THE UNIVERSITY
of EDINBURGH

Disease-associated Brd4 mutation – linking chromatin binding and the DNA damage response

Gabrielle Olley

THE UNIVERSITY *of* EDINBURGH

PhD by Research

University of Edinburgh

2018

Table of contents

Table of figures

Table of tables

Abbreviations

ΔΔ*G* – *change in free energy*

3C - chromosome conformation capture

4OHT - 4-Hydroxytamoxifen

4sU-seq – 4-thiouridine-sequencing

Å – ångström

A - alanine

aa – amino acid

Ac – acetylation

AID - activation induced cytidine deamination

Ala - alanine

APH – aphidicolin

Arg/R – arginine

AS – active site

ASD - atrial septal defects

Asn/N – asparagine

AT - ataxia telangiectasia

BD – bromodomain

BET – bromodomain and extraterminal domain

BETi – BET inhibitors

BRCT - BRCA1 carboxy-terminal

CDK - cyclin dependent kinase

CdLS – Cornelia de Lange Syndrome

cDNA – complementary DNA

CHD – congenital heart defect

CSS - Coffin-Siris syndrome

ChIP-seq – chromatin immunoprecipitation-sequencing

CO - crossovers

Co-IP - co-immunoprecipitation

convT – convergent transcription

CRE – cis regulatory element

CRISPR – *Clustered Regularly Interspaced Short Palindromic Repeats*

CSR – class switch recombination

CTCF - CCCTC-binding factor

CTD – C-terminal domain

Cys/C – cysteine

DDD - Deciphering Developmental Disorders

DDR – DNA damage response

DNA – deoxyribonucleic acid

dNTPs - deoxyribonucleotides

DSB – double strand break

DSE - double-strand end

DSIF - DRB sensitivity-inducing factor

E – embryonic stage

EPPF – Edinburgh Protein Production Facility

eRNA – enhancer RNA

ET – extraterminal

FA – Fanconi Anaemia

FISH – fluorescent in situ hybridisation

FPKM - Fragments per kilobase mapped

GAR -glycine and arginine rich

GC – gene conversion

GFP – green fluorescent protein

Gln – glutamine

GO – gene ontology

gRNA – guide ribonucleic acid

HAT - histone acetyltransferase

HDR - homology directed repair

His/H - histidine

HJ – holliday junction

HR – homologous recombination

HRD – HR deficient

hpf - hour postfertilisation

HU – hydroxyurea

IF – immunofluorescence

Ile/I – isoleucine

Ig - immunoglobulin

IGMM – Institute of Genetics and Molecular Medicine

IP – immunoprecipitation/immunoprecipitates

IR – ionising radiation

IRIF - irradiation induced foci

Kb - kilobase

LCL - lymphoblastoid cell line

Leu/L – leucine

LiDS - DNA ligase IV deficiency syndrome

LIF - leukemia inhibitory factor

LFQ – label free quantification

LOF – loss of function

Lys/K - lysine

M – molar

MACS - Model-based Analysis of ChIP-Seq data

Mb – megabase

Me – methylation

MEF – mouse embryonic fibroblast

mESC – mouse embryonic stem cell

MNase - Micrococcal Nuclease

MO - morpholino

mRNA - messenger RNA

MS – mass spectrometry

NCBRS - Nicolaides–Baraitser syndrome

NCS – neocarzinostatin

NELF – negative elongation factor

NHEJ – non-homologous end joining

NIPBL – Nipped-B-like

NLS - nuclear localisation sequence

nM – nanomolar

NMC - NUT midline carcinoma

NMR – nuclear magnetic resonance

OD - oligomerisation domain

OE1 – 1st oxygen atom of the carbonyl group

OH – hydroxyl

OPT - OCT1/PTF/transcription

-P - phosphorylation

PAM - protospacer adjacent motif

PCS – premature chromatid separation

PDB – Protein Data Bank

Phe/F - phenylalanine

PIC – pre-initiation complex

PIKKs - phosphoinositide-3-kinase-related protein kinases

PSF - point spread function

P-TEFb - positive transcription elongation factor B

PTM - post-translational modifications

Q - glutamine

qPCR – quantitative polymerase chain reaction

RBS - Roberts Syndrome

RIPA - radioimmunoprecipitation assay

RNA - ribonucleic acid

RNAPII – RNA polymerase II

RP10M – reads per 10 million

RPKM – reads per kilobase per million mapped reads

RS-SCID - radiosensitive severe combined immunodeficiency

RT – room temperature

RT-qPCR – reverse transcription-qPCR

SDSA - synthesis dependent strand annealing

SE- super enhancer

Ser – serine/S

SEM – standard error of the mean

SSA - single strand annealing

SMC - structural maintenance of chromosomes

snRNA – small nuclear ribonucleic acid

snRNP – small nuclear ribonucleoprotein

SS – Seckel Syndrome

SSB – single strand break

ssDNA – single stranded DNA

TAD - topologically associated domains

TAM – tamoxifen

TBS(-T) – tris-buffered saline(-tween)

TE – typical enhancer

Thr – threonine/T

TOP – topoisomerase

TPR - tetratricopeptide repeat

Trp - triptolide

TSS – transcription start site

Tyr/Y - tyrosine

UCSC - University of California Santa Cruz

UDR - ubiquitylation-dependent recruitment

UV – ultra-violet

Val/V - valine

WDSTS - Wiedemann-Steiner syndrome

µg – microgram

µl – microliter

µm – micrometre

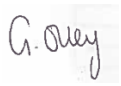
Declaration

I declare that this thesis is an original report of my research, has been written by me and has not been submitted for any previous degree. The experimental work is almost entirely my own work; the collaborative contributions have been indicated clearly and acknowledged. Due references have been provided on all supporting literatures and resources.

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification. Data and conclusions from

Parts of this work have been published in Olley *et al.*, 2018.

Gabrielle Olley



Acknowledgements

Firstly, I would like to thank my supervisors, Professor Wendy Bickmore, for having me in her lab, for her unwavering enthusiasm – especially at times when mine was running low –, and her wealth of knowledge and ideas, and Professor David FitzPatrick, for his help and support throughout, but particularly in publishing my first paper.

I am extremely grateful to my 'in lab' supervisors; Pradeepa Madapura, for welcoming me into the lab and showing me the ropes, and Charlène Boumendil, without whom my project would have turned out very differently. Her vast knowledge of the DNA damage and repair field, and her unbreakable faith in me and my science, has been invaluable.

The Bickmore Lab, past and present, have been a constant source of support, cake, WhatsApp messages, laughter, and more cake. There are too many to mention them all, but special shout-outs must go to Shelagh Boyle - for keeping the lab running smoothly, and of course, for the cake - and to Rob Illingworth - for help with all things ChIP, 4sU, bioinformatic and beyond, and for never acting like my questions were stupid (excepting the Highland Coo one).

For my sanity, I thank Abby, Jane, Zoe and Luke, who had the unique position of always being in the same boat as me. Thanks for being my housemates, workmates, travel buddies, drinking partners and therapists (and providing the occasional scientific insight).

Finally, but most importantly, a huge thank you to my parents, Karyn and Mike, and my Nan, Cilla. Who have not only supported me emotionally and financially for the last 26 years, but also taught me that I could achieve whatever I wanted in life, without ever once trying to tell me what that should be.

Lay summary of thesis

Cornelia de Lange Syndrome (CdLS) is a rare developmental genetic disorder. It causes a wide range of symptoms, including limb abnormalities, delayed growth, and intellectual disability.

The majority of CdLS patients have a mutation in one of a group of proteins that form a protein complex called cohesin. Cohesin is important during cell division and also has roles in the organisation of chromosomes and the repair of damaged DNA. However, cohesin mutations only make up about 70% of CdLS cases, leaving a significant number of patients for whom the underlying mutation is unknown.

In this thesis I describe the identification and characterisation of a mutation in a protein called BRD4, which is unrelated to the cohesin complex, in a patient with a CdLS-like syndrome. I recreated this mutation in mouse embryonic stem cells, allowing me to study its effects on cell behaviour. Surprisingly, I found that this mutation, although affecting the ability of BRD4 to bind correctly to genomic regions that regulate the expression of genes, does not alter gene regulation. Instead I found an increase in DNA damage and a slower cell cycle in the mutated cells, which could explain the CdLS-like symptoms in the patient

Abstract

Acetylation of lysine residues is a histone modification associated with active chromatin. The modified residues provide docking sites for the epigenetic reader BRD4, which binds to the acetylated lysines via its two bromodomains. BRD4 is known to be involved in RNA polymerase II activation, maintaining the pluripotency of embryonic stem cells and in DNA damage response signalling.

In this thesis, I describe a newly identified missense mutation (Y430C) in *BRD4* in a patient with a Cornelia de Lange syndrome (CdLS)-like phenotype. CdLS is a neurodevelopmental disorder that can cause a range of symptoms including upper limb malformations, craniofacial abnormalities and intellectual disability and is usually associated with mutations in components of the cohesin complex and the cohesin loader NIPBL. How these mutations may cause CdLS is currently unknown, but the assumption has been that they cause altered gene regulation during development.

Using a mouse embryonic stem cell (mESC) line, engineered through CRISPR-Cas9 technology to be homozygous for the patient mutation in *BRD4*, I show that the mutation decreases the affinity of BRD4 for acetylated lysines. This causes a loss of BRD4 binding in the genome, most noticeable at cis-regulatory elements. However, surprisingly I found no evidence for altered gene expression in the cells with the *Brd4* mutation. Instead I identify increased G2/M checkpoint activation in the Y430C mESCs compared to wild-type cells, and my research suggests that this is attributable to an increase in DNA damage signalling

Chapter 1: Introduction

1.1 Cornelia de Lange Syndrome

1.1.1 Clinical phenotype and inheritance

Cornelia de Lange Syndrome (CdLS) is a rare neurodevelopmental disorder, estimated to affect 1:10,000-1:50,000 new-borns (Beck and Fenger, 1985; Liu and Baynam, 2010). CdLS is characterised by a range of symptoms, that vary vastly in severity between patients (Newkirk *et al.*, 2017). The most diagnostic symptom is the distinctive facial dysmorphism (Figure Ch1-1A-D); patients often present with microbrachycephaly, synophrys, long philtrum, thin lips, crescent shaped mouth, depressed nasal bridge with anteverted nares and cleft palate (Rohatgi *et al.*, 2010). Other symptoms include growth retardation, intellectual and psychomotor delay, upper limb malformations (Figure Ch1-1E-J), hirsutism, and cardiac and gastrointestinal issues (Liu and Baynam, 2010; Noshir Mehta and Bhatia, 2013). A large percentage of patients die before the age of 2, due to pneumonia or cardiac, respiratory and gastrointestinal complications (Beck and Fenger, 1985) however, in the absence of these problems, patients often survive to adulthood (Badoe, 2006).

Figure Ch1-1. CdLS is a genetically heterogeneous syndrome that causes a spectrum of phenotypes. Characteristic facial phenotype in two patients with NIPBL mutations **(A&B)** and two patients with SMC1A mutations **(C&D).** Spectrum of upper limb involvement is shown with decreasing severity **(E-J).** Taken from Dorsett and Krantz, 2009.

Most cases of CdLS are caused by de novo mutations. The low probability of an affected individual to reproduce may limit the level of familial inheritance, although some cases have been reported and point towards an autosomal dominant mode of inheritance. (Russell *et al.*, 2001). There have been rare cases of phenotypically unaffected parents having multiple children with CdLS, suggesting that germline mosaicism may occur (Gillis *et al.*, 2004).

1.1.2 De novo gene mutations causing CdLS

1.1.2.1 *NIPBL*

In 1933 Cornelia de Lange first described CdLS as a distinct syndrome (Badoe, 2006), but it wasn't until 2004 that the first causal gene for CdLS was reported in back-to-back publications. The breakpoints of a de novo balanced translocation, found in a patient with CdLS, were mapped to 5p13.1 and 13q12.1. This identified a new gene - named Nipped-B-like (*NIPBL*) -, found to also carry point mutations in a number of CdLS patients. In parallel to this, genome wide linkage exclusion analysis in 12 families with CdLS identified mutations in *NIPBL* in 6 of the patients studied (Krantz *et al.*, 2004).

NIPBL mutations are not identified in all CdLS patients – only about half of all CdLS cases are caused by *NIPBL* loss of function (LOF) mutations (Zuin *et al.*, 2014), although the total percentage may be higher due to the presence of somatic mosaicism for *NIPBL* mutations in around 23% of mutation negative individuals (Huisman *et al.*, 2013; Ansari *et al.*, 2014). This, combined with the variability of the CdLS phenotype, suggests genetic heterogeneity (Revenkova *et al.*, 2009).

1.1.2.2 Cohesin subunits

NIPBL is the ortholog of Scc2-type sister chromatid cohesion proteins in yeast and Nipped-B in D. melanogaster (Tonkin *et al.*, 2004). Scc2 regulates the loading and unloading of the cohesin and condensin complex (Deardorff *et al.*, 2007). This led others to look for mutations in other members of the cohesin complex.

Mutations (largely missense) in the cohesin components *SMC1A* (Musio *et al.*, 2006), *SMC3* (Deardorff *et al.*, 2007) and the kleisin subunit *RAD21* (Deardorff, Wilde, *et al.*, 2012), and the cohesin regulator *HDAC8* (Deardorff, Bando, *et al.*, 2012) were also found to cause CdLS-like phenotypes and are responsible for ~10-15% of cases.

1.1.2.3 Transcriptional regulators

Recently, a number of patients have been reported that have CdLS-like phenotypes and mutations in genes that are involved in transcriptional regulation, including *AFF4, ANKRD11*, *KMT2A* and *TAF6*. This is discussed in more detail in section 1.2.8.4.

1.1.2.4 Genotype-phenotype correlation

Mutations in *NIPBL* cause the most typical and severe CdLS phenotypes, with mutations in the other genes contributing to a smaller number of milder cases – often lacking the major limb malformations and showing atypical growth or facial phenotypes (Ansari *et al.*, 2014)(Figure Ch1-1). A genotype-phenotype correlation also exists within *NIPBL* mutation positive cases; the most common *NIPBL* mutations in CdLS are nonsense, splice site, or frame shifts that result in truncation of the protein and NIBPL haploinsufficiency. These are associated with the more severe phenotypes. Missense mutations also occur, but are less frequent and cause milder phenotypes with the absence of limb abnormalities and less severe developmental delay (Gillis *et al.*, 2004; Mannini *et al.*, 2013; Boyle *et al.*, 2015).

Figure Ch1-2. Phenotypic spectrum of NIPBL mutations. Patients with missense mutations have the mildest symptoms, whilst those with frameshift/nonsense mutations are most severe. Taken from Boyle *et al.*, 2015.

1.2 The cohesin complex

1.2.1 Structural maintenance of chromosome complexes

Structural maintenance of chromosomes (SMC) complexes are ring like structures, conserved from bacteria to humans, which can entrap DNA. The most well-known SMC complexes are cohesin and condensin, which regulate DNA loop formation (van Ruiten and Rowland, 2018), mitotic chromosome condensation, sister chromatid cohesion, DNA repair, and transcriptional regulation (Murayama *et al.*, 2018).

The cohesin complex was first discovered in yeast for its role in sister chromatid cohesion, and this is thought of as its canonical role. However it also has non-canonical roles in the regulation of gene expression, chromatin organisation and the DNA damage response (DDR) (Bettini *et al.*, 2018).

1.2.2 Structure

The cohesin complex (Figure Ch1-3) is composed of four subunits, known in vertebrates as SMC1, SMC3, RAD21 and STAG-1/-2 (Solomon, Kim and Waldman, 2014). The SMC proteins fold in half at 'hinge' domains, so that their N- and C-termini meet at one end, forming globular ATPase domains (Revenkova *et al.*, 2009). These so called 'head' domains are separated from the hinge by antiparallel coiled coils. SMC1 and SMC3 hetero-dimerise by sandwiching two ATP molecules between them. The head domains are connected by RAD21, which binds to the SMC3 head at its N-terminal domain and the SMC1 head at its C-terminal domain (Makrantoni and Marston, 2018) (Figure Ch1-3). This forms a tripartite ring-structure that is thought to 'embrace' DNA strands (Zuin *et al.*, 2014).

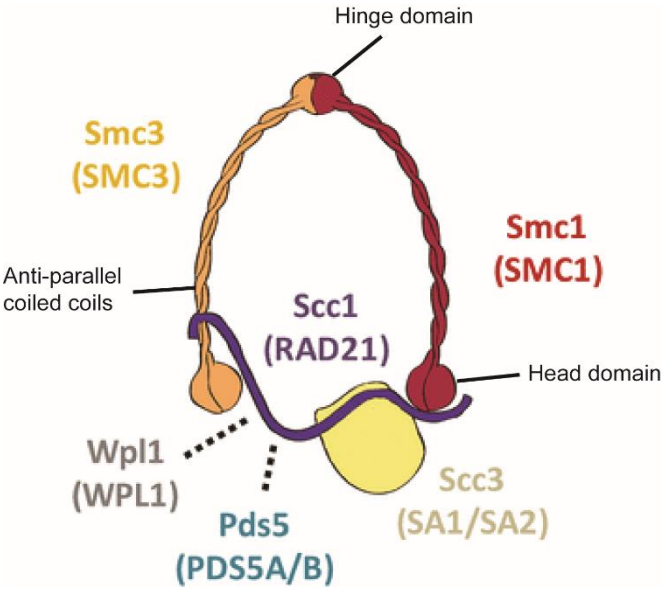


Figure Ch1-3. Cartoon of the cohesin complex in human somatic cells. SMC1 and SMC3 fold in half at a 'hinge' domain, where they interact to form a heterodimer. The hinge domain is separated from the N- and C-terminal globular 'head' domains by anti-parallel coiled coils. RAD21 interacts with the head domain of SMC1 and the coiled-coil of SMC3, forming a tripartite ring-structure, and with SA1/SA2. PDS5 and WAPL are involved in the regulation of cohesin-chromatin binding. Adapted from (Villa-Hernández and Bermejo, 2018).

1.2.3 DNA binding

Cohesin is loaded onto chromatin during G1 phase, and this requires the loading factors NIPBL and MAU2, which together form a complex known as Kollerin (Zuin *et al.*, 2014; Elbatsh *et al.*, 2016). The binding of the RAD21 C-terminus to the SMC1 head domains induces a rearrangement of the SMC1a active site (AS1) (Figure Ch1-4A). This allows ATP hydrolysis, and in turn induces the ATPase activity of the SMC3 head domain (AS2) (Figure Ch1-4B). Once ATP hydrolysis has occurred at both active sites (Figure Ch1-4C), the ATPase heads separate and allow DNA to pass through the open structure (Figure Ch1-4D) (Marcos-Alcalde *et al.*, 2017).

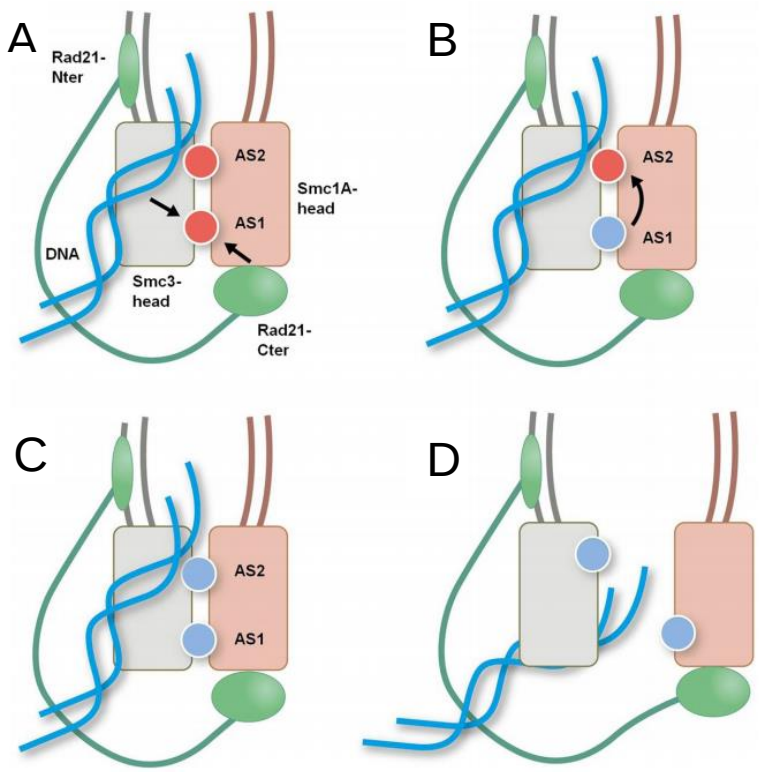


Figure Ch1-4. Schematic model for ATP hydrolysis-driven head opening. Binding of the Rad21 C-terminus to the Smc1 head domain allows hydrolysis at AS1 **(A).** This induces a rearrangement and activation of AS2 **(B)**, leading to ATP hydrolysis **(C).** ATP hydrolysis at AS1 and AS2 facilitates the separation of the ATPase head domains **(D).** Taken from Marcos-Alcalde *et al.*, 2017.

In the simplest 'ring' model of cohesin's interaction with DNA, there are three functional states that may occur – 1. The non/pseudo-topological configuration (Figure Ch1-5A), in which DNA is 'extruded' through the cohesin ring until it reaches boundary sites (thought to be CCCTC-binding factor (CTCF) binding sites) (Makrantoni and Marston, 2018) (see section 1.2.7.1). This directs chromosomal organisation by forming DNA loops and can regulate gene expression by bringing linearly distal genomic regions into close proximity (Bouwman and de Laat, 2015). 2. The one-DNA-topological configuration (Figure Ch1-5B). This configuration posits that the cohesin ring entraps a single dsDNA, thought to be the product of cohesin loading and the precursor for establishment of sister chromatid cohesion. 3. The two-DNAs-topological configuration (Figure Ch1-5C), where the cohesin ring entraps two dsDNA molecules, providing cohesion between them (Makrantoni and Marston, 2018).

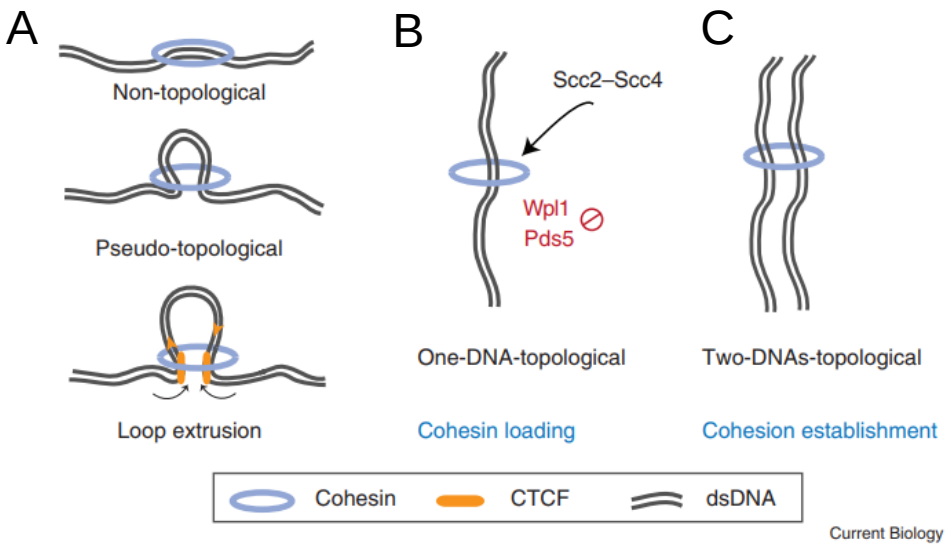


Figure Ch1-5. Schematic model for three functional states of cohesin-DNA interactions. **A)** The 'non/pseudo-topological' cohesin-DNA interaction is thought to mainly drive loop extrusion, **B)** The 'one-DNA-topological' configuration is thought to be the product of cohesin loading, **C)** The 'two-DNAs-topological' form is thought to be the final product of cohesion. Taken from Makrantoni and Marston, 2018.

The 'two ring handcuff' model has also been suggested, but remains to be thoroughly tested (Wendt, 2017). In this model two RAD21 molecules move into antiparallel orientation and are paired and tethered through their interactions with STAG1/2. Each ring entraps one sister chromatid, holding them together (N. Zhang *et al.*, 2008).

1.2.4 NIPBL

1.2.4.1 Structure

NIPBL consists of 47 exons that encodes a large multi-domain protein (Mannini *et al.*, 2013). In higher eukaryotes, there are two isoforms - the canonical isoform (Isoform A) which is 316 kDa and the slightly smaller 304 kDa isoform (Isoform B) (Jahnke *et al.*, 2008). NIPBL's domains include HEAT repeats, a HP1 interacting motif, a bipartite nuclear localisation sequence (NLS), a undecapeptide repeat, a coiled-coil, a glutamine rich region and a MAU2 (Scc2) interacting region (Figure Ch1-6) (Mannini *et al.*, 2013). HEAT repeats are flexible arrays of helices, made up of both hydrophobic and hydrophilic parts. They are found in many eukaryotic proteins and enable interactions with many different binding partners (Yoshimura and Hirano, 2016). The C-terminal HP1 interacting motif and HEAT repeats are required for the recruitment of NIPBL to sites of DNA damage (Oka *et al.*, 2011) whilst

NIPBL's interaction with MAU2 is essential for the loading of cohesin at the G1 phase of the cell cycle (Zuin *et al.*, 2014).

CdLS-causing mutations occur throughout NIPBL, although the majority are found in coding regions. Exon 10 (marked on Figure Ch1-6), encoding the coiled-coil region and the undecapeptide repeat, harbours the largest number of CdLS mutations (49/278 as of 2013) (Mannini *et al.*, 2013).

1.2.4.2 Kollerin (NIPBL-MAU2)

NIPBL forms the Kollerin complex with MAU2 (Scc2-Scc4), and this is essential for loading cohesin onto chromatin (Wendt, 2017). Surprisingly, *MAU2* heterozygous mice are indistinguishable from WT, suggesting that it may play a different role to NIPBL in gene regulation, or is less sensitive to changes in dosage (Singh and Gerton, 2015). MAU2 is a tetratricopeptide repeat (TPR) superhelix, which entraps the N-terminus of NIPBL. The crystal structure of the Scc2-Scc4 complex shows that elongated Scc4 twists around a fragment of Scc2, which emerges at both ends of the structure (Hinshaw *et al.*, 2015). The complex is made up of a head, a body and a hook domain.

Scc2 displaces Pds5 from cohesin, in order to bind to Scc1 (RAD21) and stimulate cohesin loading onto chromatin. Scc2 also stimulates the ATPase activity of the SMC head domains, catalysing translocation of cohesin along DNA, and may continue to regulate the ATPase activity after loading (Petela *et al.*, 2018).

Kollerin may also play a role in the loading of the condensin complex, which contains the SMC proteins SMC2 and SMC4, onto chromatin, although this is likely indirect (Visnes *et al.*, 2014).

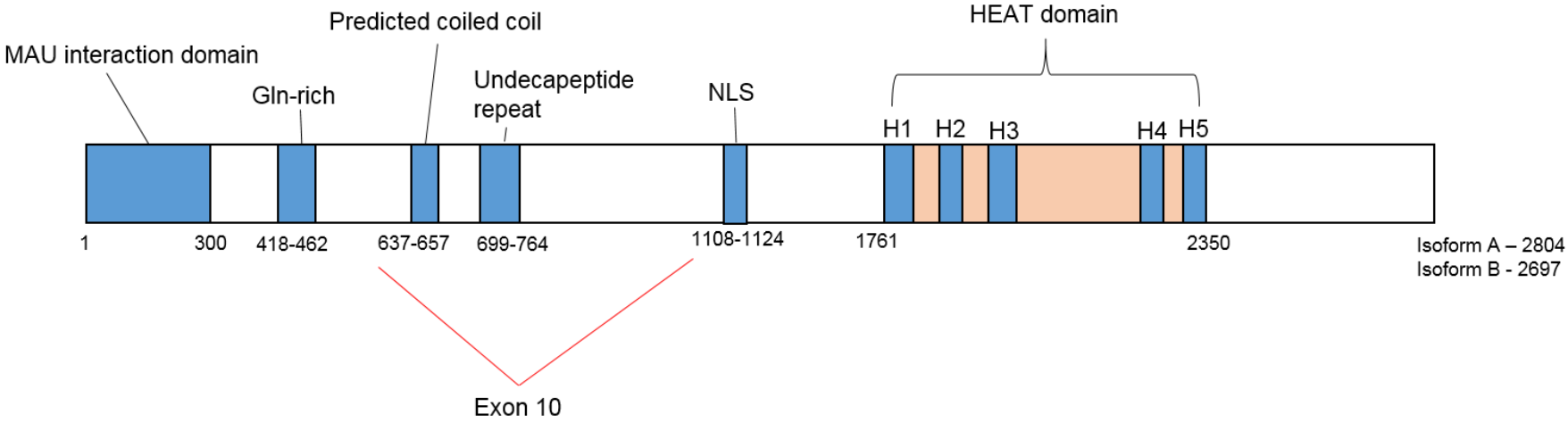


Figure Ch1-6. NIPBL is a multi-domain protein that is often mutated in CdLS. Schematic of human NIPBL protein structure. Domains are labelled: MAU interaction domain (1–300), glutamine rich domain (Gln-rich, 418–462, predicted coiled coil (637–657), undecapeptide repeat (699–764), NLS (1,108–1,124), HEAT domain (1,767–2,350) consisting of five repeats (H1: 1,767–1,805, H2: 1,843–1,881, H3: 1,945–1,984, H4: 2,227–2,267, H5: 2,313–2,351). Location of exon 10 (in which the majority of CdLS-causing mutations occur) is labelled.

1.2.5 Genomic distribution of NIPBL/cohesion

1.2.5.1 Drosophila and Yeast

Many studies have aimed to determine the genomic localisation of Kollerin, and thus of cohesin loading. This appears to differ between species however, and the exact location of cohesin loading remains unknown (Wendt, 2017). In Drosophila, Nipped-B and cohesin colocalise throughout the genome, preferentially at transcribed regions and overlapping with RNA polymerase II (RNAPII) (Misulovin *et al.*, 2008). This contrasts with yeast, where cohesin is loaded at Scc2 sites and pushed along chromosomes by transcribing RNAPII to accumulate at convergently oriented genes (Borrie *et al.*, 2017).

1.2.5.2 Mammals

In mammalian cells cohesin positioning is thought to be determined by CTCF - an insulator protein that blocks enhancer activity and/or inhibits the spread of heterochromatic domains (Ball, Chen and Yokomori, 2014), – transcription, and WAPL (Busslinger *et al.*, 2017).

RAD21 ChIP-seq showed that around 50% of cohesin peaks overlap with CTCF (Wendt, 2017), and are thought to be involved in the regulation of higher order chromatin structure. CTCF independent cohesin peaks are found at cis-regulatory elements (CRE) of active genes and overlap with RNAPII, transcription factors (TFs), the Mediator complex and NIPBL (Kagey *et al.*, 2010; Mannini *et al.*, 2015) (Figure Ch1-7A).

NIPBL is required for the loading of cohesin onto chromatin (Wendt, 2017), but is often not present at CTCF-cohesin sites (Figure Ch1-7B), suggesting that cohesin is repositioned after DNA binding. Loss of CTCF causes accumulation of cohesin at the transcription start sites (TSS) of active genes, whilst in the absence of both CTCF and WAPL – a protein involved in removing cohesin from chromatin - cohesin was shown to accumulate at the 3' end of active genes. These data suggest that mammalian cohesin can be relocated by transcription, similarly to yeast cohesin, and removed from active genes by WAPL mediated release (Busslinger *et al.*, 2017).

In mouse embryonic stem cells (mESCs) NIPBL was originally found to co-localise with cohesin (Kagey *et al.*, 2010) (Figure Ch1-7A), however ChIP-sequencing in human cells found that this is not always the case; a subset of low-enrichment NIPBL peaks overlapped

with those of cohesin, but a larger subset of 'major' peaks occurred at promoters of actively transcribed genes that were devoid of cohesin (Figure Ch1-7B) (Zuin *et al.*, 2014).

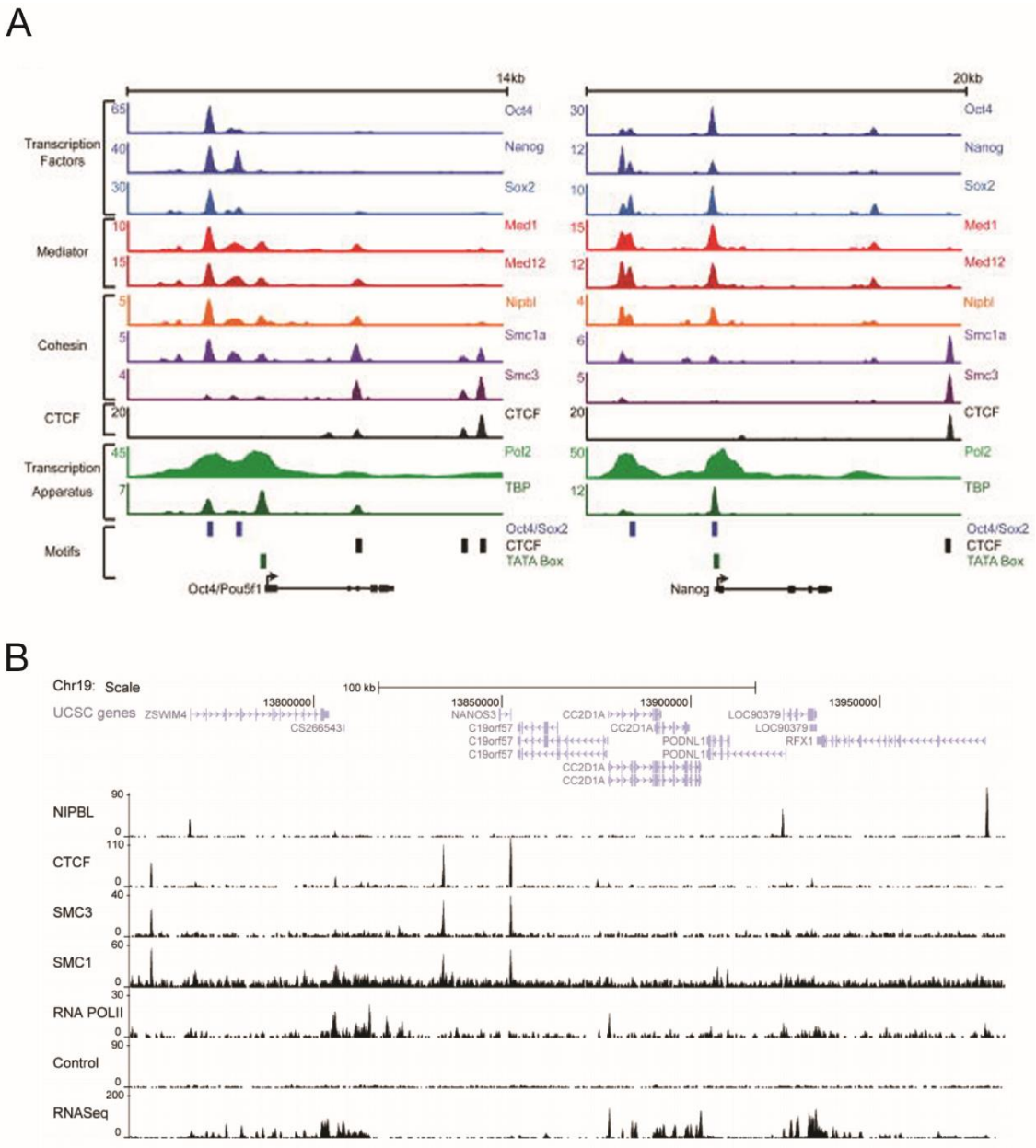


Figure Ch1-7. NIPBL binding at CTCF and non-CTCF cohesin sites. A) mESC ChIP-sequencing tracks from Kagey *et al.* 2010 show the co-localisation of NIPBL, SMC3, SMC1A, Mediator, TFs and the transcriptional apparatus at non-CTCF cohesin sites. Some co-localisation of these factors is observed at CTCF cohesin sites, but it is much decreased. **B)** ChIP-sequencing tracks from Zuin *et al.* 2013 show a lack of co-localisation between NIPBL and cohesin at CTCF bound cohesin sites in human breast endothelial cell line HB2. NIPBL instead binds at cohesin free gene promoters.

1.2.6 Sister chromatid cohesion during the cell cycle

1.2.6.1 Sister chromatid cohesion during mitosis

During mitosis sister chromatids must be correctly distributed between the resulting daughter cells (Uhlmann, 2004). Efficient loading of sister chromatids onto the mitotic spindle (Uhlmann, 2004), prevention of premature chromatid separation (PCS), and accurate segregation of chromatids into new daughter cells, all require the physical tethering of sister chromatids to each other, a process mediated by the cohesin complex (Zheng and Yu, 2015).

Cohesin binding is enriched at centromeres and nearby pericentromeric regions. Its targeted loading at centromeres requires phosphorylation of a kinetochore component, yeast Ctf19 (CCAN in humans), by DDK, a cell cycle regulated kinase. Phosphorylated Ctf19 provides a binding site for Scc2/4 at kinetochores by interacting directly with a conserved region of Scc4 that is required to recruit Scc2/4 to centromeres (Hinshaw *et al.*, 2017).

When cohesin rings are not in their cohesive state, they transiently associate with chromatin through the continuous entrapment and release of DNA (Elbatsh *et al.*, 2016). At this stage, the action of cohesin loading is actively opposed by PDS5-WAPL mediated release, keeping the chromatin bound cohesin in a highly dynamic state (Zheng and Yu, 2015). During S phase, the ESCO2 dependent acetylation of SMC3 antagonises the activity of PDS5-WAPL, resulting in stably bound cohesin that mediates cohesion between sister chromatids (Elbatsh *et al.*, 2016). In vertebrates the displacement of WAPL-PDS5 by Sororin also contributes to the inhibition of cohesin release from chromatin (Nishiyama *et al.*, 2010).

Regulated removal of cohesin from the sister chromatids is essential for successful segregation during mitosis (Zheng and Yu, 2015). In prophase, cohesin is first removed from the chromosome arms by the WAPL-mediated opening of the SMC3-RAD21 gate (Beckouët *et al.*, 2016). The SMC1 ATPase domain appears to be required for this removal, with mutations in this domain causing stable cohesion even in the absence of ESCO2 (Elbatsh *et al.*, 2016). Centromeric cohesin is maintained by the protective role of Shugoshin (Haarhuis *et al.*, 2013) and is important for the correct orientation of sister chromatids on the mitotic spindle during metaphase (Tanaka *et al.*, 2000).

Correct attachment of the kinetochores to the microtubules of the spindle silences the spindle checkpoint and activates a protein called Separase. At the same time, Shugosin is redistributed, exposing the centromeric cohesin and allowing Separase to cleave RAD21

(Hauf, Waizenegger and Peters, 2001; Zheng and Yu, 2015). This leads to the removal of centromeric cohesin and triggers the onset of anaphase (Hauf, Waizenegger and Peters, 2001.) HDAC8 is responsible for the deacetylation of the SMC3 subunit, imperative for the recycling of cohesin so it can be reloaded in the next cell cycle (Dasgupta *et al.*, 2016).

1.2.6.2 Sister chromatid cohesion during meiosis

Cohesin is also essential for meiosis. Mammalian meiotic cells contain four meiosis specific cohesin subunits – the kleisin subunits REC8 and RAD21L, the SMC protein SMC1B, and STAG3 – in addition to the five ubiquitously expressed subunits. Varying combinations of subunits are known to form at least six different meiosis specific cohesin complexes, but the precise differences in their roles are not well understood (Biswas *et al.*, 2016).

In meiotic cells, cohesin along the chromosome arms, and physical connections between homologous chromosomes known as chiasma, generate tension between homologous chromosomes and ensures their faithful segregation. At anaphase I, chromosome arm cohesion is destroyed, causing the homologous chromosomes to separate to opposite poles of the cell. Pericentromeric and centromeric cohesin is protected at this stage, ensuring that the physical linkage of sister chromatids is maintained until anaphase of meiosis II (Kitajima, Kawashima and Watanabe, 2004). Furthermore the presence of cohesin around the centromeres acts to repress crossover recombination in these regions, which could otherwise result in the non-disjunction of sister chromatids in meiosis II (Vincenten *et al.*, 2015).

1.2.6.3 Sister chromatid cohesion in CdLS

Developmental disorders caused by mutations in cohesin or its accessory factors are known as cohesinopathies (Zakari, Yuen and Gerton, 2015). Aside from CdLS the most classic example of a cohesinopathy is Roberts Syndrome (RBS) (OMIM 268300). Patients with RBS have mutations in both alleles of *ESCO2,* and present with symptoms similar to those of CdLS, such as craniofacial anomalies, growth retardation and limb malformation (Vega *et al.*, 2005). Loss of cohesion is known to cause PCS, mis-segregation, and aneuploidy in resulting daughter cells (Peters and Nishiyama, 2012). Whilst these cellular perturbations are features of RBS, they have not been observed in CdLS patients; metaphase spread analysis showed that PCS occurred at a similar frequency in CdLS patients and controls (Castronovo *et al.*, 2009).

Nipbl transcripts levels are reduced by only ~30% in *Nipbl* heterozygous mice (Newkirk *et al.*, 2017), and in *NIPBL* mutation positive CdLS patients (Zuin *et al.*, 2017), most likely due to upregulation of the intact allele (Newkirk *et al.*, 2017). This appears to be sufficient to maintain the cohesin fraction responsible for robust sister chromatid cohesion (Remeseiro *et al.*, 2013), indicating that the canonical function of cohesin is not affected by the mutations that cause CdLS.

1.2.7 The role of cohesin in chromatin structure and organisation

1.2.7.1 Genome compartmentalisation

Hi-C, a high throughput variant of the chromosome conformation capture (3C) techniques, revealed that the genome is organised into areas of active and inactive chromatin, known respectively as A and B compartments, and further compartmentalised into areas of highly interacting chromatin known as topologically associating domains (TADs) (Gong *et al.*, 2018). TADs can range from a few 100 Kb to 5 Mb in size, have sharp boundaries, and are highly conserved between cell types and throughout evolution, suggesting that they represent the fundamental unit of physical organisation of the genome (Rocha *et al.*, 2015). Peaks (or loops) are visible as focal enrichments in contact frequency between pairs of loci and often form at the corners of TADs (Schwarzer *et al.*, 2017).

Compartmentalisation is thought to be established by the local chromatin state, whilst loop extrusion, regulated by CTCF and cohesin, is the preferred hypothesis for TAD formation (Nuebler *et al.*, 2018). In the loop extrusion model, cohesin is loaded onto chromatin at sites distinct from CTCF (Barrington, Finn and Hadjur, 2017) and translocates along DNA until it reaches CTCF. CTCF is proposed to function as a boundary element that acts as a barrier, impeding further translocation (Barrington, Finn and Hadjur, 2017). This provides time for WAPL-mediated removal of cohesin; in the absence of WAPL, cohesin continues to translocate past TAD boundaries (Nuebler *et al.*, 2018).

Multiple mechanisms exist for how the translocation of cohesin forms loops, and these are compatible with both the simple ring model of cohesin binding and the two-ring handcuff model (see section 1.2.3). Cohesin may bind at a small genomic region and translocate in two directions until it reaches CTCF sites (Figure Ch1-8A). Cohesin may also bind to regions that are physically proximal but linearly distant, resulting in the formation of larger domains (Figure Ch1-8B). However, the combination of these two alternatives is most likely to produce the observed multiple domains of genome structure – long-range TADs that are subdivided into smaller loops (Barrington, Finn and Hadjur, 2017) (Figure Ch1-8C). The translocation of cohesin is likely mediated by ATP hydrolysis, stimulated by interaction of cohesin with NIPBL (Petela *et al.*, 2018).

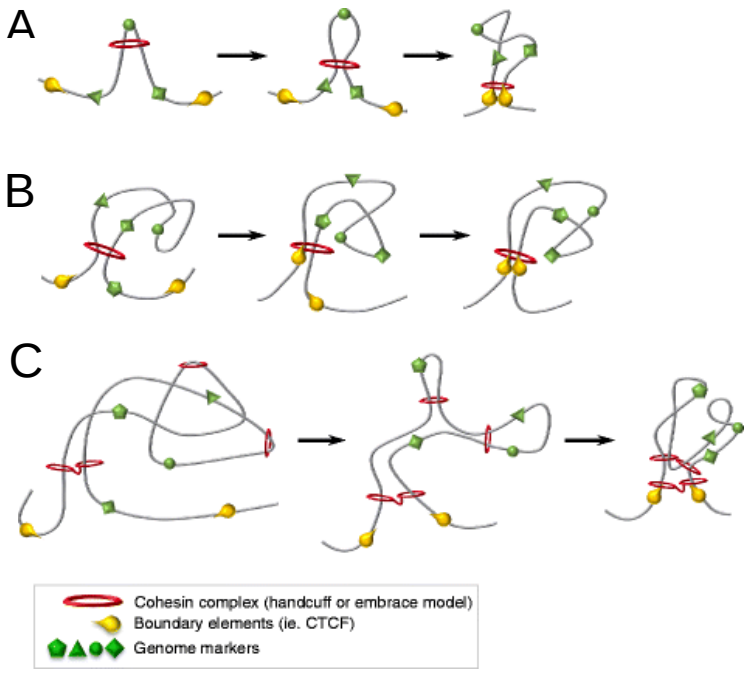


Figure Ch1-8. Possible models for domain formation via loop extrusion. A) Cohesin binds within a small genomic region and translocates in opposing directions until it reaches CTCF. **B)** Cohesin binds to two spatially proximal but genomically distant regions, resulting in larger domains. **C)** A combination of **A** and **B** occurs, forming both small and large domains. Taken from Barrington, Finn and Hadjur, 2017

CTCF and cohesin co-localise at the anchors of TAD boundaries (Rao *et al.*, 2017). The loss of NIPBL (and consequently chromatin bound cohesin) leads to a decrease in local TAD patterns (Figure Ch1-9A), without affecting CTCF binding, suggesting that the roles of cohesin and CTCF in loop extrusion are distinct (Schwarzer *et al.*, 2017). Auxin induced

degradation of CTCF causes the loss of TAD insulation – increasing inter-TAD interactions and decreasing intra-TAD interactions (Figure Ch1-9B), supporting the role of CTCF as a boundary element (Nora *et al.*, 2017). WAPL depletion leads to an extension of chromatin loops and the accumulation of contacts at TAD corners (Figure Ch1-9C). This suggests that chromatin loops are formed through the processive enlargement of smaller loops - with TADs reflecting polyclonal populations of loops forming between two CTCF sites - and removal of cohesin by WAPL allows TADs to stay dynamic (Haarhuis *et al.*, 2017). A and B compartmentalisation is not affected by NIPBL or CTCF depletion (Nora *et al.*, 2017; Schwarzer *et al.*, 2017), but is slightly less defined in WAPL deficient cells; Loss of WAPL increased the switching between A and B compartments, indicating that cohesin acts to limit nuclear compartmentalisation (Haarhuis *et al.*, 2017).

