different TFs

Estrogen induces expression of thousands of genes, through activation of the transcriptional regulator ER. Interestingly, full expression of many estrogen-target genes relies on A3B, and its cytidine deaminase activity (Periyasamy *et al.*, 2015). A3B binds at EREs together with ER, and through deamination of cytosines to uracil, and subsequent DNA repair at these sites through the BER pathway, transcription of these genes is facilitated (Periyasamy *et al.*, 2015). Several DNA repair factors appear to be involved in transcriptional regulation. γ H2AX leads to recruitment SWI/SNF chromatin remodelling complex catalytic subunit BRG1, which is also involved in activating transcription (Lee *et al.*, 2010; Kadam and Emerson, 2003). DNA repair factors are also thought to relieve torsional stress generated by supercoiling of DNA during the process of transcription (Gerasimova *et al.*, 2016). The estrogen-induced increase in transcription leads to formation of R-loops (Stork *et al.*, 2016; Zhang *et al.*, 2022). A3B binds ssDNA substrates, and many of the A3B

binding sites in presence of estrogen are at the ssDNA opposite of the RNA-DNA hybrid structure at sites known to form R-loops (Zhang *et al.*, 2022).

A3B, ER and γ H2AX ChIP-seq files generated by Periyasamy *et al* were analysed using bioinformatic tools updated or developed since the original findings were published. In agreement with previous findings, a large increase in A3B binding events was detected in estrogen-induced cells, around half of which overlap with ER binding regions. A3B binding sites include several known estrogen-target genes. In line with previous findings that A3B deamination leads to γ H2AX formation (marking DSBs), an increase in γ H2AX was seen at some – but not all – of these genes. In fact, only a minority of γ H2AX binding sites overlap with regions co-occupied by A3B and ER. A3B-mediated cytidine deamination and subsequent BER was previously found to be required for full expression of estrogen-target genes, and estrogen leads to an increase in DSBs (Stork *et al.*, 2016; Zhang *et al.*, 2022). However, a reliance on DNA damage induction for expression of (potentially) thousands of genes, where use of error-prone DNA repair mechanisms could introduce mutations in genes involved in regulating processes which may be harmful if dysregulated (including cell proliferation) is a very high risk method of regulating gene expression (Yang *et al.*, 2017).

Around 50% of A3B binding sites in estrogen-induced cells overlap with ER binding sites, in agreement with previous findings (Periyasamy *et al.*, 2015). As expected, sites co-occupied by A3B & ER were enriched for the half- and full length palindromic ERE motif. Interestingly, sites bound by A3B-only were enriched for the half-length ERE motif, although ER was not detected at these sites. A transient ER binding to these half-length EREs could potentially be sufficient to recruit A3B to these sites, without binding stably enough to be detected by ChIP-seq. Both the half and full-length EREs contain a cytosine preceded by thymine and followed by adenine (TCA), which is favoured for A3B cytidine deamination (Zou *et al.*, 2017).

In presence of estrogen, the majority of A3B binding regions are located at introns, followed by exons and promoters, as previously discovered (Periyasamy *et al.*, 2015). Interestingly, both binding sites occupied by A3B-only and A3B & ER overlap significantly with TSS, suggesting a role for A3B in transcriptional regulation also at sites where ER occupancy is not detected. Moreover, both genes bound by A3B-only and A3B & ER together are enriched for estrogen-response related pathways. The enrichment for ERE motifs, and binding at genes involved in estrogen-related pathways, suggests that A3B-only binding sites might also be involved in regulation of estrogen-target gene expression.

Only around 20% of ER binding sites overlap with A3B, in presence of estrogen, and the genes bound by ER-only and A3B & ER differ. Although both ER-only and A3B & ER bound genes are involved in the estrogen response pathway, ER-only also bound genes involved in transcription/translation. Interestingly, genes bound by A3B & ER were recognised as transcriptionally regulated by ER, in addition to the pioneer factor FOXA1 and oncogenic TF JUN. The fact that A3B & ER binding regions exhibit significant enrichment for ERE motifs and estrogen-regulated genes further supports the hypothesis that expression of ER's well-characterised targets is dependent on A3B recruitment (Periyasamy *et al.*, 2015). Genes bound by ER-only were recognised as bound by TFs involved in transcriptional inhibition, but not recognised as ER-regulated genes, suggesting that ER's involvement in regulation of these genes is less studied.

Both A3B and ER were enriched at promoters and enhancers in estrogen-induced MCF7 cells, in agreement with previously reported results (Periyasamy *et al.*, 2015). A3B was recently found to bind R-loops in an estrogen-dependent manner in the ER+ T-47D breast cancer cell line (Zhang *et al.*, 2022). It would be interesting to investigate whether the A3B binding sites found here overlap with R-loops. A3B only binds ssDNA; however, ER does not (Mason *et al.*, 2010). Stork *et al* suggests a model

wherein ER recruits A3B indirectly, from binding sites adjacent to the R-loop. However, many of the A3B/ER binding sites identified here and by Periyasamy *et al* are directly overlapping, and how/whether these sites serve as single-stranded substrates for A3B is unknown. A3B-only binding sites are enriched for EREs and estrogen-response related pathways, and these sites may coincide with R-loop formation. However, it is important to note that these CHIP-seq experiments are performed on pools of cells, and temporal co-occupancy at binding sites can only be determined by using single-cell techniques.

It would be interesting to investigate whether A3B is also a co-activator for other transcription factors, or whether this mechanism has evolved for regulation of ER target genes specifically. Androgen receptor (AR) is another hormone-regulated transcriptional regulator, which, upon testosterone binding, translocates to the nucleus where it regulates expression of its target genes. ER and AR evolved through duplications of an ancestral steroid receptor (Baker, 1997). CHIP-seq experiments in cells regulated by AR would reveal whether A3B binding sites overlap with AR binding sites, and whether A3B may play a role in regulation of AR-target genes as well.