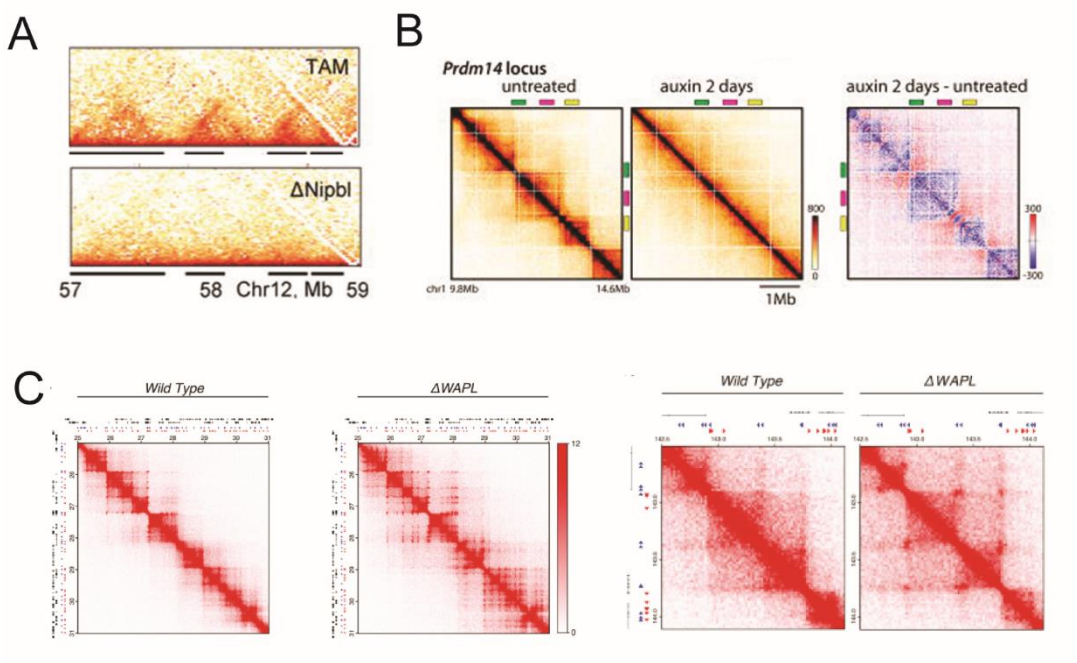


Figure Ch1-9. Effect of NIPBL, CTCF and WAPL loss on TAD patterning. **A)** NIPBL depletion (through a tamoxifen (TAM) inducible Cre driver) causes loss of local TAD patterns compared to a TAM control. Hi-C map shows a 2Mb region (top – TAM control, bottom – NIPBL deleted). Black bars show WT TADs. Taken from Schwarzer *et al.*, 2017. **B)** Auxin induced CTCF degradation causes a loss of TAD insulation. Left - Snapshot of Hi-C data at the *Prmd14* locus in untreated (left) and auxin treated (right) cells. Right - differential contact map showing more inter-TAD (red) and fewer intra-TAD (blue) Hi-C signal after auxin treatment. Taken from Nora *et al.*, 2017. **C)** Loss of WAPL causes the extension of loops (left) and the accumulation of contacts at TAD corners (right). Hi-C contact matrices for a zoomed in region on Chr7 (left) and Chr2 (right) for WT and ΔWAPL. Contact matrices are normalised to 100 million contacts, shown resolution is 20 kb. Taken from Haarhuis *et al.*, 2017.

1.2.7.2 Chromatin structural changes in CdLS

Given the role of cohesin and NIPBL in chromatin interactions, it could be expected that mutations in the proteins that cause CdLS, might affect chromatin structure. Indeed, fluorescent in situ hybridisation (FISH) assays revealed a visible decompaction of chromatin in LCLs from CdLS patients positive for *NIPBL* mutations, and upon NIPBL siRNA (Nolen *et al.*, 2013) - consistent with Hi-C data after NIPBL depletion (Schwarzer *et al.*, 2017). This decompaction was wide-spread, but correlated with gene density and the severity of the *NIPBL* mutations. Knockdown of SMC3 had a smaller effect on decompaction however, suggesting that the decompaction may result from perturbation of a cohesin independent role of NIPBL (Nolen *et al.*, 2013).

Overlap between the clinical phenotypes of CdLS and syndromes caused by mutations in chromatin remodelling complexes also suggest that changes in chromatin structure could underlie CdLS. Coffin-Siris syndrome (CSS, OMIM #135900) and Nicolaides-Baraitser syndrome (NCBRS, OMIM #601358) have similar phenotypes to CdLS and are caused by mutations in the SWI/SNF complex, a chromatin remodeller that shares some functions of cohesin. In yeast, Scc2 is known to interact with the SWI/SNF complex to maintain nucleosome free regions (Lopez-Serra *et al.*, 2014). Furthermore, the analysis of seven patients with CdLS-like phenotypes but no mutation in the known causative genes, found that 4/7 patients had mutations in the SWI/SNF complex (Parenti *et al.*, 2017).

1.2.8 Cohesin and the regulation of gene expression

Cohesin regulates gene expression in a number of species, including mammals, drosophila and yeast, with depletion of cohesin and NIPBL resulting in changes in transcription (Lopez-Serra *et al.*, 2014; Zuin *et al.*, 2014; Y. Wu *et al.*, 2015; Mills *et al.*, 2018).

1.2.8.1 TADs

The formation of TADs, regulated by cohesin and CTCF (see section 1.2.7), is thought to be of functional importance for regulation of gene expression. Sequences within TADs interact with high frequency and show similarities in expression and histone marks (Bouwman and de Laat, 2015). Furthermore TADs are thought to regulate cohesin mediated interactions between enhancers and promoters, by preventing interactions across TAD boundaries (Lupiáñez *et al.*, 2015).

However, loss of TADs upon NIPBL/cohesin depletion does not correlate with changes in gene expression; Hi-C maps show that TAD patterns vanish equally in regions with unchanged gene expression, or with up- and down-regulated transcription, when NIPBL is depleted (Schwarzer *et al.*, 2017). Furthermore, whilst cohesin deletion eliminates all loop domains, it only affects the expression of a subset of genes. If cohesin regulated gene expression via TADs, one might expect to see a global change in gene expression, caused by ectopic interactions between enhancers and targets from other loop domains, (Rao *et al.*, 2017). These results suggest that regulation of gene expression by cohesin is largely independent of TAD patterning.

1.2.8.2 Enhancer-promoter looping

Cohesin-mediated looping regulates gene expression independently of CTCF. CTCF independent cohesin was found to colocalise with the Mediator complex at the anchoring sites of enhancer-promoter interactions of active genes (Kagey *et al.*, 2010). Mediator is a large multi-subunit cofactor complex, which controls the transcriptional programs of most genes in eukaryotic cells, through interactions with RNAPII and other pre-initiation complex (PIC) components (Allen and Taatjes, 2015).

Mediator had been shown to bind at enhancers, and transiently interact with promoters. However, recent evidence that a single Mediator complex occupies the enhancer and promoter of any given gene, has led to the idea that Mediator interacts with the PIC bound at promoters, and forms transient chromatin loops that link enhancers to core promoters (Petrenko *et al.*, 2016). Cohesin localisation at these sites may act to stabilise the loops (Allen and Taatjes, 2015) (Figure Ch1-10).

Other models for transcriptional activation through promoter-enhancer interactions have been proposed, such as the tracking model, where enhancer-bound transcription complexes, including active RNAPII, move towards the target promoter in a unidirectional manner. These enhancer-bound proteins may not leave the enhancer, bringing the enhancer to the promoter. It is thought that different mechanisms of enhancer-promoter interactions may regulate expression of different genes (Meng and Blaine, 2017)

Cohesin and Mediator localisation differs between different cell types, suggesting that they may be important for cell-type specific gene expression. For instance, in mESCs cohesin and Mediator occupy the promoters of genes such as Oct4 and Nanog, which are essential for pluripotency. Loss of cohesin or Mediator caused reduced OCT4 protein levels, loss of ES cell colony morphology, and changes in transcription, consistent with differentiation, suggesting they are important in maintaining a stem-cell like state (Kagey *et al.*, 2010).

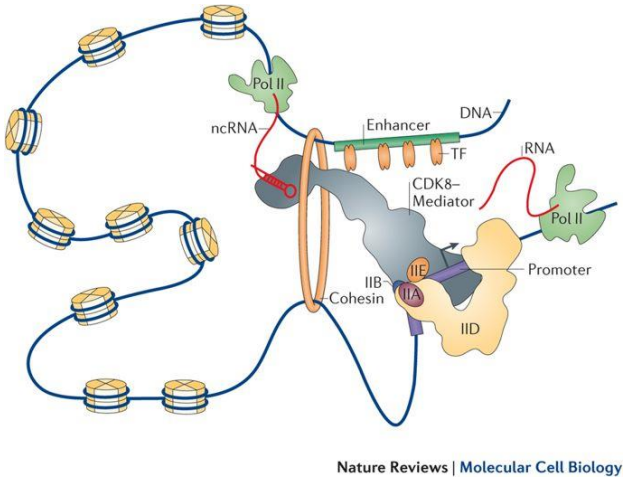


Figure Ch1-10. Mediator regulates the formation of transient chromatin loops that link enhancers to core promoters. Interactions between Mediator TFs and cohesin stabilise enhancer–promoter loops. Taken from Allen and Taatjes, 2015.

1.2.8.3 NIPBL and gene regulation

NIPBL co-localises with Mediator and cohesin at many sites, suggesting that it is also involved in gene expression through enhancer-promoter looping. However it is possible that NIPBLs role at these sites is simply to load cohesin (Kagey *et al.*, 2010). As discussed in section 1.2.5.2, NIPBL also binds at a number of cohesin-free sites. Interestingly, depletion of NIPBL affected the transcription of the genes associated with these sites, whereas depletion of SMC3 had little effect, suggesting that NIPBL can regulate gene expression independently of its role in cohesin loading (Zuin *et al.*, 2014). This may explain the more severe CdLS phenotype seen with *NIPBL* mutations (Figure Ch1-1)(Newkirk *et al.*, 2017); in *NIPBL*-deficient patients both cohesin dependent and independent mechanisms of transcriptional regulation may be perturbed.

1.2.8.4 Transcriptional changes in CdLS

Dysregulation of gene expression is now the favoured mechanism for CdLS, and has been implicated more generally as a cause of cohesinopathies (Bose and Gerton, 2010). Many studies have shown altered gene expression profiles in cells from CdLS patients. These studies use mainly differentiated cell lines such as LCLs and report subsets of dysregulated genes (~300-500 genes), rather than global changes, and a correlation with cohesin binding (Liu *et al.*, 2009; Mannini *et al.*, 2015; Yuan *et al.*, 2015; Boudaoud *et al.*, 2017; Mills *et al.*, 2018). Mills *et al.*, (2018) found that the genes dysregulated in *NIPBL*$^{+/-}$ patient-derived cardiomyocytes were involved in processes such as chromatin modifications, nucleosome assembly and transcriptional regulation, and reduced immunological functions, regulation of apoptosis and proteasome dynamics (Mills *et al.*, 2018).

The profiles of transcriptional dysregulation have been shown to correlate with patient phenotype severity, and similar transcriptional profiles were observed between NIPBL truncating mutations from CdLS patients, and two patients with RBS, consistent with both disorders being caused by mutations in the cohesin complex (Liu *et al.*, 2009).

Given the proposed roles of cohesin and NIPBL in chromatin loop formation, and the loss of loop domains in the absence of cohesin (Rao *et al.*, 2017), I would expect that these changes in transcription occur through decreased interactions between promoters and their distal enhancers. Indeed, it's been shown that mutant cohesin reduces both RNAPII recruitment to promoters and the levels of elongating RNAPII, leading to decreased transcription initiation and elongation (Mannini *et al.*, 2015). Furthermore, the transcriptional changes that occur upon NIPBL deletion were found to be concentrated within regions that form larger TADs in WT, consistent with defective long-range regulatory interactions in the mutant cells (Schwarzer *et al.*, 2017).

Recent studies have identified a number of transcription related genes whose mutation can also cause CdLS-like phenotypes. Missense mutations in *AFF4* were identified in three unrelated patients with a phenotype overlapping that of CdLS. AFF4 is a component of the super elongation complex - a complex that strongly activates transcription elongation by RNAPII and is known to interact with the cohesin complex (Izumi *et al.*, 2015).

Three patients with CdLS-like phenotypes were found to have de novo mutations in *ANKRD11*. These consist of a frameshift mutation, an intragenic deletion and a

heterozygous LOF mutation (Ansari *et al.*, 2014). ANKRD11 is an ankyrin repeat-containing cofactor that represses ligand-dependent transcriptional activation and mutations in this gene are known to cause KBG syndrome (Sirmaci *et al.*, 2011).

There is also some overlap between CdLS and Wiedemann-Steiner syndrome (WDSTS, MIM # 605130), a syndrome caused by mutations in the histone H3 methyltransferase, KMT2A. Methylation mediated by KMT2A is associated with epigenetic transcriptional activation important for the regulation of embryogenesis, cell fate determination, cell-cycle progression and stem-cell function. Mutations in *SMC1/SMC3* contribute to phenotypes resembling WDSTS, whilst a mutation in *KMT2A* was identified in a patient with CdLS (Yuan *et al.*, 2015), indicating a possible role of histone modifications in CdLS.

1.3 Experimental models of CdLS

1.3.1 Mouse

1.3.1.1 Viability

The core cohesin subunits have all been shown to be indispensable for early embryonic development in mice. Homozygous knockout of *Smc3* or *Rad21* are embryonic lethal prior to E14.5, whilst *Smc1a* mutant mice have not been reported, suggesting that they are not viable. *Smc3* heterozygous mice show phenotypes similar to CdLS patients, including reduced size and craniofacial anomalies. However, *Rad21* heterozygous mice have defects in DNA repair, but no overt developmental phenotypes (Singh and Gerton, 2015).

1.3.1.2 *Nipbl*$^{+/-}$

A mouse model for *NIPBL* haploinsufficiency (*Nipbl*$^{+/-}$), the most common cause of CdLS, was created by introduction of a gene trap into intron 1 of the *Nipbl* gene. *Nipbl*$^{+/-}$ mice present with some of the phenotypic features observed in CdLS patients, including reduced body size (Figure Ch1-11), microbrachycephaly, reduced body fat, craniofacial anomalies and heart defects. These symptoms arise despite *Nipbl* transcript levels decreasing by only ~30%, indicating a large sensitivity to small changes in NIPBL activity during development (Kawauchi *et al.*, 2009).

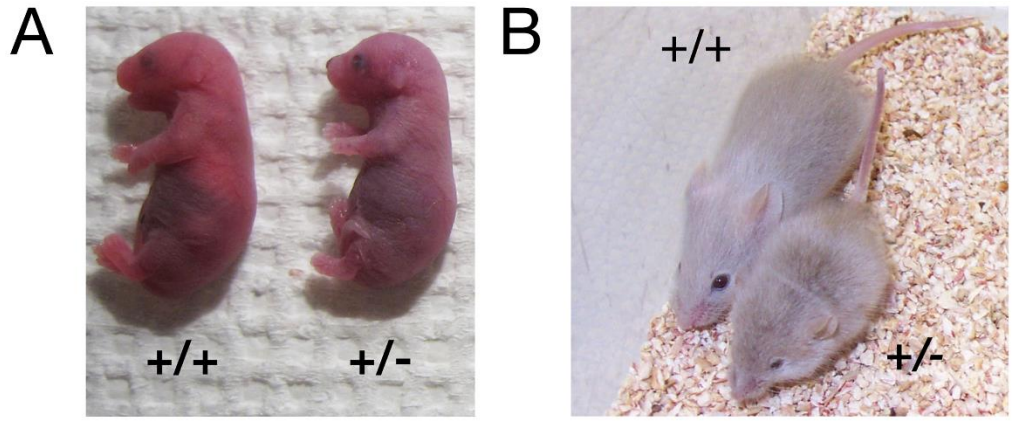


Figure Ch1-11. Reduced body size and growth in Nipbl+/- mice. *Nipbl*$^{+/-}$ mice are smaller than their WT litter-mates at birth **(A)** and at 4 weeks **(B).** Taken from Kawauchi *et al.*, 2009

Using mouse embryonic fibroblasts (MEFs) from *Nipbl*$^{+/-}$ mice, it was shown that *Nipbl* haploinsufficiency causes a global decrease in cohesin binding. Cohesin-bound genes were disproportionately downregulated in the mutant MEFs, accounting for more than half of the significantly dysregulated transcripts, and there was evidence of reduced promoter-enhancer interactions, suggesting that transcription may be affected through disruption of chromatin interactions. Downregulated genes were enriched for those involved in development, providing a direct link between decreased NIPBL/cohesin and the dysregulation of development, which is likely to contribute to the CdLS phenotype (Newkirk *et al.*, 2017).

Nipbl$^{+/-}$ mice have significant reductions in body fat compared to their WT litter-mates, and defective adipogenic differentiation (Kawauchi *et al.*, 2009). Analysis of adipogenesis genes in *Nipbl*$^{+/-}$ MEFs, found that these are often bound by cohesin, and dysregulated compared to WT cells. This suggests that multiple adipogenesis genes are sensitive to *Nipbl* haploinsufficiency (Newkirk *et al.*, 2017), and may explain the reduced body fat phenotype found in patients with CdLS (Kawauchi *et al.*, 2009).

Atrial septal defects (ASD), one of the congenital heart defects (CHDs) found in CdLS patients, were observed in around half of *Nipbl*$^{+/-}$ embryos (Kawauchi *et al.*, 2009). Santos *et al.*, 2016 used the targeted trapping allele, Flip-Excision, to conditionally knockout *Nipbl* in specific tissues. Loss of *Nipbl* in the developing myocardium or endoderm caused ASD at an incidence of ~30%, the same as that caused by global *Nipbl*-deficiency, suggesting non-additive effects of *Nipbl*-deficiency in different cardiac developmental lineages. Conditional *Nipbl* expression in either the developing myocardium or the endoderm, in the context of

global *Nipbl* knockout, rescued ASD, despite the lack of expression in the other tissue. Therefore, *Nipbl* deficiency in just one of these tissues was sufficient for ASD in a *Nipbl*-WT embryo, whilst expression of *Nipbl* in the other tissue was sufficient to rescue ASD in a *Nipbl*-deficient embryo. This suggests that the risk of ASD, a defect of the heart, is actually determined by factors in the rest of the embryo. This study thus identified the presence of complex interactions between different lineages in the aetiology of CHDs - the risk conferred by a mutation in one lineage is modified by the status of other lineages – and may help explain the diverse cardiac phenotypes seen in CdLS patients (Santos *et al.*, 2016).

Not all of the phenotypic features of CdLS are recapitulated in the mouse model. For example, limb abnormalities occur in 20-30% of CdLS cases, but were originally not noted in *Nipbl*$^{+/-}$ mice (Kawauchi *et al.*, 2009). More recently 25-30% of *Nipbl*$^{+/-}$ mice were found to have postaxial polydactyly (Lopez-Burks *et al.*, 2016) and modest changes in the expression of genes involved in normal limb development, such as *Shh*, *Hox* genes, and genes of the *Fgf*, *Bmp* and *Wnt* signalling pathways (Muto *et al.*, 2014). Reduced *Bmp* and *Hox* expression were found to enhance the polydactyly observed in *Nipbl*$^{+/-}$ mice to around 96% and 63% respectively, suggesting that these pathways are directly involved in limb abnormalities in CdLS. The limb reduction phenotypes, that are seen in some CdLS patients however, were still not observed.

1.3.2 Zebrafish

Translation blocking morpholinos (MOs) were used to knockdown the expression of nipbla and nipblb, the two zebrafish nipbl genes, and to create Nipbl-deficient zebrafish embryos (*nipbl* morphants). The zebrafish nipbl morphants exhibit phenotypes common to CdLS patients, such as growth retardation and heart deficits, by 24 hours postfertilisation (hpf). These phenotypes become more pronounced over time and the morphants die in the larval stages (Figure Ch1-12) (Kawauchi *et al.*, 2016).

nipbl morphant larvae exhibited a range of defects in early heart development, resulting in impaired heart function, circulation defects and pericardial edema (Figure Ch1-12) later in development, and ultimately, premature death.

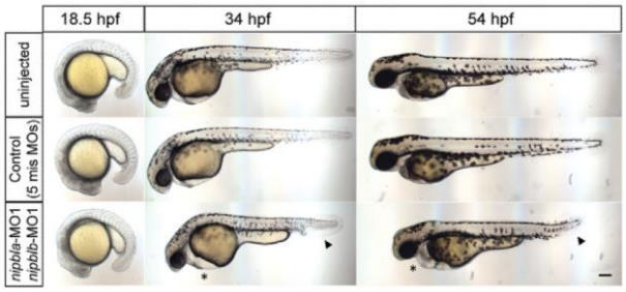


Figure Ch1-12. Reduced body size and growth in Nipbl+/- mice. *Nipbl* morphant (nipbla-MO1/nipblb-MO1) zebrafish larvae resemble WT (uninjected) and control at 18.5 hpf but show pericardial edema (asterisk) and tail defects (arrow) by 34 hpf and these progress over time. Scale bar = 25 mm. Takgen from Kawauchi *et al.*, 2016

nipbl morphants also have malformations of the developing gut, including reduced size, patterning defects and absence of the associated organs. Gastrointestinal reflux is a hallmark of CdLS, which causes death in a number of patients. Transcriptomic analysis of mRNA from embryos at 6 hpf showed small but significant changes in expression upon nipbl reduction, consistent with what has been seen for other model organisms. Dysregulated genes included regulators of endoderm development, such as *sox17* and *foxa2*, which are important during the development of the heart and visceral organs (Kawauchi *et al.*, 2016).

Genes involved in the regulation of vertebrate limb growth and patterning, such as fgfs, shh and hox genes, are also among those dysregulated in nipbl morphant embryos. This dysregulation precedes size reduction and patterning defects of the pectoral fin (forelimb). Knockdown of the mediator complex subunit Med12, mimicked both the gene expression changes and structural changes seen upon nipbl reduction, and combined reduction had a synergistic effect. Furthermore, 3D-FISH in zebrafish fin buds revealed that nipbl and Med12 modulate chromatin organisation near the hoxda cluster. These results suggest NIPBL and Mediator together regulate the expression of genes needed for correct growth and patterning of the vertebrate limb, by modulating long-range enhancer-promoter interactions (Muto *et al.*, 2014). Interestingly, *MED12* mutations in humans cause an X-linked recessive intellectual disability syndrome characterised by dysmorphic features (Yamamoto and Shimojima, 2015), but do not cause CdLS. This suggests that disruption of NIPBL and Mediator regulated gene expression alone is not sufficient to cause CdLS.

1.3.3 Drosophila

In Drosophila Melanogaster, Nipped-B is critical for transcription and development. It was discovered during a screen for factors that aid expression of the *cut* homeobox gene in the developing wing margin; Nipped-B facilitates interactions of the *cut* promoter with a distal upstream enhancer that drives it's expression (Rollins, Morcillo and Dorsett, 1999). Heterozygous Nipped-B LOF mutant (Nipped-B 407/+) adult flies display similar phenotypic features to CdLS patients, including reduced growth, learning and memory, relative to matched WT flies.

The growth defects were found to be caused by a decreased number and size of cells, rather than differences in systemic growth control or developmental timing. Over-expression of dm (the drosophila C-MYC ortholog), a gene regulated by cohesin, rescued this phenotype, suggesting that the reduced size is caused, at least in part, by decreased dm expression. Furthermore RNA-seq revealed a modest decrease in the expression of a large number of genes in Nipped-B mutant flies.

The deficiencies in learning and short-term memory in *Nipped-B* ^{407}/+ flies was caused by abnormal brain structures; 48% of *Nipped-B* ^{407}/+ flies had abnormal mushroom body (MB) morphology, compared to just 4% of WT flies. MBs are the primary brain structures involved in associative learning and memory in Drosophila and aberrant axon pruning of the Drosophila MB γ-neurons has also been seen in homozygous deficiencies of *SMC1* and *Rad21 (Y. Wu et al., 2015).*

1.3.4 Yeast

In S. *cerevisiae*, cohesin has been implicated in chromatin condensation during mitosis. Scc2-Scc4 mutant cells have chromatin condensation defects that were mediated by the loss of cohesin deposition, rather than the loss of condensin. Cohesin promotes the condensation of chromatin through cis tethering interactions that are formed during prophase (Shen and Skibbens, 2017). Recreation of RBS and CdLS mutations in the homologous yeast genes (*eco1-W216G* and *scc2-D730V respectively) lead to* chromosomal decondensation, indicating that perturbations in cohesin cause changes in chromatin structure that may contribute to CdLS (Gard *et al.*, 2009). This is consistent with reports of chromatin decompaction in CdLS patient cells (Nolen *et al.*, 2013).

The RSC chromatin remodelling complex recruits Scc2-Scc4 to promoters that are characterised by large nucleosome free regions. Scc2-Scc4 was shown to help maintain these regions and thus regulate gene expression. This suggests that changes in transcription underlying CdLS could be the result of nucleosome positioning defects, rather than changes in long-range chromatin interactions. Mutations in the human ortholog of RSC (SWI/SNF) cause CSS, which has a similar phenotype to CdLS, supporting this idea (Lopez-Serra *et al.*, 2014)

In another study, the inactivation of Scc2 was found to alter the overall transcriptional profile, but this did not correlate with changes in cohesin binding, differing from what has been seen in CdLS patient cells. When a DSB was induced, many genes involved in cellular responses to DNA damage, DNA repair and DNA recombination were downregulated in *Scc2 deficient* cells compared to WT cells, revealing that Scc2 influences the transcriptional response caused by induction of a DSB. Furthermore, in Scc2 deficient cells there was concurrent activation of genes and decreased cohesin binding, in a 50kb region surrounding a DSB. This indicates that cohesin may be needed for the silencing of DSB proximal genes (Lindgren *et al.*, 2014).

1.4 BRD4

1.4.1 Acetylated lysine recognition

Histones are subject to post-translational modifications (PTMs). These include phosphorylation, methylation, ubiquitination and acetylation, which can impact the interactions between histones and DNA, affecting DNA-templated processes such as transcription, DNA replication, DNA repair, and recombination (Su and Denu, 2016). In histone acetylation, the addition of an acetyl group to a lysine residue by a histone acetyltransferase (HAT), neutralises the positive charge of histones, decreasing their affinity for negatively charged DNA. This relaxes chromatin structure, promoting a conformation more accessible to regulatory factors (Koprinarova, Schnekenburger and Diederich, 2015) and associated with active transcription (Struhl, 1998). Additionally, specific PTMs, or combinations of multiple PTMs, provide docking sites for histone 'readers' – chromatin binding proteins that contain domains which recognise and bind the modified residues

(Gillette and Hill, 2015). Acetyl-lysines can be recognised by a number of domains. These include the YEATS domain, found in just four human proteins (Hsu *et al.*, 2018), tandem PHD fingers (Ali *et al.*, 2012) and the bromodomains (BD) (Jung *et al.*, 2014).

BDs are found in 42 different human proteins (Fujisawa and Filippakopoulos, 2017), including nearly all known HATs (Zeng and Zhou, 2002), in a number of chromatin remodellers such as SWI/SNF (Clapier and Cairns, 2009a), and in BD and extra-terminal domain (BET) proteins (Gillette and Hill, 2015). Chromatin-bound BD proteins act as scaffolds for the assembly of protein complexes, function as TFs or catalyse various chromatin modifications (Fujisawa and Filippakopoulos, 2017).

The BET protein family is made up of BRD2, BRD3, BRD4 and BRDT, all of which contain two N-terminal BDs (BD1 and BD2) and an extra-terminal (ET) domain. BRD4 also has a unique C-terminal domain (CTD) that can recruit proteins for transcriptional activation (Decker *et al.*, 2017) (Figure Ch1-11).

1.4.2 BRD4

BRD4 is the best studied of the BETs, and is known to have many functions, including transcriptional regulation (Hajmirza *et al.*, 2018), cell cycle regulation (Maruyama *et al.*, 2002) and DNA repair (Li *et al.*, 2018). It has been proposed that BRD4 also regulates higher order chromatin structure, however the evidence for this is contradictory; on the one hand BRD4 has been shown to condense chromatin (Wang *et al.*, 2012; Floyd *et al.*, 2013), whilst others have found that it inhibits heterochromatin formation and leads to decondensation (Pongas *et al.*, 2017). Given the role of BRD4 in transcriptional activation, I would expect BRD4 to promote an open chromatin structure. BRD4 null mice are embryonic lethal, highlighting the importance of BRD4 for proper embryonic development (Houzelstein *et al.*, 2002).

1.4.2.1 Structure of BRD4

Alternate splicing yields three different mRNA isoforms from the *Brd4* gene (Figure Ch1-13). The N-terminal BDs and the ET domain are common to all three isoforms, meaning that they are all able to bind acetylated histones. The CTD however, important for transcriptional regulation (Rahman *et al.*, 2011), is absent in the two shorter isoforms (B&C). In isoform B, the CTD is replaced with a 75 amino acid (aa) unique sequence, differentiating it from isoform C. This sequence may be important for BRD4s role in the DDR (Floyd *et*

al., 2013) (see section 1.5.5). In mice and humans, the majority of BRD4 is isoform A, and this longer isoform accounts for much of BRD4s biological activity (Wu and Chiang, 2007).

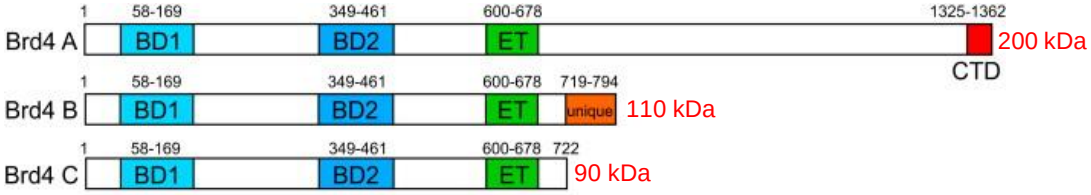


Figure Ch1-13. Three isoforms of BRD4. Schematics of BRD4 isoforms A, B and C. Domains are labelled. BD= bromodomain. ET = extraterminal domain. CTD= c-terminal domain. Isoform B contains a unique 75aa sequence at the C-terminus. Predicted molecular weights are shown in red. Taken from Shi and Vakoc, 2014

1.4.2.2 Acetyl lysine binding by BRD4

The crystal structures of BRD4 BD1 and BD2 show that they exhibit the classical BD fold (found in all BD proteins) (Filippakopoulos *et al.*, 2012) - a left-handed four-helix bundle (αZ, αA, αB, αC) with long intervening loops between the αZ and αA helices (ZA loop) and the αB and αC helices (BC loop) (Figure Ch1-14A, left) (Vollmuth, Blankenfeldt and Geyer, 2009). The loop regions form a hydrophobic cavity for acetyl-lysine binding, in which many interactions take place (Figure Ch1-14A, right); acetyl-lysines are anchored to the BD via hydrogen binding to a conserved Asn residue (Figure Ch14-6B, highlighted in green), and this is enhanced by several residues within the loop regions (Figure Ch1-14B, highlighted in purple). A second interaction is formed between the acetyl carbonyl oxygen atom (OE1) and a conserved Tyr (Figure Ch1-14A, right). Finally a hydrophobic region of the BC loop known as the "WPF shelf" (W81, P82, F83 in BD1) (Figure Ch1-14A&B, highlighted in green) is also important for acetyl-lysine binding (Z. Liu *et al.*, 2017).

In BRD4, BD1 and BD2 were found to interact most strongly with di- and tetra- acetyl-lysines in H4 N-terminal tails (Jung *et al.*, 2014). However, despite their sequence similarity, there are some differences between BD1 and BD2's interaction preferences; BD1 binds to the H4K5AcK8Ac mark, recruiting BRD4 to promoters and enhancers, whilst BD2 interacts with H3K4AcK9Ac. BRD4-BD2 has also been shown to recruit non-histone proteins such as TWIST (Z. Liu *et al.*, 2017).

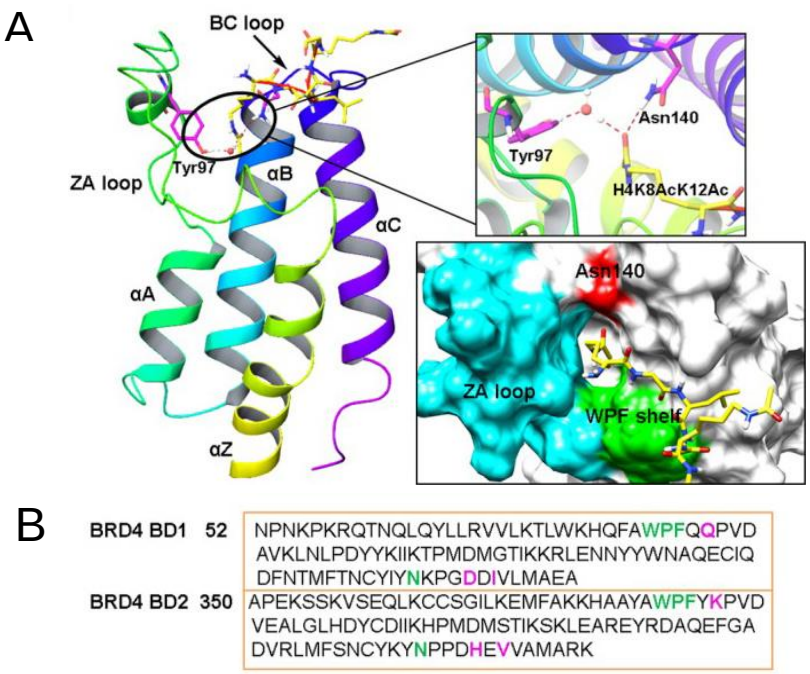


Figure Ch1-14 Conserved regions of BD1 and BD2. A) (Left) Ribbon representation of BRD4 BD1 (PDB IF: 3UW9). (Right) Surface representation of the acetyl-lysine binding pocket of BRD4 BD1. Asn140 is highlighted in red, WPF shelf is highlighted in green and ZA loop highlighted in blue. **B)** Conserved sequences of BRD4 BD1 and BD2. Conserved Asn residues and WPF shelf are highlighted in green. Loop residues that may contribute to acetyl-lysine binding are highlighted in purple. Adapted from Liu *et al.*, 2017

1.4.2.3 BET inhibitors

BET proteins, particularly BRD4, have been implicated in cancer. For instance, in NUT midline carcinoma (NMC) a translocation-induced fusion between BRD4 and NUT causes hyperacetylated nuclear foci and transcriptional dysregulation (Alekseyenko *et al.*, 2015). This made BET proteins an attractive therapeutic target, and led to the development of the first BET inhibitors (BETi), JQ1 and IBET762, in 2010 (Ferri, Petosa and McKenna, 2016). Many more BETi now exist, and a number of these have entered clinical trials (Galdeano and Ciulli, 2016).

All BETi compete with acetylated peptides for binding to the BD pocket, displacing BET proteins from chromatin (Ferri, Petosa and McKenna, 2016). BRD4 inhibitors have head moieties that form critical hydrogen bonds with the conserved Asn and Tyr residues (Asn140 and Tyr97 in BD1), can interact with the WPF shelf and often have small hydrophobic groups that occupy the base of the acetyl-lysine binding pocket (Z. Liu *et al.*, 2017) (Figure Ch1-15A).

One drawback of most BETi is the lack of selectivity between BDs and BETs. Caused by the high conservation of BDs, especially at the residues important for acetyl-lysine binding, this limits the conclusions that can be drawn from BETi studies (Galdeano and Ciulli, 2016). Development of more selective inhibitors is ongoing, and would help elucidate some of the functions specific to each BD (Galdeano and Ciulli, 2016).

JQ1 (Figure Ch1-15B), is the most well studied of the BETi. It is selective for the BET family with a K_D of 50 nM for BRD4-BD1 and of 60–190 nM for the other family members (Ferri, Petosa and McKenna, 2016), is effective against NMC (Filippakopoulos *et al.*, 2010) and decreases the transcription of BRD4 target genes (Horne *et al.*, 2015; Korb *et al.*, 2015). Furthermore the existence of the stereoisomer (−)-JQ1 (Figure Ch1-15C), which shows no significant interactions with BDs, provides a useful negative control for validating the biological effects of (+)-JQ1 (Liu *et al.*, 2017).

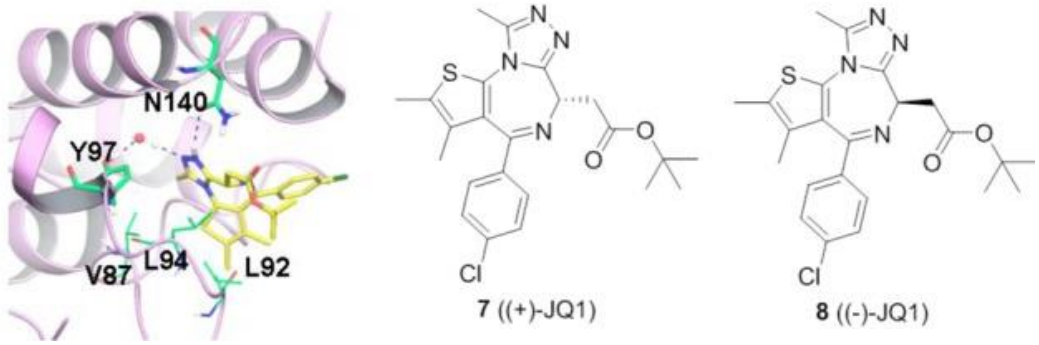


Figure Ch1-15. JQ1 inhibits BRD4 by mimicking acetyl-lysine binding to BDs. A) Co-crystal structure of (+)-JQ1 with BRD4 BD1 (PDB ID:3MXF). Residues Asn140 (N140), Tyr97 (Y97), Leu94 (L94), Val87 (V87) and Leu92 (Leu92) are highlighted. **B)** Chemical structure of (+)-JQ1. **C)** Chemical structure of (-)-JQ1. Adapted from Liu *et al.,* 2017

1.4.3 BRD4 and gene regulation

ChIP-seq experiments have shown that BRD4 is bound to enhancers and promoters of its target genes. BRD4 at these regions is important for the activation and elongation of transcription via interactions with Mediator and positive transcription elongation factor B (P-TEFb) (Hajmirza *et al.*, 2018).

1.4.3.1 BRD4 at enhancers and super enhancers

Enhancers are CREs that recruit trans-acting factors such as TFs and other components of the transcription apparatus (Sabari *et al.*, 2018), and interact with target gene promoters to regulate the expression of cell type–specific genes (Zabidi and Stark, 2016). Active enhancers are identified by the accumulation of specific histone PTMs. These include H3K4me1, histone tail acetylations - H3K27ac and H4K16ac – and histone globular domain acetylations - H3K64ac and H3K122ac (Pradeepa *et al.*, 2016). Lysine acetylation at enhancers results in the binding of BRD4. BRD4 recruits a number of trans-activators to enhancers (Rahnamoun *et al.*, 2018), including the Mediator complex (Bhagwat *et al.*, 2016). Mediator is thought to regulate gene expression by promoting the formation of transient chromatin loops that bring distal enhancers into contact with promoters (see section 1.2.8.2). Inhibition of BRD4 causes the loss of Mediator subunits at enhancers and the repression of target genes (Bhagwat *et al.*, 2016).

BRD4, MED12 and MED1 are most highly enriched at large clusters of enhancers, termed super enhancers (SEs) (Di Micco *et al.*, 2014). SEs, first identified in ESCs (Whyte *et al.*, 2013), can span 10s of Kb (McKeown *et al.*, 2017), are bound by very high levels of Mediator and master regulatory factors, and regulate the transcription of genes involved in cell identity (Whyte *et al.*, 2013; Klein *et al.*, 2017; Brown *et al.*, 2018). BETi treatment causes preferential loss of BRD4 at SEs, and preferentially affects genes associated with SEs (Chapuy *et al.*, 2013; Lovén *et al.*, 2013).

Recently it's been proposed that BRD4 is involved in the formation of SEs (Winter *et al.*, 2017). BRD4 and Mediator are thought to form phase-separated condensates that compartmentalise and concentrate the transcription apparatus at SE regulated genes, ensuring robust gene expression (Sabari *et al.*, 2018).

Whether SEs represent a new paradigm – a functional element that is more than the sum of its parts –, or whether their transcriptional effect is simply the additive effect of their constituent enhancers, has been debated (Hay *et al.*, 2016; Shin *et al.*, 2016). Personally, I believe that regardless of whether individual enhancers function additively or synergistically within SEs, their clustering, and recruitment of large amounts of regulatory factors, is enough to render them important. This is supported by substantial evidence that SEs are enriched

for disease-associated sequence variations (Hnisz *et al.*, 2013; Vahedi *et al.*, 2015; W. Sun *et al.*, 2018).

1.4.3.2 RNAPII pause release

At a large proportion of active genes, RNAPII transcribes between 20 and 100 nucleotides before pausing proximal to the promoter (Winter *et al.*, 2017). Pausing involves the association of two negative factors - DRB sensitivity-inducing factor (DSIF) and negative elongation factor (NELF) – with the initiated RNAPII (Lu *et al.*, 2016). RNAPII pausing is released through the recruitment of P-TEFb – made up of CDK9 and Cyclin T1 - to promoters, where it mediates the phosphorylation of Ser2 in the RNAPII CTD, and of DSIF and NELF (Lu *et al.*, 2016). It has been suggested that BRD4 itself has a kinase activity that phosphorylates Ser2 of RNAPII (Devaiah *et al.*, 2012), but there is little further evidence to substantiate these claims.

Promoter-proximal pausing is a rate-limiting step in the transcription of many genes, and the activity of P-TEFb is tightly regulated (Lu *et al.*, 2016). The majority of P-TEFb is sequestrated in an inactive 7SK small nuclear ribonucleoprotein (snRNP) complex that contains 7SK snRNA, HEXIM1/2, MePCE and LARP7. Release of P-TEFb from this complex allows for the release of paused RNAPII (Figure Ch1-16)(Flynn *et al.*, 2016). The BRD4 CTD interacts with both Cyclin T1 and CDK9 components of P-TEFb, whilst BD2 can also interact with an acetylated region of Cyclin T1 (Hajmirza *et al.*, 2018). BRD4 recruits P-TEFb to enhancers, thus regulating the pause release of RNAPII.

1.4.3.3 Chromatin remodelling

BRD4's ET domain is also involved in transcriptional regulation through interactions with histone modifiers such as JMJD6 and the ATP-dependent chromatin remodellers SWI-SNF and CHD2 (Hajmirza *et al.*, 2018).

JMJD6 immunoprecipitates with BRD4 in complex with P-TEFb, and knockdown of JMJD6 increases the RNAPII occupancy at promoter-proximal regions, suggesting it is also involved in RNAPII pause release (Taniguchi, 2016) (Figure Ch1-16). ChIP-seq locates BRD4 and JMJD6 at enhancer regions, suggesting that looping between the enhancer and promoter is also important for pause release (Liu *et al.*, 2013).

During transcription elongation, histone exchange and nucleosome shifting is required to allow progression of RNAPII. This is achieved through chromatin remodelling by SWI-SNF and CHD2 (Cucinotta and Arndt, 2016), and BRD4 itself has been reported to act as a histone chaperone and support the progression of RNAPII through hyperacetylated nucleosomes. This activity is dependent on BD association with acetylated histones, but exactly how it is achieved remains to be elucidated (Figure Ch1-16) (Kanno *et al.*, 2014).

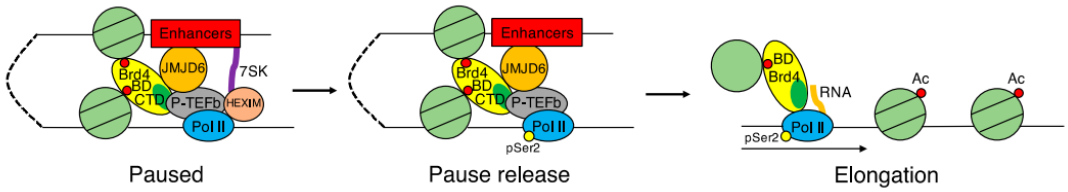


Figure Ch1-16. Transcriptional regulation by BRD4. In the promoter-proximal region, RNAPII pauses due to sequestration of P-TEFb in the 7SK-snRNP complex. Enhanced recruitment of P-TEFb by BRD4 causes Ser2 phosphorylation in RNAPII, leading to RNAPII release. The pause release is also supported by the interaction of P-TEFb with BRD4 and JMJD6 associated with distal enhancers. BRD4 may also promote RNA synthesis along the gene by interactions between the BD and hyperacetylated nucleosomes. Red circle – BDs, Green circle – CTD. Taken from Taniguchi, 2016

1.4.3.4 Enhancer RNA transcription

Most promoters in mammalian genomes are inherently bi-directional. Similarly, transcription emanating from enhancers yields short, unstable enhancer RNAs (eRNAs). eRNAs are often a marker of active enhancers and their transcription generally increases with activation of specific TFs (Azofeifa *et al.*, 2018). Whilst the importance of these eRNAs is not clear, it has been shown that knockdown of eRNAs can cause the downregulation of the enhancer-associated genes (Meng and Blaine, 2017).

eRNAs have been found to promote interactions between promoters and distal enhancers (Hsieh *et al.*, 2014; Yang *et al.*, 2016), suggesting that they might form bridges between the two elements to facilitate DNA looping at genes (Meng and Blaine, 2017). eRNAs may also be involved in RNAPII pause release; upon induction of immediate early genes in neuronal cells, eRNAs directly bind to NELF and facilitate its release from paused RNAPII. This mechanism may be common to other genes and cell types (Schaukowitch *et al.*, 2014).

BRD4 inhibition by JQ1 inhibits eRNA synthesis, suggesting that BRD4 facilitates eRNA transcription, as well as normal mRNA transcription (Kanno *et al.*, 2014). However it has also been shown that BRD4 functions to suppress pervasive eRNA transcription; BRD4

recruits a complex of 7SK snRNA and the chromatin remodelling complex BAF (distinct from the 7SK snRNP complex found at promoters) to enhancers, where BAF-dependent nucleosome remodelling decreases eRNA transcription. This activity is important for the inhibition of convergent transcription (convT) – the transcription of one region on both the sense and antisense strands at the same time – that can arise from excessive eRNA transcription (Flynn *et al.*, 2016).

1.5 Double strand breaks and their repair

The integrity of DNA can be threatened by both endogenous and exogenous sources. Together these can induce up to 10^5 lesions in every cell every day. Unrepaired, these lesions can lead to mutations or larger genomic aberrations, resulting in decreased cell or organism fitness (Hoeijmakers, 2009). Double strand breaks (DSBs) are considered the most dangerous of the damage induced lesions, with the highest probability of causing cell death, mutation or a carcinogenic transformation (Schipler and Iliakis, 2013).

1.5.1 Endogenous causes of DSBs

1.5.1.1 Replication stress

Slowing or stalling of the replication fork is known as replication stress, and often leads to the exposure of long stretches of ssDNA and activation of the replication checkpoint (Berti and Vindigni, 2016). Prolonged replication stress can lead to fork collapse - replisome components prematurely dissociate from the fork, and it can no longer support DNA synthesis. Nucleolytic processing of the ssDNA at collapsed fork results in DSBs (Toledo, Neelsen and Lukas, 2017).

Endogenous impediments to the progression of the replication fork include head-to-tail collisions of adjacent forks (Alexander and Orr-Weaver, 2016), transcribing RNA polymerases, tightly-bound protein-DNA complexes, oncogene activation and unusual DNA structures (Berti and Vindigni, 2016) such as R-loops.

R-loops are stable, three-stranded nucleic acid structures, that arise during transcription through the hybridisation of the newly transcribed RNA with the DNA template, and exposure of the non-template DNA strand as a loop of ssDNA. Dysregulation or inefficient

removal of R-loops poses a barrier to replication fork progression (Chang and Stirling, 2017), and they have been shown to increase replication fork-stalling and replication-dependent DSBs (Gan *et al.*, 2011).