ER+ breast cancers exhibit high levels of C>T transitions and C>G transversions at TpC dinucleotides due to cytidine deamination (Nik-Zainal *et al.*, 2012; Alexandrov *et al.*, 2013; Burns *et al.*, 2013). This is accompanied by high expression of A3B, which has been proposed to be a key driver of these mutations (Burns *et al.*, 2013; Taylor *et al.*, 2013). New A3B binding sites have been identified, where A3B binds together with ER to regulate estrogen-target gene expression (Periyasamy *et al.*, 2015; Zhang *et al.*, 2022). It would be interesting to investigate whether the cytidine deaminase-associated mutational landscapes seen in breast cancer patients is enriched at regions bound by A3B and ER in these cell lines. This would suggest that regions bound by A3B, which is necessary for full activation of ER-regulated genes, can also serve as substrates resulting in high levels of somatic mutations seen in breast

cancer. In addition, a dependence on A3B cytidine deamination for expression of estrogen target genes, would be expected to result in C-T/G mutations in the vicinity of A3B/ER binding sites, possibly leading to depletion of TC sites. It would be interesting to compare these genomic regions in humans, with a closely related species not expressing A3B, such as mice. Unlike humans, mice only express one A3, which can deaminate cytosines preceded by thymine, cytosine or guanine (TC, CC or GC) (Renard *et al.*, 2010). Even if mouse A3 is involved in regulation of ER-target gene expression, a TC-specific depletion would not be expected to be seen. The dependence on cytidine deamination for transcriptional regulation of estrogen-target genes, which can lead to TC depletion, will result in a loss of the A3B substrates required for the transcriptional regulation. However, TC depletion will only take place if repair of the deaminated cytosine does not take place prior to replication. These processes may be kept separate in the cell cycle, and it would be interesting to investigate further whether the involvement of A3B in estrogen-target gene expression is avoided during S-phase. However, BER-factors are commonly mutated in various cancers, which could affect the repair of cytidine deamination by A3B, leading to accumulation of C-T/G mutations in cancers (Wallace, Murphy and Sweasy, 2012). Interestingly, many people in South-East Asia have no A3B protein due to the A3A_B deletion polymorphism (Kidd *et al.*, 2007). Due to deletions, the A3A mRNA contains the A3B 3'UTR, and is therefore regulated differently (Kidd *et al.*, 2007). It would be interesting to investigate whether another A3 compensates for the lack of A3B in regulation of estrogen-target gene expression, or whether A3A_B is capable of interacting with ER and EREs. How ER-regulated gene expression is affected in individuals missing A3B, and how the A3B substrates are reserved through evolution requires further investigation.

7.0 References

- Abe, H., Alavattam, K. G., Hu, Y. C., Pang, Q., Andreassen, P. R., Hegde, R. S. and Namekawa, S. H. (2020) 'The Initiation of Meiotic Sex Chromosome Inactivation Sequesters DNA Damage Signaling from Autosomes in Mouse Spermatogenesis', *Curr Biol*, 30(3), pp. 408-420.e5.
- Aguilera, A. and García-Muse, T. (2012) 'R loops: from transcription byproducts to threats to genome stability', *Mol Cell*, 46(2), pp. 115-24.
- Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., Bignell, G. R., Bolli, N., Borg, A., Børresen-Dale, A. L., Boyault, S., Burkhardt, B., Butler, A. P., Caldas, C., Davies, H. R., Desmedt, C., Eils, R., Eyfjörd, J. E., Foekens, J. A., Greaves, M., Hosoda, F., Hutter, B., Ilcic, T., Imbeaud, S., Imielinski, M., Jäger, N., Jones, D. T., Jones, D., Knappskog, S., Kool, M., Lakhani, S. R., López-Otín, C., Martin, S., Munshi, N. C., Nakamura, H., Northcott, P. A., Pajic, M., Papaemmanuil, E., Paradiso, A., Pearson, J. V., Puente, X. S., Raine, K., Ramakrishna, M., Richardson, A. L., Richter, J., Rosenstiel, P., Schlesner, M., Schumacher, T. N., Span, P. N., Teague, J. W., Totoki, Y., Tutt, A. N., Valdés-Mas, R., van Buuren, M. M., van 't Veer, L., Vincent-Salomon, A., Waddell, N., Yates, L. R., Zucman-Rossi, J., Futreal, P. A., McDermott, U., Lichter, P., Meyerson, M., Grimmond, S. M., Siebert, R., Campo, E., Shibata, T., Pfister, S. M., Campbell, P. J., Stratton, M. R., Initiative, A. P. C. G., Consortium, I. B. C., Consortium, I. M.-S. and PedBrain, I. (2013) 'Signatures of mutational processes in human cancer', *Nature*, 500(7463), pp. 415-21.
- Ambrosini, G., Groux, R. and Bucher, P. (2018) 'PWMScan: a fast tool for scanning entire genomes with a position-specific weight matrix', *Bioinformatics*, 34(14), pp. 2483-2484.
- Amemiya, H. M., Kundaje, A. and Boyle, A. P. (2019) 'The ENCODE Blacklist: Identification of Problematic Regions of the Genome', *Sci Rep*, 9(1), pp. 9354.

Andrews, S. (2010) *FastQC: a quality control tool for high throughput sequence data*. Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.

Arnould, C., Rocher, V., Finoux, A. L., Clouaire, T., Li, K., Zhou, F., Caron, P., Mangeot, P. E., Ricci, E. P., Mourad, R., Haber, J. E., Noordermeer, D. and Legube, G. (2021) 'Loop extrusion as a mechanism for formation of DNA damage repair foci', *Nature*, 590(7847), pp. 660-665.

Auerbach, R. K., Euskirchen, G., Rozowsky, J., Lamarre-Vincent, N., Moqtaderi, Z., Lefrançois, P., Struhl, K., Gerstein, M. and Snyder, M. (2009) 'Mapping accessible chromatin regions using Sono-Seq', *Proc Natl Acad Sci U S A*, 106(35), pp. 14926-31.

Aydin, Ö., Vermeulen, W. and Lans, H. (2014) 'ISWI chromatin remodeling complexes in the DNA damage response', *Cell Cycle*, 13(19), pp. 3016-25.

Aymard, F., Bugler, B., Schmidt, C. K., Guillou, E., Caron, P., Briois, S., Iacovoni, J. S., Daburon, V., Miller, K. M., Jackson, S. P. and Legube, G. (2014) 'Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks', *Nat Struct Mol Biol*, 21(4), pp. 366-74.

Bailey, T. L., Johnson, J., Grant, C. E. and Noble, W. S. (2015) 'The MEME Suite', *Nucleic Acids Res*, 43(W1), pp. W39-49.

Baker, M. E. (1997) 'Steroid receptor phylogeny and vertebrate origins', *Mol Cell Endocrinol*, 135(2), pp. 101-7.

Bannister, A. J. and Kouzarides, T. (2011) 'Regulation of chromatin by histone modifications', *Cell Res*, 21(3), pp. 381-95.

Bassing, C. H., Chua, K. F., Sekiguchi, J., Suh, H., Whitlow, S. R., Fleming, J. C., Monroe, B. C., Ciccone, D. N., Yan, C., Vlasakova, K., Livingston, D. M., Ferguson, D. O., Scully, R. and Alt, F. W. (2002) 'Increased ionizing radiation sensitivity and genomic instability in the absence of histone H2AX', *Proc Natl Acad Sci U S A*, 99(12), pp. 8173-8.

Bassing, C. H., Suh, H., Ferguson, D. O., Chua, K. F., Manis, J., Eckersdorff, M., Gleason, M., Bronson, R., Lee, C. and Alt, F. W. (2003) 'Histone H2AX: a dosage-dependent suppressor of oncogenic translocations and tumors', *Cell*, 114(3), pp. 359-70.

Beacon, T. H., Delcuve, G. P., López, C., Nardocci, G., Kovalchuk, I., van Wijnen, A. J. and Davie, J. R. (2021) 'The dynamic broad epigenetic (H3K4me3, H3K27ac) domain as a mark of essential genes', *Clin Epigenetics*, 13(1), pp. 138.

Bednar, J., Horowitz, R. A., Grigoryev, S. A., Carruthers, L. M., Hansen, J. C., Koster, A. J. and Woodcock, C. L. (1998) 'Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin', *Proc Natl Acad Sci U S A*, 95(24), pp. 14173-8.

Beischlag, T. V., Prefontaine, G. G. and Hankinson, O. (2018) 'ChIP-re-ChIP: Co-occupancy Analysis by Sequential Chromatin Immunoprecipitation', *Methods Mol Biol*, 1689, pp. 103-112.

Berkovich, E., Monnat, R. J. and Kastan, M. B. (2007) 'Roles of ATM and NBS1 in chromatin structure modulation and DNA double-strand break repair', *Nat Cell Biol*, 9(6), pp. 683-90.

Bewersdorf, J., Bennett, B. T. and Knight, K. L. (2006) 'H2AX chromatin structures and their response to DNA damage revealed by 4Pi microscopy', *Proc Natl Acad Sci U S A*, 103(48), pp. 18137-42.

Bhatia, V., Barroso, S. I., García-Rubio, M. L., Tumini, E., Herrera-Moyano, E. and Aguilera, A. (2014) 'BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2', *Nature*, 511(7509), pp. 362-5.

Bonner, W. M., Redon, C. E., Dickey, J. S., Nakamura, A. J., Sedelnikova, O. A., Solier, S. and Pommier, Y. (2008) 'GammaH2AX and cancer', *Nat Rev Cancer*, 8(12), pp. 957-67.

Boque-Sastre, R., Soler, M., Oliveira-Mateos, C., Portela, A., Moutinho, C., Sayols, S., Villanueva, A., Esteller, M. and Guil, S. (2015) 'Head-to-head antisense

transcription and R-loop formation promotes transcriptional activation', *Proc Natl Acad Sci U S A*, 112(18), pp. 5785-90.

Brandsma, I. and Gent, D. C. (2012) 'Pathway choice in DNA double strand break repair: observations of a balancing act', *Genome Integr*, 3(1), pp. 9.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. and Greenleaf, W. J. (2013) 'Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position', *Nat Methods*, 10(12), pp. 1213-8.

Burns, M. B., Lackey, L., Carpenter, M. A., Rathore, A., Land, A. M., Leonard, B., Refsland, E. W., Kotandeniya, D., Tretyakova, N., Nikas, J. B., Yee, D., Temiz, N. A., Donohue, D. E., McDougle, R. M., Brown, W. L., Law, E. K. and Harris, R. S. (2013) 'APOBEC3B is an enzymatic source of mutation in breast cancer', *Nature*, 494(7437), pp. 366-70.

Campbell, S. J., Edwards, R. A. and Glover, J. N. (2010) 'Comparison of the structures and peptide binding specificities of the BRCT domains of MDC1 and BRCA1', *Structure*, 18(2), pp. 167-76.

Carlile, G. W., Smith, D. H. and Wiedmann, M. (2004) 'Caspase-3 has a nonapoptotic function in erythroid maturation', *Blood*, 103(11), pp. 4310-6.

Caron, P., Choudjaye, J., Clouaire, T., Bugler, B., Daburon, V., Aguirrebengoa, M., Mangeat, T., Iacovoni, J. S., Álvarez-Quilón, A., Cortés-Ledesma, F. and Legube, G. (2015) 'Non-redundant Functions of ATM and DNA-PKcs in Response to DNA Double-Strand Breaks', *Cell Rep*, 13(8), pp. 1598-609.

Carroll, J. S. (2016) 'Mechanisms of oestrogen receptor (ER) gene regulation in breast cancer', *Eur J Endocrinol*, 175(1), pp. R41-9.

Carroll, J. S., Liu, X. S., Brodsky, A. S., Li, W., Meyer, C. A., Szary, A. J., Eeckhoute, J., Shao, W., Hestermann, E. V., Geistlinger, T. R., Fox, E. A., Silver, P. A. and Brown, M. (2005) 'Chromosome-wide mapping of estrogen receptor binding

reveals long-range regulation requiring the forkhead protein FoxA1', *Cell*, 122(1), pp. 33-43.

Celeste, A., Difilippantonio, S., Difilippantonio, M. J., Fernandez-Capetillo, O., Pilch, D. R., Sedelnikova, O. A., Eckhaus, M., Ried, T., Bonner, W. M. and Nussenzweig, A. (2003) 'H2AX haploinsufficiency modifies genomic stability and tumor susceptibility', *Cell*, 114(3), pp. 371-383.

Celeste, A., Petersen, S., Romanienko, P. J., Fernandez-Capetillo, O., Chen, H. T., Sedelnikova, O. A., Reina-San-Martin, B., Coppola, V., Meffre, E., Difilippantonio, M. J., Redon, C., Pilch, D. R., Oлару, A., Eckhaus, M., Camerini-Otero, R. D., Tessarollo, L., Livak, F., Manova, K., Bonner, W. M., Nussenzweig, M. C. and Nussenzweig, A. (2002) 'Genomic instability in mice lacking histone H2AX', *Science*, 296(5569), pp. 922-7.

Chang, H. H. Y., Pannunzio, N. R., Adachi, N. and Lieber, M. R. (2017) 'Non-homologous DNA end joining and alternative pathways to double-strand break repair', *Nat Rev Mol Cell Biol*, 18(8), pp. 495-506.

Chatterjee, N. and Walker, G. C. (2017) 'Mechanisms of DNA damage, repair, and mutagenesis', *Environ Mol Mutagen*, 58(5), pp. 235-263.

Chen, Q., Bian, C., Wang, X., Liu, X., Ahmad Kassab, M., Yu, Y. and Yu, X. (2021) 'ADP-ribosylation of histone variant H2AX promotes base excision repair', *EMBO J*, 40(2), pp. e104542.

Cheng, Y., Geng, H., Cheng, S. H., Liang, P., Bai, Y., Li, J., Srivastava, G., Ng, M. H., Fukagawa, T., Wu, X., Chan, A. T. and Tao, Q. (2010) 'KRAB zinc finger protein ZNF382 is a proapoptotic tumor suppressor that represses multiple oncogenes and is commonly silenced in multiple carcinomas', *Cancer Res*, 70(16), pp. 6516-26.