1.5.1.2 Transcription

Transcription can cause the formation of R-loops which, as explained above, can result in DSBs by impeding the replication fork. DSBs may also arise through the head-on collisions of two RNA polymerases. In mammalian cells, antisense noncoding transcripts and genes embedded within other genes (and transcribed in the opposing direction) are prevalent, and cause convT. ConvT can trigger recruitment of activation-induced cytidine deamination (AID) - the conversion of deoxycytidine residues into deoxyuracil, causing a UG mismatch that results in a DSB through base-excision repair or mismatch repair (Khan and Ali, 2017) - and RNAPII collisions, which may also lead to DSBs (Meng *et al.*, 2014).

1.5.1.3 Programmed DSBs

1.5.1.3.1 Transcription

As transcription proceeds, there is a necessity for RNAPII to travel along the helical DNA template (Liu and Wang, 1987). This results in negative DNA supercoiling (under-twisting) upstream of RNAP and positive supercoiling (over-twisting) downstream (Figure Ch1-17) (Ma and Wang, 2016). The resolution of such topological changes requires regulated DNA damage (Calderwood, 2016); Topoisomerase (TOP) enzymes catalyse the breaking and rejoining of DNA strands to relieve the torsional stress associated with supercoiling (Figure Ch1-17) (Wang, 2002). TOP1 introduces single strand breaks (SSBs), allowing the broken strand to rotate around the intact strand (Pommier *et al.*, 2016), whilst homodimeric TOP2 enzymes induce DSBs and catalyse the ATP-dependent passage of one DNA double helix through another (Wang, 2002).

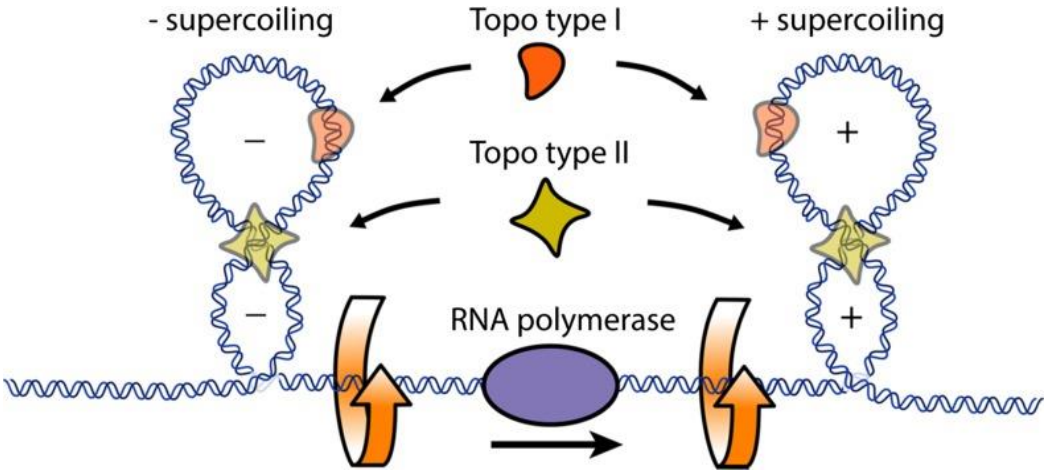


Figure Ch1-17. Transcription produces torsional stress in the DNA double helix. Illustration of the torsion created by RNAP progression along the DNA template. Negative supercoiling is produced behind the RNAP, and positive supercoiling is produced ahead of RNAP. Supercoiling is relieved by the action of the topoisomerases (Topo type I and Topo type II). Taken from Ma and Wang, 2016.

This function of TOP enzymes is required for transcriptional elongation of certain genes (Pommier *et al.*, 2016). TOP2B induces programmed DNA breaks at promoters. These DSBs are critical for RNAPII pause release and transcriptional activation caused by heat shock, serum induction and responses to nuclear receptors (Ju *et al.*, 2006; Bunch *et al.*, 2015; Calderwood, 2016). TOP1 is bound at promoters and becomes fully active upon the BRD4-mediated phosphorylation of RNAPII CTD and subsequent RNAPII pause release. The transcriptional machinery regulates TOP1 activity throughout the transcription cycle, relaxing torsional stress in gene bodies - that would otherwise impede transcription elongation -, whilst maintaining negative supercoiling at TSS - to assist in DNA melting and transcription initiation (Baranello *et al.*, 2016).

Whilst controlled induction of DSBs promotes transcription, unregulated breaks are known to inhibit transcription. In mammals, RNAPII is degraded by the proteasome upon DNA damage, preventing the expression of aberrant transcripts from damaged loci (D'Alessandro and d'Adda di Fagagna, 2017).

1.5.1.3.2 Immunological functions of DSBs

Programmed DSBs create the diverse repertoire of B and T cells that function in immune responses in mammalian lymphoid cells (Prochazkova and Loizou, 2016). V(D)J recombination allows lymphoid cells to generate enormous antigen-receptor diversity by recombining their germline variable (V), diversity (D), and joining (J) exon gene segments in

various combinations. DSBs are crucial for this process, marking the correct areas for recruitment of the recombinational machinery (Khan and Ali, 2017).

After V(D)J, further programmed DSBs enhance antibody functionality of mature B cells through class-switch recombination (CSR) (Qiao *et al.*, 2017). In CSR, DSBs initiate recombination of immunoglobulin (Ig) heavy constant chains, resulting in a switch in Ig isotypes. This allows B cells to produce a new class of antibodies, which retain the original antigen specificity of antibodies but have distinct effector functions. DSBs in CSR are generated by AID (Khan and Ali, 2017).

1.5.1.3.3 Meiosis

During meiosis I, crossovers (CO) occur between homologous chromosomes. At least one CO per chromosome is required for proper dysjunction. COs are formed through the programmed induction of hundreds of DSBs along the chromosome and their repair by homologous recombination (HR) (Borde and de Massy, 2013) (see section 1.5.4.2 for description of HR).

1.5.2 Exogenous sources of DSBs and studying DSBs

DSBs can also be caused by exogenous agents. These may occur naturally, such as ionising radiation (IR) from high energy cosmic rays and radioactive materials in the earth's crust (Bibbo and Piotto, 2014). However, DSBs can also be induced artificially through a range of damaging agents. A number of these, described in detail below, can be utilised to study DSB repair and cellular mechanisms of DDR.

1.5.2.1 IR and radiomimetic drugs

IR can lead to a variety of lesions, including DSBs. IR leads to extensive base damage and produces reactive oxygen species that attack the sugar-phosphate backbone and produce SSBs (Mehta and Haber, 2014). The frequency of DSBs increases linearly with increasing IR (Lomax, Folkes and O'Neill, 2013) as, with higher doses of IR, the likelihood of two SSBs occurring in close proximity in complementary DNA strands also increases (Mehta and Haber, 2014).

Radiomimetic drugs mimic the effects of IR by generating free radicals that damage DNA and result in DSBs. They can be added to culture medium to induce nonspecific DSBs in

cells. Bleomycin is clinically available as an anti-tumor drug, whilst others such as neocarzinostatin (NCS) are solely used for research purposes (Andros *et al.*, 2015).

1.5.2.2 Replication inhibitors

Replication inhibitors that impair the progress of the replication fork, resulting in the fork collapse, can also be used to induce DSBs.

Hydroxyurea (HU) causes replication fork stalling through the inhibition of ribonucleotide reductase, the enzyme required for the production of deoxyribonucleotides (dNTPs), and therefore inhibits DNA synthesis (Singh and Xu, 2016). Aphidicolin (APH) specifically inhibits B-family DNA polymerases by binding at, or near, the dNTP binding site of the polymerase and inhibiting the incorporation of specific dNTPs (Baranovskiy *et al.*, 2014).

1.5.2.3 Targeted endonucleases

Site specific endonucleases can induce DSBs at defined sites in the genome (Bryant and Johnston, 1993), giving them an advantage over other exogenous DSB inducing agents, where the site of DSBs is largely unknown. This has been utilised to generate stable cell lines in which multiple site specific DSBs can be induced; Iacovoni *et al.,* fused the *Asi*SI restriction enzyme to a modified oestrogen receptor hormone-binding domain that responds only to 4-hydroxy tamoxifen (4OHT). Expression of the fusion protein in U2OS cells allows the induction of DSBs with 4OHT treatment (Iacovoni *et al.*, 2010).

Recently the DNA endonuclease Cas9, which can be targeted to a specific DNA sequence by a small "guide" RNA (gRNA), has emerged as a powerful molecular tool. Cas9 is routinely used for genome editing (Paix *et al.*, 2017) and the recruitment of proteins to specific genomic loci (Tian *et al.*, 2017), but also has the potential to inform us about DNA repair. It's ability to induce site-specific DSBs has been utilised to induce several breaks throughout the genome, and allow the mapping of DDR factors in response to breaks (Iacovoni *et al.*, 2010), whilst the addition of a repair template can be used to compare the efficiencies of repair by non-homologous end joining (NHEJ) and HR (Zaboikin *et al.*, 2017).

1.5.2.4 Others

There are a multitude of other drugs that can induce DSBs, including: TOP inhibitors – which trap covalently linked topoisomerase-DNA cleavage complexes, preventing the religation of breaks (Mehta and Haber, 2014) – and DNA-alkylating agents - which introduce

an alkyl group into a nucleophilic site of DNA and cause inter- or intra-strand cross-linking (Siddik, 2002).

1.5.3 The DDR to DSBs

DSBs can cause the mis-segregation of chromosomes during mitosis (Bakhoum *et al.*, 2017) and impair replication (Berti and Vindigni, 2016), leading to genomic instability. To avoid this, the cell employs several checkpoints that temporarily prevent progression through the cell cycle (Barnum and O'Connell, 2014). The DDR is a complex signal transduction pathway that co-ordinates this cell-cycle regulation with the sensing, signalling and repair of DNA damage (Figure Ch1-18) (Jackson and Bartek, 2010).

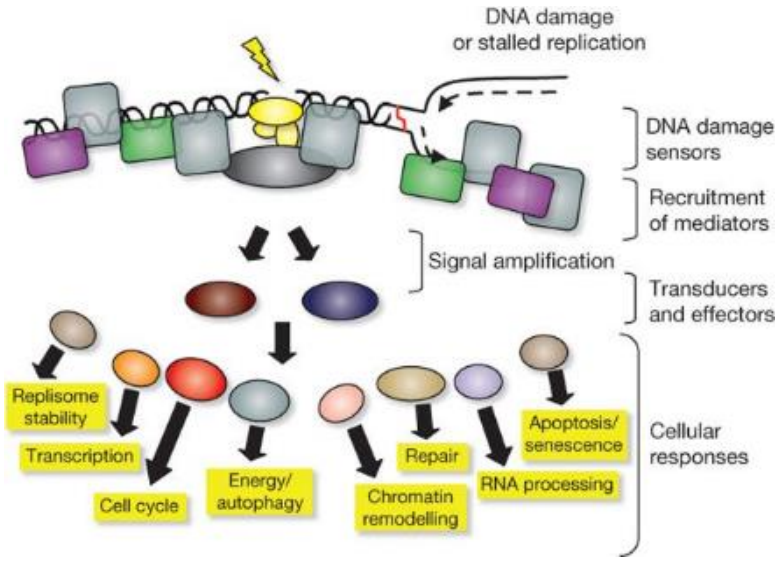


Figure Ch1-18. Model of the DDR. Illustration showing the key steps in the DNA damage response. DNA damage or stalled replication are detected by DNA damage sensors, initiating a signalling pathway that involves recruitment of mediating proteins and signal amplification and can result in a number of cellular responses. Taken from Jackson and Bartek, 2010.

1.5.3.1 Sensing of DNA damage

The first stage of the DDR is the recognition of the DSB. Several sensor proteins recognise DSBs and result in the recruitment of different downstream proteins and different mechanisms of repair. Tight regulation is required to ensure the correct response to a DSB, and it is possible that interplay between different sensor proteins may be involved (Hartlerode *et al.*, 2015).

The MRN complex is made up of MRE11, RAD50 and NBS1; the ATP binding globular domains of a RAD50 dimer interact with two molecules of MRE11, and these molecules directly bind DSBs and recruit NBS1 (Lavin *et al.*, 2015). RAD50 recognizes the ends of the DSB and holds them in close proximity (Yang *et al.*, 2018a). MRE11 has endo- and exonuclease activities that facilitate DSB end-resection, a pre-requisite for repair through HR (Goldstein and Kastan, 2015). NBS interacts with the downstream kinase ATM, to activate the DDR (Lavin *et al.*, 2015), and serves as a scaffold for the recruitment of DNA damage repair proteins including MDC1, BLM, BRCA1 and CtIP (Schiller *et al.*, 2014). Mutations in MRN components causes chromosomal instability, sensitivity to DNA damage and cell cycle defects in patients (Lavin *et al.*, 2015).

The Ku complex is a heterodimer of Ku70 and Ku80 (Ku70/80). Through recent super-resolution fluorescent particle tracking, Ku70/80 has been identified as one of the earliest sensors to localise at DSBs (Yang *et al.*, 2018b). It forms a ring around DNA, that achieves high affinity binding without sequence specificity (Walker, Corpina and Goldberg, 2001). Once bound it recruits and activates DNA-PKcs, which facilitate DNA repair by NHEJ (Yang *et al.*, 2018b).

The RPA complex, composed of RPA1, RPA2 and RPA3, recognises ssDNA that arises through the processing of DSB ends, and so is likely to be recruited later than the other DSB sensors (Yang *et al.*, 2018b). Once bound, it mediates the activation of the signal transducing kinase ATR. RPA is important in DNA repair through HR, as it recognises ssDNA that results from DSB end-resection (Chen, Lisby and Symington, 2013).

The PARP family of enzymes catalyse PARylation of proteins (including histones). PARylation is the addition of ADP ribose units, forming long poly ADP ribose (PAR) polymers (Schuhwerk *et al.*, 2017). PARP1 senses SSBs and DSBs through N-terminal Zinc finger motifs (Yang *et al.*, 2018b). It is rapidly recruited to breaks and PARylation is one of the earliest DNA damage responses (Schuhwerk *et al.*, 2017). PARylation at DSBs mediates the recruitment of numerous factors involved in DNA repair (Yang *et al.*, 2018b) through a variety of PAR-binding modules, that range from completely folded domains, to disordered stretches of sequences that make multivalent interactions with PAR (Teloni and Altmeyer, 2016).

1.5.3.2 Signal transduction

DSB sensors recruit three kinases of the phosphoinositide-3-kinase-related protein kinases (PIKKs) - ATM, ATR and DNA-PKcs -, the most upstream kinases in the DDR (Maréchal and Zou, 2013). The PIKKs mainly target substrates proximal to the site of the DSB, but their targets are vast and can often diffuse freely, meaning that the signal is transduced to distant substrates (Maréchal and Zou, 2013).

ATM is recruited to DSBs as an inactive dimer, by the MRN complex, and its retention at these sites relies partially on its interactions with MDC1 (see section 1.5.3.4). Upon recruitment to DSBs, the kinase domains of each ATM molecule phosphorylate Ser1981 in the FAT domain of the other. This causes them to undimerise, resulting in two active ATM monomers. Autophosphorylation on Ser1893, Ser19881 and Ser2996, and acetylation by Tip60 on Lys3016, are also important for ATM's functions in DNA damage signalling (Guleria and Chandna, 2016). ATM phosphorylates a plethora of downstream effector proteins and further kinases, which themselves become active and phosphorylate their own repertoire of target proteins. For instance, ATM phosphorylates CHK2, a diffusible signal transducing kinase that regulates cell cycle progression (see section 1.5.3.5.3) (Shiloh and Ziv, 2013).

Inactivation of the ATR kinase in mice results in chromosome fragmentation and early embryonic lethality (Brown and Baltimore, 2000; De Klein *et al.*, 2000). ATR exists in a complex with ATRIP that is recruited to ssDNA (Saldivar, Cortez and Cimprich, 2017). ssDNA is an intermediate in many damage pathways, meaning that unlike ATM that responds only to DSBs (Guleria and Chandna, 2016), ATR can respond to a broad spectrum of damage (Saldivar, Cortez and Cimprich, 2017).

A conformational change in ATR (as yet undefined) is required for its activation. This change is triggered by the binding of an activator protein. Two ATR activator proteins have been identified in vertebrates – TOPBP1 and ETAA1 – and are thought to be recruited to damage independently of ATR. Similarly to ATM, ATR has a large number of target proteins. These include FANCI – which helps ensure robust replication and protect against replication stress (Awasthi, Foiani and Kumar, 2016) – and CHK1, which is involved in cell cycle checkpoint activation (see section 1.5.3.5.3).

Whilst ATM and ATR are thought to be the master DDR transducers, phosphorylating >900 sites in over 700 proteins (Matsuoka *et al.*, 2007), the DNA-PKcs regulates a small number of targets (Maréchal and Zou, 2013). DNA-PKcs is recruited to DSBs through interactions with the Ku70/80 heterodimer, forming the DNA-PK holoenzyme which boosts the kinase activity 5-10 times (Jette and Lees-Miller, 2015).

The main target of DNA-PKcs is itself; Autophosphorylation of DNA-PKcs results in inactivation of its kinase activity and dissociation from Ku70/80 at the DSB, and is important for NHEJ (see section 1.5.4.1) (Davis, Chen and Chen, 2014). Cells lacking DNA-PKcs or Ku70/80 have problems with V(D)J recombination and increased sensitivity to DSB induction, caused by impaired NHEJ (Jette and Lees-Miller, 2015). DNA-PK is also involved in other aspects of the DDR. For instance, in response to replication stress DNA-PK phosphorylates Ser4/Ser8 of RPA32, a subunit of RPA. This regulates activation of the G2/M checkpoint and supresses HR (Ashley *et al.*, 2014).

Whilst each PIKK interacts with a distinct DSB sensing protein, their recruitment to DNA damage involves a common mechanism. An evolutionarily conserved C-terminal motif was first identified in NBS1, which when deleted resulted in the loss of ATM recruitment to DSBs. This motif was later found in the C-termini of Ku70/80 and ATRIP, where it serves to recruit their respective PIKK interaction partners, potentially through interactions with PIKK HEAT repeat containing regions (Falck, Coates and Jackson, 2005) (Figure Ch1-19). Most DSB signal transduction pathways are thought to be regulated by the co-ordinated action of ATM and ATR, with DNA-PKc being dispensable for DSB signalling and the resulting checkpoint activation (Callén *et al.*, 2009).

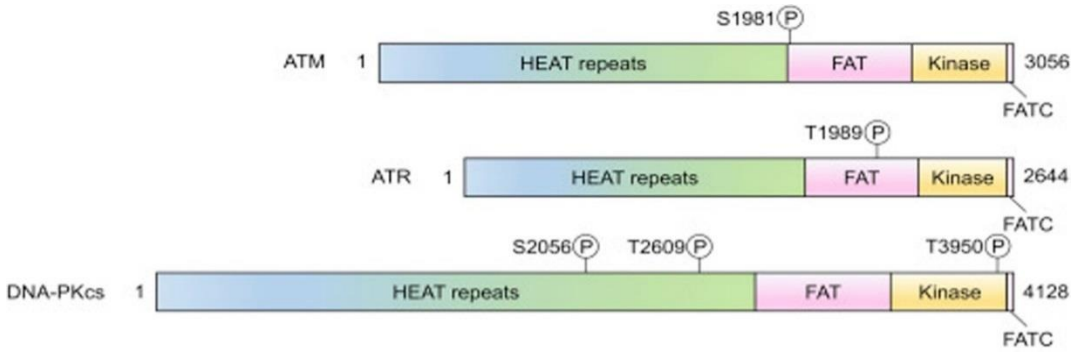


Figure Ch1-19. PIKKs share domain organisation and various structural features. PIKK domain organization. Coloured boxes represent different domains and numbers represent amino acid residues. P - Major sites of phosphorylation sites Taken from Blackford and Jackson, 2017.

1.5.3.3 Chromatin modifications during the DDR

DSBs occur within the complex structure of chromatin, and certain chromatin contexts can hinder their efficient detection, signalling and repair (Price and D'Andrea, 2013). DDR kinases promote chromatin remodelling and histone modifications to produce a more conducive environment for repair (Sirbu and Cortez, 2013), and provide platforms for the recruitment of repair proteins. There is significant evidence for cross-talk between histone modifications – for instance phosphorylation of a specific histone residue can stimulate the acetylation of a different residue within that histone (Lee *et al.*, 2010a) - suggesting that order in which these modifications occur could be important for proper DDR signalling and repair (House, Koch and Freudenreich, 2014).

1.5.3.3.1 Acetylation and phosphorylation

Nucleosome remodellers use energy from ATP hydrolysis to modulate chromatin structure and facilitate the DDR (Seeber, Hauer and Gasser, 2013). For instance p400, an ATPase of the INO80 family, exchanges H2A for H2AZ in response to DSBs, creating large domains of H2AZ containing nucleosomes around the DSB. This exchange precedes, and is required for, H4 acetylation by TIP60, which in turn leads to a more open chromatin structure (Figure Ch1-20). H2A to H2AZ exchange has been found to be required for loading of both Ku70/80 and BRCA1, implicating it in NHEJ and HR repair pathways. The loss of p400 activity causes genomic instability (Xu *et al.*, 2012).

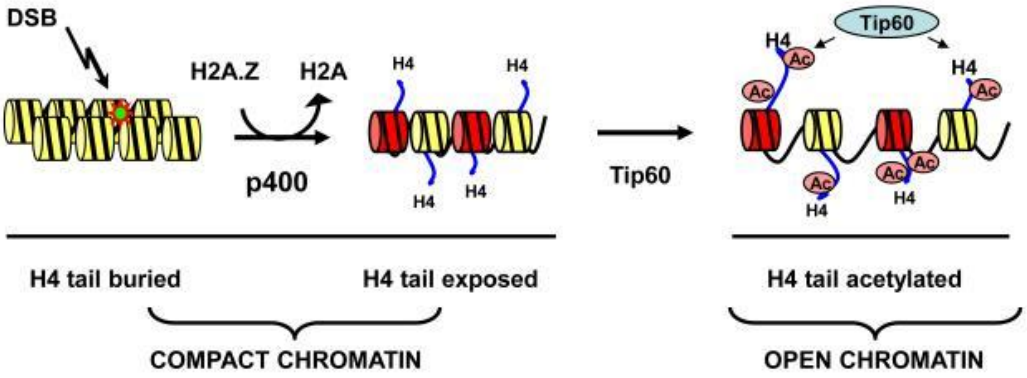


Figure Ch1-20. H2A.Z exchange drives H4 acetylation. P400 catalyses the exchange of H2A for H2AZ in response to DSBs. This expose the H4 N-terminal tail, allowing its acetylation by TIP60. The combination of H2A.Z exchange and H4 acetylation (Ac) shifts chromatin into the open, relaxed conformation required for DSB repair. Taken from Price and D'Andrea, 2013.

The chromatin remodeller CHD1, is essential for end resection during HR. CHD1 opens the chromatin at DSBs, potentially via nucleosome sliding or removal, and facilitates the binding of CtIP (Kari *et al.*, 2016). Other members of the CHD family were also shown to be important for DSB repair (Rother and van Attikum, 2017).

One of the earliest steps in the DDR is the S139 phosphorylation of H2AX (γ-H2AX) by the PIKKs. γ-H2AX recruits HATs such as GCN5 to acetylate H3 within the same nucleosome, and in the surrounding nucleosomes (Lee *et al.*, 2010b). The BD containing SWI/SNF family of nucleosome remodelling complexes, recognise and bind acetyl-lysine residues (Clapier and Cairns, 2009b) and in turn facilitate further phosphorylation of S139, most likely by increasing the accessibility of the neighbouring nucleosomes. This results in a positive feedback loop between SWI/SNF, γ-H2AX and H3 acetylation, and an accumulation of γ-H2AX around the DSB (Iacovoni *et al.*, 2010; Lee *et al.*, 2010b) (Figure Ch1-21).

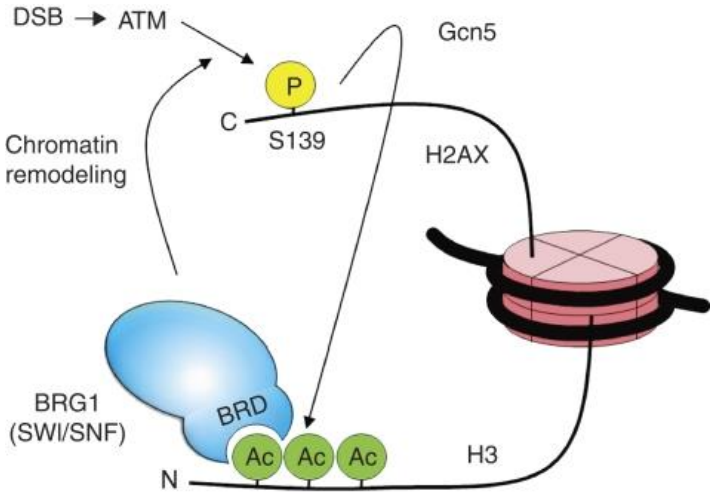


Figure Ch1-21. Positive feedback loop between SWI/SNF, S139P and H3ac results in high levels of γ-H2AX. DSB-activated ATM initiates phosphorylation of S139, triggering the acetylation of H3 within the same and/or neighbouring nucleosomes. SWI/SNF interacts with the acetylated H3 via its BDs and facilitates ATM-mediated S139-P. Taken from Lee *et al.*, 2010

Histone H4 is also acetylated by TIP60-TRRAP in response to DNA damage, increasing the accessibility of chromatin to the DDR complex (Murr *et al.*, 2006). CBP/p300 HATs are recruited to sites of DSBs where they acetylate multiple lysines on both H3 and H4 and facilitate the recruitment of Ku70/80 (Ogiwara *et al.*, 2011).

1.5.3.3.2 Methylation

A complex containing KAP-1, HP1 and the SUV39H1 methyltransferase is also recruited to DSBs, where SUV39H1 methylates H3K9 in the surrounding nucleosomes. HP1 recognises H3K9me3, leading to further loading of the KAP-1/HP1/SUV39H1 complex. This cycle of methylation and loading causes the spreading of H3K9me3 tens of Kb away from the DSB, transiently forming domains of repressive chromatin. These domains act to stabilise the chromatin and activate TIP60 acetyltransferase (Ikura *et al.*, 2007a).

The constitutive histone modification, H4K20me2, plays an important role in the DDR through the recruitment of the effector protein 53BP1. H4K20me2 is usually masked by proteins such as L3MBTL1 (Acs *et al.*, 2011), JMJD2A (Mallette *et al.*, 2012) and TIRR (Drané *et al.*, 2017), but becomes accessible to the tandem tudor domains of 53BP1 upon DNA damage (see section 1.5.3.5.1) (Hartlerode *et al.*, 2012). This association can be limited by TIP60-dependent H4K16 acetylation (Tang *et al.*, 2013).

1.5.3.3.3 Ubiquitination

In response to DSBs, TIP60 acetyltransferase associates with the ubiquitin-conjugating enzyme UBC13, inducing the acetylation and ubiquitination of H2AX, and release of H2AX from damaged chromatin. The reorganisation of chromatin via histone eviction and histone variant exchange, helps to recruit DDR factors to facilitate DNA repair (Ikura *et al.*, 2007b).

Ubiquitination can also directly recruit DDR proteins; Phosphorylated MDC1 is bound by the ubiquitin ligase RNF8, which extends pre-existing ubiquitin on H1 and stimulates the recruitment of RNF168 to DSBs (Mandemaker *et al.*, 2017).. With UBC13, RNF168 initiates ubiquitination of H2A/H2AX at K13/15 (Bohgaki *et al.*, 2013). This ubiquitination cascade results in the recruitment of the effector proteins 53BP1 and BRCA1 to DSBs (Citterio, 2015). Mutations in RNF168 cause RIDDLE syndrome, a immunodeficiency and radiosensitivity disorder, likely through the perturbation of BRCA1 and 53BP1 binding at DSBs (Stewart *et al.*, 2009).

ATM-dependent monoubiquitination of H2B by RNF20-RNF40 is also induced by DSBs, and is required for the accumulation of NHEJ and HR proteins (Moyal *et al.*, 2011).

1.5.3.4 Amplification and spreading of signal

Following initial recruitment and activation, the PIKKs trigger a cascade of DDR signalling on the chromatin surrounding the site of damage (Maréchal and Zou, 2013).

Activation of ATM or DNA-PKcs at DSBs causes the rapid S139 phosphorylation of H2A.X (Caron *et al.*, 2015), and initiates the DDR (Ljungman, 2010). γ-H2AX function is mainly mediated by MDC1 (Caron *et al.*, 2015). Two positive feedback loops involving MDC1 result in the spread of γ-H2AX up to 3 Mb from the break (Rogakou *et al.*, 1999; Iacovoni *et al.*, 2010), amplifying the original signal and providing a platform for the recruitment of further proteins: 1. γ-H2AX must be dephosphorylated on Y142 (a constitutively phosphorylated

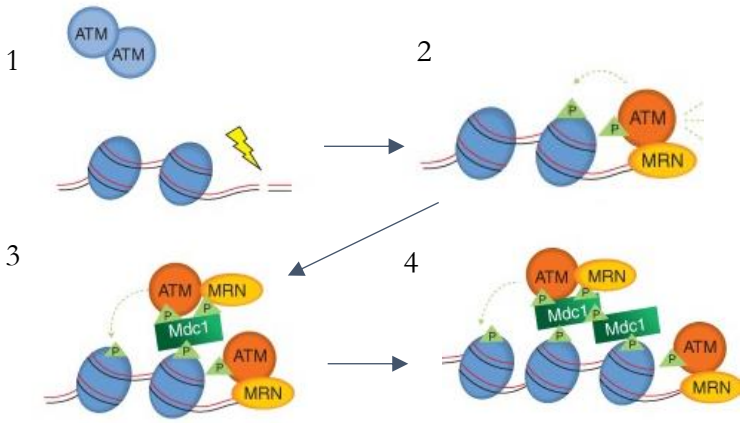


Figure Ch1-22. Recognition of γH2AX by MDC1 triggers a feed-forward loop that spreads ATM and γH2AX over large chromatin domains. Illustration of γH2AX spreading. MRN recognises DSBs and activates and recruits ATM. ATM phosphorylates γH2AX, which is then recognised by MDC1. MDC1 recruits more ATM, which phosphorylates H2AX in adjacent nucleosomes. Modified from Maréchal and Zou, 2013

residue), before MDC1 can bind. ZNF506 is recruited to DSBs in an ATM dependent manner, and recruits the protein phosphatase EYA, which carries out Y142 dephosphorylation. MDC1 then facilitates the recruitment of further ZNF506 (Nowsheen *et al.*, 2018). 2. γ-H2AX bound MDC1 recruits ATM, which phosphorylates adjacent H2AX molecules, MDC1 and ZNF506. MDC1 recruits more ATM to DSBs, resulting in a positive feedback loop that maintains ATM at the original DSB and facilitates spreading of the signal along the DNA (Maréchal and Zou, 2013) (Figure Ch1-22).

The spread of the DDR signal results in microscopically visible foci, named irradiation induced foci (IRIF), made up of DDR proteins (Vignard, Mirey and Salles, 2013) (Figure Ch1-23). IRIF are a useful tool for the study of DNA damage and repair. These foci are

routinely imaged through fluorescence microscopy and, when coupled with fluorescent protein tags such as GFP, they allow in vivo visualisation of breaks and the estimation of the rate of turnover of proteins at their natural binding sites (Nagy and Soutoglou, 2009). This method can be used to assay endogenous breaks, or breaks can be induced in a random or site specific manner (see section 1.5.2) (Nagy and Soutoglou, 2009).

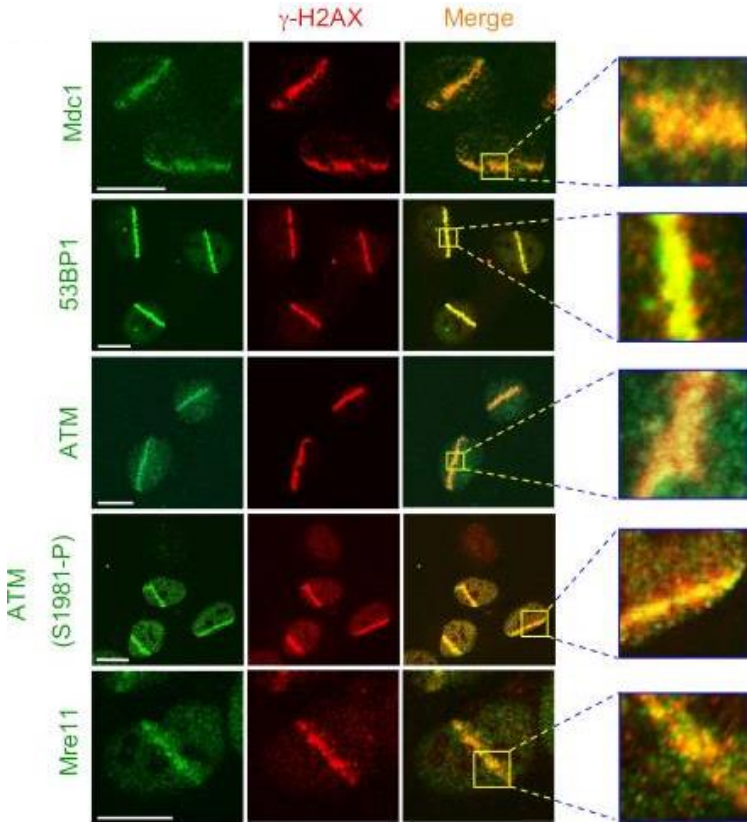


Figure Ch1-23. DDR proteins can spread Mb from the site of a DSB. Exponentially growing U20S cells were sensitised with BrdU and micro-irradiated with UV-A lasers. After 1 hour, cells were fixed and stained with the indicated antibodies. Insets show higher magnifications. Taken from Bekker-Jensen *et al.,* 2006

1.5.3.5 Effector proteins of the DDR

1.5.3.5.1 53BP1

53BP1 is a large (350 kD) protein, that has no enzymatic activity of its own, but contains multiple domains (Figure Ch1-24) that allow both its own recruitment to DSBs, and the

further recruitment of DSB responsive proteins (Panier and Boulton, 2014). 53BP1 recognises and binds H2AK15ub through its ubiquitination-dependent recruitment (UDR) domain (Kocyłowski *et al.*, 2015). RNF168 also interacts with 53BP1 before it is recruited to DSBs, and ubiquitinates it independently of γ-H2A.X–MDC1–RNF8. This ubiquitination is important for the initial localisation of 53BP1 to DSB sites (Bohgaki *et al.*, 2013). The BRCT domain recruits 53BP1 to DSBs through direct binding to γ-H2AX (Baldock *et al.*, 2015), even in the absence of RNF8 and RNF168 (Kocyłowski *et al.*, 2015). The tandem Tudor domains of 53BP1 recognise H4K20me2 (Hartlerode *et al.*, 2012) (see section 1.4.2.1). However, these modifications are present even in the absence of DSBs. It may be that changes in the higher order chromatin structure upon DNA damage are required to expose methylated histones for 53BP1 binding (Huyen *et al.*, 2004). Once at DSBs, 53BP1 can form microscopically visible nuclear foci. 53BP1 focus formation is mediated by the minimal focus forming region, made up of an oligomerisation domain (OD), a Glγ- and Arg-rich motif (GAR) and the UDR and Tudor domains (Panier and Boulton, 2014) (Figure Ch1-24).

The key role of 53BP1 is in the inhibition of DSB end resection, which inhibits HR and promotes NHEJ, and this is mediated through its interactions with RIF1 and PTIP (Figure Ch1-24) (see section 1.5.3.3. for more detail). 53BP1 is also important in the response to replication stress; 53BP1 is recruited to stalled replication forks (Her *et al.*, 2018) where it can regulate the accumulation of BLM - a helicase that prevents replication fork collapse and resolves HR intermediates (Sengupta *et al.*, 2004)- and recruit TOPB1- which activates CHK1 in response to stalled replication forks (Kim *et al.*, 2005). Loss of 53BP1 leads to defective CHK1 signalling and increased cell death in response to replication stress (Her *et al.*, 2018).

Furthermore, 53BP1 forms large nuclear bodies in G1 cells which are thought to represent DNA lesions that are generated when under-replicated DNA undergoes mitosis (Harrigan *et al.*, 2011; Lukas *et al.*, 2011). These foci are resolved during early S phase, raising the possibility that they are repaired by HR (Fernandez-Vidal, Vignard and Mirey, 2017). If this is the case, it's possible that 53BP1 plays a protective role - sequestering the DNA lesions and shielding them from error-prone repair by NHEJ in G1, so that they can be repaired by HR in S phase (Harrigan *et al.*, 2011). Finally, bivalent interactions of the 53BP1 BRCT domain with p53 and the ubiquitin-specific protease USP28, allow 53BP1 to directly

modulate p53's transcriptional activities in response to multiple stimuli (Cuella-Martin *et al.*, 2016).

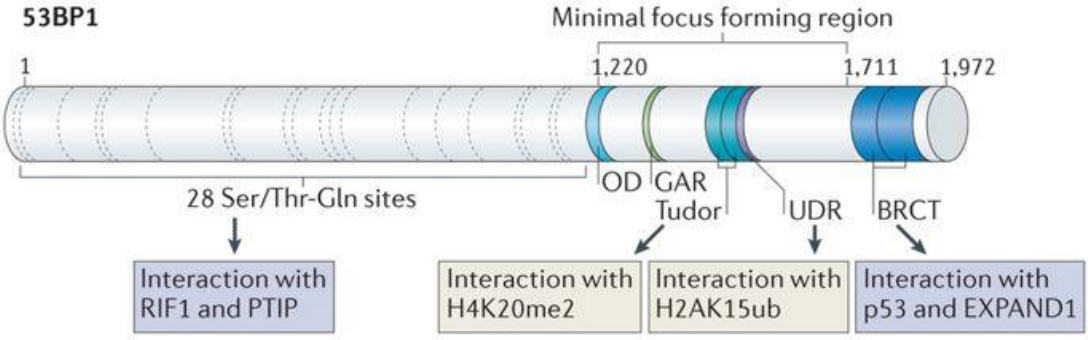


Figure Ch1-24. Domain structure and interaction partners of 53BP1. 53BP1 contains 28 N-terminal Ser/Thr-Gln sites that are phosphorylated by ATM to promote RIF1 and PTIP binding. The 53BP1 C-terminus contains tandem BRCT (BRCA1 carboxγ-terminal) domains that bind to p53 and EXPAND1 and recruits 53BP1 to γH2AX. The minimal focus-forming region contains an oligomerization domain (OD), a Glγ- and Arg-rich (GAR) motif, a tandem Tudor motif that binds to H4K20me2 and an ubiquitination-dependent recruitment (UDR) motif that interacts with H2AK15ub. Amino acid positions are indicated. Taken from Panier and Boulton, 2014.

1.5.3.5.2 BRCA1

BRCA1 is a tumour suppressor gene that is mutated in a large percentage of hereditary breast and ovarian cancers. It codes for a protein of around 220 kDa, made up of multiple functional domains (figure Ch1-25). BRCA1 is recruited to DSBs through ubiquitinated K13/15 of H2A/H2AX. The ubiquitin chains act as docking sites for RAP80, which binds via its ubiquitin-interacting motifs (Li *et al.*, 2017). RAP80 interacts with ABRA1, which in turn forms a complex with BRCA1 through the BRCT domains (Wang *et al.*, 2013), thus recruiting BRCA1 to sites of DNA damage.

Cells lacking BRCA1 are sensitive to a broad range of DNA damaging agents. BRCA1's RING domain is an E3 ubiquitin ligase that ubiquitinates a number of proteins to promote DNA repair. For instance the ubiquitination of H2A provides a binding site for the chromatin remodeller SMARCAD1, which promotes the repositioning of 53BP1 at the periphery of repair foci, counteracting the 53BP1-mediated block on resection (see section 1.5.5.3) (Densham and Morris, 2017). ATM-dependent phosphorylation of Ser1423 in BRCA1 is also important for the activation of the G2-phase cell cycle checkpoint (Xu *et al.*, 2001).

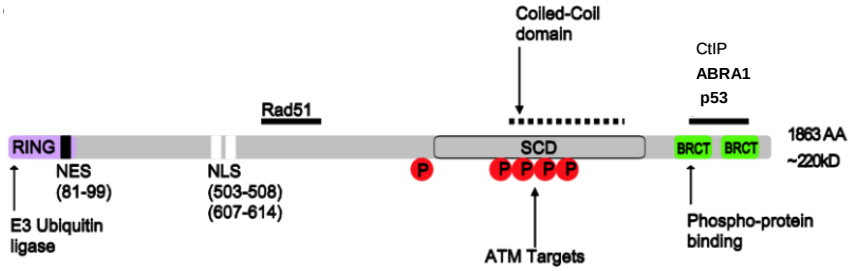


Figure Ch1-25. Domain structure and binding partners of BRCA1. Domain map of BRCA1. RING, serine containing domain (SCD), and BRCT domains are indicated. NES and NLS sequences are also depicted. Horizontal solid black lines indicate protein binding domains for the listed binding partners. Red circles mark phosphorylation sites. Adapted from Clark *et al.*, 2012

1.*5.3.5.3 Checkpoint effector kinases*

Cyclin dependent kinases (CDKs) are a family of kinases that promote progression through the cell cycle. Activity of the CDKs is partly controlled by the cell-cycle regulated accumulation of their cyclin partners, which are required for their enzymatic activity. However, a second level of regulation is achieved through the cell division cycle 25 (Cdc25) family of phosphatases, made up of Cdc25A, Cdc25B and Cdc25C. These phosphatases remove inhibitory phosphatase groups from phosphor-Thr or phosphor-Tyr residues of the CDKs, inducing their activation (Donzelli and Draetta, 2003).

In response to DNA damage, the cell cycle needs to be stalled until damage is resolved, to ensure that the damaged DNA is not inherited by subsequent generations of cells. This is achieved through the activation of G1-S, intra-S and G2-M cell cycle checkpoints by the DSB sensors ATM and ATR. This regulation is mainly carried out by the activation of the checkpoint effector kinases CHK1 and CHK2, however the complete story is complex, involves the activation of many downstream substrates, and varies depending on the type of damage detected and the stage of the cell cycle (Manic *et al.*, 2015).

In response to damage acquired during G1, ATM or ATR phosphorylate and activate CHK2 and CHK1 respectively, although there is evidence of cross-talk between the ATM CHK2 and ATR-CHK1 signalling pathways, and neither checkpoint effector kinase can be assigned unambiguously to a particular pathway (Cuadrado *et al.*, 2006). These kinases in turn phosphorylate Cdc25A phosphatase, inactivating its activity. This leads to the accumulation of the phosphorylated, inactive form of CDK2 and the inhibition of progression into S

phase. The p53-p21 pathway is also important; phosphorylated p53 induces p21, which binds both CDK2 and CDK4, inhibiting their kinase activities and resulting in G1-S checkpoint signalling (Figure Ch1-26) (He *et al.*, 2005).

Cyclin B accumulates during S/G2 and binds to its CDK partner CDK1, which promotes progression into mitosis at the end of G2. However Cyclin B is not enough to fully activate CDK1; Cdc25C is also required to remove the inhibitory phosphatases deposited by the kinases WEE1 and MYT1 (Sherr and Bartek, 2017). Phosphorylation of Cdc25C by either CHK1 or CHK2 creates binding sites for proteins of the 14-3-3 family, which act to sequester Cdc25C in the cytoplasm and inhibit its ability to dephosphorylate and activate CDK1, leading to activation of the G2-M checkpoint (Stark and Taylor, 2006)(Figure Ch1-26)

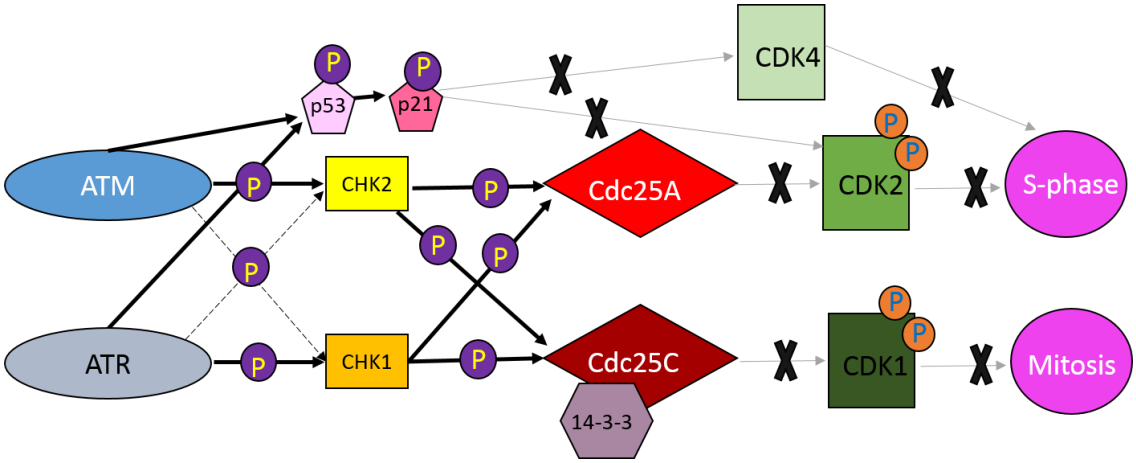


Figure Ch1-26. Schematic checkpoint signalling during G1. ATM and ATR phosphorylate and activate CHK1 and CHK2, which in turn phosphorylate Cdc25A and Cdc25C. This results in CDKs remaining phosphorylated and preventing progression of the cell cycle. The P53-P21 pathway can also be activated by ATM and ATR, preventing the activation of CDK4 and CDK2.

Despite the efforts of the G1-S checkpoint, some damaged DNA may remain in S-phase, and thus an intra-S checkpoint is employed to minimise any deleterious effects. ATR is active during unperturbed S-phases, where it regulates origin firing, and recognises RPA-ssDNA complexes which can occur when replication fork stalling exposes stretches of ssDNA, making it the more crucial of the two DNA damage sensors for intra-S checkpoint activation. During S-phase the phosphatase activity of Cdc25A is required for activation of CDK2 and the subsequent binding of Cdc45 to the pre-replicative complex and initiation of replication.

ATR-mediated CHK1 phosphorylation of Cdc25A in response to DNA damage inhibits this pathway, and thus inhibits origin firing (Iyer and Rhind, 2017). (Figure Ch1-27)

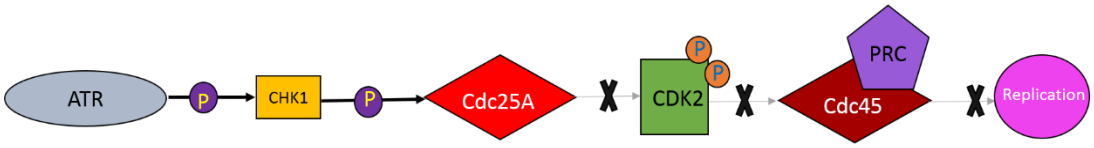


Figure Ch1-27. Schematic of intra-S checkpoint signalling. ATR phosphorylates and activate CHK1. Activated CHK1 phosphorylates Cdc25A, leading to its degradation. Loss of Cdc25A results in loss of CDK2 activation, which in turn inhibits the binding of Cdc45 to the PRC, activating the intra-S checkpoint.