Chiappetta, C., Kirkland, J. L., Loose-Mitchell, D. S., Murthy, L. and Stancel, G. M. (1992) 'Estrogen regulates expression of the jun family of protooncogenes in the uterus', *J Steroid Biochem Mol Biol*, 41(2), pp. 113-23.

Cirillo, L. A., Lin, F. R., Cuesta, I., Friedman, D., Jarnik, M. and Zaret, K. S. (2002) 'Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4', *Mol Cell*, 9(2), pp. 279-89.

Clouaire, T., Rocher, V., Lashgari, A., Arnould, C., Aguirrebengoa, M., Biernacka, A., Skrzypczak, M., Aymard, F., Fongang, B., Dojer, N., Iacovoni, J. S., Rowicka, M., Ginalski, K., Côté, J. and Legube, G. (2018) 'Comprehensive Mapping of Histone Modifications at DNA Double-Strand Breaks Deciphers Repair Pathway Chromatin Signatures', *Mol Cell*, 72(2), pp. 250-262.e6.

Consortium, E. P. (2012) 'An integrated encyclopedia of DNA elements in the human genome', *Nature*, 489(7414), pp. 57-74.

Cook, P. J., Ju, B. G., Telese, F., Wang, X., Glass, C. K. and Rosenfeld, M. G. (2009) 'Tyrosine dephosphorylation of H2AX modulates apoptosis and survival decisions', *Nature*, 458(7238), pp. 591-6.

Corujo, D. and Buschbeck, M. (2018) 'Post-Translational Modifications of H2A Histone Variants and Their Role in Cancer', *Cancers (Basel)*, 10(3).

Cowell, I. G., Sunter, N. J., Singh, P. B., Austin, C. A., Durkacz, B. W. and Tilby, M. J. (2007) 'gammaH2AX foci form preferentially in euchromatin after ionising-radiation', *PLoS One*, 2(10), pp. e1057.

Crosetto, N., Mitra, A., Silva, M. J., Bienko, M., Dojer, N., Wang, Q., Karaca, E., Chiarle, R., Skrzypczak, M., Ginalski, K., Pasero, P., Rowicka, M. and Dikic, I. (2013) 'Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing', *Nat Methods*, 10(4), pp. 361-5.

Dhuppar, S., Roy, S. and Mazumder, A. (2020) 'gammaH2AX in the S Phase after UV Irradiation Corresponds to DNA Replication and Does Not Report on the Extent of DNA Damage', *Mol Cell Biol*, 40(20).

Fong, Y. W., Cattoglio, C. and Tjian, R. (2013) 'The intertwined roles of transcription and repair proteins', *Mol Cell*, 52(3), pp. 291-302.

- Furey, T. S. (2012) 'ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions', *Nat Rev Genet*, 13(12), pp. 840-52.
- Gao, T. and Qian, J. (2020) 'EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species', *Nucleic Acids Res*, 48(D1), pp. D58-D64.
- Gel, B., Díez-Villanueva, A., Serra, E., Buschbeck, M., Peinado, M. A. and Malinverni, R. (2016) 'regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests', *Bioinformatics*, 32(2), pp. 289-91.
- Gerasimova, N. S., Pestov, N. A., Kulaeva, O. I., Clark, D. J. and Studitsky, V. M. (2016) 'Transcription-induced DNA supercoiling: New roles of intranucleosomal DNA loops in DNA repair and transcription', *Transcription*, 7(3), pp. 91-5.
- Giacinti, C. and Giordano, A. (2006) 'RB and cell cycle progression', *Oncogene*, 25(38), pp. 5220-7.
- Ginno, P. A., Lim, Y. W., Lott, P. L., Korf, I. and Chédin, F. (2013) 'GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination', *Genome Res*, 23(10), pp. 1590-600.
- Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I. and Chédin, F. (2012) 'R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters', *Mol Cell*, 45(6), pp. 814-25.
- Gorski, J. J., Kennedy, R. D., Hosey, A. M. and Harkin, D. P. (2009) 'The complex relationship between BRCA1 and ERalpha in hereditary breast cancer', *Clin Cancer Res*, 15(5), pp. 1514-8.
- Groh, M., Lufino, M. M., Wade-Martins, R. and Gromak, N. (2014) 'R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome', *PLoS Genet*, 10(5), pp. e1004318.

- Gruber, C. J., Gruber, D. M., Gruber, I. M., Wieser, F. and Huber, J. C. (2004) 'Anatomy of the estrogen response element', *Trends Endocrinol Metab*, 15(2), pp. 73-8.
- Gruosso, T., Mieulet, V., Cardon, M., Bourachot, B., Kieffer, Y., Devun, F., Dubois, T., Dutreix, M., Vincent-Salomon, A., Miller, K. M. and Mechta-Grigoriou, F. (2016) 'Chronic oxidative stress promotes H2AX protein degradation and enhances chemosensitivity in breast cancer patients', *EMBO Mol Med*, 8(5), pp. 527-49.
- Hatimy, A. A., Browne, M. J. G., Flaus, A. and Sweet, S. M. M. (2015) 'Histone H2AX Y142 phosphorylation is a low abundance modification', *International Journal of Mass Spectrometry*, 391, pp. 139-145.
- He, H., Sinha, I., Fan, R., Haldosen, L. A., Yan, F., Zhao, C. and Dahlman-Wright, K. (2018) 'c-Jun/AP-1 overexpression reprograms ER α signaling related to tamoxifen response in ER α -positive breast cancer', *Oncogene*, 37(19), pp. 2586-2600.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H. and Glass, C. K. (2010) 'Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities', *Mol Cell*, 38(4), pp. 576-89.
- Henderson, S. and Fenton, T. (2015) 'APOBEC3 genes: retroviral restriction factors to cancer drivers', *Trends Mol Med*, 21(5), pp. 274-84.
- Hentges, L. D., Sergeant, M. J., Cole, C. B., Downes, D. J., Hughes, J. R. and Taylor, S. (2022) 'LanceOtron: a deep learning peak caller for genome sequencing experiments', *Bioinformatics*.
- Hewitt, S. C., Winuthayanon, W. and Korach, K. S. (2016) 'What's new in estrogen receptor action in the female reproductive tract', *J Mol Endocrinol*, 56(2), pp. R55-71.

Ho, J. W., Bishop, E., Karchenko, P. V., Nègre, N., White, K. P. and Park, P. J. (2011) 'ChIP-chip versus ChIP-seq: lessons for experimental design and data analysis', *BMC Genomics*, 12, pp. 134.

Hoopes, J. I., Cortez, L. M., Mertz, T. M., Malc, E. P., Mieczkowski, P. A. and Roberts, S. A. (2016) 'APOBEC3A and APOBEC3B Preferentially Deaminate the Lagging Strand Template during DNA Replication', *Cell Rep*, 14(6), pp. 1273-1282.

Huang, W., Peng, Y., Kiselar, J., Zhao, X., Albaqami, A., Mendez, D., Chen, Y., Chakravarthy, S., Gupta, S., Ralston, C., Kao, H. Y., Chance, M. R. and Yang, S. (2018) 'Multidomain architecture of estrogen receptor reveals interfacial cross-talk between its DNA-binding and ligand-binding domains', *Nat Commun*, 9(1), pp. 3520.

Hurtado, A., Holmes, K. A., Ross-Innes, C. S., Schmidt, D. and Carroll, J. S. (2011) 'FOXA1 is a key determinant of estrogen receptor function and endocrine response', *Nat Genet*, 43(1), pp. 27-33.

Hustedt, N. and Durocher, D. (2016) 'The control of DNA repair by the cell cycle', *Nat Cell Biol*, 19(1), pp. 1-9.

Iacovoni, J. S., Caron, P., Lassadi, I., Nicolas, E., Massip, L., Trouche, D. and Legube, G. (2010) 'High-resolution profiling of gammaH2AX around DNA double strand breaks in the mammalian genome', *EMBO J*, 29(8), pp. 1446-57.

Jeffery, N. N., Davidson, C., Peslak, S. A., Kingsley, P. D., Nakamura, Y., Palis, J. and Bulger, M. (2021) 'Histone H2A.X phosphorylation and Caspase-Initiated Chromatin Condensation in late-stage erythropoiesis', *Epigenetics Chromatin*, 14(1), pp. 37.

Jeggo, P. A. and Downs, J. A. (2014) 'Roles of chromatin remodellers in DNA double strand break repair', *Exp Cell Res*, 329(1), pp. 69-77.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K. and Felsenfeld, G. (2009) 'H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions', *Nat Genet*, 41(8), pp. 941-5.

Kadam, S. and Emerson, B. M. (2003) 'Transcriptional specificity of human SWI/SNF BRG1 and BRM chromatin remodeling complexes', *Mol Cell*, 11(2), pp. 377-89.

Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., Ahmad, K. and Henikoff, S. (2019) 'CUT&Tag for efficient epigenomic profiling of small samples and single cells', *Nat Commun*, 10(1), pp. 1930.

Kidd, J. M., Newman, T. L., Tuzun, E., Kaul, R. and Eichler, E. E. (2007) 'Population stratification of a common APOBEC gene deletion polymorphism', *PLoS Genet*, 3(4), pp. e63.

Kim, J. A., Kruhlak, M., Dotiwala, F., Nussenzweig, A. and Haber, J. E. (2007) 'Heterochromatin is refractory to gamma-H2AX modification in yeast and mammals', *J Cell Biol*, 178(2), pp. 209-18.

Krishnan, N., Jeong, D. G., Jung, S. K., Ryu, S. E., Xiao, A., Allis, C. D., Kim, S. J. and Tonks, N. K. (2009) 'Dephosphorylation of the C-terminal tyrosyl residue of the DNA damage-related histone H2A.X is mediated by the protein phosphatase eyes absent', *J Biol Chem*, 284(24), pp. 16066-16070.

Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B. E., Bickel, P., Brown, J. B., Cayting, P., Chen, Y., DeSalvo, G., Epstein, C., Fisher-Aylor, K. I., Euskirchen, G., Gerstein, M., Gertz, J., Hartemink, A. J., Hoffman, M. M., Iyer, V. R., Jung, Y. L., Karmakar, S., Kellis, M., Kharchenko, P. V., Li, Q., Liu, T., Liu, X. S., Ma, L., Milosavljevic, A., Myers, R. M., Park, P. J., Pazin, M. J., Perry, M. D., Raha, D., Reddy, T. E., Rozowsky, J., Shores, N., Sidow, A., Slattery, M., Stamatoyannopoulos, J. A., Tolstorukov, M. Y., White, K. P., Xi, S., Farnham, P. J., Lieb, J. D., Wold, B. J. and Snyder, M. (2012) 'ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia', *Genome Res*, 22(9), pp. 1813-31.

Lawrence, M., Daujat, S. and Schneider, R. (2016) 'Lateral Thinking: How Histone Modifications Regulate Gene Expression', *Trends Genet*, 32(1), pp. 42-56.

- Lee, H. S., Park, J. H., Kim, S. J., Kwon, S. J. and Kwon, J. (2010) 'A cooperative activation loop among SWI/SNF, gamma-H2AX and H3 acetylation for DNA double-strand break repair', *EMBO J*, 29(8), pp. 1434-45.
- Li, G. and Zhu, P. (2015) 'Structure and organization of chromatin fiber in the nucleus', *FEBS Lett*, 589(20 Pt A), pp. 2893-904.
- Li, X. and Heyer, W. D. (2008) 'Homologous recombination in DNA repair and DNA damage tolerance', *Cell Res*, 18(1), pp. 99-113.
- Li, X. and Li, X. D. (2021) 'Integrative Chemical Biology Approaches to Deciphering the Histone Code: A Problem-Driven Journey', *Acc Chem Res*, 54(19), pp. 3734-3747.
- Lieber, M. R. (2010) 'The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway', *Annu Rev Biochem*, 79, pp. 181-211.
- Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S. and Dekker, J. (2009) 'Comprehensive mapping of long-range interactions reveals folding principles of the human genome', *Science*, 326(5950), pp. 289-93.
- Lindahl, T. (1993) 'Instability and decay of the primary structure of DNA', *Nature*, 362(6422), pp. 709-15.
- Liu, S., Hua, Y., Wang, J., Li, L., Yuan, J., Zhang, B., Wang, Z., Ji, J. and Kong, D. (2021) 'RNA polymerase III is required for the repair of DNA double-strand breaks by homologous recombination', *Cell*, 184(5), pp. 1314-1329.e10.
- Lu, B., Dong, L., Yi, D., Zhang, M., Zhu, C., Li, X. and Yi, C. (2020) 'Transposase-assisted tagmentation of RNA/DNA hybrid duplexes', *Elife*, 9.
- Lukas, J. and Bartek, J. (2009) 'DNA repair: New tales of an old tail', *Nature*, 458(7238), pp. 581-3.

Maciejowski, J., Chatzipli, A., Dananberg, A., Chu, K., Toufektchan, E., Klimczak, L. J., Gordenin, D. A., Campbell, P. J. and de Lange, T. (2020) 'APOBEC3-dependent kataegis and TREX1-driven chromothripsis during telomere crisis', *Nat Genet*, 52(9), pp. 884-890.