1.5.3.6 Outcomes of the DDR

Activation of the cell cycle checkpoints by CHK1 and CHK2 activation, can have three outcomes – cell cycle arrest, apoptosis or senescence. Transient cell cycle arrest, coupled with DNA repair, is the main outcome of the DDR, giving cells time to resolve the damage before the cell cycle is resumed. If the DNA cannot be repaired however, the cells may undergo apoptosis or become senescent. The determinants of the choice between senescence and apoptosis are not well understood, but may include cell type, the nature of the DNA damage, and the intensity and duration of the DDR. For instance, senescence appears to be associated with large DDR foci and prolonged signalling (D'Adda Di Fagagna, 2008).

1.5.4 DSB repair

As well as cell cycle checkpoint activation, recognition of DNA damage promotes DNA repair (Sirbu and Cortez, 2013). DSBs can be repaired by two main pathways – NHEJ and HR (Rodgers and Mcvey, 2016).

1.5.4.1 NHEJ

NHEJ mediates the direct ligation of the two DNA ends at a DSB, repairing the break without the need for an intact sister chromatid as a template. NHEJ can therefore occur at any stage of the cell cycle (Price and D'Andrea, 2013), and is now thought to predominate throughout (Chang *et al.*, 2017a). The core NHEJ machinery consists of the Ku70/80 heterodimer, DNA-PKcs, XRCC4, DNA ligase IV, Artemis and XLF (Figure Ch1-28) (Davis and Chen, 2013a)(Weterings and Van Gent, 2004), however various NHEJ proteins exist and the exact complement appears to depend on the structure of the DNA ends (Chang *et al.*, 2017b).

NHEJ initiates with DNA-PK binding at DSB ends (Davis and Chen, 2013b). Cryo-EM suggests that Ku70/80 and the bound DNA causes a conformational change in the DNA-PKc, which enhances its Ser/Thr kinase activity (Sharif *et al.*, 2017) – a requirement for successful NHEJ (Kurimasa *et al.*, 1999).

Minimal DSB end processing is required during NHEJ (relative to HR), but some nuclease activity is needed to ensure the compatibility of the two DNA ends. This can be through minor degradation of 5' or 3' overhangs or through the generation of a small region of homology, termed microhomology, between the strands. Removal of overhangs is primarily carried out by Artemis which, when activated by DNA-PK, has 5' and 3' endonuclease activity (Chang *et al.*, 2017b). In most human cells, DNA Pol μ and Pol λ interact with Ku70/Ku80 and create microhomology between the strands through template independent nucleotide addition to the DNA ends (Pannunzio, Watanabe and Lieber, 2017).

XRCC4 and Ligase IV are recruited to the DSB as a complex, through interactions with Ku70/80, and ligate the two ends of the DSB. Ligation is promoted by XLF, which stimulates the activity of DNA Ligase IV towards mismatched and non-cohesive ends (Davis and Chen, 2013a).

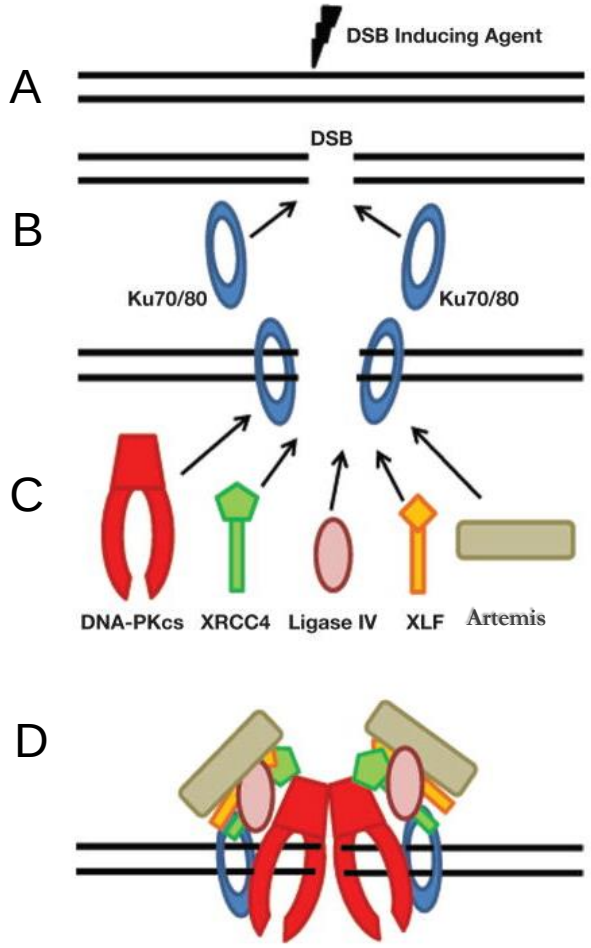


Figure Ch1-28. Illustration outlining the key steps of NHEJ. (A) A DSB is induced and **B)** recognised by Ku70/80. **(C)** Ku70/80 binds the broken DNA ends and serves as a scaffold for the recruitment of DNA-PKcs, XRCC4, DNA ligase IV, XLF and Artemis. **(D)** This results in the formation of a stable complex at the DSB. Modified from Davis and Chen, 2013b.

1.5.4.2 HR

In contrast to NHEJ, HR repairs DSBs by utilising the redundancy of genetic information that exists in the form of sister chromatids or homologous chromosomes, restricting its use to S and G2 in human cells (Wright, Shah and Heyer, 2018a).

HR begins with the recognition of DSBs by the MRN complex. MRN recruits CtIP (Schiller *et al.*, 2014), and this complex initiates DSB end resection with the endonucleolytic cleavage of the 5' internal ends, creating short 3' ssDNA tails (Figure Ch1-29). A more processive

resection is then carried out by the 5'-3' exonuclease EXO1 and the action of the BLM-DNA2 helicase-endonuclease complex (Symington, 2014).

The resulting 3'-ssDNA filaments are initially stabilised by the loading of RPA. As the resection matures, RPA is replaced with RAD51 (Ochs *et al.*, 2016), forming the RAD51-ssDNA filament (Wright, Shah and Heyer, 2018a). RAD51 regulates the stretching of the ssDNA within the filament, increasing its length by up to 50% (Klapstein, Chou and Bruinsma, 2004), a process that is critical for the subsequent homology search (Wright, Shah and Heyer, 2018a).

One strand of the genomic dsDNA is destabilised, opening the helix and allowing the RAD51-ssDNA filament to scan for a homologous template sequence from which to initiate DNA synthesis (Wright, Shah and Heyer, 2018b). Once a homologous region has been identified, RAD51 facilitates the formation of the synaptic complex, made up of the RAD51-ssDNA and both strands of the homologous sister chromatid.

The 3' end of the broken strand must then invade the homologous duplex DNA template and intertwine with its complementary donor strand, forming a heteroduplex DNA known as the D-loop. The 3' end of the invading strand acts as a primer for DNA synthesis by DNA polymerase δ, using the complementary strand as a template (Wright, Shah and Heyer, 2018b).

There are multiple possible outcomes of HR, which can result in integration of the template DNA with the broken DNA (CO) or no integration (non-CO) (Potts, Porteus and Yu, 2006). In somatic cells, the majority of DSBs are repaired through disruption of the D-loop and annealing of the new DNA with the other end of the DSB (Figure Ch1-29). This is known as synthesis dependent strand annealing (SDSA) and avoids CO, preventing loss of heterozygosity (Wright, Shah and Heyer, 2018).

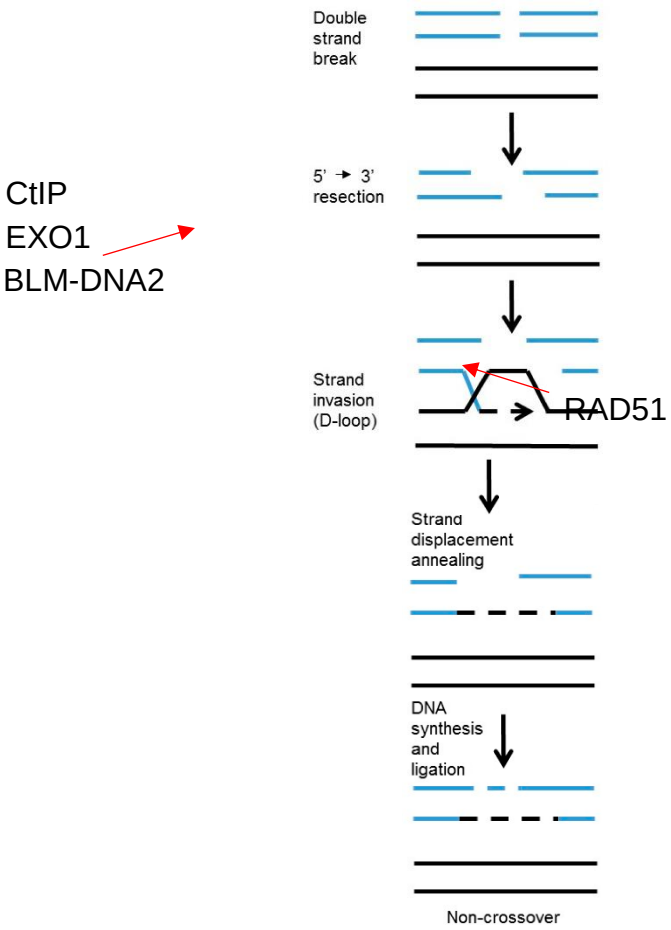


Figure Ch1-29. The key steps of HR. 5'-3' DSB resection exposes 3' ssDNA at DSB ends. RAD51-ssDNA filaments promote homology search and strand invasion into the homologous duplex DNA template, leading to D-loop formation. D-loops can be resolved through SDSA, forming non-crossover products. Red arrows indicate key proteins at each step. Adapted from Hanamshet, Mazina and Mazin, 2016.

1.5.4.3 DNA repair pathway choice

1.5.4.3.1 Cell cycle

The need for a homologous DNA template restricts HR to the S and G2 phases of the cell cycle (Wright, Shah and Heyer, 2018a), but its peak activity appears to occur during mid-S phase, whilst NHEJ predominates in G1 and G2. Cell-cycle regulation of repair pathway choice is mediated largely by CDK dependent phosphorylation of CtIP (Her and Bunting, 2018) and RECQL4 (Lu *et al.*, 2017).

During S phase, phosphorylated CtIP promotes HR by stimulating end resection (see section 1.5.4.2.1), whilst in G1 phase CtIP is degraded by the proteasome (Her and Bunting, 2018).

RECQL4, on the other hand, can promote both HR and NHEJ, but functions differently throughout the cell cycle to co-ordinate the two; in G1 - when overall CDK activity is low - RECQL4 interacts with Ku70 and promotes NHEJ. In S/G2 – when CDK levels are higher - RECQL4 is phosphorylated, enhancing its interaction with MRE11 and its recruitment to DSBs, and activating a helicase activity that contributes to end-resection (Lu *et al.*, 2017).

1.5.4.3.2 Chromatin state

The genome is divided into distinct chromatin states, identifiable by their specific combinations of histone modifications and chromatin protein occupancies. These chromatin signatures are diagnostic of different functional genomic elements such as promoters and enhancers, and their regulatory state. Across the genome, DSBs are not necessarily repaired by the same pathway and it is thought that the choice of pathway is directed by the chromatin state, with many studies supporting the idea of 'histone code driven' repair pathway choice (Clouaire and Legube, 2015).

A good example of this is the favoured use of HR in transcriptionally active chromatin, mediated by the enriched H3K36me3 mark at these regions. On DSB induction, (LEDGF)/p75 recognises H3K36me3 and recruits the resection promoting protein, CtIP. Depletion of the H3K36me3 methyltransferase, SETD2, impaired HR at these breaks, indicating the requirement for H3K36me3 (Clouaire and Legube, 2015).

1.5.4.3.3 Antagonistic roles of BRCA1 and 53BP1

53BP1 and BRCA1 act at the junction of NHEJ and HR, and regulate pathway choice (Aparicio, Baer and Gautier, 2014).

53BP1 recruits RIF1 via its N-terminal Ser/Thr repeats (Panier and Boulton, 2014), and this localises the shieldin complex (C20orf196, FAM35A, CTC-534A2.2 and REV7) at DSBs through interactions with the REV7 subunit. The FAM35A subunit binds ssDNA via OB-fold domains, protecting them from interactions with the resection promoting RPA. Loss of shieldin, or perturbation of the OB-fold domain, impairs NHEJ, suggesting that the binding of shieldin to ssDNA is essential for 53BP1 mediated NHEJ (Noordermeer, Adam, Setiaputra, Barazas, Pettitt, Ling, Olivieri, Álvarez-Quilón, Moatti, Zimmermann, Annunziato, Krastev, Song, Brandsma, Frankum, Brough, Sherker, Landry, Szilard, Munro, McEwan, Goullet de Rugy, *et al.*, 2018).

BRCA1 on the other hand promotes DSB end resection, shifting the balance towards HR (Zámborszky *et al.*, 2017). BRCA1 stabilises PALB2 and BRCA2 at sites of DNA damage, which in turn facilitates RAD51-ssDNA filament formation. Loss of BRCA1 abolishes RAD51's localisation at breaks and impairs HR. Furthermore BRCA1 interacts with the MRN complex during HR, suggesting it may participate directly in end resection (Wu, Lu and Yu, 2010).

$BRCA1^{-/-}$ cells show increased genome instability and sensitivity to DNA damage (Zámborszky *et al.*, 2017), a phenotype which has been shown to be dependent on both 53BP1 and DNA ligase IV (Bunting *et al.*, 2010). This suggests that when BRCA1 is absent, breaks that are normally repaired by the high-fidelity HR pathway are instead aberrantly repaired by NHEJ.

This antagonistic relationship between BRCA1 and 53BP1 regulates DNA repair pathway choice throughout the cell cycle; during G1, 53BP1 blocks recruitment of BRCA1 to DSB ends to supress resection and promote NHEJ, whilst in the S and G2 stages, BRCA1 is able to resect DNA ends for HR (Aparicio, Baer and Gautier, 2014).

1.5.5 The role of BRD4 in the DDR

BRD4 is well characterised as a regulator of transcription, but recently its overexpression in cancers, and evidence that its targeted inhibition can sensitise cancer cells to DNA damaging agents, has led to its identification as a key player in the DDR (Pongas *et al.*, 2017; C. Sun *et al.*, 2018; Zhang *et al.*, 2018). BRD4 is recruited to DSBs via H4 tail lysine residues that become acetylated in response to damage (Dhar *et al.*, 2017; Li *et al.*, 2018), and regulates many steps of the DDR.

BRD4 interacts with components of the DNA pre-replication complex, including CDC6, to regulate activation of the G2/M checkpoint in response to replication stress. Combined BETi and HU treatment (to induce replication stress) in U2OS cells caused a CDC6 dependent decrease in CHK1 phosphorylation compared to HU treatment alone, resulting in decreased G2/M checkpoint activation and an increased sensitivity to replication stress (Zhang *et al.*, 2018).

BRD4 regulates DNA repair through both NHEJ and HR, and loss of BRD4 has been shown to result in defective repair of induced DSBs. BRD4 promotes NHEJ directly - by recruiting certain NHEJ DNA repair proteins, including 53BP1, to DSBs, forming a stable

DNA repair complex – and indirectly – through the regulation of NHEJ DNA repair genes expression (Li *et al.*, 2018). BRD4 also regulates the expression of genes involved in HR, and inhibition of BRD4 induces a HR-deficient (HRD) gene expression signature. One of the HRD genes most significantly downregulated by BETi is CtIP, which stimulates DSB resection via the creation of ssDNA tails (see section 1.5.4.2), and thus promotes DNA repair by HR (C. Sun *et al.*, 2018).

Finally, BRD4 has been implicated in the regulation of the chromatin structure surrounding a DSB, altering the accessibility of chromatin to DNA damage signalling and repair proteins. However, reports regarding this function have been contradictory; BRD4 isoform B is thought to inhibit DDR signalling through the recruitment of the condensin II complex. The condensin complex compacts chromatin, promoting a more 'closed' structure that is less permissive to the binding of DDR proteins (Floyd *et al.*, 2013). Conversely, BRD4 has been shown to inhibit heterochromatin formation and facilitate the DNA damage response, through the repression of HP1 gene expression and subsequent decrease in HP1 protein levels (Pongas *et al.*, 2017). HP1 binds H3K9me3 histones, a hallmark of heterochromatin, and maintains the heterochromatic state (Machida *et al.*, 2018).

1.5.6 Cohesin and the DDR

1.5.6.1 Cohesin recruitment to DSBs

Cohesin accumulates at DSBs (Countryman *et al.*, 2018) and at stalled replication forks (Frattini *et al.*, 2017; Countryman *et al.*, 2018; Villa-Hernández and Bermejo, 2018), where it is important in the regulation of cell cycle checkpoints and DNA repair.

In S. *cerevisiae*, the WAPL homolog (Wpl1/Rad61) promotes efficient DNA repair through modulation of the Smc3p-Mcd1p (SMC3-RAD21) interface, which results in the removal of cohesin from DNA. Deletion of Wpl1 was shown to cause G2/M cell cycle delays, and it was suggested that it is needed for repair of multiple types of DNA damage induced during S-phase. Cohesin is loaded onto DNA prior to S-phase in order to establish cohesion, but is loaded de novo at sites of DNA damage and stalled replication forks. Wpl1's regulation of cohesin removal may therefore be important to remove DNA-bound cohesin, allowing its recruitment to breaks (Bloom, Koshland and Guacci, 2018).

1.5.6.2 Cohesin in cell cycle checkpoint activation and DNA repair

Cohesin binding at DSBs and phosphorylation of the SMC1 and SMC3 subunits recruits 53BP1 (Watrin and Peters, 2009). Sororin, required for the establishment of cohesion by cohesin (Nishiyama *et al.*, 2010), is dispensable for checkpoint activation, suggesting that cohesin's role in cell cycle checkpoint activation is independent of its function in sister chromatid cohesion (Watrin and Peters, 2009).

In contrast, cohesin's ability to regulate DNA damage repair relies on cohesion between sister chromatids. In S and G2, cohesin promotes the close association of sister chromatids, facilitating the repair of DSBs through HR (Frattini *et al.*, 2017). In fission yeast, cohesin is methylated on two evolutionarily conserved lysine residues, K536 and K1200 of Pms1, the ortholog of SMC1. Mutations that inhibit this methylation render cells sensitive to DNA-damaging agents. Evidence of an interaction between cohesin and Mus81-Eme1 - an endonuclease involved in the resolution of intermediate structures during HR - suggested that methylation of Pms1 promotes DNA repair through HR (Sanyal *et al.*, 2018). Furthermore SA2 has been found to directly facilitate HR-mediated DSB repair (Countryman *et al.*, 2018).

Cohesin has also been implicated in the regulation of NHEJ. The accuracy of end-joining decreases with increased distance between double-strand ends (DSEs). This is a particular problem during S phase, where replication stress can generate one-ended DSEs, which require ligation with another, distant, one-ended DSE for NHEJ. Cohesin protects the genome against rearrangements by inhibiting the ligation of distant DSEs but maintaining the ligation of proximal DSEs (Gelot *et al.*, 2016).

1.5.6.3 NIPBL and the DDR in CdLS

NIPBL may mediate the DDR simply through its role as a cohesin loader. However, two independent mechanisms of NIPBL recruitment to DNA damage have been shown, suggesting that it regulates the temporal and spatial specificity of cohesin binding, relative to the type of DNA damage. In response to endonuclease induced DSBs, NIPBL is recruited through an interaction with HP1γ, whilst in response to laser microirradiation induced damage, it is recruited through its HEAT repeats and redundant ATM/ATR activity (Bot *et al.*, 2017).

Consistent with the roles of the cohesin complex and NIPBL in DNA repair, increased sensitivity to DNA damage has been seen in some CdLS patient derived cell lines (Vrouwe *et al.*, 2007; Enervald *et al.*, 2013). Fibroblast and LCLs derived from CdLS patients, both with and without identified NIPBL mutations, show reduced survival after exposure to the DNA interstrand cross-linking agent mitomycin C. Furthermore, upon exposure to X-rays in G2, CdLS cells undergo more chromatid breaks and chromatid exchanges than WT cells. This increase was not observed in G1 (Vrouwe *et al.*, 2007), when DSBs are mainly repaired by NHEJ (Wright, Shah and Heyer, 2018b), suggesting a defect in HR-mediated DSB repair.

Enervald *et al.*, (2013) saw a similar increase in DNA damage sensitivity when they exposed NIPBL-deficient CdLS patient cells to γ-irradiation, although they attribute this to impaired NHEJ rather than HR; sequencing of CSR junctions from patient derived B-cells, which have been shown to reflect the repair pathway used for the re-joining of the broken DNA ends, showed a pattern resembling those of patients with known NHEJ defects (characterised by a lack of "direct end joining") (Enervald *et al.*, 2013).

It has been shown that the reduction of cohesion to 30% of wild-type levels compromises its role in DNA repair, without affecting sister chromatid cohesion or chromosome segregation (Heidinger-Pauli *et al.*, 2010). This could explain how mutations in cohesin can cause disorders such as CdLS, without perturbing cell division to a great enough extent to cause embryonic lethality.

1.5.7 Developmental disorders and the DDR

Since 2005, several congenital disorders that are caused by mutations in components of the DDR have been identified (O'Driscoll, 2012).

1.5.7.1 Shared phenotypic features

Syndromes caused by disorders of the DNA damage response share a number of symptoms, some of which are also present in patients with CdLS.

One of the most common is a predisposition to cancer (O'Driscoll and Jeggo, 2008); impaired cell cycle checkpoints or damage repair pathways may lead to the accumulation of genetic alterations that confer growth and survival advantages to a cell (Broustas and Lieberman, 2014).

DDR defective syndromes also cause many symptoms characteristic of developmental abnormalities, such as characteristic facial features and skeletal abnormalities, microcephaly, growth delay, and immune and neurological deficits. These symptoms are also seen in patients with CdLS. Symptoms such as growth delay and microcephaly could be explained a decrease in cell number, caused by increased apoptosis due to the inability to repair DNA damage (O'Driscoll and Jeggo, 2008). Immunodeficiency may be caused by impaired CSR, V(D)J recombination and/or somatic hypermutation, which require NHEJ to generate diversity in immune cells (Hwang, Alt and Yeap, 2015).

1.5.7.2 Mutations in NHEJ proteins

Mutations in components of the NHEJ repair pathway cause primary immunodeficiency syndromes, through the impairment of the processes required for correct lymphocyte development (X. Liu *et al.*, 2017).

DNA ligase IV deficiency syndrome (LiDS) patients present with a range of phenotypes including normal to severe combined immunodeficiency, sensitivity to IR, primordial growth failure, severe microcephaly and learning difficulties (Altmann and Gennery, 2016) (see table Ch1-1).

1.5.7.3 Mutations in HR proteins

The Fanconi Anaemia (FA) pathway is important for the repair of interstrand cross-links and the response to replication stress. There are 22 FA proteins that function in this pathway. Mutations in these proteins can cause FA - a chromosomal instability disorder characterised by progressive bone marrow failure, cancer predisposition and multiple developmental phenotypes such as small size and heads, and cardiac and skeletal anomalies (Nepal *et al.*, 2017). The HR proteins BRCA1, BRCA2 and RAD51 have also been shown to function in the FA pathway, and mutations in these proteins cause phenotypes overlapping those of FA, leading to their assignation as FA proteins (Michl, Zimmer and Tarsounas, 2016).

1.5.7.4 Mutations in ATM and ATR

Mutations in ATM and ATR also cause developmental syndromes. As these proteins are master regulators of the DNA damage response, their mutations can affect cell cycle checkpoint activation, DNA repair and metabolic responses to DSBs (Liu *et al.*, 2016).

Mutations that truncate or destabilise ATM can result in *Ataxia-telangiectasia (AT)* (Liu *et al.*, 2016), characterised by a diverse range of symptoms including cancer predisposition, neuro-degeneration, and immune dysfunction (Hartlerode *et al.*, 2015). However, the loss of the ATM protein particularly affects cerebellar cells, causing the progressive loss of cerebellar neurons and debilitating ataxia (Liu *et al.*, 2016).

Hypomorphic mutations in ATR alleles cause Seckel syndrome (SS) (Saldivar, Cortez and Cimprich, 2017). SS is a rare developmental disorder, characterised by primordial dwarfism, intellectual disability, microcephaly and severe craniofacial dysmorphism. SS was first coined 'bird-headed dwarfism' due to the characteristic facial phenotype – prominent beaked nose, large eyes and dysplastic ears. In addition, cardiovascular and central nervous system abnormalities have been described (Ramalingam *et al.*, 2012).

1.5.7.5 Perturbation of histone modifications

As described in section 1.5.3.3 histone modifications are important regulators of the DDR. Monoallelic missense mutations affecting H4K91 were found in three patients with syndromic symptoms of growth delay, microcephaly, distinct facial features and intellectual disability. Recapitulation of these mutations in zebrafish found that the mutations affect H4K91 ubiquitination, which normally plays a protective role against DNA damage. Loss of this ubiquitination caused increased DNA damage, which led to abnormal cell cycle progression and apoptosis, suggesting that a perturbed DDR is responsible for the syndromic phenotype of the patients (Tessadori *et al.*, 2017).

RIDDLE is a developmental syndrome characterised by radiosensitivity, immunodeficiency, dysmorphic features and learning difficulties (Stewart *et al.*, 2007). RIDDLE is caused by mutations in the E3 ubiquitin ligase, RNF168, which binds ubiquitinated H2A and promotes further ubiquitination, resulting in the recruitment of 53BP1 and BRCA1 (Stewart *et al.*, 2009). Cells derived from a patient with RIDDLE syndrome are hypersensitive to IR, and have defective cell cycle checkpoints and aberrant CSR (Stewart *et al.*, 2007). This suggests that the RNF168 mutations cause RIDDLE syndrome through decreased accumulation of 53BP1 and BRCA1 at DSBs and the consequent impairment of DNA repair.

Mutation	Syndrome	Phenotype	
		Shared with CdLS	**Not in CdLS**
NHEJ proteins			
DNA ligase IV	LiDS	Normal to severe combined immunodeficiency, sensitivity to IR, growth failure, severe microcephaly and learning difficulties	
HR proteins			
BRCA1	FA	Short stature, microcephaly, developmental delay, dysmorphic face	Early onset cancer, increased sensitivity to DNA damage
RAD51	FA	Developmental delay, skeletal anomolies	Increased sensitivity to DNA damage
Signal transducing proteins			
ATM	AT	Immune dysfunction	Cancer predisposition, neuro-degeneration,
ATR	SS	Intellectual disability, microcephaly, severe craniofacial dysmorphism, cardiovascular abnormalities	Primordial dwarfism, and central-nervous system abnormalities
Histone modifications			
H4K91		Growth delay, microcephaly, distinct facial features and intellectual disability.	
RNF168	RIDDLE	Radiosensitivity, immunodeficiency, dysmorphic features and learning difficulties	

Table Ch1-1: Overview of DDR defective syndromes described in the text

1.6 PhD aims

The current literature regarding the mechanism of CdLS is moving away from the idea that CdLS is simply a cohesinopathy. Instead it focusses on the identification of causative genes that are not components of the cohesin complex, the transcriptional dysregulation observed in many patients, and the overlap between CdLS and other syndromes caused by regulators of transcription or chromatin structure.

David FitzPatrick's lab at the MRC human genetics unit have recently identified mutations in the *BRD4* gene in three patients with CdLS-like syndromes. BRD4 is not a member of the cohesin complex, or involved in its regulation, and is best characterised as a transcriptional regulator. Given the current focus of the field, understanding how these mutations affect BRD4 and how this might cause CdLS is of importance.

The aim of my PhD is therefore to determine the effects of one of these BRD4 mutations – Tyr430Cys (Y430C) – on BRD4 function by:

- Recapitulating the Y430C mutation in mESCs
- Identifying binding partners of WT and Y430C BRD4
- Mapping the genome wide chromatin binding of WT and Y430C BRD4
- Analysing the effects of the Y430C mutation on transcription

Chapter 2: Materials and Methods

2.1 Stock solutions, reagents and buffers

4% paraformaldehyde – 20 g paraformaldehyde was added to 500 ml PBS and heated to 60°C in a water-bath overnight. Dissolved 4% paraformaldehyde was aliquoted and frozen.

4sU washing buffer: 100 mM Tris HCl pH 7.5, 10 mM EDTA, 1 M NaCl, 0.1% Tween 20

5X DNA Loading Buffer: 50% glycerol; 5 mM EDTA pH8; 0.3% Orange G (v/v).

ChIP elution buffer: (filtered 0.2-0.45 micron filter unit) 1% SDS, 0.1 M $NaHCO_3$

ChIP RIPA buffer: (filtered 0.2-0.45 micron filter unit) 1X PBS; 1% NP-40; 0.5% sodium deoxycholate; 0.1% SDS; Complete Mini EDTA-free protease inhibitor* (Roche)

DNA lysis buffer: 10 mM Tris-HCl (pH 7.5); 10 mM EDTA; 0.5% SDS; 10 mM NaCl; 1 mg/ml Proteinase K*

PBS: PBS was made by technical services at MRC HGU, except for PBS used in culturing cells. 160 mM NaCl; 3 mM KCl; 8 mM Na_2HPO_4; 1 mM KH_2PO_4 from tablets (Oxoid, ThermoScientific, cat# BR0014).

NaCl/EtOH: 0.3 M NaCl in 100% EtOH

Farnham lysis buffer: (filtered 0.2-0.45 micron filter unit) 5 mM PIPES pH 8.0; 85 mM KCl; 0.5% NP-40; Complete Mini EDTA-free protease inhibitor* (Roche)

LiCl wash buffer: (filtered 0.2-0.45 micron filter unit) 100 mM Tris pH 7.5, 500 mM LiCl, 1% NP-40, 1% Sodium deoxycholate

RIPA buffer for nuclear IP: 50 mM Tris pH 7.5, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, benzonase and Complete Mini EDTA-free protease inhibitor (Roche)

RIPA buffer for protein extraction: 150 mM sodium chloride; 1.0% NP-40; 0.5% sodium deoxycholate; 0.1% SDS; 50 mM Tris, pH 8.0

Swelling buffer: 10 mM Hepes; pH 7.9, 1.5 mM $MgCl_2$; 10 mM KCl; 0.5 mM DTT*; Complete Mini EDTA-free protease inhibitor* (Roche)

TAE buffer: 10mM Tris base, 20mM acetic acid, 1mM EDTA (prepared by technical services)

TBE buffer: 40 mM Tris base, 20 mM boric acid, 1 mM EDTA (prepared as 20X stock by technical services)

TE buffer: 10 mM Tris-HCl pH7.6, 1 mM EDTA (prepared by technical services)

TBS: 50mM Tris-HCl pH7.4, 150mM NaCl. (prepared by technical services)

TBS-T: TBS with 0.1% tween

Trypsin: 10X Trypsin (Sigma; 59427C), aliquoted and stored long term at -20ºC, and short-term at 1X concentration at 4ºC.

Western Transfer Buffer (wet transfer): 25 mM Tris base, 200 mM glycine, 20% methanol (v/v), 0.02% SDS (v/v). Made fresh prior to use.

*added fresh

2.2 Cell culture

2.2.1 mESCs

BRD4-Y430C mutant and corresponding WT mESCs were generated by CRISPR Cas9 genome editing in 46C mESCs (Ying *et al.*, 2003)(*RRID:CVCL_Y482)*, as described below. 46C cells were cultured on 0.1% gelatin coated tissue culture flasks (Corning) in GMEM medium (GIBCO; 11710035) supplemented with 10% Fetal Calf Serum (FCS), 5% penicillin-streptomycin, 1 mM sodium pyruvate (GIBCO; 11360070), 1X non-essential amino acids (GIBCO; 11140050), 50 μM 2-Mercaptoethanol (GIBCO; 31350010), 2 mM L-glutamine and 500 U/ml Leukaemia Inhibitory Factor (LIF) (in house). Media was replaced every 24 hours and cells were passaged every 48-72 hours, seeding at a density of $2x10^5$ cells/ml. Cells were grown at 37°C in a 5% CO_2 humidified atmosphere. Cells were frozen at ~$3x10^6$ cells/ml in 80% mESC media, 10% FCS, 10% DMSO and stored in liquid nitrogen.

2.2.2 Lymphoblastoid cells

I1206del (Tonkin *et al.*, 2004) and R2298H (Gillis *et al.*, 2004) LCLs from CdLS patients and cultured in RPMI 1640 medium (GIBCO; 11875093) supplemented with 15% FCS and 2

mM L-glutamine. Cells were grown in suspension and passaged every 72-96 hours, seeding at a density of $2x10^5$ cells/ml. Cells were grown at 37°C in a 5% CO_2 humidified atmosphere. Cells were frozen at ~$5x10^6$ cells/ml in RPMI 1640 with 20% FBS and 6% DMSO and stored in liquid nitrogen.

2.2.3 Drosophila Schneider 2 (S2) cells

S2 cells were cultured in Schneider's Drosophila Medium (Invitrogen; 11720-034), supplemented with 10% heat-inactivated FCS and 5% penicillin-streptomycin. Cells were passaged once they reached a density of ~$2x10^7$ cells/ml and seeded at a density of ~$4x10^6$. Cells were grown at 28°C in a 5% CO_2 humidified atmosphere. Cells were frozen at a density of ~$1x10^7$ cells/ml in 45% conditioned Schneider's Drosophila Medium media (containing 10% FCS), 45% fresh Schneider's Drosophila Medium supplemented with 10% FCS, and 10% DMSO, and stored in liquid nitrogen.

2.2.4 Induction of DNA damage

Cells were incubated with mESC media supplemented with either neocarzinostatin (NCS) (Sigma; N9162), to a final concentration of 25 ng/ml, or with hydroxyurea (HU) (Acros Organics; 151680250), to a final concentration of 50 mM for 15 mins at 37°C.

2.2.5 Aphidicolin treatment

Aphidicolin (APH) (Sigma; A0781) was resuspended in DMSO to a concentration of 1.5 mM and added to mESC media to a final concentration of 0.2 μM. mESCs were plated on gelatanised coverslips and incubated at 37°C for 24 hours, with APH supplemented media. Control cells were incubated with media containing equal percentage of DMSO.

2.2.6 Transcriptional inhibition

Cells were incubated with mESC media supplemented with 500 nM triptolide (Cayman; CAYM11973-5) (Trp) for 2 hours at 37°C, 20 μM alpha amanitin (Sigma; A2263) for 9 hours or 5 μg/ml actinomycin D (Fisher Scientific; BP60610) for 6 hours.

2.3 Cell cycle analysis

2.3.1 Growth assay

mESCs were seeded in 4 wells of a 6-well plate (1 x 10^4 cells/well). At 24, 48, 72 and 96 hours post seeding, the cells from 1 well were trypsinised, centrifuged and resuspended in media. Cell counting was carried out manually using a haematocytometer. The addition of trypan blue dye (final concentration 0.32%) allowed for the exclusion of dead cells.

2.3.2 Flow cytometry

2 million mESCs or LCLs were fixed in 70% ethanol (in PBS) at 4°C for 1 hour. Fixed cells were centrifuged at 931 g at 4°C for 5 min, washed twice with PBS and resuspended in 500 µl PBS. 20 µg RNase A was added and cells were incubated at 37°C for 10 min. Cells were stained with propium iodide at a final concentration of 50 µg/ml. Acquisition was carried out on a BD LSRFortessa cell analyser, collecting 25,000 events per sample. BD FACSDivasoftware (Becton Dickinson, Version 8.0.1) was used for instrument control and data analysis. Gated cells were manually categorized into cell cycle stages G0/G1, S and G2/M.

2.4 Genome editing in mESCs

BRD4-Y430C mESCs and their WT counterparts were generated using the CRISPR-Cas9 nickase system.

2.4.1 CRISPR-Cas9 construct design and cloning

CRISPR plasmids were designed and cloned by Hemant Bengani of the FitzPatrick lab. Briefly gRNAs were designed across Y430 using the online tool DNA 2.0. The WT and mutant repair templates (chr17: 32,220,150–32,220,271; GRCm38) (Table Ch2-1) were synthesized by IDT as 122-bp UltramerssODNs bearing the desired sequence change. gRNAs 1 and 2 (Table Ch2-1) were cloned into PX461 (Addgene; 48140) (Figure Ch2-1A) and PX462 (Addgene; 62987) (Figure Ch2-1B) respectively. Plasmid DNA was purified using the QIAGEN Plasmid *Plus* Maxi Kit (Qiagen; 12963) and diluted to 50 ng/µl with DNAse/RNAse free water.

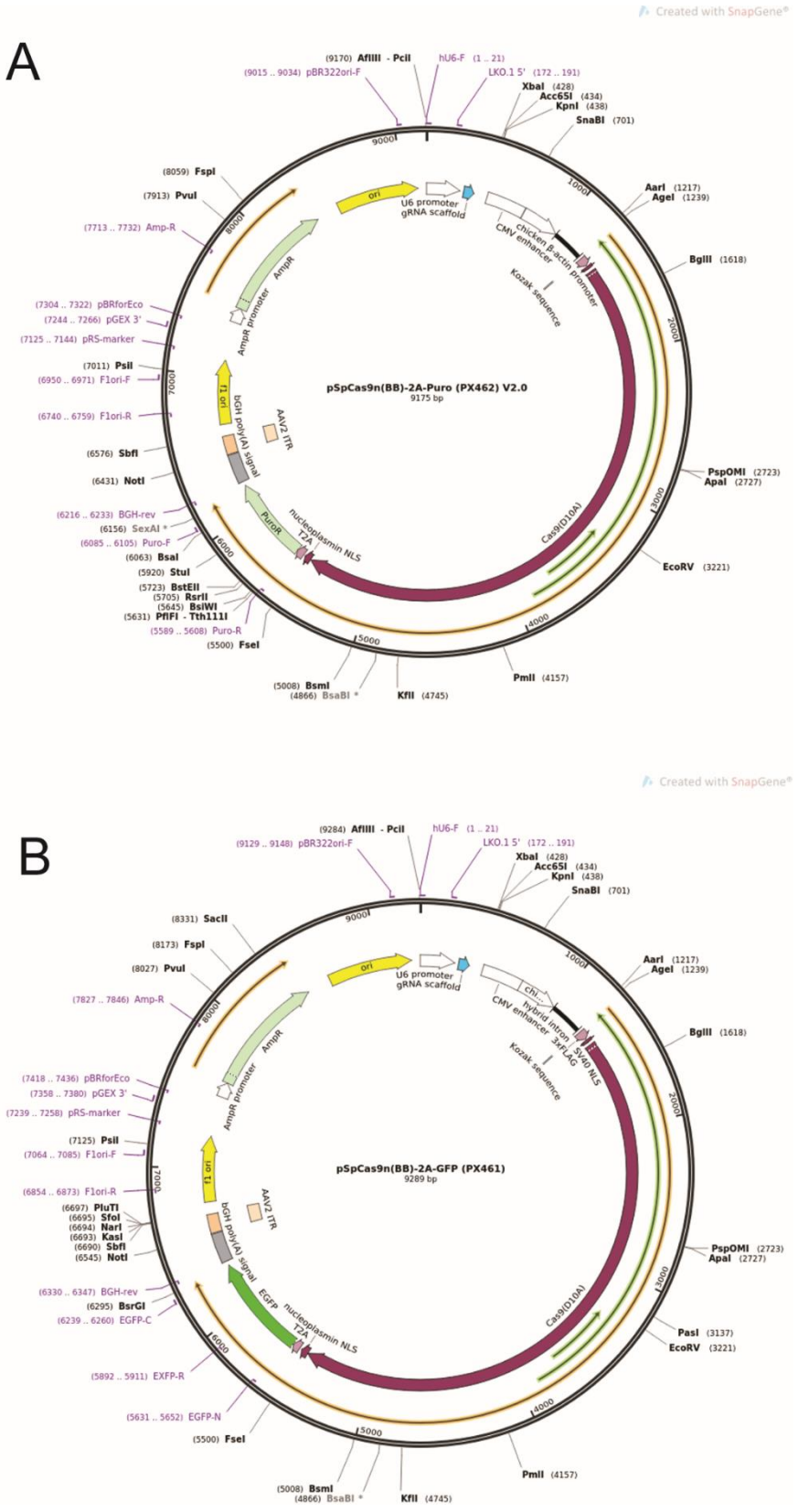


Figure Ch2-1. Full sequence map for CRISPR plasmids. **A)** PX461 and **B)** PX462

2.4.2 Transfections

46C mESCs were co-transfected with the two resulting plasmids and either WT or Y40C mutant repair templates using Lipofectamine® 3000 Reagent (ThermoFisher Scientific; L3000001) as per the manufacturer's instructions. Briefly, $1x10^6$ ESCs were transfected in a 6-well plate with 1 μg of each plasmid, 0.5 μg of repair template, 5 μl of lipofectamine and 2 μl P3000 reagent. After 48 hours, double transfected cells were selected for, first by resistance to puromycin (PX462) and subsequently by FACS based on GFP expression (PX461). Surviving cells were plated at $5x10^3$ cells/10 cm^2 and grown for 1 week. Colonies were picked and plated in duplicate at 1 colony/well of a 96-well plate. Genomic DNA was extracted from the colonies and sequenced by Sanger sequencing, using Y430C genotyping primers (Table Ch2-1). Confirmed WT and p.Tyr430Cys homozygous clones were expanded and frozen for later use.

Y430C-genotyping	**Forward**	**Reverse**
	GAAGCTACCTTGTTTCCCAT	TGCTCAGCCTGATCTATGTT
Cas9 gRNAs	**1**	**2**
	CAACCCCCCTGACCATGAAG	AGCAGTTGGAGAACATCAAT
Repair templates	**Wild type**	**Y430C mutant**
	GAGAGTACAGAGATGCCCAG GAATTTGGTGCTGATGTCCGA TTGATGTTCTCCAACTGCTGC AAGTACAACCCCCCTGACCAT GAAGTGGTAGCCATGGCTCG AAAACTCCAGGATGTGTT	GAGAGTACAGAGATGCCCAGGA ATTTGGTGCTGATGTCCGATTGA TGTTCTCCAACTGCTACAAGTACA ACCCCCCTGACCATGAAGTGGTA GCCATGGCTCGAAAACTCCAGGA TGTGTT

Table Ch2-1: Oligonucleotides used for mESC genome editing

2.5 Preparation and handling of DNA and RNA

2.5.1 DNA extraction

For 96 well plates, DNA was extracted once cells were highly confluent and media had turned yellow. Wells were rinsed 2X with PBS. 50 µl DNA lysis buffer was added per well. Plates were sealed with parafilm and incubated overnight in a humid chamber at 55°C. Plates were allowed to cool to room temperature (RT) and lysed cells were transferred to a microcentrifuge tube. DNA was purified by phenol-chloroform extraction and recovered by ethanol precipitation. DNA was resuspended in 50 µl TE.

For all other plates DNA was extracted using the Qiagen DNeasy blood and tissue kit following the manufacturer's instructions, with the addition of RNase A.

2.5.2 Resolution of DNA on agarose gels

DNA was resolved on agarose gels made to the appropriate percentage; UltraPure™ Agarose (ThermoFisher Scientific; 16500500) was dissolved in TAE and ethidium bromide was added at 0.5 µg/ml. 5X DNA loading buffer was added to the sample to give a final concentration of 1X. 1 Kb Plus DNA Ladder (Invitrogen; 10787018) was loaded as a size marker and DNA was resolved by subjecting to 150 volts. Gels were visualised under UV light (BioDoc-It System, UVP).

2.5.3 Quantification of DNA

Routinely DNA was quantified using the NanoDrop 8000 spectrophotometer (ThermoFisher Scientific; ND-8000-GL) as per the manufacturer's instructions. ChIP-DNA was quantified with the Qubit dsDNA HS Assay Kit (ThermoFisher Scientific; Q32851) and Qubit 4 fluorometer (ThermoFisher Scientific; Q33226) following the manufacturer's instructions. Briefly the Qubit dsDNA HS reagent was diluted 1:200 in Qubit dsDNA HS buffer. 10 µl of each Qubit standard was added to 190 µl diluted buffer, and 1 µl of each DNA sample was added to 199 µl diluted buffer. Samples were vortexed for 2-3 seconds and incubated at RT for 2 min. Standards were read on the Qubit 4 fluorometer, and samples were then assayed.

2.5.4 RNA extraction and cDNA synthesis

RNA was extracted from cells using the RNeasy Mini Kit (Qiagen; 74104) using spin technology, with an additional on-column DNA digestion using the RNase-Free DNase Set (Qiagen; 79254). cDNA was synthesised from 1 µg RNA using SuperScript II Reverse Transcriptase (ThermoFisher Scientific; 18064-014) as per manufacturer's instructions.

2.6 Expression analysis

2.6.1 Expression analysis by RT-qPCR

cDNA was diluted 1 in 25 for qPCR analysis. SYBR-green based qPCR reactions were performed (Table Ch2-2) in a final volume of 20 µl containing diluted cDNA, SYBR select master mix (ThermoFisher Scientific; 4472908) and 0.5 µM/l of region specific intron-spanning primer pairs (Table Ch2-3).

Stage	**Temperature (ºC)**	**Time (min:sec)**	**Ramp speed (ºC/s)**	**Acquisition**	**# cycles**
Pre-incubation	50	02:00	4.4	None	1
	95	02:00			
Amplification	95	00:15	4.4	None	50
	60	00:50	2.2		
	60	00:10	4.4	Single	
Melt	95	00:05	4.4	None	1
	65	01:00	2.2		
	97	00:30	0.11	Continuous, 5 acquisitions/ ºC	
Cooling	40	00:30	2.2	None	1

Table Ch2-2: qPCR experiment details

	Forward	Reverse
Klf4	GTGCAGCTTGCAGCAGTAAC	AGCGAGTTGGAAAGGATAAAGTC
Myc	CCCTAGTGCTGCATGAGGA	CGTAGTTGTGCTGGTGAGTG
Sox2	AAGAAAGGAGAGAAGTTTGGAGCC	GAGATCTGGCGGAGAATAGTTGG
Oct4	CGAGAACAATGAGAACCTTC	CCTTCTCTAGCCCAAGCTGAT
Nanog	TGGTCCCCACAGTTTGCCTAGTTC	CAGGTCTTCAGAGGAAGGGCGA
Brd4	CCTCCTCTCCTCTCATTCCCA	CATTCTTCACCAGGCACTCCAT
Gapdh	TGCGACTTCAACAGCAACTC	CTTGCTCAGTGTCCTTGCTG

Table Ch2-3: Primers for expression analysis

2.6.2 JQ1 treatment

1 mM BRD4 inhibitor (+)-JQ1, or its inactive form (-)-JQ1 (Merck; 500586) (diluted in DMSO), was added to mESC media at a final concentration of 300 nM JQ1+/-. mESCs were incubated at 37°C with JQ1+/- supplemented media for 48 hours. For qPCR analysis concentration of JQ1+ cDNA was calculated relative to JQ1- (arbitrarily set to 1).

2.6.3 4-thiouridine (4sU) labelling

4sU RNA was generated and isolated as in Pintacuda *et al.*, 2017, with some minor changes; 4-thiouridine (4sU, Sigma, T4509) was added to mESC media to a final concentration of 500 μM and the media was warmed to 37°C. ~10 million cells were incubated at 37°C with 4sU-supplemented medium for 20 min. Cell culture medium was rapidly removed from cells and total RNA was extracted with Trizol (Invitrogen) and treated with DNase using the DNA-free DNase treatment kit (Ambion; AM1906). After addition of inactivation reagent, RNA and beads were transferred to QIAshredder columns (Qiagen; 79656) and centrifuged at 1000 g for 1 min. Flow-through (RNA) was transferred to a new tube. For each μg of total RNA, 2 μl Biotin-HPDP, diluted to 1 mg/ml in DMF, and 1 μl 10xBiotinylation buffer (100 mM Tris HCl pH 7.4, 10 mM EDTA), were added. Reaction volume was made up to

10 µl per µg of RNA used. The reaction was incubated with rotation for 1.5 hours at RT. RNA was transferred to Phase Lock Gel Heavy Tubes (Eppendorf), and an equal volume of chloroform was added. After vigorously mixing, tubes were left incubating for 3 min at 25°C and then centrifuged at 16,089 g for 5 min at 4°C. The upper phase was transferred to new Phase Lock Gel Heavy Tubes, and the process was repeated. The upper phase was again transferred to Phase Lock Gel Heavy Tubes. 1/10 reaction volume of 5M NaCl was added, followed by 1X reaction volume of isopropanol. Tubes were mixed to invert and incubated at RT for 10 min. Tubes were centrifuged at 10,000 g for 20 min at RT. The pellet was washed in 80% ethanol, centrifuged at 10,000 g for 10 min at 4°C and resuspended in 100 µl water. RNA was dissolved by heating to 40ºC for 10 min in a shaker, and then put on ice for 5 min.

Biotinylated 4sU-RNA was recovered using the µMacs Streptavidin Kit (Miltenyi; 130-074-101), with a modified protocol and buffers. 1 µl of streptavidin beads was added per µg of recovered biotinylated 4sU-RNA, in a total volume of 200 µl, and samples were incubated with rotation for 15 min at 25°C. µMacs columns supplied with the µMacs Streptavidin Kit were equilibrated in 900 µl of washing buffer at 65°C. Samples were added to the columns that were then washed 3 times with 900µl washing buffer at 65°C and three times with 900 µl washing buffer at 25°C. RNA was eluted by adding 100 µl Elution Buffer (100 mM DTT) to the column and flow-through was collected. Elution was repeated 3 min later and flow-throughs combined. RNA was further purified using the RNAeasy MinElute (Qiagen; 74204) according to the manufacturer's guidelines, eluting in 20 µl water.

2.6.4 Generation of spike-in RNA

mESCs and S2 cells were harvested and counted. 0.2 million S2 cells were mixed with 10 million mESCs, and RNA was extracted using the RNeasy Mini Kit (Qiagen; 74104) using spin technology, with an additional on-column DNA digestion using the RNase-Free DNase Set (Qiagen; 79254).

2.7 Cross-linked Chromatin Immunoprecipitation (ChIP)

30 million cells were harvested by trypsinising and fixed with 1% formaldehyde (Thermo Fisher Cat. 28906) in media (25°C, 10 min). This reaction was quenched with 0.125 M Glycine. Cross-linked cells were re-suspended in Farnham lysis buffer for 30 min and centrifuged at 228 g for 5 min at 4°C. Nuclei were resuspended in RIPA buffer and sonicated using a Bioruptor® Plus sonicator (Diagenode) at full power for 60 min (30 seconds on, 30 seconds off) to produce fragments of 100-500 bp. 5 µg of each antibody (Table Ch2-4) were incubated with Protein A Dynabeads (ThermoFisher Scientific; 10001D) in 5 mg/ml BSA in PBS on a rotating platform at 4°C for two hours. An arbitrary concentration of 200 µg chromatin was incubated with antibody bound Dynabeads in a rotating platform at 4°C for 16 hours. 10% input was removed, prior to incubation on beads, and frozen for later use.

Name	Catalogue number	Dilution
BRD4	Bethyl; A301-985A	5 µg/ml
53BP1	Novus; NB100-904	5 µg/ml

Table Ch2-4: ChIP Antibodies

After incubation, beads were washed 5 times (5 min each) on a rotating platform with cold LiCl wash buffer and two times with RT TE buffer. ChIP complexes were eluted in 100 µl of ChIP elution buffer and shaken at 37°C for 15 min. The pH of eluted immunoprecipitates (IP) was reduced by the addition of 6 µl 2M Tris-HCl pH 6.8. 20 µg RNaseA was added to IP and 10% input and all samples were incubated at 65°C for 1 hour. 20 µg Proteinase K was added and samples were incubated at 55°C for a further 12 hours. DNA was purified using QIAquick PCR Purification Kit (Qiagen; 28104), eluting in 50 µl elution buffer, and quantified by Q-bit high sensitivity DNA assay.

For qPCR analysis ChIP DNA was diluted 10-fold in ddH20, and qPCR was carried out as described (section 2.6.1), using primers designed for ChIP-qPCR (Table Ch2-5). Concentrations of IPs were calculated relative to 1% input.

Name	Forward	Reverse
Sox2 SE	TAGAGGAAGGAGCTGGAGGA	AAGGAAAGAAGGAGGGACGG
Klf4 SE	CACAATGCCAGCTATGCGAT	TCCTGCCCAAATGTGAGGAT
Nanog SE	GTGAAGGTAGTTTGCTGGGC	GGTCCTTTCCCACCCTCTAC
Oct4 SE	CCTTCGTTCAGAGCATGGTG	GAGCCTACCCTGAACTTCCC
Nanog TE	AGTGATAGTGTGGCAGCAGT	CCACCATTGTTACCACTGCC
C-myc promoter	GACTCGCTGTAGTAATTCCAG	GCAAAGCCCCTCTCACTCCA
C-myc SE	CAGAGCAGCCACGAAAGATC	ACTGCAGAGGAATGTGAGCT
Gsc promoter	CTTCTCGGCGTTTTCTGACT	AGCCAAGTGGAGACGACAG

Table Ch2-5: ChIP primers

2.8 Library preparation and sequencing

2.8.1 ChIP-seq

DNA libraries for ChIP-sequencing were prepared from 1 ng of total sheared chromatin (input) and ChIP-enriched fragments using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina (NEB; E7645S), following the manufacturer's instructions. Optional size selection steps were carried out, using SPRIselect (Beckman Coμlter; B23319). Libraries were indexed using NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1) (NEBnext; E7335) and amplified by PCR for 12 (Input) or 15 (IP) cycles. Library concentration and correct size distribution (100-500 bp) was confirmed on the Agilent 2100 Bioanalyser with the DNA HS Kit. For BRD4-ChIP sequencing was carried out at BGI (Hong Kong; 50-base single-end reads) using the HiSeq 4000 system (Illumina). For 53BP1 ChIP sequencing was carried out at WTCRF (Edinburgh; 75-base paired-end reads) using NextSeq 500/550 High-Output v2 (150 cycle) Kit (# FC-404-2002) on the NextSeq 550 platform (Illumina Inc, #SY-415-1002).

2.8.2 4sU-seq and spike-in RNA-seq

To make 4sU sequencing and RNA-seq libraries, RNA was first depleted of rRNA using the Low Input Ribominus Eukaryotic System V2 (ThermoFisher Scientific; A15027) as per the manufacturer's instructions. 600 ng of RNA was used as input and eluted in 5 μl RNase free water. All the resulting rRNA free RNA was then used to prepare 4sU sequencing libraries, using NEBnext ultra Directional RNA library prep kit of Illumina (NEB; E7420). RNA fragmentation was carried out at 94°C for 15 minutes, as suggested for intact RNA. Libraries were indexed with multiplex Oligos for Illumina® (Index Primers Set 1) (NEBnext; E7335) and amplified by PCR for 13 cycles. Library concentration and correct size distribution (100-500 bp) were confirmed on the Agilent 2100 Bioanalyser with the DNA HS Kit. Libraries were sequenced at BGI (Hong Kong; 100-base paired-end reads) using the HiSeq 4000 (Illumina).

2.9 Protein extraction and analysis

2.9.1 Protein extraction

Ice-cold RIPA buffer was added to adherent cells growing on plates (1 ml/10^7 cells). Cells were scraped, and the suspension transferred into a pre-chilled microcentrifuge tube. Tubes were incubated with shaking at 4°C for 30 min before centrifugation at 16,089 g for 15 min. Supernatant was retained and quantified.

2.9.2 Protein quantification

The Bradford assay was performed to determine the concentration of protein samples. 5 μl of each sample was added to 1ml 1X Bradford dye and the absorption at 595 nm was measured using a spectrophotometer. The absorbance was compared to a standard curve drawn using BSA samples of known concentration (0.1, 0.2, 0.4, 0.6, 0.8 and 1.0 mg/ml) to calculate protein concentration.

2.9.3 Co-immunoprecipitation (Co-IP)

30 x 10^6 WT and Y430C BRD4 mESCs were trypsinised, pelleted and resuspended in 5 ml ice-cold swelling buffer for 5 min on ice. Nuclei were pelleted by centrifugation at 931 g for

5 min at 4°C. The resulting nuclear pellets were sonicated in 2 ml RIPA buffer, using a Bioruptor® Plus sonicator (Diagenode) at 4°C, 30 seconds on, 30 seconds off. It was noted that prolonged (1 hour) exposure to the detergents in RIPA buffer affected the interactions of BRD4 as measured by mass spectrometry (MS). Nuclear extracts were cleared by centrifugation at 16,089 g for 10 min at 4°C. Protein A Dynabeads (Life Technologies) were blocked prior to antibody coupling by washing 3 times with 5% BSA in PBS. Antibodies (Table Ch2-6) were coupled to the beads at 5 mg/ml by rotation for 1 hour at 4°C. Equivalent nuclear protein amounts were incubated with antibody coupled beads for 1 hour at 4°C. Beads were washed and pulled down proteins analysed by MS or western blot.

Name	Catalogue number	Host	Subtype	Dilution
NIPBL	Bethyl; A301-779A	Rabbit	IgG	5 µg/ml
SMC3	Bethyl; 0300-060A	Rabbit	IgG	5 µg/ml
BRD4	Bethyl; A301-985A	Rabbit	IgG	5 µg/ml
Normal rabbit IgG	Santa Cruz; sc-2025	/	/	5 µg/ml

Table Ch2-6: Antibodies used for co-IP

2.9.4 Western blotting

For western blot analysis beads (containing primary antibody and bound protein) were washed 5 times with RIPA buffer, bound proteins eluted by boiling in 1X NuPage LDS buffer (ThermoFisher Scientific) with 1X NuPage reducing agent (ThermoFisher Scientific) for 5 minutes and separated on a 3-8% tris-acetate gel (ThermoFisher Scientific; EA0375BOX) (reciprocal BRD4/SMC3/NIPBL IPs and MEF cell lysates) or 4-12% bis-tris gel (BRD4 IPs for acetylated histone binding) (ThermoFisher Scientific; NP0322BOX). Following electrophoresis, proteins were transferred to nitrocellulose membranes (ThermoFisher Scientific) using an iBlot 2 Dry Blotting System (ThermoFisher Scientific) for 7 min (when probing for proteins <250 kDa only) or to PVDF membranes by wet transfer for 90 min (when probing for proteins >250 kDa) and incubated with primary antibodies (Table Ch2-7) overnight at 4°C. Membranes were washed 3 times in tris-buffered saline-tween (TBS-T) and probed with HRP-conjugated secondary antibodies (Table Ch2-7) for 1 hour at RT. After 3 more washes in TBST, membranes were incubated with Pierce

ECL Western Blotting Substrate (ThermoFisher Scientific) for 5 min and imaged using an Image-quant machine (GE Healthcare).

Acetylated histone western blots were quantified using ImageQuantTL, 1D gel analysis (GE Healthcare). Input bands were set to an intensity of 1, and IP bands calculated as a percentage of input.

Name	Catalogue number	Host	Subtype	Dilution
BRD4	Bethyl; A301-985A	Rabbit	IgG	1:3000
CHK1	Abcam; ab47574	Rabbit	IgG	1:1000
CHK1-p	Cell signalling technologies; 2348	Rabbit	IgG	1:500
PCNA	Santa Cruz; sc-56	Mouse	IgG_{2a}	1:3000
NIPBL	Bethyl; A301-779A	Rabbit	IgG	1:1000
SMC3	Bethyl; 0300-060A	Rabbit	IgG	1:1000
ACTIN-B	Abcam; ab8229	Goat	IgG	1:500
H3	Abcam; ab1791	Rabbit	IgG	1:5000
H3K9ac	Abcam; ab10812	Rabbit	IgG	1:500
H4K8ac	Abcam; ab15823	Rabbit	IgG	1:1000
H3K27ac	Genetex; GTX128944	Rabbit	IgG	1:1000
Anti rabbit	Santa cruz; sc2030	Goat	IgG-HRP	1:10,000
Anti mouse	Santa cruz; sc2031	Goat	IgG-HRP	1:10,000

Table Ch2-7: Antibodies for western blot

2.9.5 MS

For analysis by MS, beads (from co-IP) were washed 3 times with TBS and excess buffer was removed. The beads were taken to the IGMM MS facility for further analysis. Briefly, immunoprecipitations were digested on beads, desalted and analysed on a Q-Exactive plus mass spectrometer as previously described (Turriziani *et al.*, 2014). Proteins were identified and quantified by MaxLFQ (Cox *et al.*, 2014) by searching with the MaxQuant version 1.5

against the Mouse proteome database (Uniprot). Modifications included C Carbamlylation (fixed) and M oxidation (variable). Bioinformatic analysis was performed with the Perseus software suite (Tyanova *et al.*, 2016).

2.9.6 Immunofluorescence

For immunofluorescence (IF) experiments mESCs were cultured on gelatinised coverslips and LCLs were grown in suspension. LCLs were harvested and resuspended in PBS to 1.8 x 10^5 cells/ml. 500 µl of cell suspension was added to a Shandon Single Cytofunnel (ThermoFisher Scientific; 5991040), with a microscope slide attached. Slides were centrifuged in a Shandon Cytospin 3 (ThermoFisher Scientific) at 800 rpm for 5 min, after which the LCLs had attached to the slide. All cells were fixed in 4% paraformaldehyde for 10 min and washed 3x 3 min in PBS. Cells were then permeabilised in 0.5% Triton (in PBS) for 10 min and washed 3x 3 min in PBS. Cells were blocked in 1% BSA in PBS for 30 min at RT, incubated with primary antibody (Table 8) diluted in 1% BSA for 1 hr at RT and washed 3x 3 min in PBS. Cells were next incubated with secondary antibody (Table Ch2-8) diluted in 1% BSA for 45 min at RT, washed 3x 3 min in PBS, incubated with DAPI in PBS (250 ng/ml) for 2 min, and washed 3x 3 min in PBS. Coverslips were mounted on slides in Vectashield mounting medium for fluorescence (Vector; H1000).

Name	Catalogue number	Host	Subtype	Dilution
OCT4	Santa-cruz; sc5279	Mouse	IgG_{2b}	1:400
ESRRB	R&D; PPH6705-00	Mouse	IgG2a	1:500
SOX2	Abcam; ab9759	Rabbit	IgG	1:1000
BRD4	Bethyl; A301-985A	Rabbit	IgG	1:100
53BP1	Novus; NB100-304	Rabbit	IgG	1:1000
RAD51	Calbiochem; PC130-100	Rabbit	IgG	1:500
Anti-Rabbit	Invitrogen; A11034	Goat	IgG-Alexa Fluor 488	1:750
Anti-Rabbit	Invitrogen; A10042	Donkey	IgG-Alexa Fluor 586	1:750
Anti-mouse	Invitrogen; A31571	Donkey	IgG-Alexa Fluor 647	1:750

Table Ch2-8: Antibodies for IF

2.10 Image capture and analysis

2.10.1 Imaging

All slides were viewed and images taken for counting/analysis using epifluorescence microscopes. Epifluorescent images were acquired using a Photometrics Coolsnap HQ2 CCD camera and a Zeiss AxioImager A2 fluorescence microscope with Plan-neofluar objectives (Carl Zeiss, Cambridge, UK), a Mercury Halide fluorescent light source (Exfo Excite 120, Excelitas Technologies) and Chroma #89014ET three colour filter set (Chroma Technology Corp., Rockingham, VT) The single excitation and emission filters are installed in motorised filter wheels (Prior Scientific Instruments, Cambridge, UK). Image capture was performed using Micromanager (https://open-imaging.com/).

Epifluorescent images were acquired using a Photometrics Coolsnap HQ2 CCD camera and a Zeiss AxioImager A1 fluorescence microscope with a Plan Apochromat 100x 1.4NA objective, a Lumen 200W metal halide light source (Prior Scientific Instruments, Cambridge, UK) and Chroma #89014ET single excitation and emission filters (Chroma Technology Corp., Rockingham, VT) with the excitation and emission filters installed in Prior motorised filter wheels. A piezoelectrically driven objective mount (PIFOC model P-721, Physik Instrumente GmbH & Co, Karlsruhe) was used to control movement in the z dimension. Hardware control, image capture and analysis were performed using Volocity (Perkinelmer Inc, Waltham, MA). FITC,Texas Red and DAPI signals were deconvolved using a calculated PSF with the iterative algorithm of Volocity (Perkinelmer Inc, Waltham MA).

2.10.2 Counting foci

Images (.tiff files) were opened in ImageJ as 2D stacks. The number of foci per cell (individual cells identified by DAPI staining) were counted, and cells were recorded as having 0, ≥1, ≥5 or ≥10 foci. Those with ≥10 foci would therefore be included in those with ≥5 and ≥1 foci. Results are plotted as a percentage of the total number of cells counted. ~50 cells were counted per experiment.

2.10.3 Foci size

The area of foci were quantified running a script (written by Ahmed Fetit, Imaging resources, IGMM) as a macro in ImageJ. Channels are split automatically. Setting the threshold for foci

detection, and selection of the regions of interest (foci), are carried out manually. The area of all the detected foci in the image are then calculated and output.

2.11 Bioinformatics

2.11.1 ChIP-seq mapping and peak calling

Fastq files were quality controlled using FastQC and mapped to the mm9 genome using Bowtie2 (parameters: default). Sam files were converted to bam files and sorted using SamTools. Homer was used to make tagdirectories (makeTagDirectory, parameters: –unique, fragLength 150) and bedgraphs (makeUCSCfile, parameters: default). For visualisation of BRD4 ChIP data, bedgraphs were uploaded to the genome browser UCSC (https://genome.ucsc.edu). For visualisation of 53BP1 ChIP data, bedgraphs were converted to bigwig files using Homer (bedGraphToBigWig) and uploaded to UCSC via CyVerse.

Peak calling was carried out using MACS2; Duplicates were filtered (filterdup, parameters:--keep-dup=1), peaks called (callpeaks, parameters: -B --nomodel -p 1e-5) and differential peaks were found (bdgdiff, parameters: -g 60 -l 250).

2.11.2 4sU-seq mapping and peak calling

Fastq files were quality controlled using FastQC and mapped to the mm9 genome using tophat (parameters: --library-type fr-firststrand -r 200). Homer was used to make tagdirectories (makeTagDirectory, parameters: -unique -sspe -flip -fragLength 150), and to make bedgraphs for visualisation on UCSC (makeUCSCfile, parameters: -strand separate -style rnaseq).

Cufflinks was used for peak calling; transcripts were assembled for individual experiments (cufflinks, parameters: –m 200 –library-type fr) and both replicates of WT and Y430C were combined to form one assembly (cuffmerge, parameters: default). Differentially expressed peaks were determined from this assembly using cuffdiff (Cuffdiff. Parameters: default).

2.11.3 Dotplots

Reads per kilobase per million mapped reads (RPKM) was quantified for all exons using Homer analyzeRepeats.pl (parameters: -count exons, --strand +, -norm 1e7, -rpkm). WT and Y430C Log2 (RPKM) were plotted against each other using custom R scripts.

2.11.4 Heatmaps

deepTools2 was used to make heatmaps; score files were made across specific genomic regions (computeMatrix, parameters: scale-regions scale regions –b 500 –a 500 –bs 50 –bl mm9 blacklist) and these were used to plot heatmaps (plotHeatmap, parameters: --colormap RdBluYl reverse).

2.11.5 Genomic region enrichment

Bioconductor package regioneR (Gel *et al.*, 2016) was used to assess the relationship between the WT and Y430C peaks called by MACS2 (section 7.10.1) and the peaks identified as differentially bound between WT and Y430C.

regioneR uses circular permutation (n=1000) to count the number of overlaps between two genomic regions. A z-score and p value is calculated between the observed number of overlaps and the random distribution of permuted regions. Only canonical, autosomal peaks and genomic regions were included in the permutation test.

Genomic regions based on a chromHMM segmentation of the mm9 genome were obtained from https://github.com/gireeshkbogu/chromatin_states_chromHMM_mm9. SE, typical enhancer and promoter regions were those defined by Flynn *et al.*, 2016.

2.11.6 Functional annotation analysis

DAVID functional annotation tool was used to identify GO (gene ontology) terms enriched in gene lists. Gene names were converted to ENTREZ_GENE_IDs using DAVIDs Gene ID conversion tool. This allows genes in the list to be recognised by DAVID. The lists were then uploaded for Functional Annotation, limiting annotations to Mus Musculus and using Mus Musculus as a background. The following GO terms - Biological processes (GOTERM_BP), cellular components (GOTERM_CC) and Molecular Function (GOTERM_MF) – and the functional category – UniProt Keywords (UP_KEYWORDS) were selected for functional annotation clustering. P-values were set to < 0.5. Results are shown in tables in the text.

Chapter 3: BRD4 interacts with NIPBL and is mutated in a CdLS-like syndrome

3.1 BRD4 haploinsufficiency causes a CdLS-like phenotype

Approximately 55% of CdLS cases are caused by mutations in the cohesin loader *NIPBL* and <10% are a result of mutations in the other cohesin related genes (Newkirk *et al.*, 2017) (see section 1.1). This means that for approximately 35% of patients, the causal mutation remains unknown. Studies have begun to identify patients with CdLS-like phenotypes that have mutations in genes unrelated to the cohesin complex (Ansari *et al.*, 2014; Yuan *et al.*, 2015), indicating the necessity for non-candidate driven approaches (section 1.2.8).

David FitzPatrick's lab carried out whole exome sequencing in a group of 92 individuals with CdLS, in whom no plausibly diagnostic variants could be identified in the known causative genes. They first identified one individual with a de novo heterozygous 1.04 Mb deletion encompassing 29-protein coding genes, including *BRD4* (Family 4198, Figure Ch3-1A). Targeted resequencing of this region in the remaining 91 individuals identified a second individual with a de novo mutation in *BRD4*, this time a missense mutation mapping to BD2 (c.1289A>G, p.(Tyr430Cys), Family 3049, Figure Ch3-1B). Recently two more affected individuals have been identified with de novo frameshift mutations in *BRD4*, the first through ongoing screening of patients (c.1224delinsCA, p.(Glu408Aspfs*4), CDL038, Figure Ch3-1B) and the second through analysis of trio whole exome sequencing generated by the Deciphering Developmental Disorders study (DDD study) (McRae *et al.*, 2017) (c.691del, p.(Asp231Thrfs*9), DDD 264293, Figure Ch3-1B).

The FitzPatrick lab reviewed the phenotypes of seven other individuals who had been reported to have heterozygous deletions encompassing *BRD4* (Figure Ch3-2). The phenotypes of these individuals showed significant overlap with CdLS, with at least two of the seven individuals (28.6%) fulfilling the diagnostic criteria for CdLS (Kline *et al.*, 2007).

Taken together, these data support *BRD4* haploinsufficiency as the genetic mechanism for the CdLS-like phenotype of the patients in Figure Ch3-1.

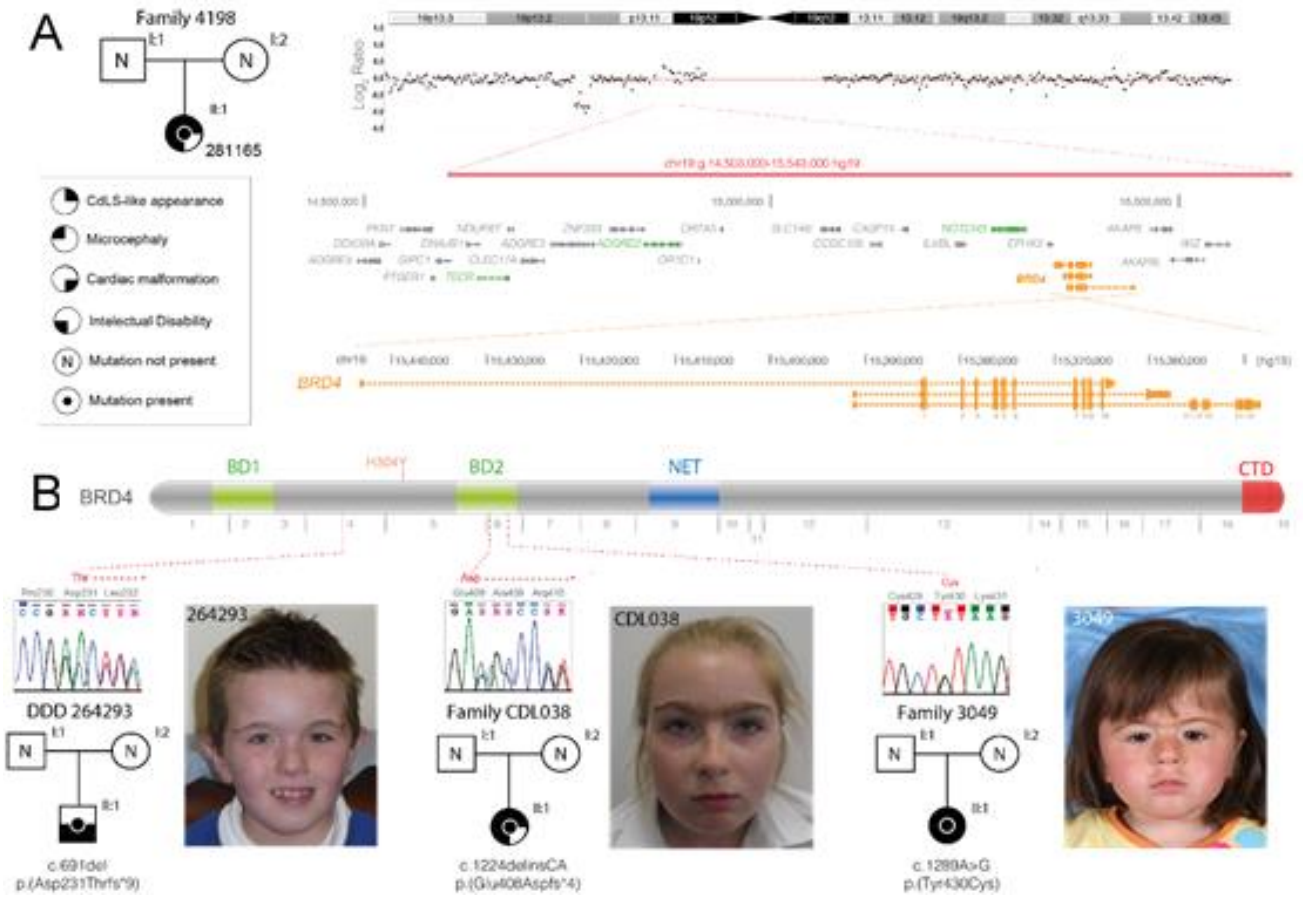


Figure Ch3-1. BRD4 mutations in patients with CdLS-like phenotypes. A. Pedigree drawing of proband 4198 with a de novo 1.04-Mb microdeletion of 19p (red bar), the location of which is shown on the $\log_2$ ratio plot of the array-based comparative genomic hybridization. Bottom left, key to the symbols used in the pedigree images. Bottom right, RefSeq gene content of the deleted region with the location of *BRD4* indicated in orange. **B.** Pedigree drawings and facial photographs of probands 264293, CDL038 and 3049 with intragenic mutations of *BRD4*. Above is a cartoon of the BRD4 protein indicating the position of each of the variants in relation to BD1, BD2, the N-terminal extraterminal domain (NET) and the C-terminal domain (CTD).

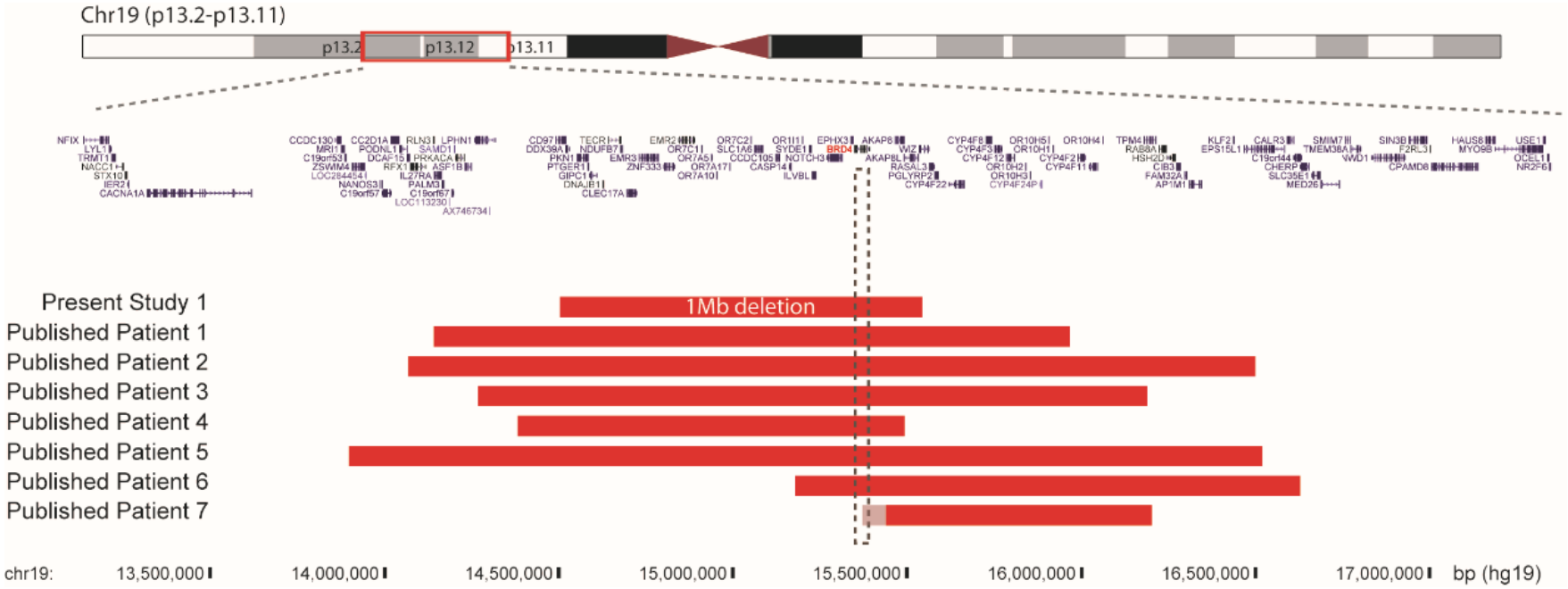


Figure Ch3-2. Summary of deletions reported in cases of BRD4 haploinsufficiency. Proband 4198 (Present Study 1) and 7 other patients - Published Patient 1 (Engels *et al.*, 2007), Published Patient 2 (Jensen *et al.*, 2009), Published Patient3 (Bonaglia *et al.*, 2010)), Published Patient 4 (Van der Aa, Vandeweyer and Kooy, 2010) , Published Patient 5 (Gallant *et al.*, 2011), Published Patient 6 (Jelsig *et al.*, 2012) and Published Patient 7 (Kosaki *et al.*, 2011) – all have deletions that affect *BRD4*. Locations of the deleted regions are shown as red blocks based on the human genome reference assembly hg19.

3.2 Experimental models of the Tyr430Cys mutation

In order to investigate the effects of *BRD4* haploinsufficiency on cell function I focussed on proband 3409, who carries the Tyr430Cys point mutation in BD2 (hereafter termed Y430C).

To assess the functional consequences of this mutation, CRISPR Cas9 (D10A) nickases were used to introduce the mutation into mouse embryos and mESCs; guide RNAs (gRNA) complimentary to the regions on either side of a target region direct the Cas9 nickases to these sites where they introduce SSBs in the DNA. Homology directed repair (HDR) is promoted by providing a repair template, which contains some homology with the region, but can also carry the specific mutation that is to be introduced (Figure Ch3-3).

3.2.1 Mouse embryos

David Fitzpatrick's lab first generated *Brd4*$^{Y430C/+}$ and *Brd4*$^{Y430C/Y430C}$ F_0 mouse embryos through injections of Cas9 reagents into single-cell mouse zygotes. At E13.5, *Brd4*$^{Y430C/Y430C}$ and *Brd4*$^{Y430C/+}$ were indistinguishable from *Brd4*$^{+/+}$ embryos (Figure Ch3-4A-C). They also generated apparently non-mosaic F_0 embryos homozygous for a 15-bp in-frame deletion (:c.1288_1302del, p.(Cys430_Asn434del) designated *Brd4*$^{C429\_N433del/C429\_N433del}$ (Figure Ch3-4D), that abolished expression of BRD4 (data not shown). *Brd4*$^{C429\_N433del/C429\_N433del}$ embryos showed significant growth restriction at E13.5 but had no other obvious phenotype. Such growth restriction has been seen by others in embryos heterozygous for a null *Brd4* allele and is thought to be attributable to a reduced proliferation rate. However, the fact that the *Brd4*$^{C429\_N433del/C429\_N433del}$ embryos survive to E13.5 is contradictory to previous reports, which concluded that homozygous knockout of BRD4 is lethal early in embryonic development, due to failure to maintain the inner cell mass (Houzelstein *et al.*, 2002).

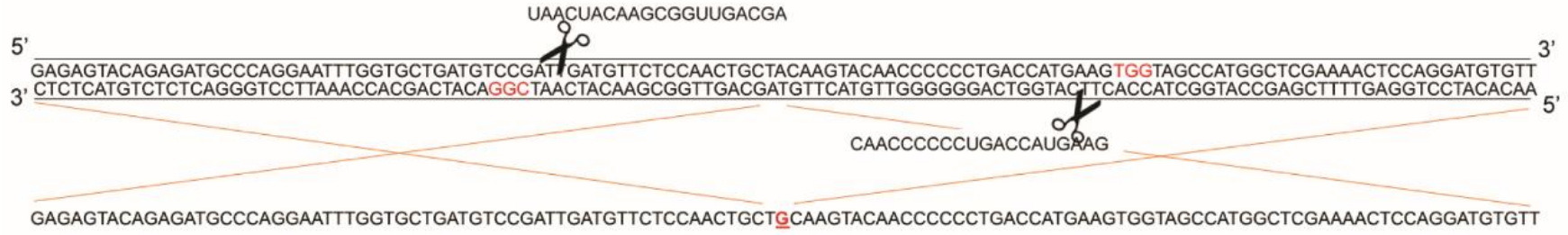


Figure Ch3-3. Y430C mutation was modelled in mouse embryos and mESCs using Cas9 (D10A). Complementary gRNAs (sequence with scissors) introduce nicks close to the PAM sites (red). The repair template contains arms of homology (orange lines) surrounding the desired point mutation (red and underlined).

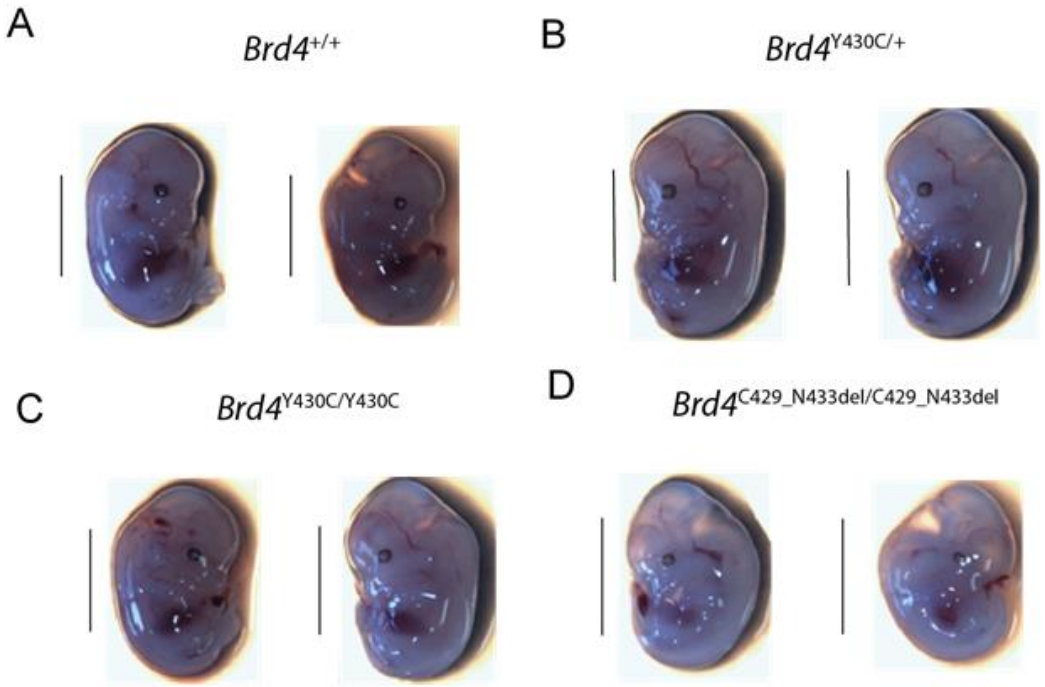


Figure Ch3-4. CRISPR targeting of Brd4 in mouse embryos. Images show E13.5 mouse embryos after CRISPR targeting of *Brd4*. Knock-in of WT or mutated repair template generated **A)** WT, **B)** *Brd4*$^{Y430C/+}$, **C)** *Brd4*$^{Y430C/Y430C}$, and **D)** *Brd4*$^{C429\_N433del/C429\_N433del}$ mouse embryos. Scale bar = 0.5 cm. (Hemant Bengani).

3.2.2 *Brd4*$^{Y430C/Y430C}$ mESCs

Pradeepa Madapura and I then individually engineered *Brd4*$^{Y430C/Y430C}$ mESCs. Cas9 reagents were transfected into 46C mESCs and successfully co-transfected cells were selected based on their puromycin resistance and positive GFP expression. We plated the resulting cells at low densities and, after 1 week, picked 96 colonies to plate in duplicate (1 colony/well of a 96-well plate). We then began extracting DNA from these colonies and sequencing it using primers that flank the target sites (Table Ch2-1, Y430C-genotyping); I sequenced 21 colonies and identified 1 homozygous clone (Figure Ch3-5A), giving a targeting efficiency of 5% (Table Ch3-1). A similar efficiency was found by Pradeepa. No heterozygotes were identified in either experiment. We expanded the two Brd4$^{Y430C/Y430C}$ clones for further use, giving me two independent cell lines.

I chose to use mESCs as these are an easily manipulated model system, for which there is extensive pre-existing epigenetic profiling datasets, both publicly available and within our lab. Furthermore, BRD4 is known to be important in the maintenance of the pluripotency and self-renewal of mESCs (Di Micco *et al.*, 2014). The mESCs were engineered to be homozygous for the Y430C mutation, in the hope that any phenotype might be more striking; *Brd4*$^{Y430C/+}$ and *Brd4*$^{Y430C/Y430C}$ embryos did not show a phenotype, suggesting that the effect of the mutation in the cells may be subtle.

Through qPCR analysis using primers for a region of the *Brd4* gene body (Table Ch2-3), and immunoblots of BRD4 protein, I determined that the Y430C mutation did not alter the levels of BRD4 protein or mRNA (Figure Ch3-5B&C). A single band of ~200 kDa was detected for BRD4 protein, indicating that the majority of BRD4 in the mESC is isoform A.

I also unsuccessfully attempted to introduce the *Brd4*$^{C429\_N433del/C429\_N433del}$ mutation into mESCs using CRISPR. A large proportion of the picked clones failed to expand after plating in 96 well plates. Those that did grow were screened by immunoblotting for BRD4 (Figure Ch3-5D), but all showed expression of BRD4. A smaller percentage of clones expanded in this experiment compared to the Y430C CRISPR experiment, and it is possible that those that failed to expand were in fact homozygous for the targeted BRD4 deletion. As others have seen that knockout of *Brd4* is embryonic lethal (Houzelstein *et*

al., 2002), I concluded that this mutation may not be viable in mESCs and decided not to pursue this further.

Experiment	Clones sequenced	Homozygous	Heterozygous	Homozygous targeting efficiency (%)
1	15	1	0	6.67
2	21	1	0	4.76
Average	18	1	0	5.56

Table Ch3-1. Results of CRISPR-Cas9 experiments

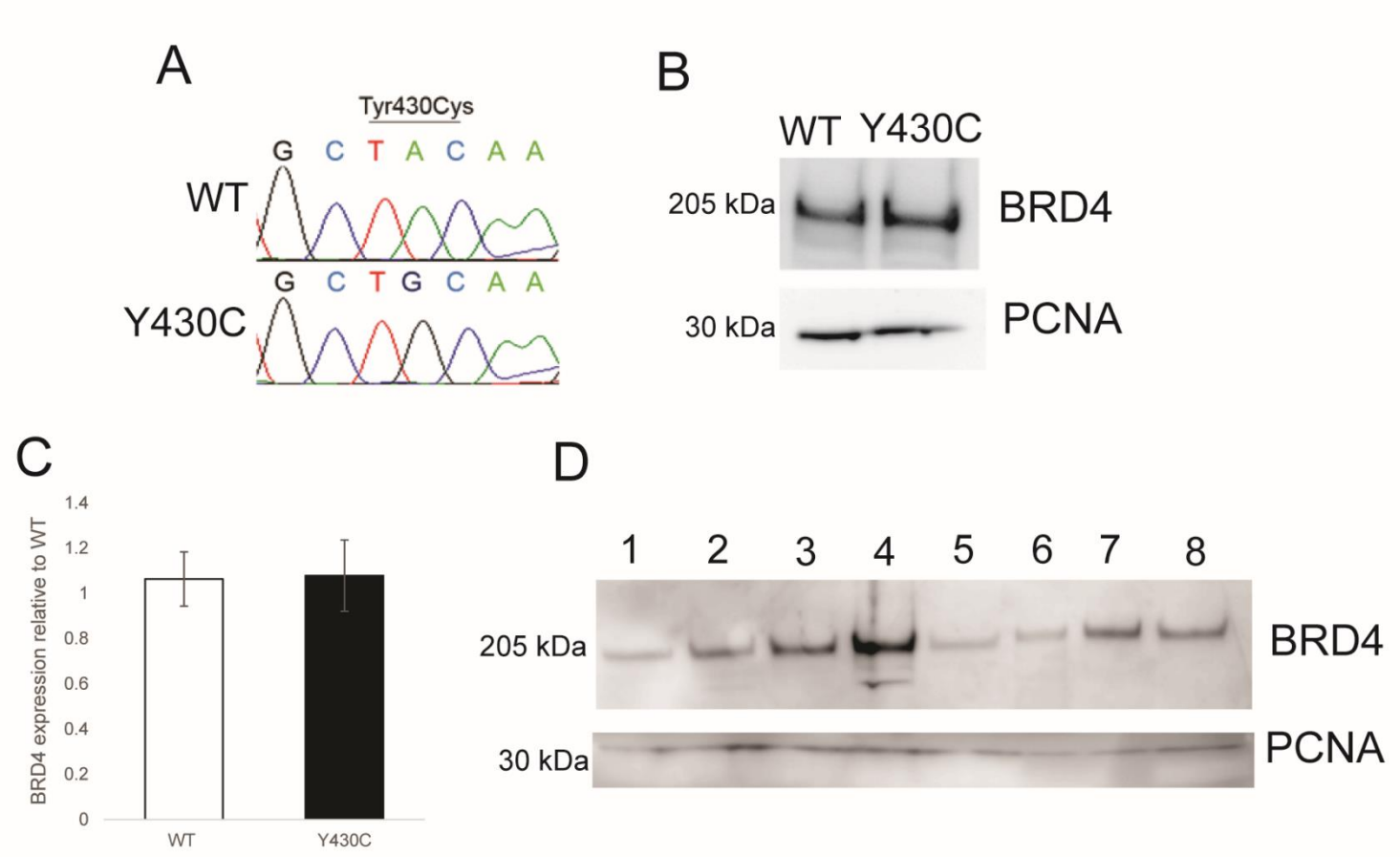


Figure Ch3-5. CRISPR targeting of Brd4 in mESCs. A. Chromatogram showing the DNA sequence of the CRISPR targeted region in WT and Y430C mESCs. **B.** Immunoblot for BRD4 and PCNA control in WT and Y430C mESCs. **C.** RT-qPCR analysis of *Brd4* mRNA levels in WT and Y430C mESCs. Error bars represent SEM from 3 technical replicates. **D.** Immunoblot for BRD4 and PCNA control in potential *Brd4*$^{C429\_N433del/C429\_N433del}$ mESC clones.

3.3 BRD4 interactions with histone and non-histone proteins

BRD4 is a member of the BET family of proteins (Rahman *et al.*, 2011). BRD4 has a number of roles in the cell, including transcriptional regulation (Hajmirza *et al.*, 2018), cell cycle regulation (Maruyama *et al.*, 2002) and DNA repair (Li *et al.*, 2018). In mESCs, BRD4 is important in the regulation of genes that maintain the pluripotent state (Di Micco *et al.*, 2014). The N-terminal BDs of BRD4 mediate interactions with acetylated histones, whilst the CTD and ET domain interact with multiple non-histone proteins (domains indicated in Figure Ch3-1B). For instance the CTD recruits P-TEFb (CDK9 and Cyclin T1) to promoters (Lu *et al.*, 2016) and the ET domain recruits JMJD6, SWI-SNF and CHD2 (Hajmirza *et al.*, 2018) (see section 1.4 for more details). I was interested to see which proteins BRD4 may interact with specifically in mESCs, and whether these interactions would differ between the WT and Y430C BRD4.

3.3.1 In silico modelling of Y430C mutation in BD2

3.3.1.1 Effect on BD stability

The homology between BDs of different proteinssuggests that the sequence and structure of the domain structure is important for its function, i.e. the binding of acetylated peptides. The Y430C mutation lies in the αB helix of BD2 of BRD4, just two amino acids away from the conserved Asn residue that mediates acetyl-lysine binding (Figure Ch1-9)(section 1.4.2.2). This proximity, and that Y430 is highly conserved between BDs, BETs and species (Figure Ch3-6) (Filippakopoulos *et al.*, 2010), led me to hypothesise that this mutation would affect the affinity of BRD4-Y430C for acetyl-lysines.

To investigate this further, Dinesh Soares, a computational structural biologist in the IGMM, modeled the Y430C mutation in silico. The crystal structure of BD2 in BRD4 has been determined many times, by both nuclear magnetic resonance (NMR) (PDB: 3JVL) and X-ray crystallography (PDB: 2YEM) and is available in both its unbound state and bound to acetylated peptides (PDB: 4KV4). Dinesh made use of these publicly available structures to determine the likely function of the Y430 side chain in the structure.

Y430 protrudes into the space between the helices. Here it supports hydrophobic interactions (within 5 Å) with side-chains of completely buried residues F426 (also in αB), V440 (αC) and A443 (αC). Additionally, the OH group of Y430 forms a H-bond with OE1 of Q447 (αC). All of these interactions are retained by the equivalent Tyr (Y137) in BD1 of BRD4 (PDB ID: 3MXF) (Figure Ch1-9) and are important in the packing of αB and αC for structural stability (Figure Ch3-7A).

When Y430 is mutated to C430, the hydrophobic interactions and putative aromatic-aromatic stacking interactions of Y430 with F426 are lost. Furthermore, the distance between this residue and the OE1 of Q447 increases, causing the loss of the important inter-helical H-bond. The overall result is a more open interface between the opposing αB and αC helices and a destabilisation of structure (Figure Ch3-7A).