Maezawa, S., Hasegawa, K., Yukawa, M., Kubo, N., Sakashita, A., Alavattam, K. G., Sin, H. S., Kartashov, A. V., Sasaki, H., Barski, A. and Namekawa, S. H. (2018) 'Polycomb protein SCML2 facilitates H3K27me3 to establish bivalent domains in the male germline', *Proc Natl Acad Sci U S A*, 115(19), pp. 4957-4962.

Mannironi, C., Bonner, W. M. and Hatch, C. L. (1989) 'H2A.X, a histone isoprotein with a conserved C-terminal sequence, is encoded by a novel mRNA with both DNA replication type and polyA 3' processing signals', *Nucleic Acids Res*, 17(22), pp. 9113-26.

Mao, Z., Bozzella, M., Seluanov, A. and Gorbunova, V. (2008) 'Comparison of nonhomologous end joining and homologous recombination in human cells', *DNA Repair (Amst)*, 7(10), pp. 1765-71.

Marabitti, V., Valenzisi, P., Lillo, G., Malacaria, E., Palermo, V., Pichierri, P. and Franchitto, A. (2022) 'R-Loop-Associated Genomic Instability and Implication of WRN and WRNIP1', *Int J Mol Sci*, 23(3).

Mason, C. E., Shu, F. J., Wang, C., Session, R. M., Kallen, R. G., Sidell, N., Yu, T., Liu, M. H., Cheung, E. and Kallen, C. B. (2010) 'Location analysis for the estrogen receptor-alpha reveals binding to diverse ERE sequences and widespread binding within repetitive DNA elements', *Nucleic Acids Res*, 38(7), pp. 2355-68.

Massip, L., Caron, P., Iacovoni, J. S., Trouche, D. and Legube, G. (2010) 'Deciphering the chromatin landscape induced around DNA double strand breaks', *Cell Cycle*, 9(15), pp. 2963-72.

Matharu, N. and Ahituv, N. (2015) 'Minor Loops in Major Folds: Enhancer-Promoter Looping, Chromatin Restructuring, and Their Association with Transcriptional Regulation and Disease', *PLoS Genet*, 11(12), pp. e1005640.

McDaniel, Y. Z., Wang, D., Love, R. P., Adolph, M. B., Mohammadzadeh, N., Chelico, L. and Mansky, L. M. (2020) 'Deamination hotspots among APOBEC3 family members are defined by both target site sequence context and ssDNA secondary structure', *Nucleic Acids Res*, 48(3), pp. 1353-1371.

Morimoto, S., Tsuda, M., Bunch, H., Sasanuma, H., Austin, C. and Takeda, S. (2019) 'Type II DNA Topoisomerases Cause Spontaneous Double-Strand Breaks in Genomic DNA', *Genes (Basel)*, 10(11).

Métivier, R., Reid, G. and Gannon, F. (2006) 'Transcription in four dimensions: nuclear receptor-directed initiation of gene expression', *EMBO Rep*, 7(2), pp. 161-7.

Nagelkerke, A., van Kuijk, S. J., Sweep, F. C., Nagtegaal, I. D., Hoogerbrugge, N., Martens, J. W., Timmermans, M. A., van Laarhoven, H. W., Bussink, J. and Span, P. N. (2011) 'Constitutive expression of γ -H2AX has prognostic relevance in triple negative breast cancer', *Radiother Oncol*, 101(1), pp. 39-45.

Negrini, S., Gorgoulis, V. G. and Halazonetis, T. D. (2010) 'Genomic instability--an evolving hallmark of cancer', *Nat Rev Mol Cell Biol*, 11(3), pp. 220-8.

Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C., Van Loo, P., Greenman, C. D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L. A., Menzies, A., Martin, S., Leung, K., Chen, L., Leroy, C., Ramakrishna, M., Rance, R., Lau, K. W., Mudie, L. J., Varela, I., McBride, D. J., Bignell, G. R., Cooke, S. L., Shlien, A., Gamble, J., Whitmore, I., Maddison, M., Tarpey, P. S., Davies, H. R., Papaemmanuil, E., Stephens, P. J., McLaren, S., Butler, A. P., Teague, J. W., Jönsson, G., Garber, J. E., Silver, D., Miron, P., Fatima, A., Boyault, S., Langerød, A., Tutt, A., Martens, J. W., Aparicio, S. A., Borg, Å., Salomon, A. V., Thomas, G., Børresen-Dale, A. L., Richardson, A. L., Neuberger, M. S., Futreal, P. A., Campbell, P. J., Stratton, M. R. and Consortium, B. C. W. G. o. t. I. C. G. (2012) 'Mutational processes molding the genomes of 21 breast cancers', *Cell*, 149(5), pp. 979-93.

Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L. B., Martin, S., Wedge, D. C., Van Loo, P., Ju, Y. S.,

Smid, M., Brinkman, A. B., Morganella, S., Aure, M. R., Lingjærde, O. C., Langerød, A., Ringnér, M., Ahn, S. M., Boyault, S., Brock, J. E., Broeks, A., Butler, A., Desmedt, C., Dirix, L., Dronov, S., Fatima, A., Foekens, J. A., Gerstung, M., Hooijer, G. K., Jang, S. J., Jones, D. R., Kim, H. Y., King, T. A., Krishnamurthy, S., Lee, H. J., Lee, J. Y., Li, Y., McLaren, S., Menzies, A., Mustonen, V., O'Meara, S., Pauporté, I., Pivot, X., Purdie, C. A., Raine, K., Ramakrishnan, K., Rodríguez-González, F. G., Romieu, G., Sieuwerts, A. M., Simpson, P. T., Shepherd, R., Stebbings, L., Stefansson, O. A., Teague, J., Tommasi, S., Treilleux, I., Van den Eynden, G. G., Vermeulen, P., Vincent-Salomon, A., Yates, L., Caldas, C., van't Veer, L., Tutt, A., Knappskog, S., Tan, B. K., Jonkers, J., Borg, Å., Ueno, N. T., Sotiriou, C., Viari, A., Futreal, P. A., Campbell, P. J., Span, P. N., Van Laere, S., Lakhani, S. R., Eyfjord, J. E., Thompson, A. M., Birney, E., Stunnenberg, H. G., van de Vijver, M. J., Martens, J. W., Børresen-Dale, A. L., Richardson, A. L., Kong, G., Thomas, G. and Stratton, M. R. (2016) 'Landscape of somatic mutations in 560 breast cancer whole-genome sequences', *Nature*, 534(7605), pp. 47-54.

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. and Kanemaki, M. (2009) 'An auxin-based degron system for the rapid depletion of proteins in nonplant cells', *Nat Methods*, 6(12), pp. 917-22.