Dinesh used FoldX (Schymkowitz *et al.*, 2005), a protein design algorithm that evaluates the effect of mutations on the stability, folding and dynamics of proteins, to predict the effects of the Y430C mutation in various BD structures from the RCSB Protein Data Bank (PDB). FoldX can be used to calculate the change in free energy (ΔΔG) between two protein conformations (eg. WT vs mutant, or bound vs unbound). ΔΔG correlates with experimentally observed changes in stability, with a higher value indicating a destabilisation of structure (Schymkowitz *et al.*, 2005). Mutation of Y137/430 was found to destabilise the BD monomer in every structure tested. The three NMR structures (yellow bars in Figure Ch3-7B) suggest comparatively moderate destabilising effects, but all high-resolution crystal structures indicate more severe loss of stability (ΔΔG ~>2.5 kcal/mol) (Figure Ch3-7B).

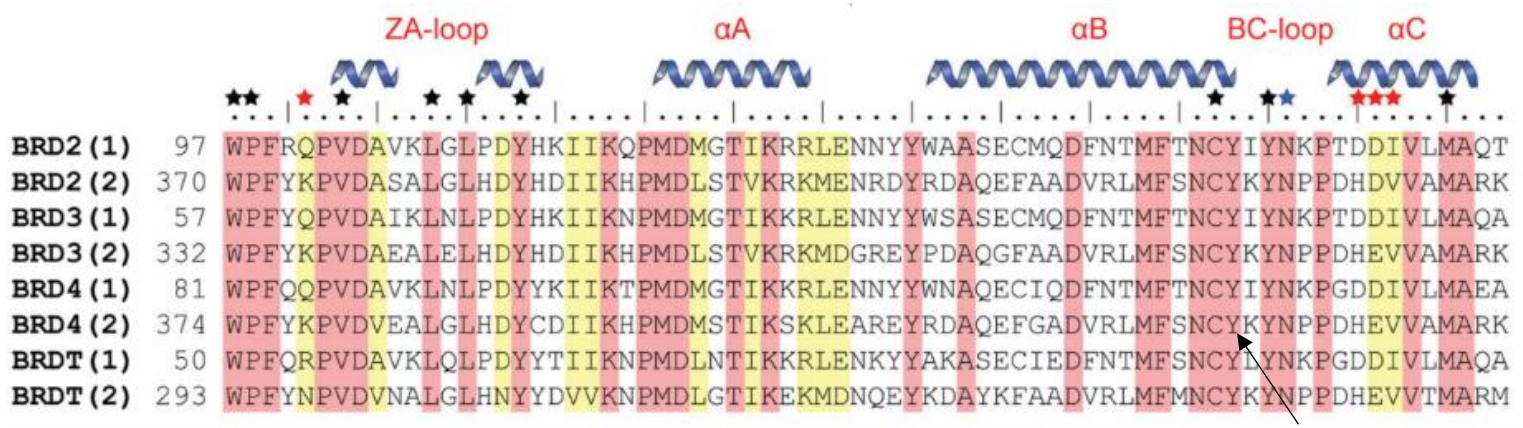
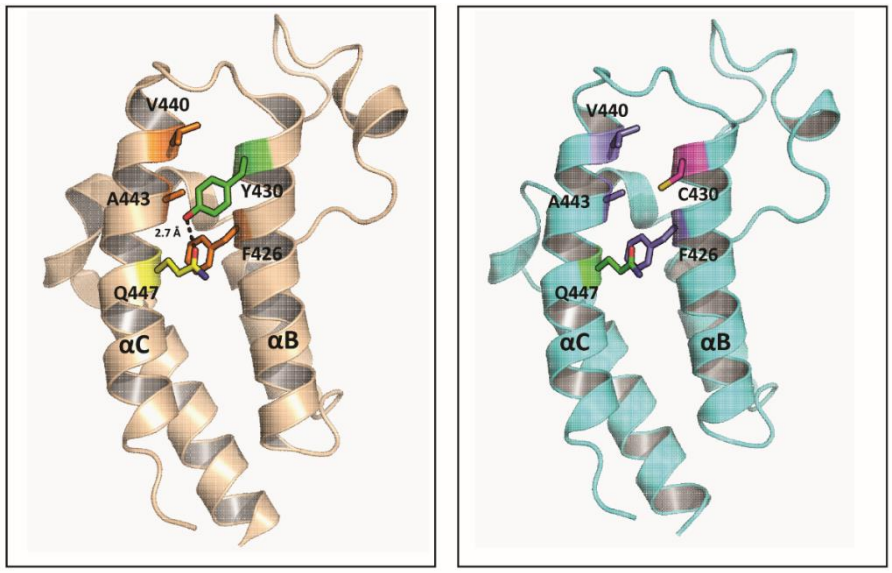


Figure Ch3-6. The BD module. A) Amino acid sequence alignments from BD1 and 2 in human BET proteins. Highly conserved regions are highlighted in pink. Similar regions are highlighted in yellow. Black arrow shows Y430 mutated in CdLS patient. *Taken from* Filippakopoulos et al., 2010

A



B

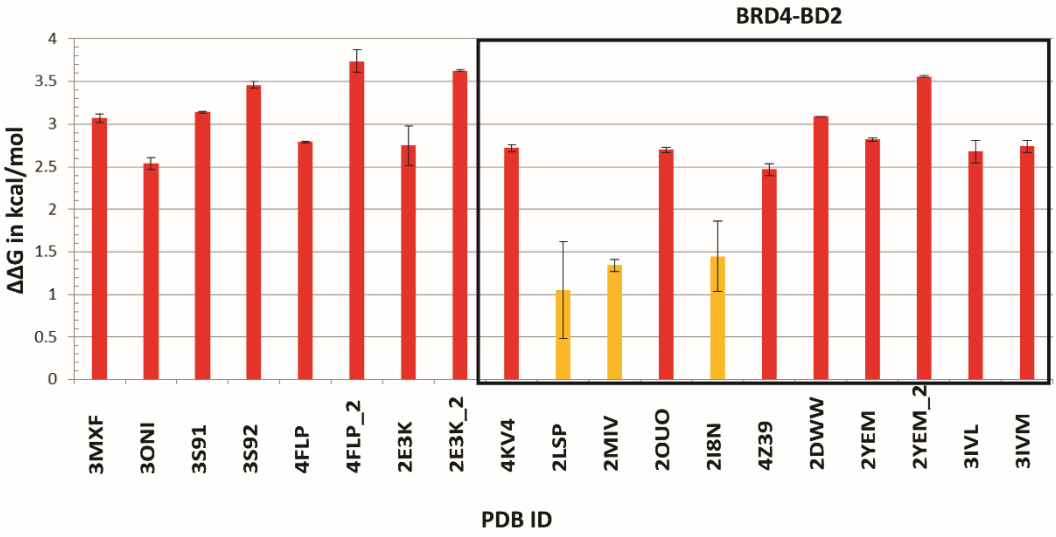


Figure Ch3-7. BRD4-BD2 in Y430 and C430. Crystal structure of BRD4-BD2 in WT (PDB ID: 2OUO (1.89 Å resolution) (from Filippakopoulos *et al.*, 2012) (left) and predicted Y430C (right) conformations. **B)** Changes in free energy between WT and Y430C conformations of various BD structures. ΔΔG was calculated using FoldX for BD structures from PBD. Black box indicates structures of BRD4-BD2 specifically. Others are various other BDs. Yellow bars - ΔΔG >0.8 to <1.6 kcal/mol: moderately destabilising. Red bars - ΔΔG >1.6 kcal/mol: severely destabilising. Average ΔΔG = 2.42 kcal/mol. structure (Schymkowitz *et al.*, 2005).

3.3.1.2 Effect on ligand binding

To model the effect of the Y430C mutation, and it's destabilisation of BD structure, on acetyl-lysine binding, Dinesh made use of available structures of BD2 in complex with the small molecule BET inhibitor, JQ1 (see section 1.4.2.3). No BRD4-BD2 structures were available however, so BRD4-BD2 in complex with JQ1 was modelled using structures of BRD4-BD2 in its unbound state (resolution 1.89 Å, 100% identity) and BRD3-BD2 bound to (+)-JQ1 (PDB ID: 3S92 resolution 1.36 Å, 79% identity) (Figure Ch3-8).

The Y430 residue was predicted not to participate directly in the binding of JQ1. However, it is located within 5 Å of N433, the residue that is involved in critical hydrogen bonding with JQ1. Consequently, the destabilisation caused by the Y430C mutation is likely to perturb the local structural environment and affect JQ1 binding.

Given the strong similarity between the binding of JQ1 and acetyl-lysines to BDs, it is likely that the mutation also disrupts BRD4's binding to acetylated histones.

3.3.2 BRD4 binding to acetylated histones

To experimentally test the hypothesis that the Y430C mutation affects binding of BRD4 to acetyl-lysines, Pradeepa Madapura compared the binding of the BD2 domain containing the Y430C substitution with that of WT BD2 and WT BD1 domains (purified by the Edinburgh Protein Production Facility). Modified histone tail peptide arrays showed that binding to a number of histone tail acetylation marks is reduced for Y430C-BD2 (Figure Ch3-9A).

I then coimmunoprecipitated BRD4 in the WT and Y430C mESCs and immunoblotted the co-precipitated proteins for various histone acetylation modifications. This revealed impaired BRD4-Y430C binding to H3K9ac and H3K27ac, confirming the in silico and in vitro results. Binding to H4K8ac was not affected (Figure Ch3-9B).

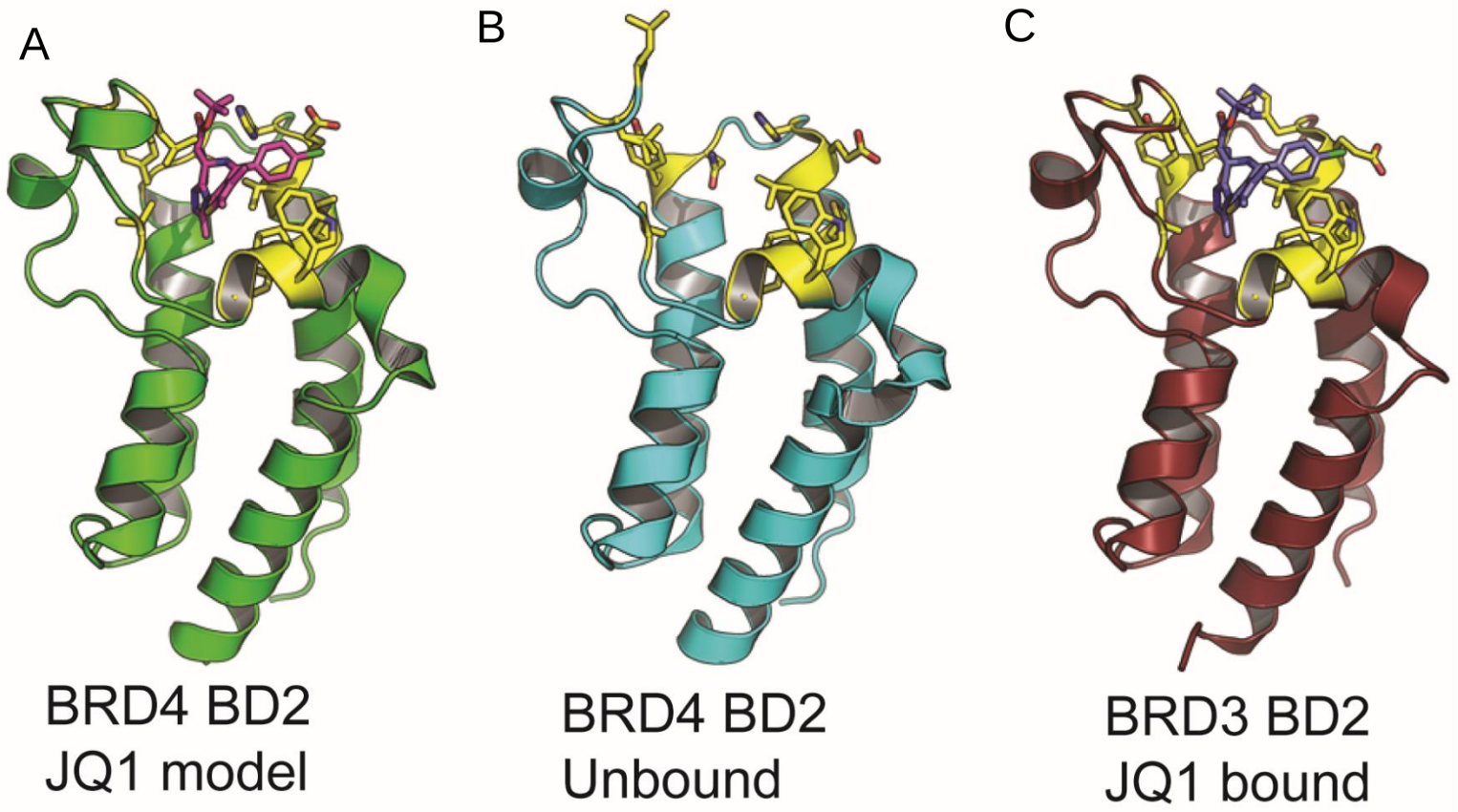


Figure Ch3-8. Modelling of JQ1 binding to BRD4 BD2. A) Structure of BRD4 BD2 in complex with JQ1, modelled using the crystal structures of **B)** BRD4 BD2 unbound (PDB ID: 2OUO, resolution: 1.89 Å) and **C)** BRD3 in complex with JQ1 (PDB ID: 3S92 (1.36 Å resolution). Work carried out by Dinesh Soares.

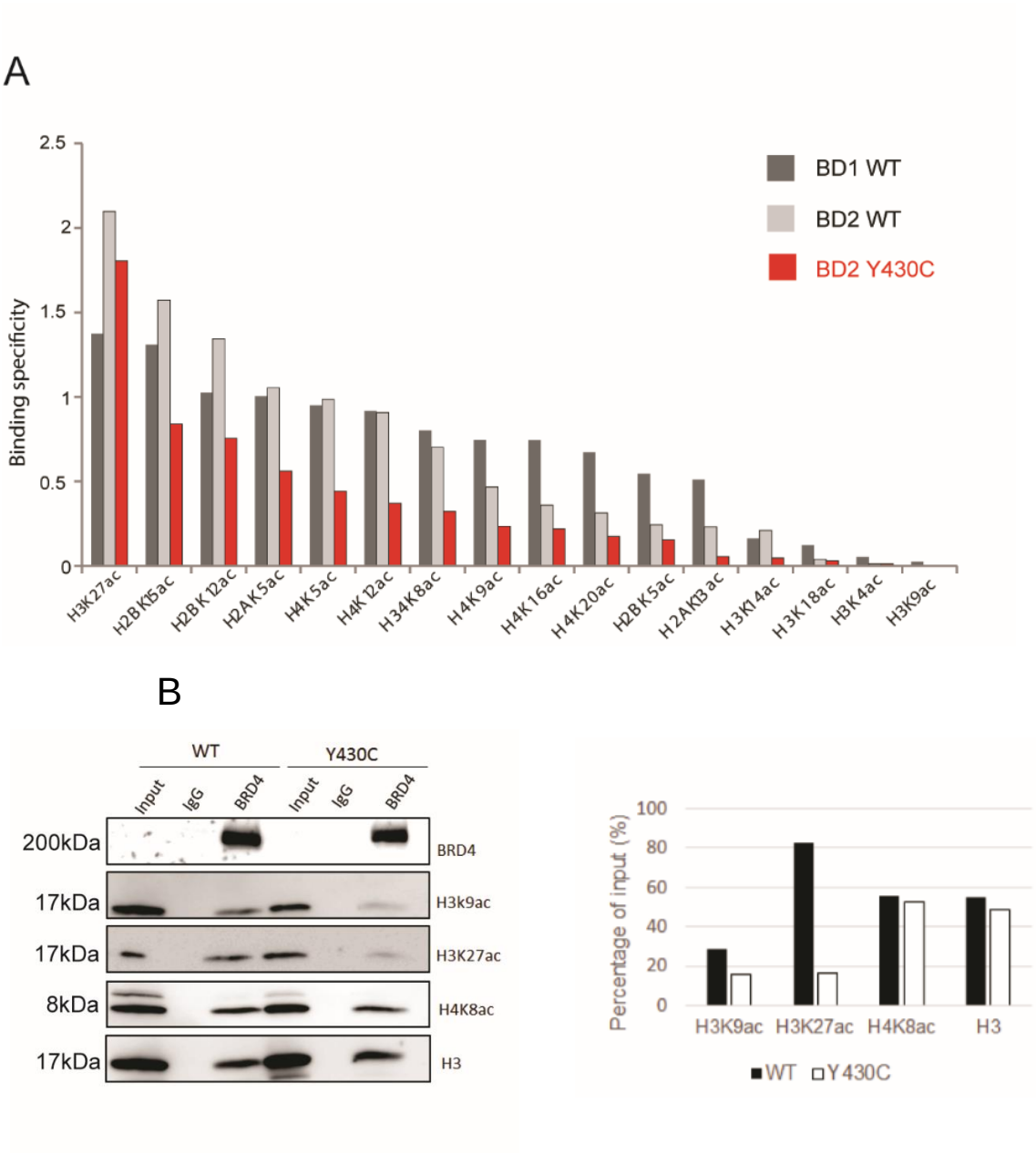


Figure Ch3-9. Binding of BRD4 to acetylated histones in WT and Y430C mESCs. **A)** Specificity of binding to arrayed acetylated histone tail peptides of WT BRD4 BD1 (BRD4 BD1 WT, dark grey bars), WT BRD4 BD2 (BRD4 BD2 WT, light grey bars) and BRD4 BD2 with the Y430C variant (BRD4 BD2 Y430C, red bars). **B)** Left - Cropped immunoblots showing co-IPs with an antibody to endogenous BRD4 and rabbit normal IgG (control) in WT and Y430C mESCs. Input is 1% of mESC nuclear extract. Antibodies used for immunoblotting are shown on the right of the images. Right - Graph shows intensity of bands in BRD4 IP lane as a percentage of input bands, calculated using ImageQuantTL.

3.3.3 BRD4 interactions with non-histone proteins

3.3.3.1 Co-immunoprecipitation-Mass spectrometry

To identify BRD4 interacting proteins more globally, I carried out a co-IP using two different antibodies against BRD4 (Bethyl; A301-985A100, Abcam; ab128874) followed by label free quantitative mass spectrometry (LFQ-MS), on protein lysates from WT and Y430C mESCs. LFQ-MS eliminates the need for a chemical label or tag, which are expensive, require extra steps for sample preparation, and can introduce variability to the experiment (Drabik, Ciborowski and Silberring, 2013). There is a lot of variability between the two replicates, suggesting that more repeats should be carried out before definitive conclusions are made.

This identified 1,082 proteins that were present in BRD4 IPs from both WT and Y430C cell lines. DAVID functional analysis (Huang, Sherman and Lempicki, 2009b, 2009a) of this protein list identified 65 enriched clusters ($p<0.5$) (data not shown), which included terms related to the known functions of BRD4. The most highly enriched cluster included ribosomal proteins and those involved in translation (score 109.72), however these proteins are known to be common MS contaminants (Mellacheruvu *et al.*, 2013). Other highly enriched clusters included mRNA processing (score 54.68), cell division (score 14.82), nucleosome regulation (score 11.87), DNA damage and repair (score 11.42) and transcription (score 11.14). Scores order the relative importance of each identified gene group.

90 of these proteins were absent in IgG controls (Figure Ch3-10), of which BRD4 was the top hit, indicating that the IP had worked well. Most proteins were present in both WT and Y430C IPs, with only 22 being differentially detected. I categorised the 90 proteins based on their known functions and found a number of common themes (Figure Ch3-10, boxed and labelled).

Interestingly, of the 90 proteins present in both WT and Y430C IPs, 4 were cohesin related – NIPBL, RAD21, ESCO2 and PDS5B (Figure Ch3-10, red). Furthermore, 5 other cohesin subunits - SMC1A, SMC3, STAG2, PDS5A and PDS5B – showed evidence of enrichment over IgG, but were not consistent between BRD4 antibodies (Figure Ch3-11). Upon replication of this experiment in the second independent Y430C mESC line, certain cohesin component-BRD4 interactions were no longer detected, despite being consistently detected by co-IP-western blot. After some optimisation, I found that it was a prolonged (>1hr) exposure to detergents in the RIPA buffer that perturbed these interactions as measured by

MS. The protocol was thus modified to limit the time in RIPA buffer and the associations were successfully replicated (see section 2.9.3).

Other categories enriched amongst these 90 proteins include proteins of the DDR (underlined) and transcriptional regulation (bold); both known functions of BRD4. The most common function I found in this group of 90 proteins was a role in splicing (Figure Ch3-10, yellow highlight). The BRD4 yeast homologue, bdf1, is involved in pre-mRNA splicing, with deletion of this gene causing a global defect in splicing (Albulescu *et al.*, 2012). However, whilst BRD4 has been shown to regulate splicing during heat shock (Hussong *et al.*, 2016; 2017), it is unclear whether BRD4 would interact with these proteins under normal conditions. Furthermore the recruitment of BRD4 to nuclear stress bodies upon heat-shock is mediated through BD acetyl-lysine binding (Hussong *et al.*, 2017), suggesting that, were these interactions specific, they would be stronger with WT-BRD4 than Y430C-BRD4.

BRD4 has been seen by others to interact with a complex of 7SK snRNA and BAF (Flynn *et al.*, 2016), and I will discuss this in greater detail later (section 4.4). BAF is a chromatin remodelling complex that can be composed of up to 15 subunits encoded by 29 genes (Panamarova *et al.*, 2016). Four of these subunits - SMARCA4, SMARCC1, SMARCB1 and SMARCA5- are pulled down by in my BRD4 co-IP, however these are excluded from the specific list due to their presence in the IgG control. The lack of specific BRD4-BAF interaction is not surprising however; the interaction between BAF and BRD4, thought to be mediated by 7SK snRNA, is likely disrupted by the addition of the endonuclease benzonase, which degrades both RNA and DNA (Oristo, Lee and Maunula, 2018), to the protein lysis buffer.

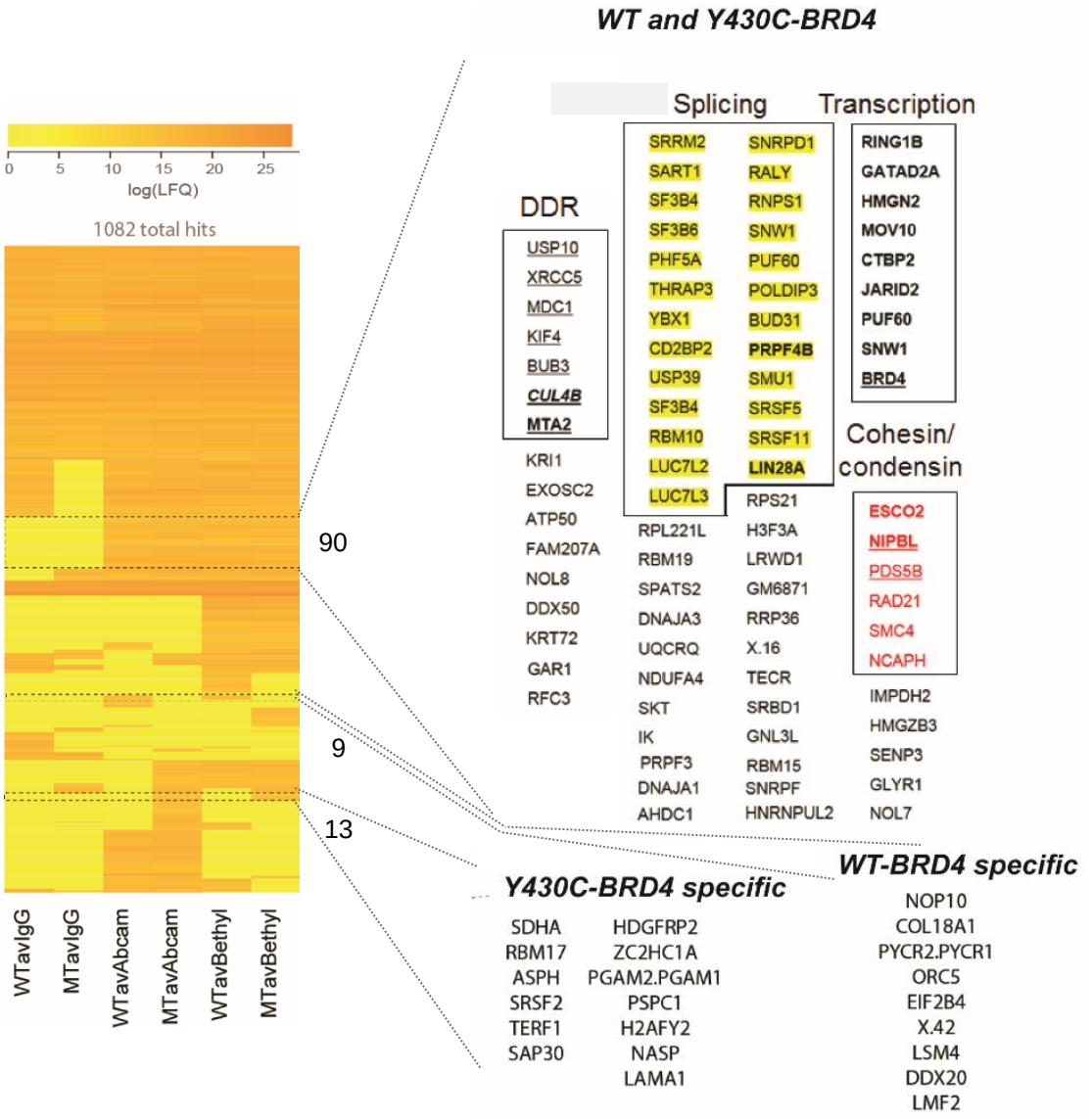


Figure Ch3-10. BRD4 protein interactions in WT and Y430C mESCs. A. Heatmap of the label-free quantitative (LFQ) mass spectrometry values assigned to each protein following BRD4 immunoprecipitation from WT and Y430C mESCs. The colour scale indicates the Log (LFQ) value calculated as an average of 3 technical replicates. Proteins that interact with WT and Y430C (top), WT only (middle) and Y430C only (bottom) are indicated by dotted lines and named. Underlined - proteins involved in the DDR. Yellow highlight – splicing factors. Bold – transcriptional regulators. Red – components of cohesin or condensing complexes

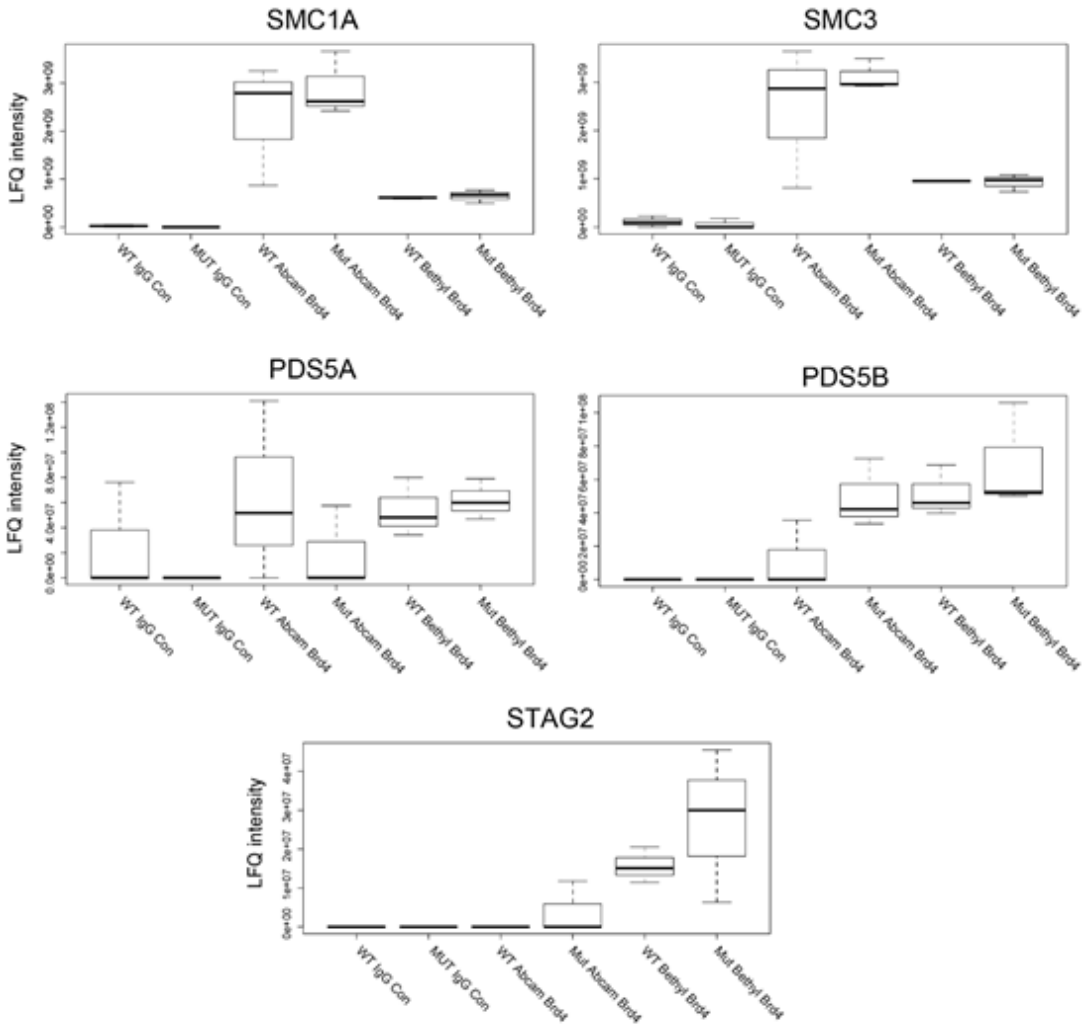


Figure Ch3-11. LFQ intensity from BRD4 co-IP MS. Box plots show details of the MS results for members of the cohesin complex that are not highlighted in Figure Ch3-6B. The y-axis represents the LFQ intensity and the x-axis represents the reagent used for the immunoprecipitation in WT and *BRD4*$^{Y430C/\ Y430C}$ (MUT) mESCs. Horizontal lines within boxes show medians, boxes are inter-quartile ranges and whiskers are range.

3.3.3.2 Western blot validation of cohesin interactions

To validate BRD4 interactions I used co-IP followed by immunoblotting for the putative interactors. Reciprocal co-IP for BRD4/NIPBL and BRD4/SMC3 confirmed that BRD4 interacts with the cohesin loader (Figure Ch3-13A&B) and core cohesin ring (Figure Ch3-12C&D) respectively.

Y430C- and WT-BRD4 show comparable levels of association with these cohesin components, both by western blot and by MS, suggesting that this interaction is unlikely to be mediated by co-binding to acetylated chromatin.

Association of BRD4 with the cohesin complex and NIPBL, indicates that there might be a common mechanism for 'classical' cohesin-related CdLS cases and those caused by mutations in BRD4. This not only strengthens my hypothesis that *BRD4* haploinsufficiency causes a CdLS-like phenotype, but may also help to identify cellular functions and pathways that are important in CdLS aetiology.

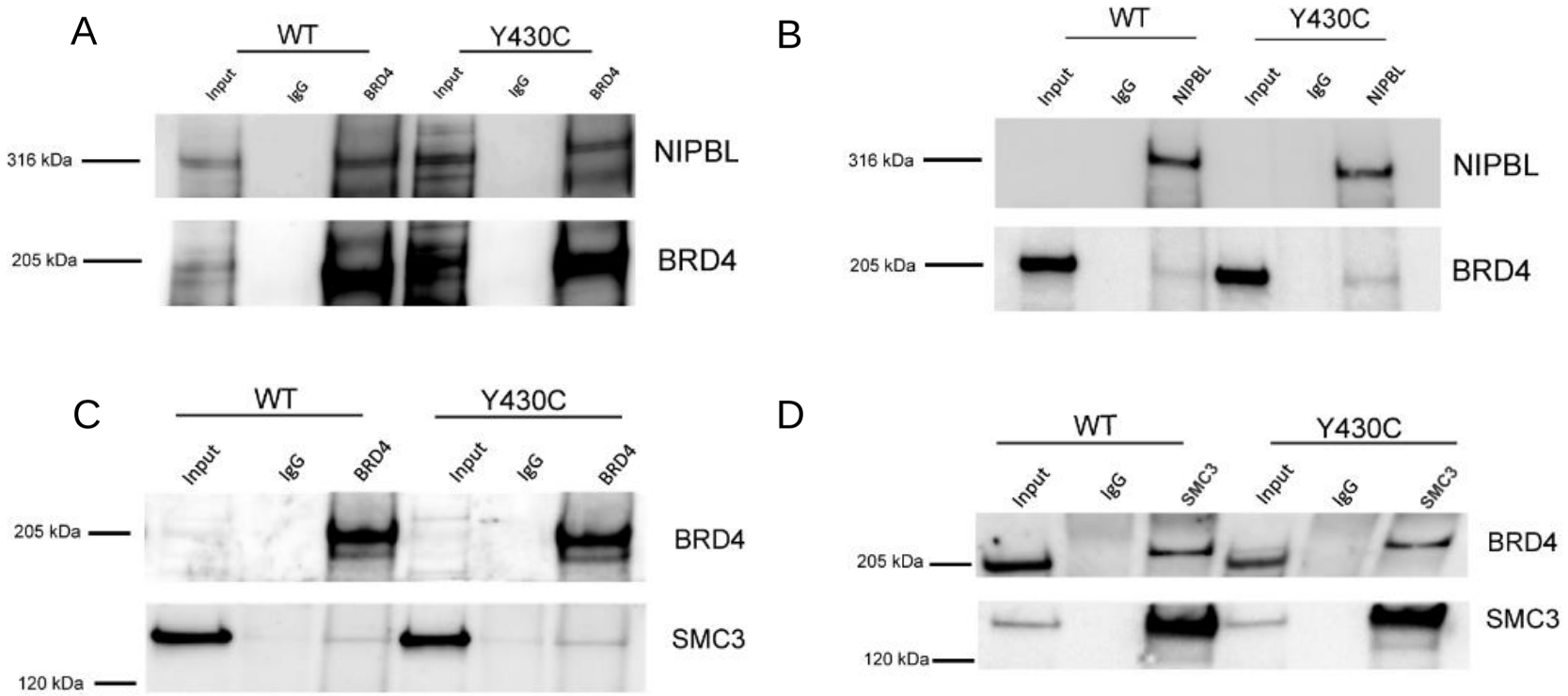


Figure Ch3-12. Reciprocal co-IPs and immunoblots validate the interactions between BRD4 and cohesin related proteins. A) BRD4 IP blot NIPBL, **B)** NIPBL IP blot BRD4, **C)** BRD4 IP blot SMC3, and **D)** SMC3 IP blot BRD4. Normal rabbit IgG is used as a control. Input =1% of mESC nuclear extract. Antibodies used for immunoblots are indicated on the right of the images.

3.3.3.3 Differentially bound proteins

Only 22 proteins fit the criteria of being identified as differentially bound between WT and Y430C by both BRD4 antibodies, and absent from the IgG control (Fig Ch3-10). No obvious trends were seen in these protein lists and these were too small in number for functional analysis. I therefore looked at those proteins that were differentially detected in WT and Y430C cells by the Bethyl BRD4 antibody. This is the antibody that I have used for all other ChIP and immunoblot experiments. This gave me a list of 62 proteins enriched in the WT-BRD4 IP, and 54 in the Y430C-BRD4 IP (Figure Ch3-13A).

I used the DAVID functional analysis tool to look for enriched GO terms and functional categories in these lists (p value <0.5). 8 and 7 clusters were identified in the WT and Y430C lists respectively (Figure Ch3-13B), with some overlaps between the two; Transcription and cell cycle related terms were enriched in both lists. The enrichment of DNA damage and repair terms, and chromatin binding and regulation, were seen in the WT protein list, but not in the Y430C list.

BRD4 is known to be involved in the DDR (Floyd *et al.*, 2013; Pongas *et al.*, 2017) and loss of binding to these proteins may be relevant to topics described later in my thesis (Chapter 5). It is possible that the loss of binding to chromatin regulators is a result of the Y430C mutation impairing the binding of BRD4 to acetylated histones.

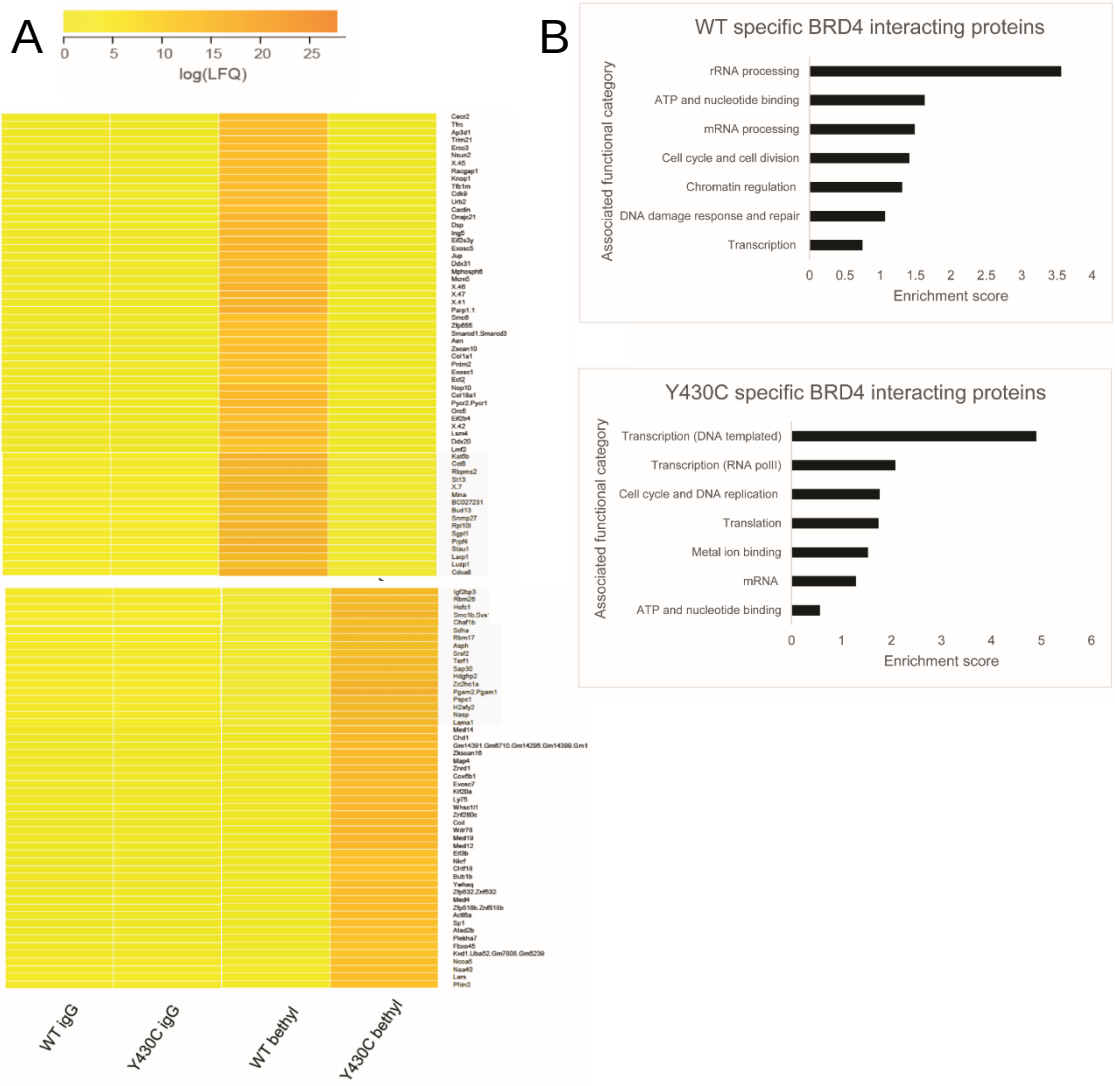


Figure Ch3-13. Proteins differentially enriched by BRD4 bethyl co-IP in WT or Y430C mESCs. A) Screenshots of heatmap shown in figure Ch3-8, highlighting proteins enriched by bethyl BRD4 antibody in WT mESCs (top) and Y430C mESCs (bottom). Colour bar = log (LFQ) intensities. **B)** Bar charts show clusters of GO terms and functional categories enriched in WT (top) and Y430C (bottom) specific protein lists from BRD4 co-IP MS. Categories were determined by DAVID ($p < 0.05$). Enrichment scores order the relative importance of each identified gene group.

3.4 Conclusions

Through targeted sequencing, David FitzPatrick's lab identified three CdLS patients with de novo mutations in *BRD4* and I've shown significant overlap between the phenotypes of these individuals and others with heterozygous deletions of *BRD4*. From these data, I conclude that *BRD4* haploinsufficiency can cause CdLS. By engineering one of these mutations, Y430C, in mESCs, I was able to show that the mutation does not affect BRD4 mRNA or protein levels. This indicates that the Y430C mutation must affect BRD4 function, rather than just decreasing the amount of protein.

I have shown that BRD4 associates with both cohesin core ring components and the cohesin loader NIPBL. This is not perturbed by the Y430C mutation, suggesting it is mediated by the ET domain or CTD, rather than via BD binding. BRD4 was also found to interact with proteins involved in mRNA processing, cell division, nucleosome regulation, DNA damage and repair, and transcription, consistent with what is already known.

The binding of BRD4 to a subset of 62 proteins was impaired by the Y430C mutation, and these proteins were enriched for terms related to DNA damage and repair, and chromatin binding and regulation. BRD4 has been shown by others to interact with, and regulate, proteins of the DDR and may be involved in their recruitment to DSBs (Li *et al.*, 2018). Therefore, a loss of interactions between BRD4-Y430C and DNA damage and repair proteins could have significant implications for the DDR in these cells.

In silico modelling strongly suggested a destabilisation of BD2 upon the Y430C mutation. Through in vivo experiments I showed that this causes a reduced association between BRD4 and acetylated H3K9 and H3K27. This raises the possibility that the Y430C mutation may preferentially affect BRD4 binding to CREs (enhancers and promoters), and this will be investigated further in the next chapter.

Chapter 4: Investigating the effect of decreased BRD4 acetyl-lysine binding in Y430C mESCs

4.1 Genome-wide chromatin and expression profiling.

In this chapter I describe the use of the genome-wide profiling methods, chromatin immunoprecipitation-sequencing (ChIP-seq) and 4 thiouridine-sequencing (4sU-seq) to explore the consequences of the *Brd4* Y430C mutation in mESCs.

4.1.1 ChIP-sequencing

A number of biological processes, including transcription; DNA replication, recombination and repair; and cell cycle progression, rely on interactions between DNA and proteins (Gade and Kalvakolanu, 2012a) and on epigenetic modifications of DNA and histones (Pillai, Dasgupta and Chellappan, 2015). Identifying and understanding these processes are essential for understanding cell and genome biology (Gade and Kalvakolanu, 2012b).

ChIP assays are routinely used to characterise DNA-protein interactions and epigenetic marks (Gade and Kalvakolanu, 2012b). For histone modification ChIP, native chromatin can be used and fragmented by enzymatic digestion, as nucleosomes are relatively stable structures (Thorne, Myers and Hebbes, 2004). However, when studying the binding of proteins such as TFs which have shorter residence times on the genome, the cells must first be treated with formaldehyde - to cross-link DNA and proteins - and the resulting nucleoprotein complexes are fragmented by sonication (Pillai, Dasgupta and Chellappan, 2015). DNA fragments are then immunoprecipitated using an antibody specific to the protein or histone modification of interest and de-crosslinked to separate the DNA and protein (Pillai, Dasgupta and Chellappan, 2015) (Figure Ch4-1A).

qPCR analysis can be used for quantification of specific DNA regions enriched in ChIP material (relative to input)(Pillai, Dasgupta and Chellappan, 2015), whilst ChIP-Chip (ChIP followed by microarray hybridisation) and ChIP-seq (ChIP followed by sequencing) technologies allow for genome-wide identification of histone modifications or DNA-protein interactions (Ren *et al.*, 2000; Johnson *et al.*, 2007). ChIP-seq is less noisy, higher resolution, requires less input DNA and provides greater coverage than ChIP-Chip. With decreasing costs of sequencing technologies, ChIP-seq is becoming the preferred tool for studying gene regulation and epigenetic mechanisms (Park, 2009).

4.1.2 4sU-sequencing

The transcriptome of a cell, that is, its total RNA population, specifies its identity and function (Mortazavi *et al.*, 2008), and therefore differs between cell types, developmental stages, pathological conditions and in response to drugs (Casamassimi *et al.*, 2017).

The study of individual transcripts dates as far back as the 1970s, but the field is now dominated by microarrays and RNA-seq, which allow the analysis of multiple transcripts at once. Microarrays can be used to analyse thousands of transcripts simultaneously, saving time and money, through the hybridisation of transcripts to an array of complementary probes (Lowe *et al.*, 2017). They are restricted however, by their reliance on prior knowledge of genome sequence, cross-hybridisation and limited dynamic range (Wang, Gerstein and Snyder, 2009).

In RNA-seq, steady state RNA is converted to cDNA and sequenced in a high-throughput manner. The resulting reads can either be mapped to a reference genome or assembled de novo, producing a whole genome transcriptional profile for the analysis of transcriptional structure and gene expression (Wang, Gerstein and Snyder, 2009). This eliminates a number of the issues associated with microarrays, but the use of total cellular RNA still limits the biological questions that can be addressed; it is impossible to determine whether changes in total RNA levels are a result of changes in synthesis, or decay. mRNA degradation is an essential step in determining mRNA abundance and can be regulated for a particular biological function or in response to intrinsic and extrinsic stimuli (Wu and Brewer, 2012).

A number of protocols have been developed to overcome this problem, by specifically detecting only newly synthesised (nascent) RNA. In 4sU-seq, cells are incubated with the uridine derivative 4sU, which is incorporated into new transcripts by RNA polymerases. RNA isolation followed by biotinylation of the 4sU fraction allows the separation of labelled (nascent) and un-labelled (pre-existing) transcripts with the use of streptavidin-coated magnetic beads. 4sU-RNA is recovered from the beads by addition of a reducing agent, which cleaves the disulphide bond between 4sU and biotin (Figure Ch4-1B).

By varying the 4sU incubation time, different aspects of transcription can be probed; short incubations (minutes) provide a snapshot of gene expression independent of downstream processes, whilst incremental incubation times reveal the kinetics of RNA processing. RNA

half-lives can also be investigated through comparisons of nascent and total RNA abundance (Rutkowski and Dölken, 2017).