Nowsheen, S., Aziz, K., Luo, K., Deng, M., Qin, B., Yuan, J., Jeganathan, K. B., Yu, J., Zhang, H., Ding, W., van Deursen, J. M. and Lou, Z. (2018) 'ZNF506-dependent positive feedback loop regulates H2AX signaling after DNA damage', *Nat Commun*, 9(1), pp. 2736.

Park, P. J. (2009) 'ChIP-seq: advantages and challenges of a maturing technology', *Nat Rev Genet*, 10(10), pp. 669-80.

Paull, T. T., Rogakou, E. P., Yamazaki, V., Kirchgessner, C. U., Gellert, M. and Bonner, W. M. (2000) 'A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage', *Curr Biol*, 10(15), pp. 886-95.

Periyasamy, M., Patel, H., Lai, C. F., Nguyen, V. T. M., Nevedomskaya, E., Harrod, A., Russell, R., Remenyi, J., Ochocka, A. M., Thomas, R. S., Fuller-Pace, F., Györfy, B., Caldas, C., Navaratnam, N., Carroll, J. S., Zwart, W., Coombes, R. C., Magnani, L., Buluwela, L. and Ali, S. (2015) 'APOBEC3B-Mediated Cytidine Deamination Is Required for Estrogen Receptor Action in Breast Cancer', *Cell Rep*, 13(1), pp. 108-121.

Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslén, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lønning, P. E., Børresen-Dale, A. L., Brown, P. O. and Botstein, D. (2000) 'Molecular portraits of human breast tumours', *Nature*, 406(6797), pp. 747-52.

Renard, M., Henry, M., Guétard, D., Vartanian, J. P. and Wain-Hobson, S. (2010) 'APOBEC1 and APOBEC3 cytidine deaminases as restriction factors for hepadnaviral genomes in non-humans in vivo', *J Mol Biol*, 400(3), pp. 323-34.

Roberts, S. A., Lawrence, M. S., Klimczak, L. J., Grimm, S. A., Fargo, D., Stojanov, P., Kiezun, A., Kryukov, G. V., Carter, S. L., Saksena, G., Harris, S., Shah, R. R., Resnick, M. A., Getz, G. and Gordenin, D. A. (2013) 'An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers', *Nat Genet*, 45(9), pp. 970-6.

Rogakou, E. P., Boon, C., Redon, C. and Bonner, W. M. (1999) 'Megabase chromatin domains involved in DNA double-strand breaks in vivo', *J Cell Biol*, 146(5), pp. 905-16.

Rogakou, E. P., Nieves-Neira, W., Boon, C., Pommier, Y. and Bonner, W. M. (2000) 'Initiation of DNA fragmentation during apoptosis induces phosphorylation of H2AX histone at serine 139', *J Biol Chem*, 275(13), pp. 9390-5.

Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. and Bonner, W. M. (1998) 'DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139', *J Biol Chem*, 273(10), pp. 5858-68.

Ross-Innes, C. S., Stark, R., Teschendorff, A. E., Holmes, K. A., Ali, H. R., Dunning, M. J., Brown, G. D., Gojis, O., Ellis, I. O., Green, A. R., Ali, S., Chin, S. F., Palmieri, C., Caldas, C. and Carroll, J. S. (2012) 'Differential oestrogen receptor binding is associated with clinical outcome in breast cancer', *Nature*, 481(7381), pp. 389-93.

Rothkamm, K., Barnard, S., Moquet, J., Ellender, M., Rana, Z. and Burdak-Rothkamm, S. (2015) 'DNA damage foci: Meaning and significance', *Environ Mol Mutagen*, 56(6), pp. 491-504.

Rouet, P., Smih, F. and Jasin, M. (1994) 'Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells', *Proc Natl Acad Sci U S A*, 91(13), pp. 6064-8.

Rowley, M. J. and Corces, V. G. (2018) 'Organizational principles of 3D genome architecture', *Nat Rev Genet*, 19(12), pp. 789-800.

Saksouk, N., Simboeck, E. and Déjardin, J. (2015) 'Constitutive heterochromatin formation and transcription in mammals', *Epigenetics Chromatin*, 8, pp. 3.

San Filippo, J., Sung, P. and Klein, H. (2008) 'Mechanism of eukaryotic homologous recombination', *Annu Rev Biochem*, 77, pp. 229-57.

Sanborn, A. L., Rao, S. S., Huang, S. C., Durand, N. C., Huntley, M. H., Jewett, A. I., Bochkov, I. D., Chinnappan, D., Cutkosky, A., Li, J., Geeting, K. P., Gnirke, A., Melnikov, A., McKenna, D., Stamenova, E. K., Lander, E. S. and Aiden, E. L. (2015) 'Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes', *Proc Natl Acad Sci U S A*, 112(47), pp. E6456-65.

Santos-Pereira, J. M. and Aguilera, A. (2015) 'R loops: new modulators of genome dynamics and function', *Nat Rev Genet*, 16(10), pp. 583-97.

Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J. and Kouzarides, T. (2002) 'Active genes are trimethylated at K4 of histone H3', *Nature*, 419(6905), pp. 407-11.

Savic, V., Yin, B., Maas, N. L., Bredemeyer, A. L., Carpenter, A. C., Helmink, B. A., Yang-lott, K. S., Sleckman, B. P. and Bassing, C. H. (2009) 'Formation of dynamic

gamma-H2AX domains along broken DNA strands is distinctly regulated by ATM and MDC1 and dependent upon H2AX densities in chromatin', *Mol Cell*, 34(3), pp. 298-310.

Sawasdichai, A., Chen, H. T., Abdul Hamid, N., Jayaraman, P. S. and Gaston, K. (2010) 'In situ subcellular fractionation of adherent and non-adherent mammalian cells', *J Vis Exp*, (41).

Serandour, A. A., Brown, G. D., Cohen, J. D. and Carroll, J. S. (2013) 'Development of an Illumina-based ChIP-exonuclease method provides insight into FoxA1-DNA binding properties', *Genome Biol*, 14(12), pp. R147.

Shi, K., Carpenter, M. A., Kurahashi, K., Harris, R. S. and Aihara, H. (2015) 'Crystal Structure of the DNA Deaminase APOBEC3B Catalytic Domain', *J Biol Chem*, 290(47), pp. 28120-28130.

Shin, Y. and Brangwynne, C. P. (2017) 'Liquid phase condensation in cell physiology and disease', *Science*, 357(6357).

Shroff, R., Arbel-Eden, A., Pilch, D., Ira, G., Bonner, W. M., Petrini, J. H., Haber, J. E. and Lichten, M. (2004) 'Distribution and dynamics of chromatin modification induced by a defined DNA double-strand break', *Curr Biol*, 14(19), pp. 1703-11.

Singh, N., Basnet, H., Wiltshire, T. D., Mohammad, D. H., Thompson, J. R., Héroux, A., Botuyan, M. V., Yaffe, M. B., Couch, F. J., Rosenfeld, M. G. and Mer, G. (2012) 'Dual recognition of phosphoserine and phosphotyrosine in histone variant H2A.X by DNA damage response protein MCPH1', *Proc Natl Acad Sci U S A*, 109(36), pp. 14381-6.

Srivastava, N., Gochhait, S., Gupta, P. and Bamezai, R. N. (2008) 'Copy number alterations of the H2AFX gene in sporadic breast cancer patients', *Cancer Genet Cytogenet*, 180(2), pp. 121-8.

Stork, C. T., Bocek, M., Crossley, M. P., Sollier, J., Sanz, L. A., Chédin, F., Swigut, T. and Cimprich, K. A. (2016) 'Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage', *Elife*, 5.

Stucki, M., Clapperton, J. A., Mohammad, D., Yaffe, M. B., Smerdon, S. J. and Jackson, S. P. (2005) 'MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks', *Cell*, 123(7), pp. 1213-26.

Swanton, C., McGranahan, N., Starrett, G. J. and Harris, R. S. (2015) 'APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity', *Cancer Discov*, 5(7), pp. 704-12.

Taylor, B. J., Nik-Zainal, S., Wu, Y. L., Stebbings, L. A., Raine, K., Campbell, P. J., Rada, C., Stratton, M. R. and Neuberger, M. S. (2013) 'DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis', *Elife*, 2, pp. e00534.

Theodorou, V., Stark, R., Menon, S. and Carroll, J. S. (2013) 'GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility', *Genome Res*, 23(1), pp. 12-22.

Turner, J. M. (2007) 'Meiotic sex chromosome inactivation', *Development*, 134(10), pp. 1823-31.

Tyteca, S., Vandromme, M., Legube, G., Chevillard-Briet, M. and Trouche, D. (2006) 'Tip60 and p400 are both required for UV-induced apoptosis but play antagonistic roles in cell cycle progression', *EMBO J*, 25(8), pp. 1680-9.

Wallace, S. S., Murphy, D. L. and Sweasy, J. B. (2012) 'Base excision repair and cancer', *Cancer Lett*, 327(1-2), pp. 73-89.

Warren, C. J., Westrich, J. A., Doorslaer, K. V. and Pyeon, D. (2017) 'Roles of APOBEC3A and APOBEC3B in Human Papillomavirus Infection and Disease Progression', *Viruses*, 9(8).

Welboren, W. J., Stunnenberg, H. G., Sweep, F. C. and Span, P. N. (2007) 'Identifying estrogen receptor target genes', *Mol Oncol*, 1(2), pp. 138-43.

Welboren, W. J., van Driel, M. A., Janssen-Megens, E. M., van Heeringen, S. J., Sweep, F. C., Span, P. N. and Stunnenberg, H. G. (2009) 'ChIP-Seq of ERalpha

and RNA polymerase II defines genes differentially responding to ligands', *EMBO J*, 28(10), pp. 1418-28.

Wu, S. J., Furlan, S. N., Mihalas, A. B., Kaya-Okur, H. S., Feroze, A. H., Emerson, S. N., Zheng, Y., Carson, K., Cimino, P. J., Keene, C. D., Sarthy, J. F., Gottardo, R., Ahmad, K., Henikoff, S. and Patel, A. P. (2021) 'Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression', *Nat Biotechnol*, 39(7), pp. 819-824.

Wysocka, J., Swigut, T., Xiao, H., Milne, T. A., Kwon, S. Y., Landry, J., Kauer, M., Tackett, A. J., Chait, B. T., Badenhorst, P., Wu, C. and Allis, C. D. (2006) 'A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling', *Nature*, 442(7098), pp. 86-90.

Xiao, A., Li, H., Shechter, D., Ahn, S. H., Fabrizio, L. A., Erdjument-Bromage, H., Ishibe-Murakami, S., Wang, B., Tempst, P., Hofmann, K., Patel, D. J., Elledge, S. J. and Allis, C. D. (2009) 'WSTF regulates the H2A.X DNA damage response via a novel tyrosine kinase activity', *Nature*, 457(7225), pp. 57-62.

Xu, Y., Ayrapetov, M. K., Xu, C., Gursoy-Yuzugullu, O., Hu, Y. and Price, B. D. (2012) 'Histone H2A.Z controls a critical chromatin remodeling step required for DNA double-strand break repair', *Mol Cell*, 48(5), pp. 723-33.

Yan, F., Powell, D. R., Curtis, D. J. and Wong, N. C. (2020) 'From reads to insight: a hitchhiker's guide to ATAC-seq data analysis', *Genome Biol*, 21(1), pp. 22.

Yang, J. A., Stires, H., Belden, W. J. and Roepke, T. A. (2017) 'The Arcuate Estrogen-Regulated Transcriptome: Estrogen Response Element-Dependent and -Independent Signaling of ER α in Female Mice', *Endocrinology*, 158(3), pp. 612-626.

Zhang, C., Lu, Y.-j., Chen, B., Bai, Z., Hervieu, A., Licciardello, M. P., Wang, M., Mitsopoulos, C., Al-Lazikani, B., Totorici, M., Rossanese, O. W., Workman, P. and Clarke, P. A. (2022) 'R-loop editing by DNA cytosine deaminase APOBEC3B determines the activity of estrogen receptor enhancers', *bioRxiv*, pp. 2022.10.21.513235.

Zhang, H., Lu, T., Liu, S., Yang, J., Sun, G., Cheng, T., Xu, J., Chen, F. and Yen, K. (2021) 'Comprehensive understanding of Tn5 insertion preference improves transcription regulatory element identification', *NAR Genom Bioinform*, 3(4), pp. lqab094.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W. and Liu, X. S. (2008) 'Model-based analysis of ChIP-Seq (MACS)', *Genome Biol*, 9(9), pp. R137.

Zhao, B., Tan, T. L., Mei, Y., Yang, J., Yu, Y., Verma, A., Liang, Y., Gao, J. and Ji, P. (2016) 'H2AX deficiency is associated with erythroid dysplasia and compromised haematopoietic stem cell function', *Sci Rep*, 6, pp. 19589.

Zheng, Y., Ahmad, K. and Henikoff, S. (2020) *CUT&Tag Data Processing and Analysis Tutorial*. Available at: https://yehengstat.github.io/CUTTag_tutorial/.

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C. and Chanda, S. K. (2019) 'Metascape provides a biologist-oriented resource for the analysis of systems-level datasets', *Nat Commun*, 10(1), pp. 1523.

Zou, J., Wang, C., Ma, X., Wang, E. and Peng, G. (2017) 'APOBEC3B, a molecular driver of mutagenesis in human cancers', *Cell Biosci*, 7, pp. 29.