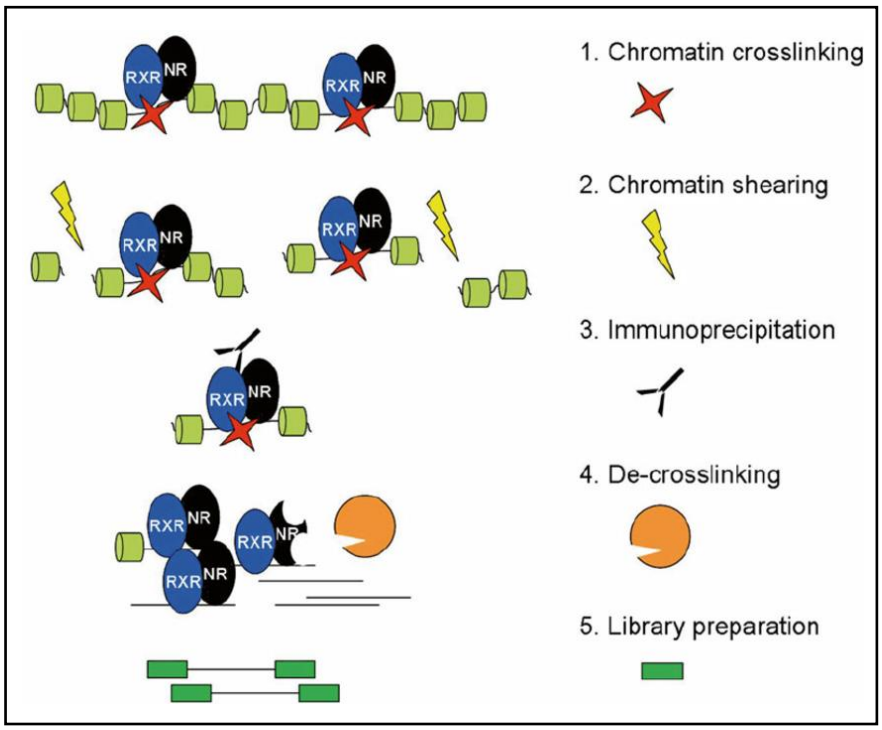
4.1.3 The need for a spike-in control

The two methods described above are often used to compare two different conditions. To do this they assume that the overall yield of DNA or RNA is the same in both conditions, and that any changes in transcription or DNA-protein binding are limited to specific loci. However, normalisation of the data based on these assumption can lead to misinterpretation of the data (Chen *et al.*, 2016).

For instance, if expression is globally increased in one RNA sample over another, normalising the reads to the same number will mask this change (Figure Ch4-2A). On the other hand, if expression at a specific genomic region is increased in one sample, this type of normalisation will cause an inappropriate reduction in the number of reads from other regions (Figure Ch4-2B).

The use of 'spike-in' controls overcomes these problems; a spike-in is the addition of a small amount of DNA or RNA from a foreign genome, prior to immunoprecipitation or RNA extraction (for ChIP or 4sU respectively). The same spike-in is added for each condition tested, ensuring that any downstream differences between spike-in levels are solely due to differences in DNA or RNA yield, and not biological. The spike-in values for each condition are normalised to each other, and the ratio between the two can be used to normalise the test data, decreasing the chance of erroneous interpretations (Figure Ch4-2) (Chen *et al.*, 2016).

A



B

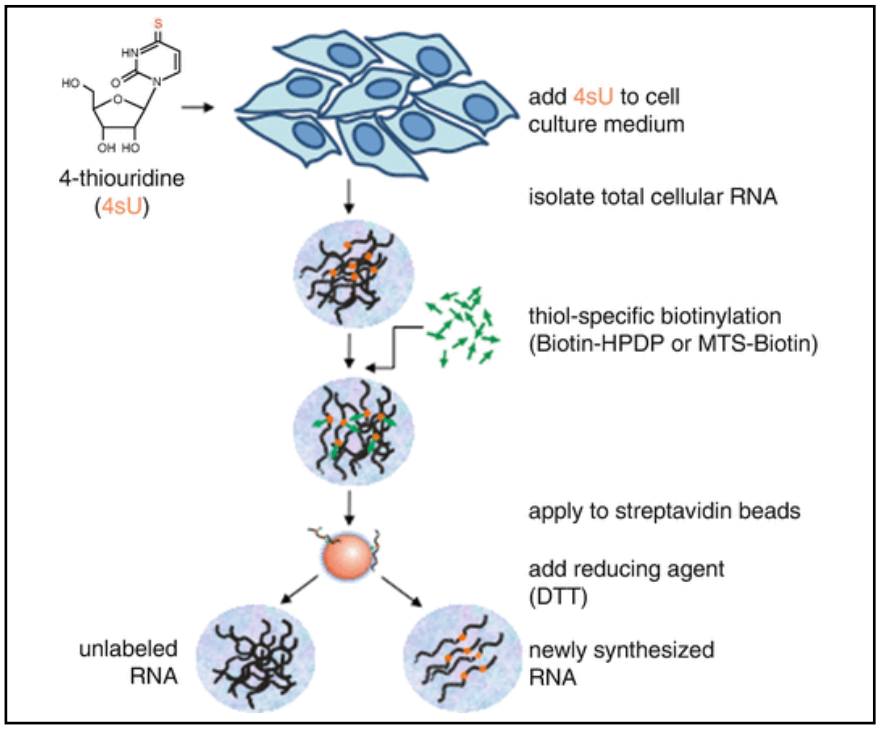


Figure Ch4-1. Chromatin and expression profiling techniques Schematics illustrating **A)** ChIP-seq (taken from Daniel *et al.*, 2014) and **B)** 4sU-seq protocols (taken from Rutkowski and Dölken, 2017).

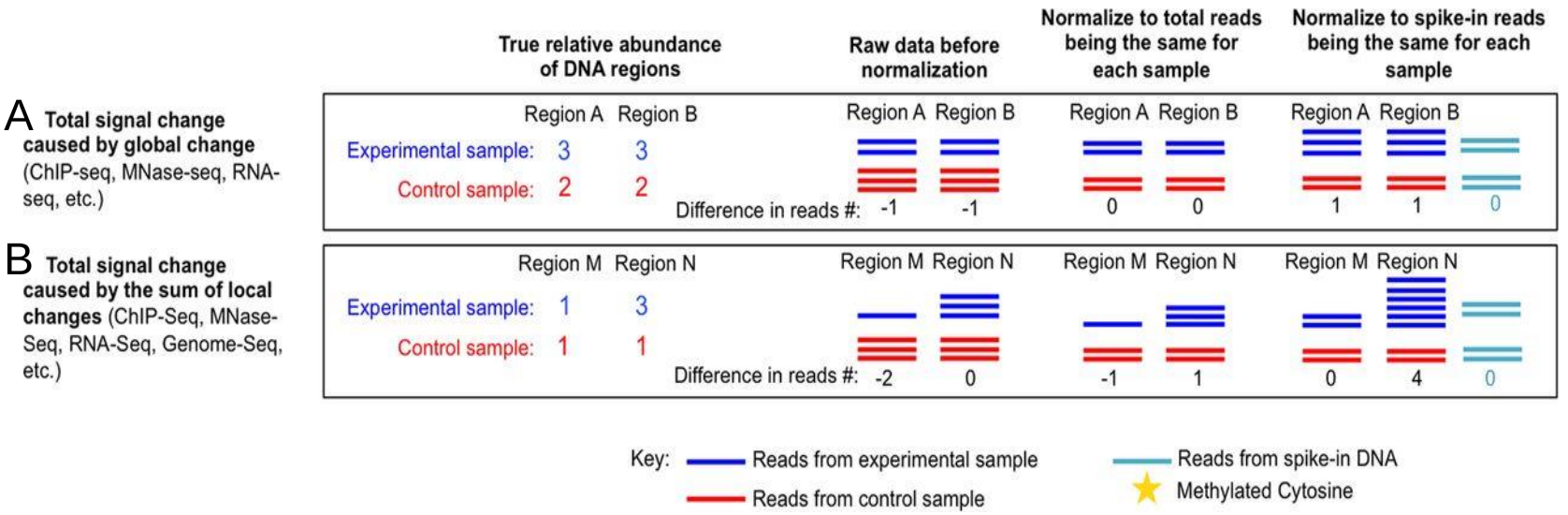


Figure Ch4-2. Effects of normalisation with and without spike-in controls. Schematic illustrating the need for a spike-in control. **A)** Shows the effects of normalisation with and without a spike-in control when a global change has occurred and **B)** shows the effects when a difference is present at a specific genomic region. Modified from Chen *et al.*, 2016

4.2 Exploring BRD4 occupancy by ChIP-seq in the genome of WT and Y430C mESCs

Having shown in Chapter 2 that the Y430C mutation reduces BRD4's association with certain acetyl-lysines of histone tails in mESCs and decreases the occupancy of BRD4 at specific target regions in MEFs, I next compared the genome-wide binding of BRD4 between WT and Y430C mESCs.

4.2.1 Choice of antibody

I chose to use the BRD4 rabbit polyclonal antibody from Bethyl lab (A301-985A) for ChIP-seq; this antibody had not been previously validated for ChIP but had been shown to work for IP and immunohistochemistry, making it a good candidate. Furthermore, I had previously shown that this antibody produces a specific band by western blot (Figure Ch3-5D) and confirmed by IP-MS that it interacts with BRD4 in mouse cells (Figure Ch3-10).

To verify that the chromatin recovered after ChIP would contain appreciably more BRD4 targets than other genomic regions, and would therefore be worth sequencing, I first conducted preliminary ChIP-qPCR experiments in the WT mESCs (Figure Ch4-3). I used primers specific for regions where I expected BRD4 to bind - just upstream of the *c-Myc* promoter, within the *c-Myc* SE, and within the *Nanog* enhancer – and at a control region where I expected little BRD4 binding – the promoter region of *Gsc*, a gene that is not expressed in mESCs (Mfopou *et al.*, 2014) (Table Ch2-5).

BRD4 ChIP samples were significantly enriched for the BRD4 target regions compared to the IgG control, indicating that the antibody efficiently binds BRD4. The *Gsc* promoter, intended as a BRD4-free control region, also showed some significant enrichment in the BRD4 ChIP over IgG. Whilst this might suggest that the antibody is not specific, it is possible that some BRD4 binding at this promoter does occur. Furthermore, the relative enrichment of all three BRD4 target regions was significantly greater than that of the *Gsc* promoter region in the BRD4 ChIP sample. Therefore, I concluded that the antibody would be suitable for ChIP-seq and proceeded with the experiment.

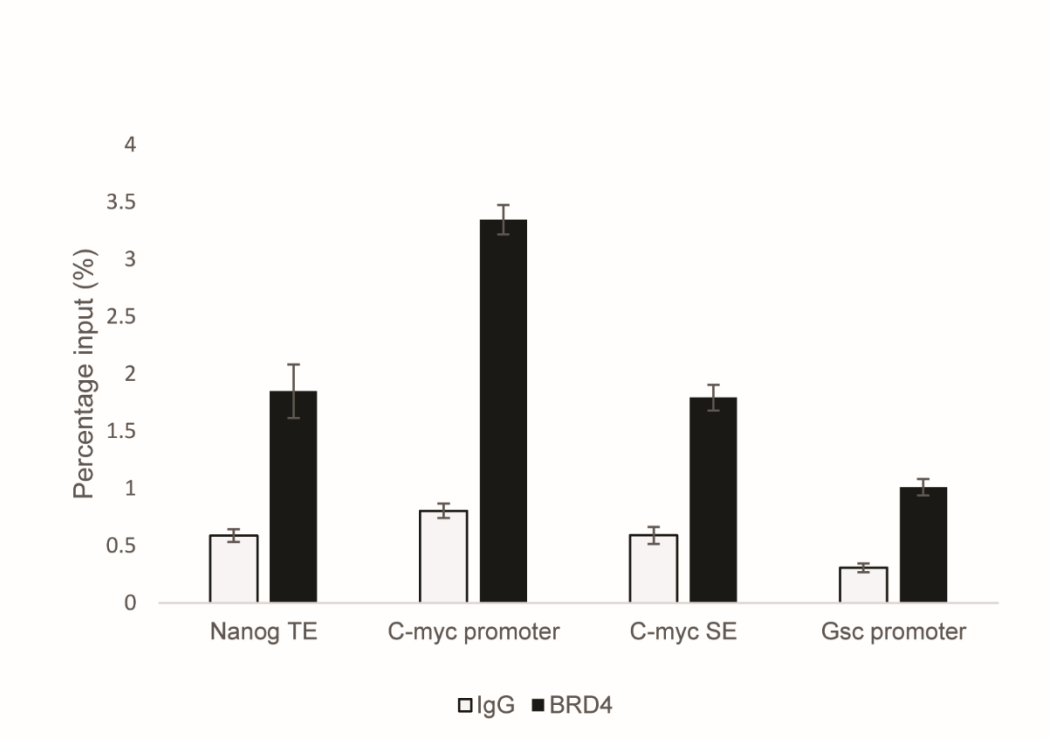


Figure Ch4-3. Specificity of BRD4 ChIP. Graph shows qPCR results using material recovered from BRD4 ChIP experiment and IgG control in WT mESCs. X-axis shows primers used. Concentrations are calculated relative to input. Data is represented as mean +/- SEM from 3 technical replicates.

4.2.2 Choice of control

I proceeded to make DNA libraries of BRD4 bound DNA and input controls (aliquot of cross-linked and fragmented chromatin prior to incubation with antibody bound beads). Biases can be introduced to ChIP-seq data at many steps of the protocol - during sonication due to non-uniform DNA breakage (Landt *et al.*, 2012); during library preparation; and in mapping the DNA to the genome. Sequencing of a control profile, such as the input DNA or an IgG 'mock' ChIP, is therefore important as it provides a background profile that the enriched ChIP signal can be normalised to (Liu, Pott and Huss, 2010). It has been suggested that spike-ins are the best control for ChIP-seq (see section 4.1.3) (Chen *et al.*, 2016). For my purposes the best organism to use would be drosophila; Drosophila cells are readily available, their genome is well studied, and sequences map to the mouse genome at a very low frequency (Orlando *et al.*, 2014). However, the lack of a BRD4 homologue in Drosophila made this unfeasible.

I chose to sequence an input control and have used this in downstream applications such as peak-calling to account for any technical biases, however this cannot replace the use of a spike-in as an appropriate control, meaning that it is difficult to make any conclusive comparisons between the two cell lines.

I made and sequenced two biological replicate ChIP-seq libraries. These show similar trends throughout the analysis, except where mentioned, but the data I have shown comes from just one replicate.

4.2.3 Quality control

I used FASTQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc) to analyse the quality of the sequenced DNA and ensure that it was suitable for downstream analyses. A phred score from 10 to 50 is assigned to each base of every read, indicating the quality of the base call; a score of 10 means the probability of an incorrect base call is 1 in 10, whilst a score of 50 means the probability of an incorrect base call is 1 in 100,000 (Illumina, 2011). The per base sequence quality for this ChIP-seq was found to be acceptable for all samples (Figure Ch4-4); the scores rarely drop below 37, indicating a high base calling accuracy. There was also shown to be no adaptor content or over-represented sequences in any of the samples.

The samples did show evidence of sequence duplication – that is when multiple sequence reads map to exactly the same region of the genome. These redundant reads may be artefacts,

caused by over amplification during the PCR step of library preparation, or they can represent true signal. There has been much debate in the field over how to treat such sequences, ie. To filter or not to filter. On the one hand, these duplicate sequences can increase false positives during peak calling, and their removal has been shown to increase the specificity of identified peaks (Chen *et al.*, 2012). On the other hand, in most peak regions these redundant peaks are thought to represent true signal (Chen *et al.*, 2012), meaning that their exclusion limits the dynamic range of ChIP signal (Carroll *et al.*, 2014). Since the removal of duplicates was found to generally remove noise in non-enriched regions, with negligible effects on peak detection sensitivity (Chen *et al.*, 2012), I chose to remove duplicates before peak-calling.

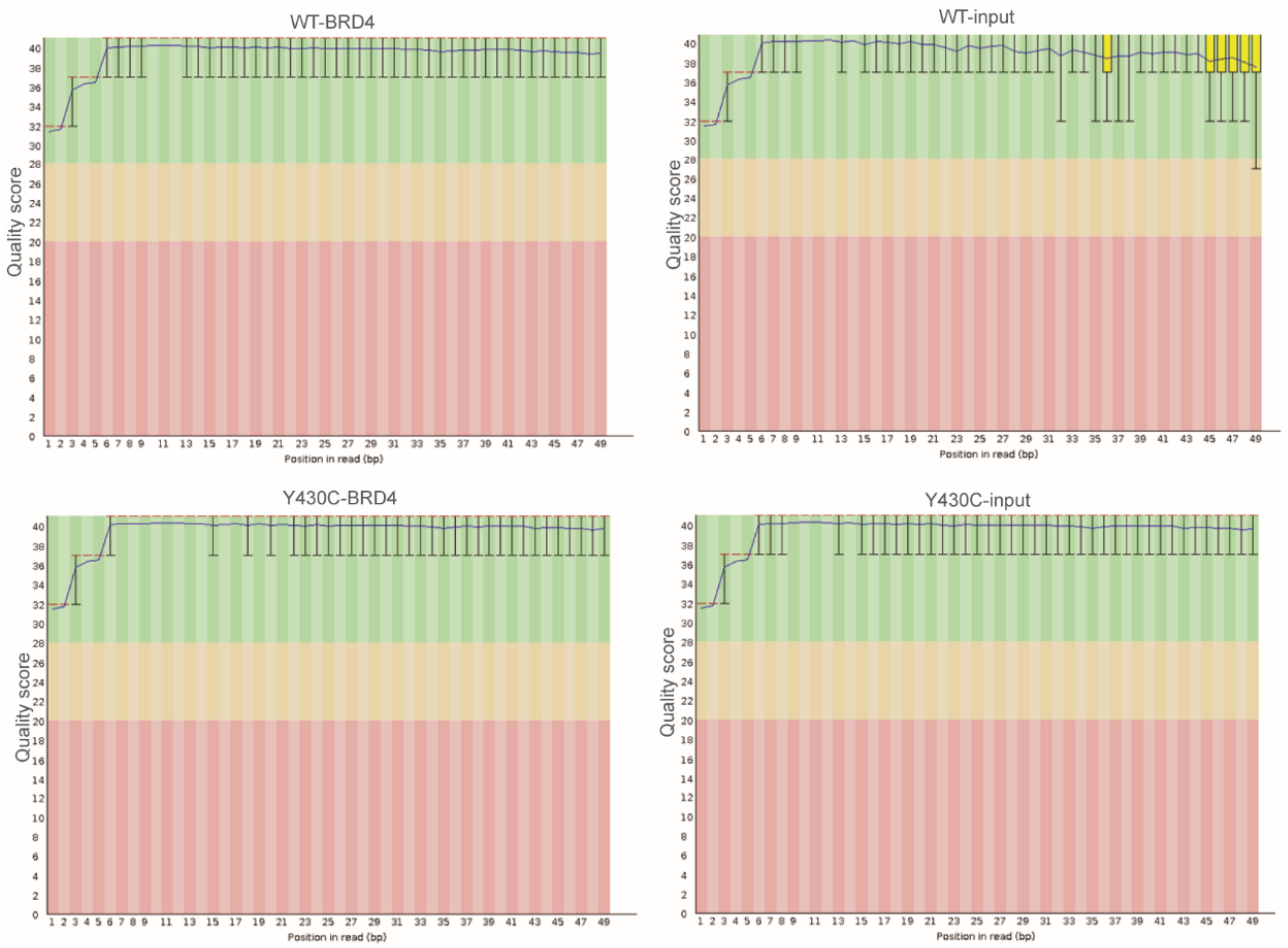


Figure Ch4-4. ChIP-seq per base sequence quality. Graphs show quality scores at each base position for the four ChIP-seq FASTQ files, as calculated by FASTQC

4.2.4 Mapping and visualisation

I mapped the sequenced DNA to the mm9 assembly of the mouse genome using Bowtie2 (Langmead and Salzberg, 2012). For a protein such as BRD4, that can show broad regions of enrichment (Jang, Shen and McBride, 2014; Najafova *et al.*, 2017), up to 60 million reads may be required per sample (Bailey *et al.*, 2013). I therefore deemed the number of mapped reads in my ChIP-sequencing to be sufficient, ranging from 57 million to 87.5 million (Table Ch4-1), with percentage of reads aligned >88% for all samples.

I converted my data to bedgraph format and visualised the data using the UCSC genome browser (http://genome.ucsc.edu). This enabled me to determine the binding profile of BRD4 over genes of interest and specific genome elements, and to compare its binding to the profiles of various histone modification marks (H3K122ac; GSM2054689, H3K9ac; GSM775313, H3K27ac; GSM594578, H3K4me1; GSM594577, DNAse I hypersensitivity signal; GSM1014154). Consistent with previous reports (Di Micco *et al.*, 2014; T. Wu *et al.*, 2015; Gonzales-Cope *et al.*, 2016), I found that WT-BRD4 occupied the promoters and enhancers of pluripotency genes such as *Sox2, Kl4* and *Nanog* (Figure Ch4-5). Consistent with my ChIP-qPCR data (Figure Ch4-3), the number of reads at these regions is greater than at the *Gsc* promoter (Figure Ch4-6A), which shows no BRD4 occupancy. There is some enrichment of BRD4 over the *c-Myc* SE and promoter regions (Ch4-6B), however this is less striking than it appeared by ChIP-qPCR (Figure Ch4-3). There is very little binding of BRD4 at *Gapdh,* which I have used as a control for qPCRs (Figure Ch4-6C).

Y430C-BRD4 showed a similar binding pattern to that of WT-BRD4, however there was a noticeable decrease in the peak heights, particularly at SEs (Figure Ch4-5&6), where BRD4 is known to preferentially bind (Di Micco *et al.*, 2014). This is illustrated by heat-maps (Figure Ch4-7A), where BRD4 read coverage is decreased in Y430C mESCs compared to WT, and I confirmed this by ChIP-qPCR in the second independent mESC line (Figure Ch4-7B).

	Total reads	Reads aligned to mm9	% Reads aligned
WT BRD4	67,828,269	61,115,388	90.10
Y430C BRD4	64,561,315	57,119,268	88.47
WT input	80,073,231	77,698,782	97.03
Y430C input	90,627,286	87,540,355	96.59

Table Ch4-1. Mapping statistics for WT and Y430C ChIP-seq FASTQ files. Files were mapped with Bowtie2 and mapping statistics calculated with Samtools

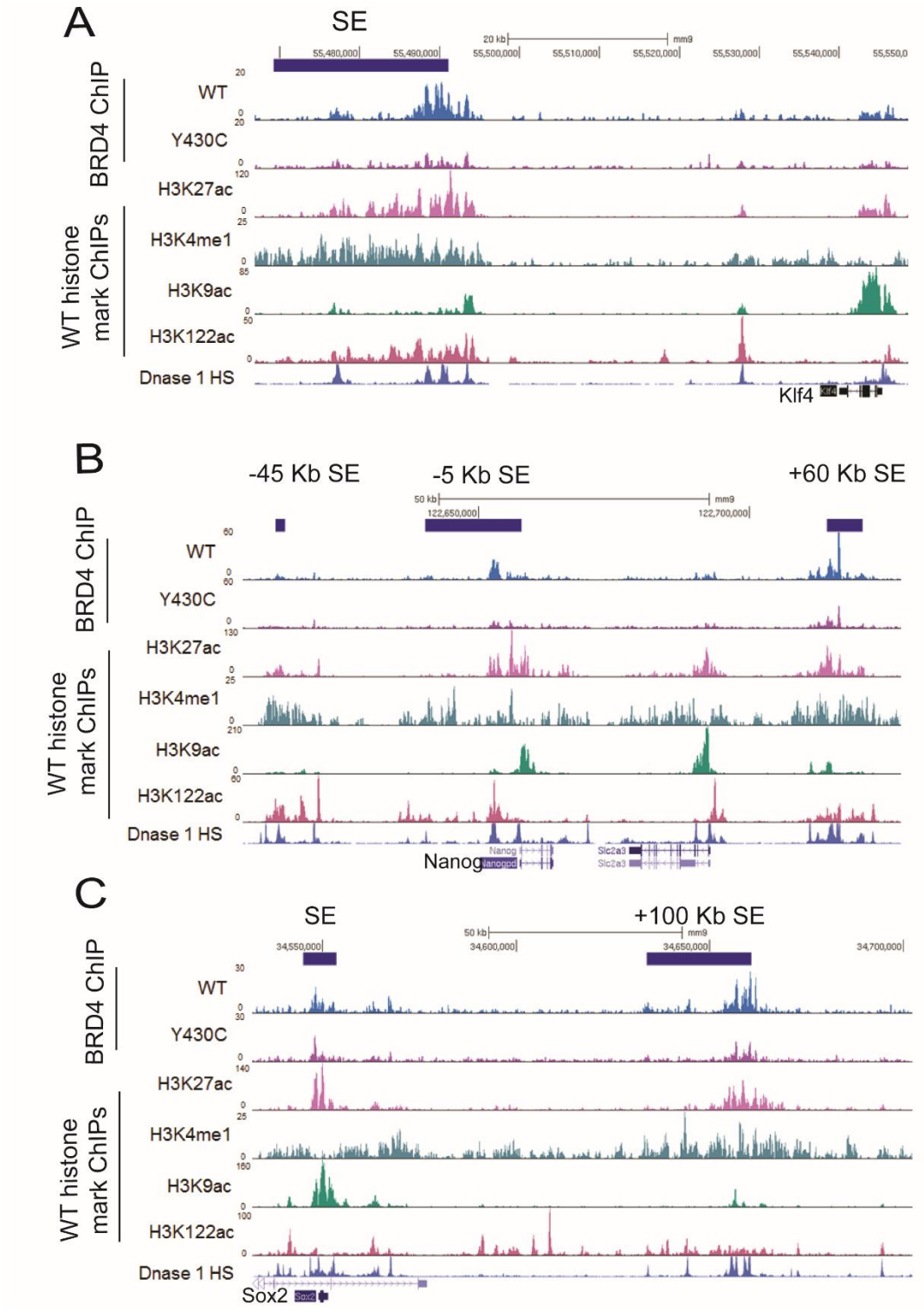


Figure Ch4-5. Genome browser view of BRD4 binding at BRD4 target genes. UCSC browser screenshot showing reads per 10 million (RP10M) over the extended **A)** *Klf4* locus, **B)** *Nanog* locus and **C)** *Sox2* locus for various ChIP-seq experiments. Tracks 1 and 2 = BRD4 ChIP-seq in WT and Y430C mESCs, tracks 3-6 = previously published ChIP-seq data for various histone modifications, track 7 = DNase I hypersensitivity.

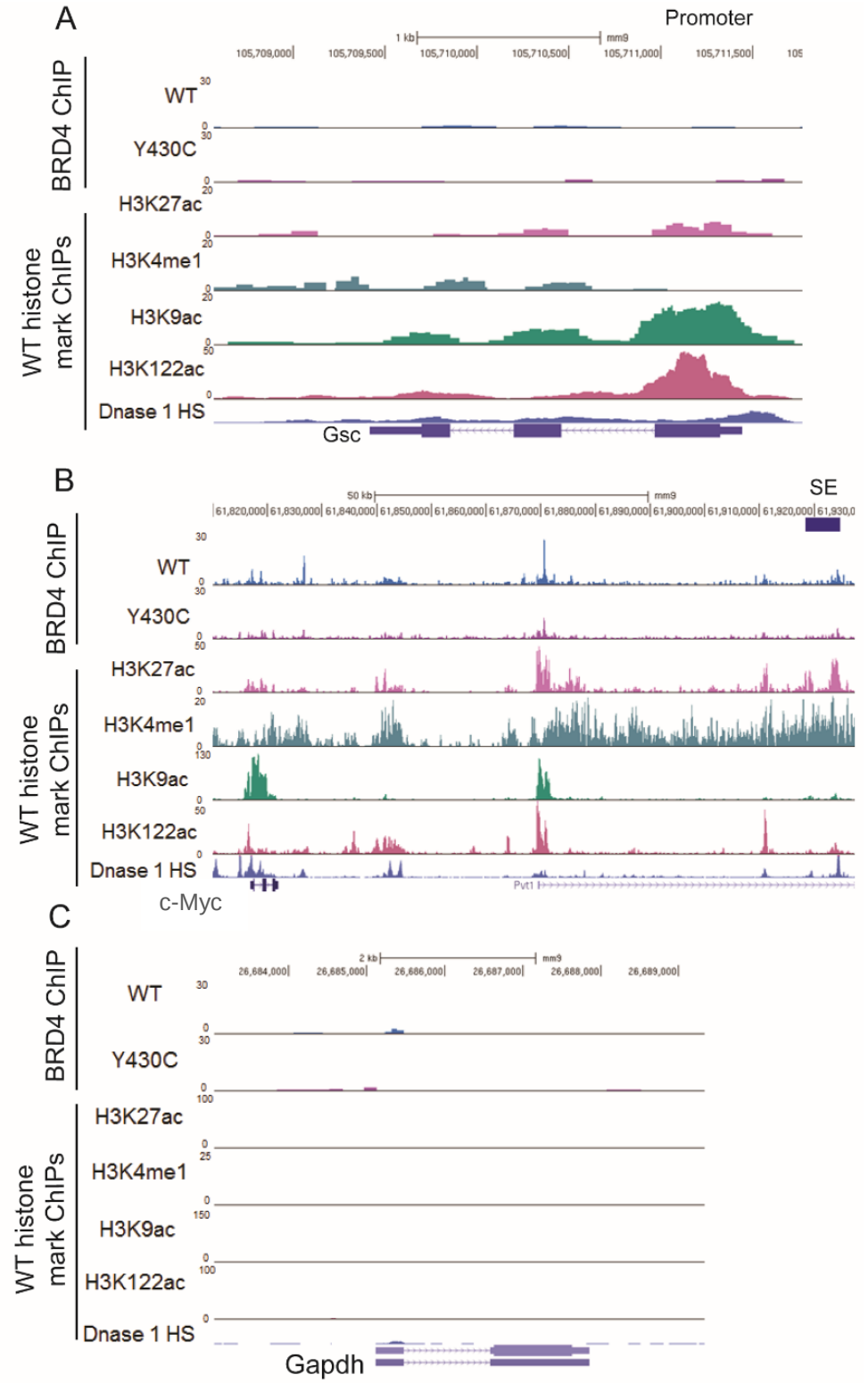


Figure Ch4-6. Genome browser view of BRD4 binding at *c-Myc* and control regions. UCSC browser screenshot showing RP10M over the extended **A)** *Gsc* locus, **B)** *c-Myc* locus and **C)** *Gapdh* locus for various ChIP-seq experiments. Tracks 1 and 2 = BRD4 ChIP-seq in WT and Y430C mESCs, tracks 3-6 = previously published ChIP-seq data for various histone modifications, track 7 = DNase I hypersensitivity.

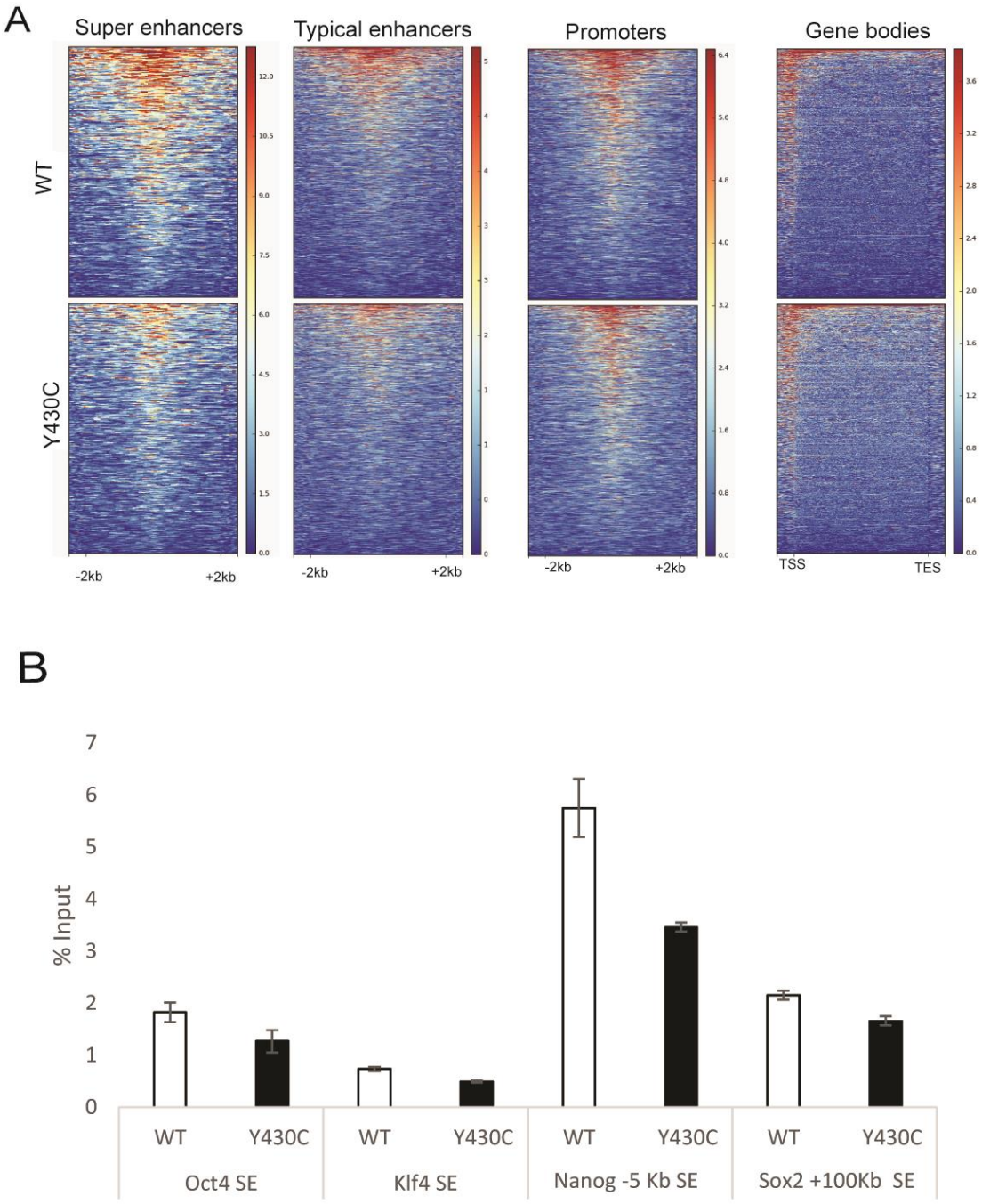


Figure Ch4-7. Genome wide binding of BRD4. A) Heatmaps show distribution of BRD4 ChIP-seq reads across SEs, typical enhancers, promoters and gene bodies in WT and Y430C cells. Colour scale represents log2 RPKM. Regions are scaled to 4 Kb. **B)** ChIP-qPCR measuring concentration of BRD4 ChIP DNA from WT (white bars) and Y430C (black bars) mESCs, relative to input across the SEs of *Oct4*, *Klf4*, *Nanog*, and *Sox2*. Data is represented as mean +/- SEM from 3 technical replicates.

4.2.5 Peak Calling

To analyse the data more quantitatively, I used the peak-calling software MACS2 (Model-based Analysis of ChIP-Seq data) to call significant peaks of BRD4 in WT and Y430C mESCs. MACS normalises the data against both local background and the provided input background, allowing for robust prediction of peaks (Y. Zhang *et al.*, 2008). As expected, I found a decrease in the total number of BRD4 peaks in Y430C cells compared to WT cells (Table Ch4-2).

Graeme Grimes (IGMM Bioinformatics Analysis Core) used the Bioconductor package regioneR (Gel *et al.*, 2016) to assess the relationship between the different peak sets and defined genomic regions. This uses circular permutation (n=1000) to count the number of overlaps between two genomic regions and calculates a z-score and a p-value between the observed number of overlaps and the random distribution of permuted regions. We compared the WT and Y430C peak files to mESC chromatin states based on a chromHMM segmentation of the mm9 genome (Ernst and Kellis, 2012) (Figure Ch4-8), and to mESC promoters, typical enhancers (TEs) and SEs defined by others (Flynn *et al.*, 2016). We found that promoters showed the highest levels of enrichment, followed by SEs and TEs, whilst heterochromatin, repressed regions and transcribed regions were the least enriched (Figure Ch4-9).

I then used the MACS2 function for differential peak calling (bdgdiff), to identify those regions that are occupied by BRD4 in the WT cells but not in the Y430C cells, and vice versa. I found that, whilst there were 1693 peaks that were specific for WT (Table Ch4-2) – ie. Where mutant BRD4 was not significantly bound - there was only one region that was specific for Y430C. This region maps to the promoter of *Snrpn,* but was not detected as a differential peak in the other BRD4 ChIP-seq replicate. GO enrichment analysis using GOrilla (Eden *et al.*, 2007, 2009), found no terms to be enriched in the list of genes associated with the BRD4 peaks specific to WT cells ($p < 0.001$).

In the absence of any enrichment in gene terms, we looked to see if the WT specific peaks were enriched for any functional genomic categories. Using regioneR we detected that WT specific peaks were enriched at TEs, SEs and active promoters, and were least enriched for weakly transcribed and repressed regions (Figure Ch4-9).

From these results I concluded that Y430C-BRD4's decreased affinity for acetyl-lysines mostly affects its binding to CREs, causing a decrease in the occupancy of BRD4 at SEs, TEs and promoters.

Peak calling results			
	Total peaks	**Specific peaks**	**% specific peaks**
WT	20,691	1693	8.2
Y430C	11,775	1	0.8×10^{-5}

Table Ch4-2. Peak calling results. MACS2 was used to call peaks in WT and Y430C BRD4 ChIP-seq data, using input data as a control (p value < 0.00001) (Total peaks). MACS2 bdgdiff was used to identify regions of differential BRD4 occupancy in WT and Y430C ChIP-seq data (Specific peaks).

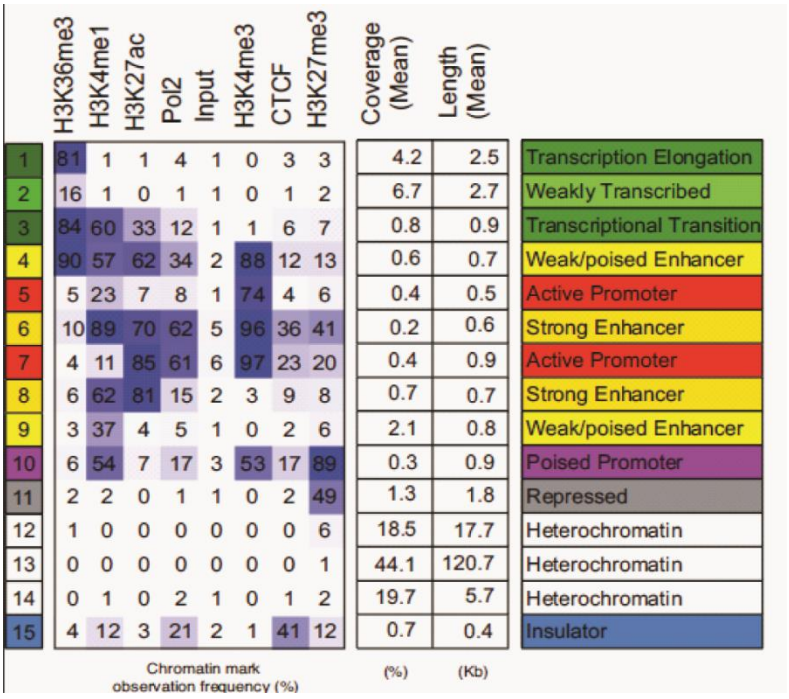


Figure Ch4-8. Combination of chromatin marks used to define different chromatin states by ChromHMM. ChIP-seq datasets for the labelled factors or histone marks were collected and a Poisson-based multivariate hidden markov model was used to identify regions or states enriched in specific combinations of histone modifications, splitting the genome into 15 different states. These were compacted into 6 states for our analysis. Taken from Bogu *et al.*, 2016

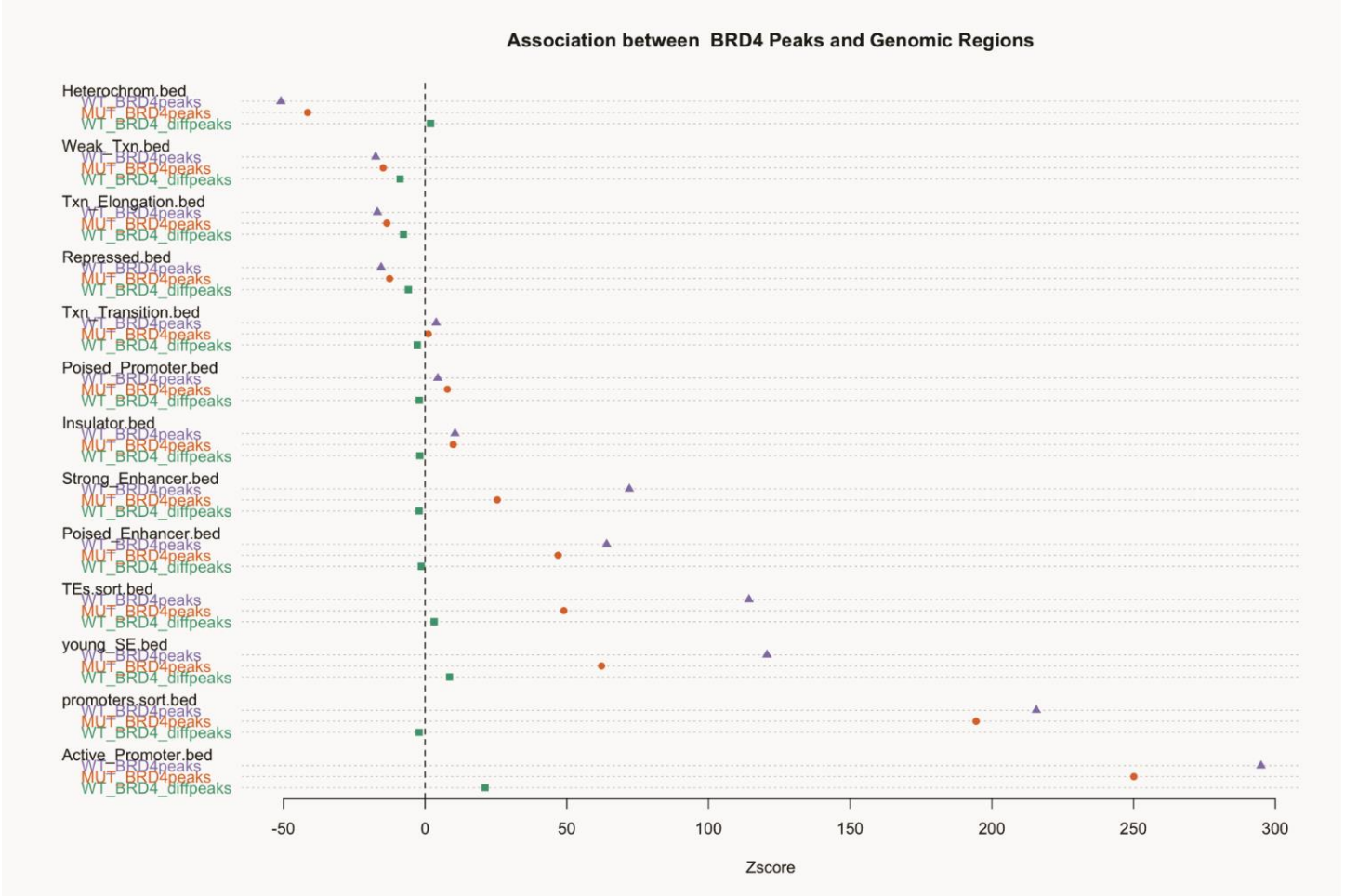


Figure Ch4-9. Enrichment of functional genomic regions in BRD4 peaks. Plots show Z-score calculated by regioneR. These represent the likelihood of BRD4 peaks overlapping with specific genomic regions. Peaks were called by MACS2 callpeaks ($p<0.00001$). Diffpeaks are those specific to WT, called by MACS2 bdgdiff. Genomic regions were defined by ChromHMM in mESCs, excluding Promoter, Typical En and Super En, which were defined by Flynn *et al.,* 2016.

4.2.6 JQ1 treatment decreases BRD4 binding to SEs

(+)-JQ1 (see section 1.4.2.3) competitively binds to bromodomains of BET proteins and has been previously shown, by ChIP-qPCR (Patel *et al.*, 2013; K. L. Cheung *et al.*, 2017) and ChIP-seq (Zhang *et al.*, 2012), to displace BRD4 from chromatin. Therefore, it provides me with a useful tool for inhibiting BRD4 binding to acetyl-lysines.

I carried out ChIP-qPCR using the BRD4 bethyl antibody in WT and Y430C cells treated with 300 nM (+)-JQ1 or (-)-JQ1 (the active and inactive enantiomers respectively) for 24 hours. The results were in accordance with what had previously been seen; (+)-JQ1 treatment significantly decreased the amount of *Oct4* and *Klf4* SE DNA in the BRD4 ChIP DNA (shown as a percentage of input DNA), compared to (-)-JQ1 treatment, in mESCs (Figure Ch4-10). The amount of *Sox2* SE DNA also decreased, but this was not significant. This corroborates the results of others, that (+)-JQ1 binding to BRD4 bromodomains inhibits BRD4 binding to SEs.

The decrease in BRD4 binding caused by (+)-JQ1 treatment was more pronounced than that which occurs due to the Y430C mutation. Furthermore, (+)-JQ1 treatment had a significant effect on both WT and Y430C BRD4 binding (Figure Ch4-10). These data suggest that the Y430C mutation has a mild effect on BRD4 binding, relative to (+)-JQ1. This is in agreement with the BRD4 ChIP-seq profiles (Figure Ch4-5&6), which show that peaks of BRD4 binding are decreased in Y430C compared to WT, but not abolished.

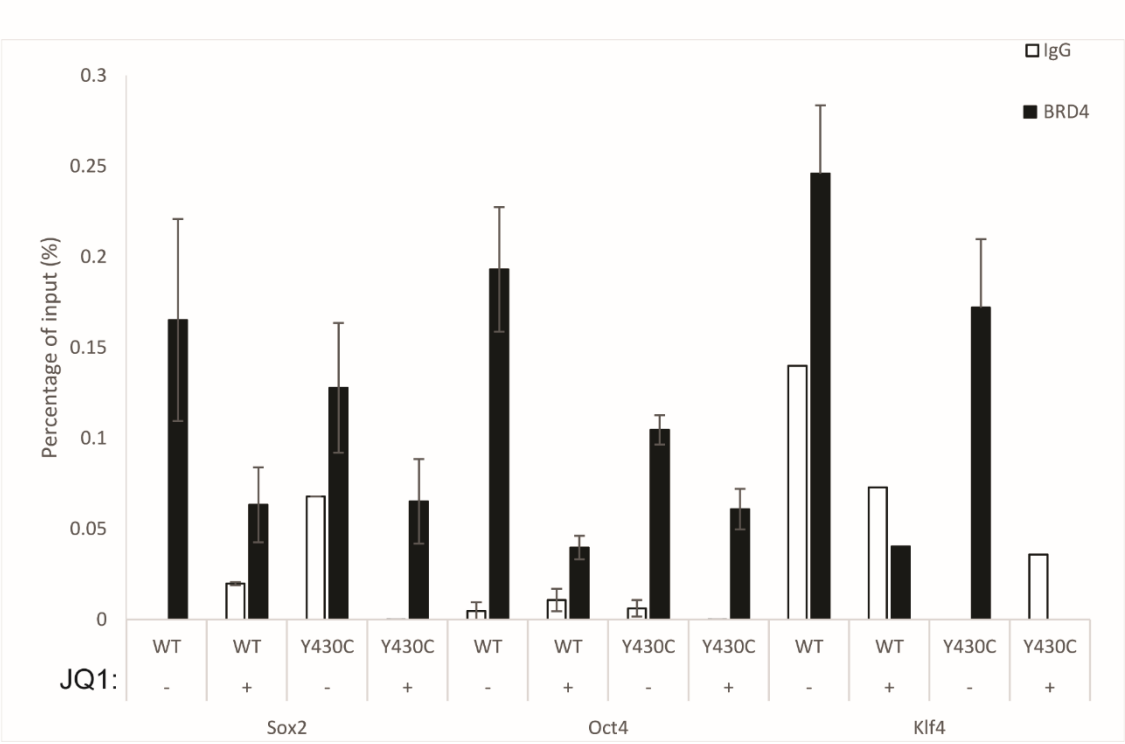


Figure Ch4-10. (+)-JQ1 treatment decreases BRD4 binding to SEs. ChIP-qPCR measuring concentration of BRD4 ChIP DNA relative to input across the SEs of *Sox2, Oct4* and *Klf4* in WT and Y430C cells after treatment with (+)-JQ1 (+) or (-)-JQ1 (-). Data is represented as mean +/- SEM from 3 technical replicates.

4.3 Assessing the effect of reduced BRD4 CRE-occupancy on transcription

4.3.1 Comparing mRNA levels of BRD4 target genes with and without JQ1 treatment

The well documented roles for BRD4 in transcriptional regulation, coupled with evidence of altered gene expression profiles in some CdLS patients (Liu *et al.*, 2009) (Boudaoud *et al.*, 2017), led me to hypothesise that the loss of BRD4 at CREs in Y430C Brd4 mutant cells would cause dysregulated expression of the associated genes. Indeed, inhibition of BRD4 acetyl-lysine binding by BETi is known to disrupt expression of target genes, and it has been suggested that genes regulated by SEs are the most sensitive to this inhibition (Lovén *et al.*, 2013). To confirm that this BD inhibition-induced transcriptional phenotype could be observed in my mESCs, I treated the WT cells with the BET inhibitor (+)-JQ1 and the inactive stereoisomer (-)-JQ1 at 300 nM for 24 hours.

RNA isolation, followed by RT-qPCR for the SE associated genes *Nanog, Myc, Klf4* and *Oct4*, confirmed that inhibition of BRD4 binding by (+)-JQ1 in WT mESCs causes a decrease in gene expression relative to treatment with (-)-JQ1. mRNA levels are normalised to the housekeeping gene *Gapdh,* which I would expect to show no change in expression in response to (+)-JQ1(Figure Ch4-11A).

4.3.2 Comparing mRNA levels of BRD4 target genes in WT and Y430C cells

I next used RT-qPCR to compare the expression of *Nanog, Klf4, Oct4* and *Sox2* between WT and Y430C cells, to see if the decreased BRD4 binding in the Y430C cells would have a similar effect on transcription as the inhibition of BRD4 binding by (+)-JQ1 did. I found that there was no significant difference in the levels of *Sox2, Nanog, Klf4* and *Oct4* mRNAs in Y430C cells compared to WT (Figure Ch4-11B), suggesting the BRD4 that remains bound at CREs of these genes in Y430C mESCs is sufficient for their transcription.

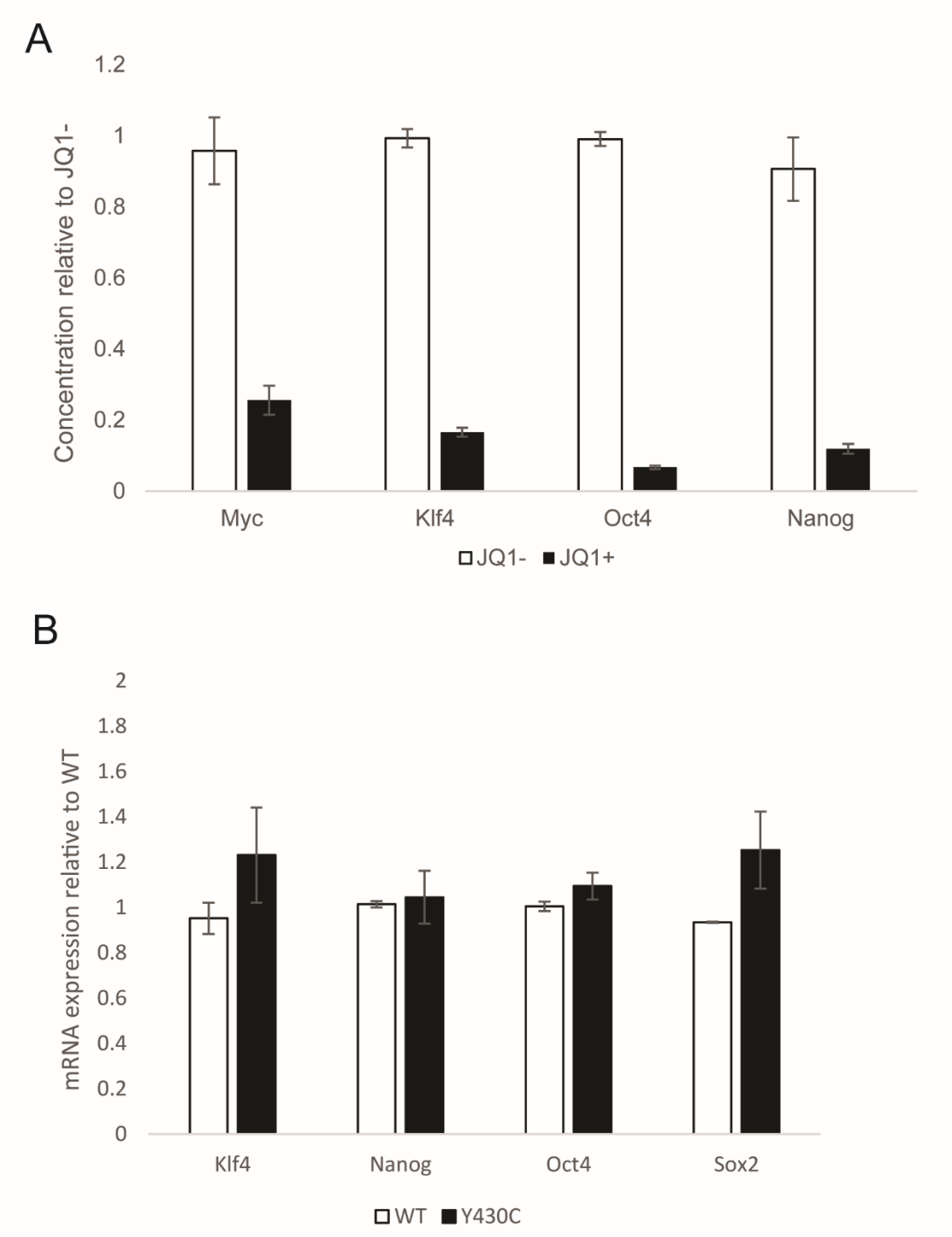


Figure Ch4-11 Effect of decreased BRD4 binding on transcription of target genes. **A)** RT-qPCR measuring mRNA concentration of BRD4 target genes after treatment with (+)-JQ1 (JQ1+) as a percentage of concentration after treatment with (-)-JQ1 (JQ1-). Data are represented as mean +/- SEM from 3 technical replicates. **B)** RT-qPCR measuring mRNA concentration for BRD4 target genes in WT and Y430C mESCs. Y430C mRNA concentration is shown relative to WT concentration, set at 1. Data are represented as mean +/- SEM from 3 biological replicates.

4.3.3 4sU sequencing

Given the decreased occupancy of Y430C-BRD4 at the CREs associated with BRD4 target genes (Figure Ch4-5&9), it was surprising that there was no change in their mRNA levels in Y430C cells. Cellular mRNA content, however, is not only dependent on its synthesis, but also on its processing and decay (Wu and Brewer, 2012). It is therefore possible that the transcription of these genes is decreased in the mutant cells, but somehow compensated for – eg. By enhanced mRNA stability. To examine this possibility, I carried out 4sU-seq to analyse nascent RNA transcripts in WT and Y430C cells. I did this for two biological replicates, and both gave comparable results. Quality control, genome browser screenshots and heatmaps are shown for just one replicate, but Cufflinks analysis and expression plots use an average of both replicates.

4.3.3.1 Quality control and mapping

As for my ChIP-seq files, I first ran FASTQC on the 4sU-seq data to check their quality. The overall per base sequence quality was slightly lower than that of the ChIP-seq, with many values in the second mate-pairs dropping as low as 30 (Figure Ch4-12). However, this is still considered to be within an acceptable range. The samples showed no evidence of adaptor content. WT pair 2 did show overrepresentation of one sequence, but this was at a very low percentage (0.1014) and didn't match any common contaminants.

Since the files all passed the quality control step, I then mapped them to the mm9 genome using TopHat. TopHat aligns reads using Bowtie, as for ChIP-seq, but also provides information about splice junctions between exons (Trapnell, Pachter and Salzberg, 2009). 100% of the reads were mapped to mm9 for both WT and Y430C, with 82.83% and 85.08% of these having corresponding paired reads (Table Ch4-3).

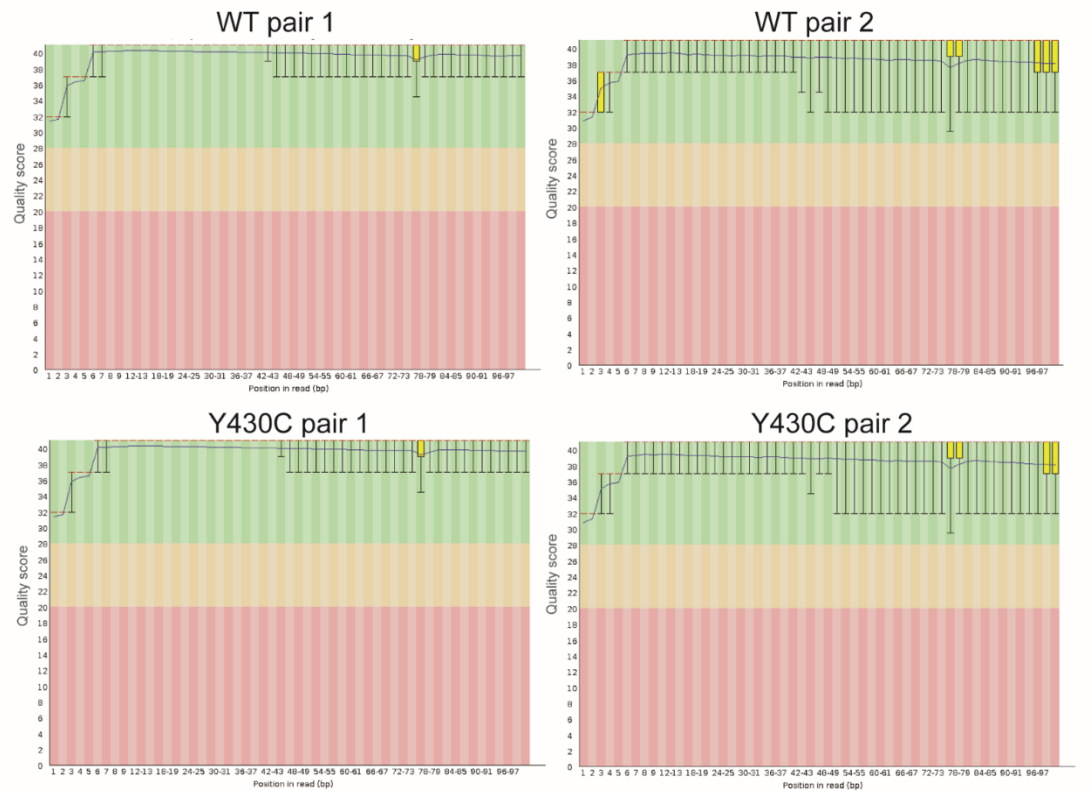


Figure Ch4-12. 4sU-seq per base sequence quality. Graphs show quality (phred) scores at each base position for the four 4sU-seq FASTQ files, as calculated by FASTQC

		Passed QC	Aligned to mm9	Properly paired
WT	**Number of reads**	355,979,124	355979124	197958934
	% of total reads	100	100	82.83
Y430C	**Number of reads**	388,181,585	388,181,585	246,358,506
	% of total reads	100	100	85.08

Table Ch4-3. Mapping statistics for WT and Y430C 4sU-seq FASTQ files. Files were mapped with Tophat and statistics taken from the alignment summaries.

4.3.3.2 Visualisation and peak-calling

I made bedgraphs from the mapped reads and visualised them on the UCSC genome browser. As expected, peaks were observed mainly over the gene bodies of genes transcribed in mESCs, and are present on either the + or – strand, consistent with strand specific sequencing.

At the BRD4 target genes, where I had seen decreased occupancy of BRD4 at the associated SEs, I found a similar number of 4sU-seq reads between WT and Y430C (Figure Ch4-13&14&16B). At *Gsc,* where there was little BRD4 occupancy at the promoter in WT and Y430C cells, I saw no transcription as expected (Figure Ch4-14). Heatmaps comparing the score distributions of WT and Y430C reads across gene bodies, showed that the similarity in transcript levels seen at BRD4 target genes is consistent across a majority of genes (Figure Ch4-15), and comparison of Y430C and WT expression profiles over exons showed that there was a good correlation overall (Figure Ch4-16A). The 4sU data therefore confirms the RT-qPCR results, suggesting there is little change in the transcription of *Klf4, Nanog*, *Sox2* and *Oct4* despite the loss of BRD4 CRE binding.

I used the Cufflinks programme (Trapnell *et al*., 2010, 2013) to assemble the aligned RNA-seq reads into parsimonious sets of transcripts. Combining the assembled transcripts from both replicates of WT and Y430C mapped reads generated a file of 313,343 transcripts corresponding to 23701 genes. Cufflinks estimates the number of cDNA fragments that originate from each transcript – the FPKM (Fragments Per Kilobase of transcript per Million mapped reads) - which is proportional to the expression of that transcript.

I then used the Cuffdiff module to identify transcripts that were differentially expressed in the WT and Y430C cells. Cuffdiff analysis showed that the majority of transcript levels were similar between WT and Y430C mESCs, corresponding with what I had observed for the SE associated genes. 40 genes were found to be expressed significantly more in WT over Y430C, whilst 114 were significantly enriched in Y430C over WT ($p<0.05$, fold enrichment >2). The DAVID functional analysis tool identified an enrichment of terms associated with 'heparin binding' in the WT enriched genes, and terms associated with GTP binding, immune response, nucleosome assembly and proteolysis in Y430C enriched genes ($p <0.05$) (data not shown). Most of these terms are not relevant to the loss of BRD4 binding in ES cells, and the genes do not correlate with regions of decreased BRD4 binding, suggesting they may be indirect targets of Y430C-BRD4.

As I will discuss in Chapter 5, the Y430C mESCs have an altered cell cycle, with a greater proportion of cells in G2/M phase and less in G1, compared to WT cells. It is possible that the transcription of these differentially expressed transcripts is coupled to the cell cycle, in which case I would expect to see an increase in the transcription of genes that are upregulated in G2/M in the Y430C cells.

The Y430C enriched term 'nucleosome assembly', represents a number of histone genes. In my 4sU-seq I see an increase in the transcription of histone cluster genes in Y430C cells (Figure Ch4-16C). Visualisation of the 4sU peaks using the UCSC genome browser, shows that transcription is increased in histone cluster 1 (Figure Ch4-17), 2 (Figure Ch4-18A&B) and 3 (Figure Ch4-18C), in Y430C cells compared to WT. This difference is most striking at histone cluster 3 genes, *Hist3h2bb-ps* and *Hist3h2a*, and BRD4 BD2 is known to bind to H3 (Z. Liu *et al.*, 2017).

A subset of histone genes are known to be highly cell cycle regulated; during each round of cell division new histones must be rapidly synthesised during S phase in order to maintain normal DNA-histone ratios (Mei *et al.*, 2017). In contrast, overexpression of the core histones outside of S phase is toxic to the cell (Kurat *et al.*, 2011). This results in 35-fold more histone mRNA in S phase compared to G1 and G2 phases (Harris *et al.*, 1991), achieved through both repression in G1 and G2 and transcriptional activation just prior to S phase (Gunjan, Paik and Verreault, 2005). Therefore, the upregulation of histone genes may be a result of the increased proportion of Y430C cells that are in S-phase (Figure Ch5-2).

Overall I concluded that the decreased BRD4 binding at CREs in Y430C mESCs does not directly affect the transcription of associated genes.

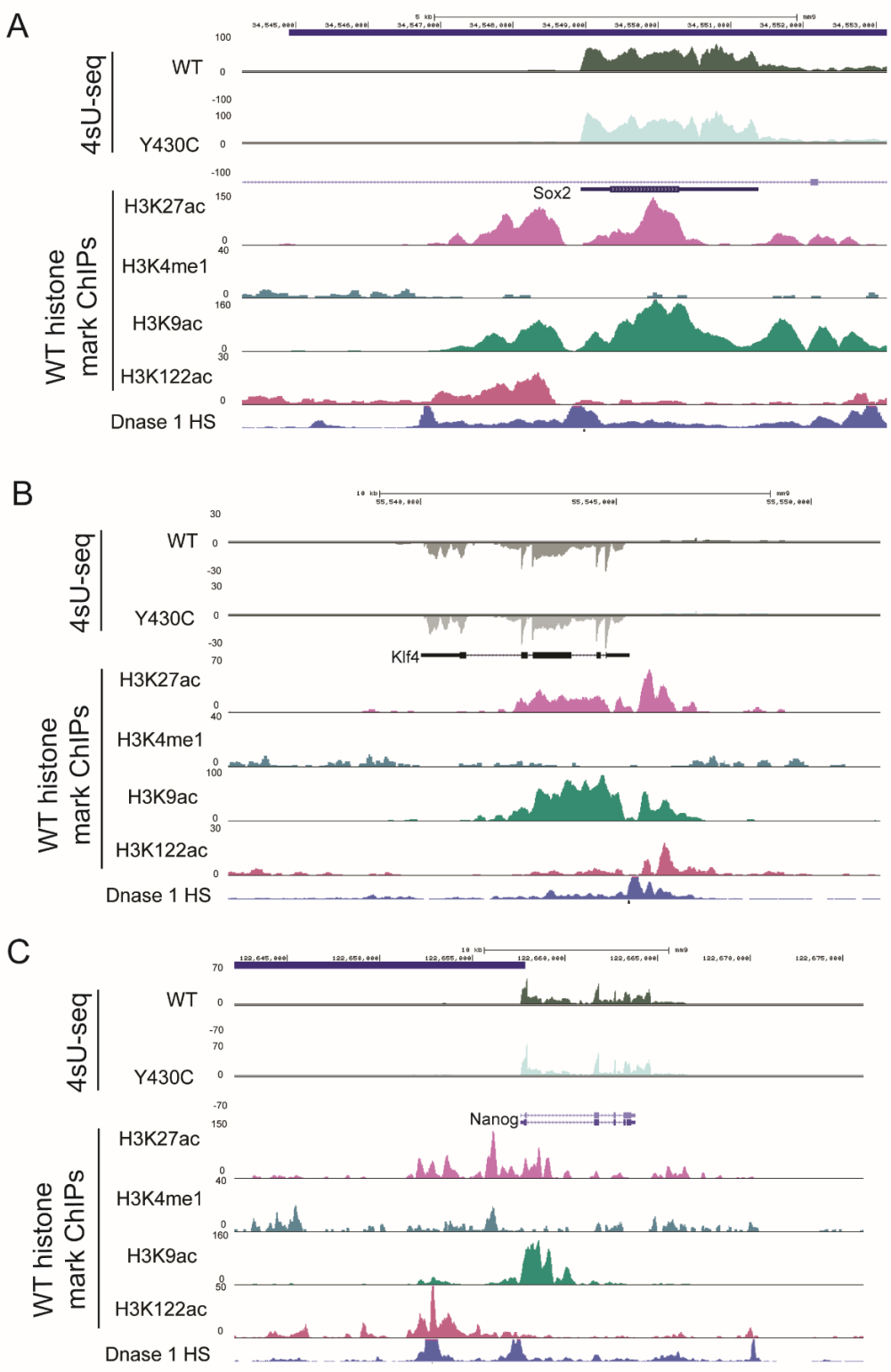


Figure Ch4-13. Genome browser view of nascent transcription at BRD4 target genes. UCSC browser screenshot showing 4sU-seq (RP10M) over **A)** *Klf4,* **B)** *Sox2* and **C)** *Nanog* genes. Tracks 1 and 2 = 4sU-seq in WT and Y430C mESCs, tracks 3-6 = previously published ChIP-seq data for various histone modifications, track 7 = DNase I hypersensitivity

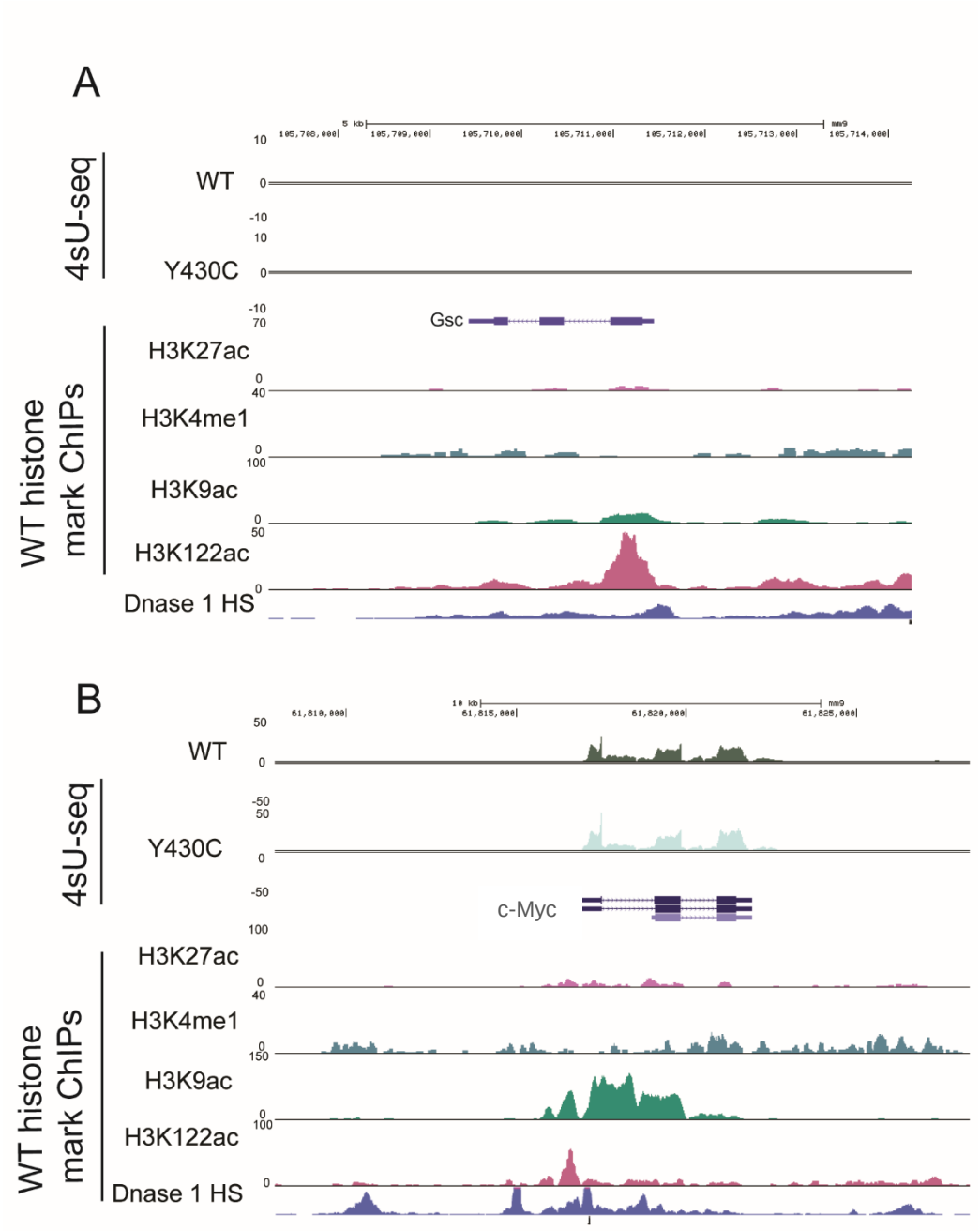


Figure Ch4-14. Genome browser view of nascent transcription at *Gsc* and *c-Myc*. UCSC browser screenshot showing 4sU-seq (RP10M) over **A)** *Gsc* and **B)** *c-Myc* genes. Tracks 1 and 2 = 4sU-seq in WT and Y430C mESCs, tracks 3-6 = previously published ChIP-seq data for various histone modifications, track 7 = DNase I hypersensitivity

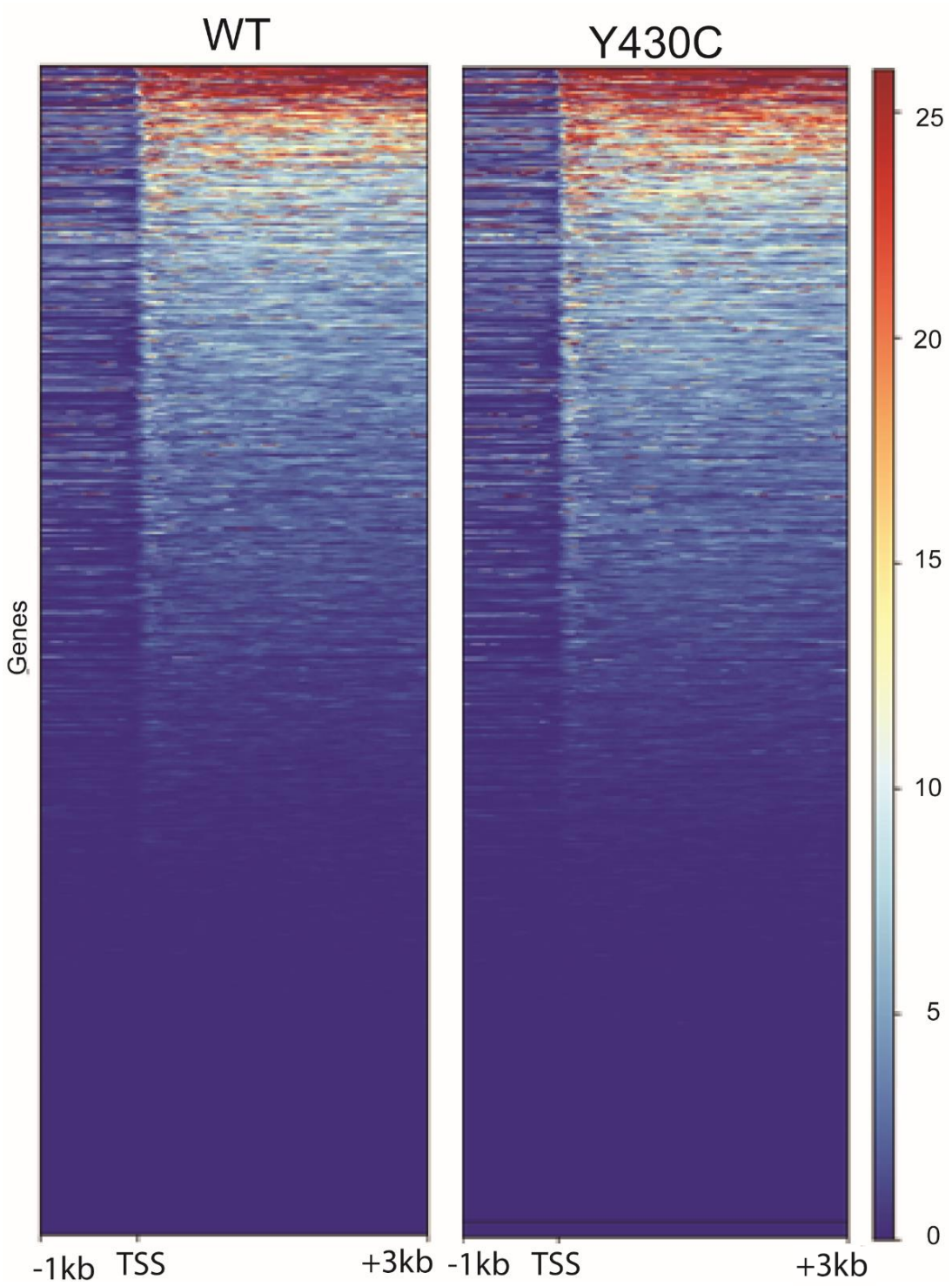


Figure Ch4-15. Nascent transcription over gene bodies. Heatmaps show distribution of reads across transcribed regions in WT and Y430C 4sU-seq data. Colour scale represents log2 read coverage normalised by RPKM. Regions 1 Kb upstream and 3 Kb downstream of the transcription start sites are plotted.

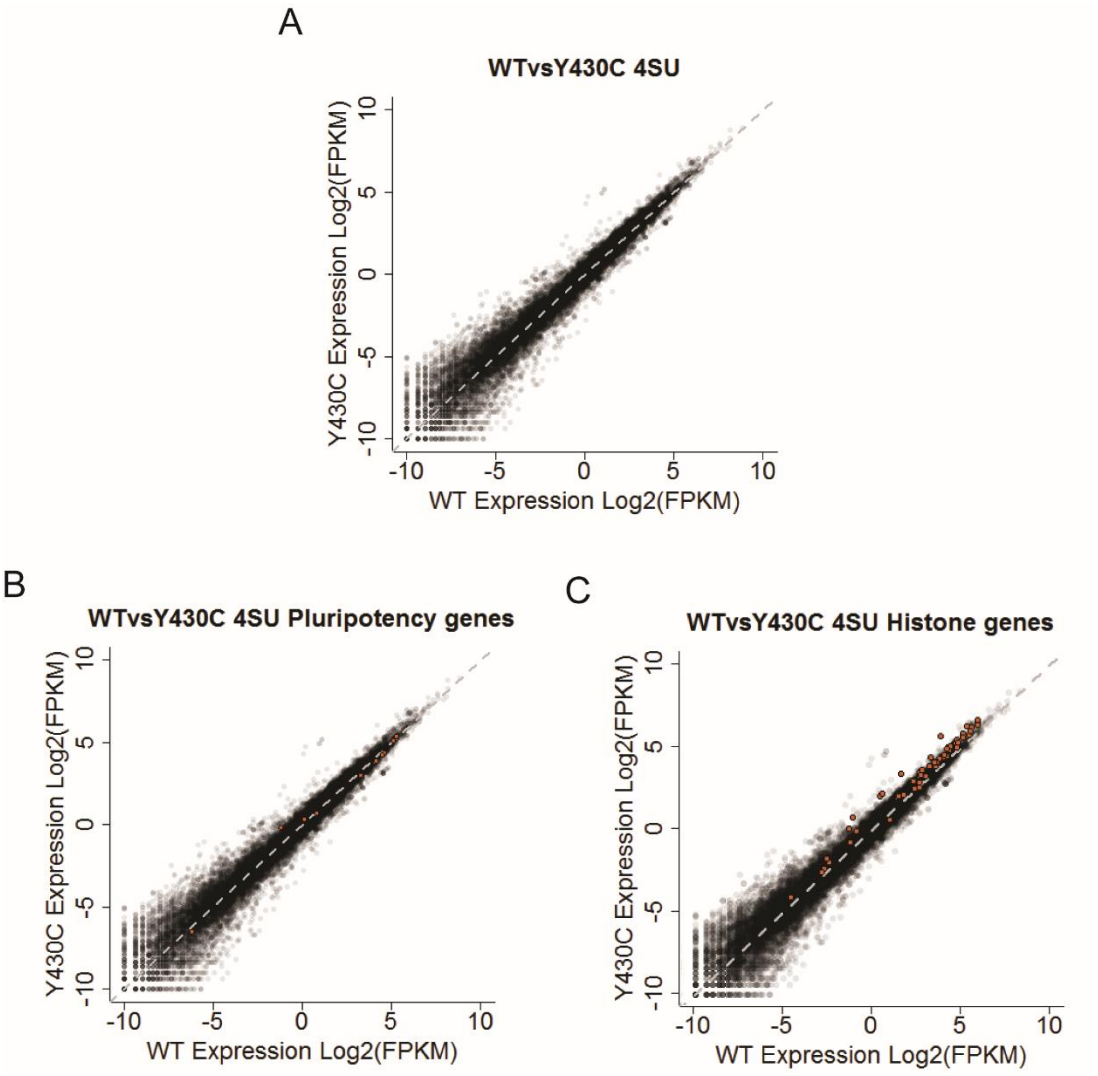
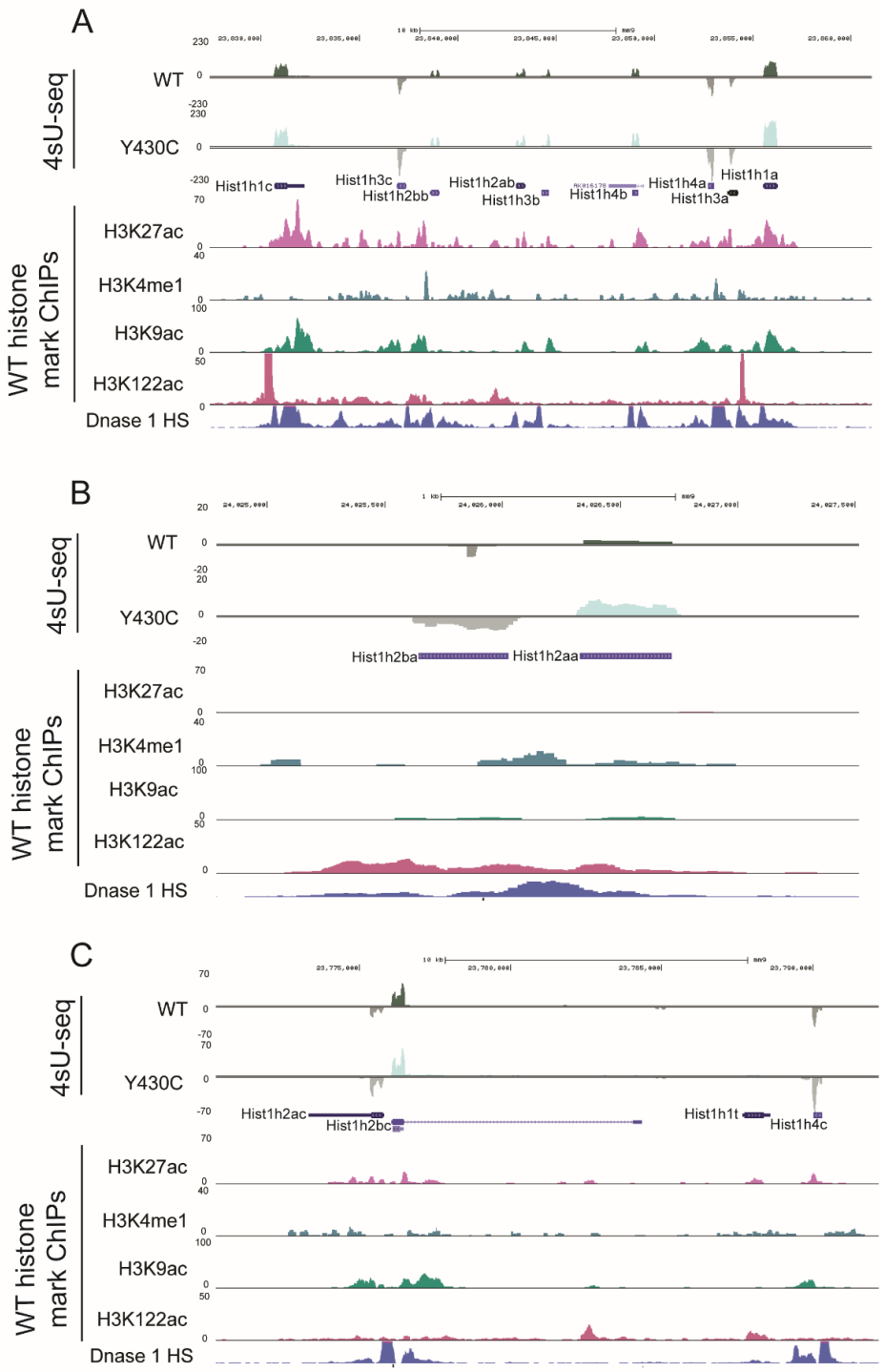


Figure Ch4-16. Nascent expression of exons in WT vs Y430C mESCs. A) Log2 (FPKM) for exons in WT 4sU data (x axis) plotted against Log2 (FPKM) for Y430C 4sU data (y-axis). **B&C)** Same plots as in **A)**, with a set of pluripotency genes and histone cluster genes respectively, highlighted in orange. Values are the average of 2 biological replicates. Grey dotted line = regression line



<u>Figure Ch4-17. Genome browser view of nascent transcription at histone cluster 1 genes.</u> UCSC browser screenshot showing 4sU-seq (RP10M) over histone cluster 1 genes. **A)** *Hist1h1c, Hist1h3c, Hist1h2bb, Hist1h2ab, Hist1h3b, Hist1h4b, Hist1h4a, Hist1h3a* and *Hist1h1a*. **B)** *Hist1h2ba* and *Hist1h2aa*. **C)** *Hist1h2ac, Hist1h2bc, Hist1h1ht* and *Hist1h4c*. Tracks 1 and 2 = 4sU-seq in WT and Y430C mESCs, tracks 3-6 = previously published ChIP-seq data for various histone modifications, track 7 = DNase I hypersensitivity

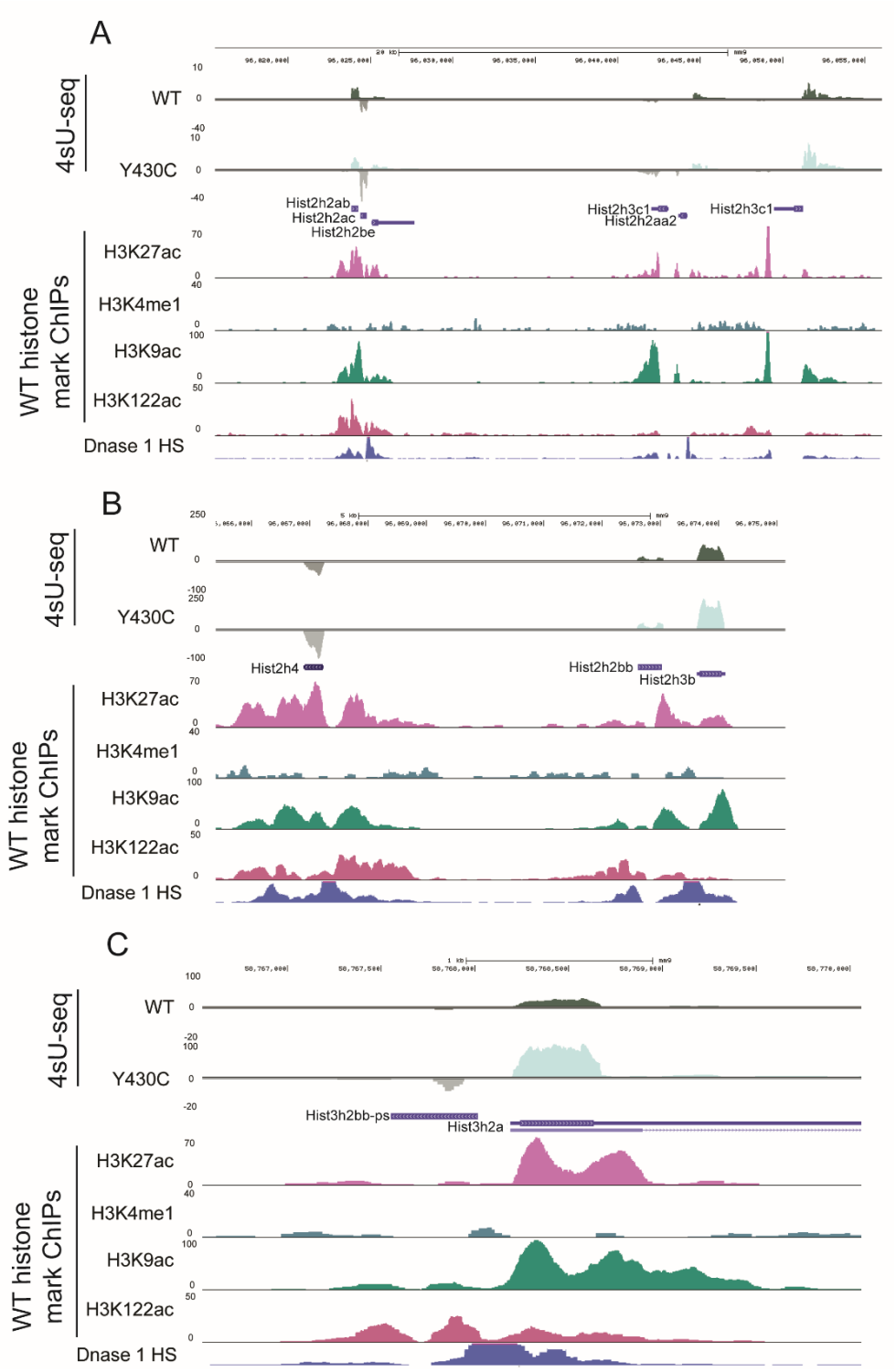


Figure Ch4-18. Genome browser view of nascent transcription at histone cluster 2 and 3 genes. UCSC browser screenshot showing 4sU-seq (RP10M) over histone cluster 2 - **A)** *Hist2h2ab, Hist2h2ac, Hist2h2be, Hist2h3c1, Hist2h2aa2, Hist2h3c1* and **B)** *Hist2h4, Hist2h2bb* and *Hist2h3b* – and Histone cluster 3 - **C)** *Hist3h2bb-ps* and *Hist3h2a.* Tracks 1 and 2 = 4sU-seq in WT and Y430C mESCs, tracks 3-6 = previously published –ChIP-seq data for various histone modifications, track 7 = DNase I hypersensitivity

4.3.4 RNA-sequencing with spike-in control

In my 4sU-seq data I have seen no difference in the levels of expression between WT and Y430C for most genomic regions. However, by normalising WT and Y430C to each other, it is possible that a global change in transcription is being masked (Chen *et al.*, 2016).

I therefore used RNA-seq to compare the overall transcription in WT and Y430C mESCs, this time spiking in a small amount of drosophila RNA. The spike-in and test samples were combined at the cell stage (200,000 drosophila cells were combined with 10,000,000 mESCs), so the samples were treated the same from this point on. I did this for three biological repeats, with similar results for each. Genome browser screen shots and heatmaps show results from one replicate, whilst expression plots represent an average of all three replicates. After normalisation to the spike-in control, I found that the results largely corroborate those of the 4sU data – there is little difference in transcription of the BRD4 target pluripotency genes (Figure Ch4-19,20&22B), or overall (Figure Ch4-21&22A), between WT and Y430C mESCs. There is a noticeable decrease in nascent *c-Myc* transcripts in the Y430C cells (Figure Ch4-20B), but this is not present in the other 2 replicates.

Plotting expression of WT vs Y430C (Figure Ch4-22A), does show more divergence in this experiment than that of the 4sU data without the spike-in (Figure Ch4-16A). These differences are in part accounted for by the expression of histone cluster genes. In the 4sU data, these show a near consistent increase in expression in Y430C over WT. This general trend is also seen in the spike-in experiment, however there are a few outliers, that show greatly differential expression between WT and Y430C (Figure Ch4-22C). Some of these are much more highly expressed in the Y430C cells than in WT, which is similar to what was seen in the 4sU data (Figure Ch4-16C,17-19), however a couple are also shown to be expressed more in the WT (Figure Ch4-22C). I checked the profiles across these differentially expressed histone cluster genes on UCSC, and this confirmed their differences in expression (Figure Ch4-23). The increased variability of expression of these genes in the total RNA-seq could be a reflection of their post-transcriptional regulation (since 4sU-seq measures only nascent transcripts). For instance these transcripts are known to be rapidly degraded at the end of S phase (Kaygun and Marzluff, 2005).

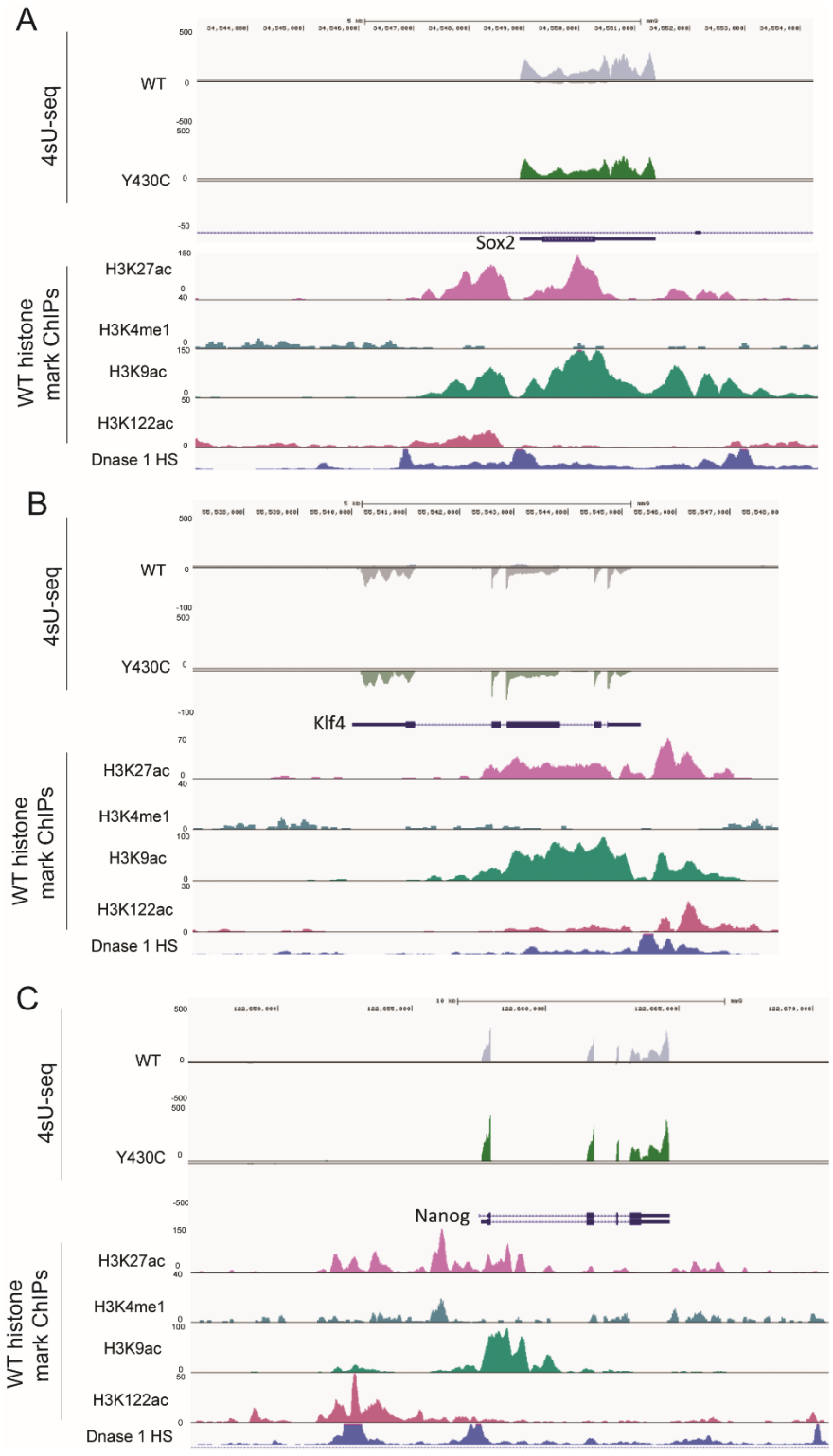


Figure Ch4-19. Genome browser view of transcription at BRD4 target genes after normalisation to spike-in. UCSC browser screenshot showing RNA-seq (RP10M) over **A)** *Sox2,* **B)** *Klf4* and **C)** *Nanog* genes. Tracks 1 and 2 = 4sU-seq in WT and Y430C mESCs, tracks 3-6 = previously published ChIP-seq data for various histone modifications, track 7 = DNase I hypersensitivity

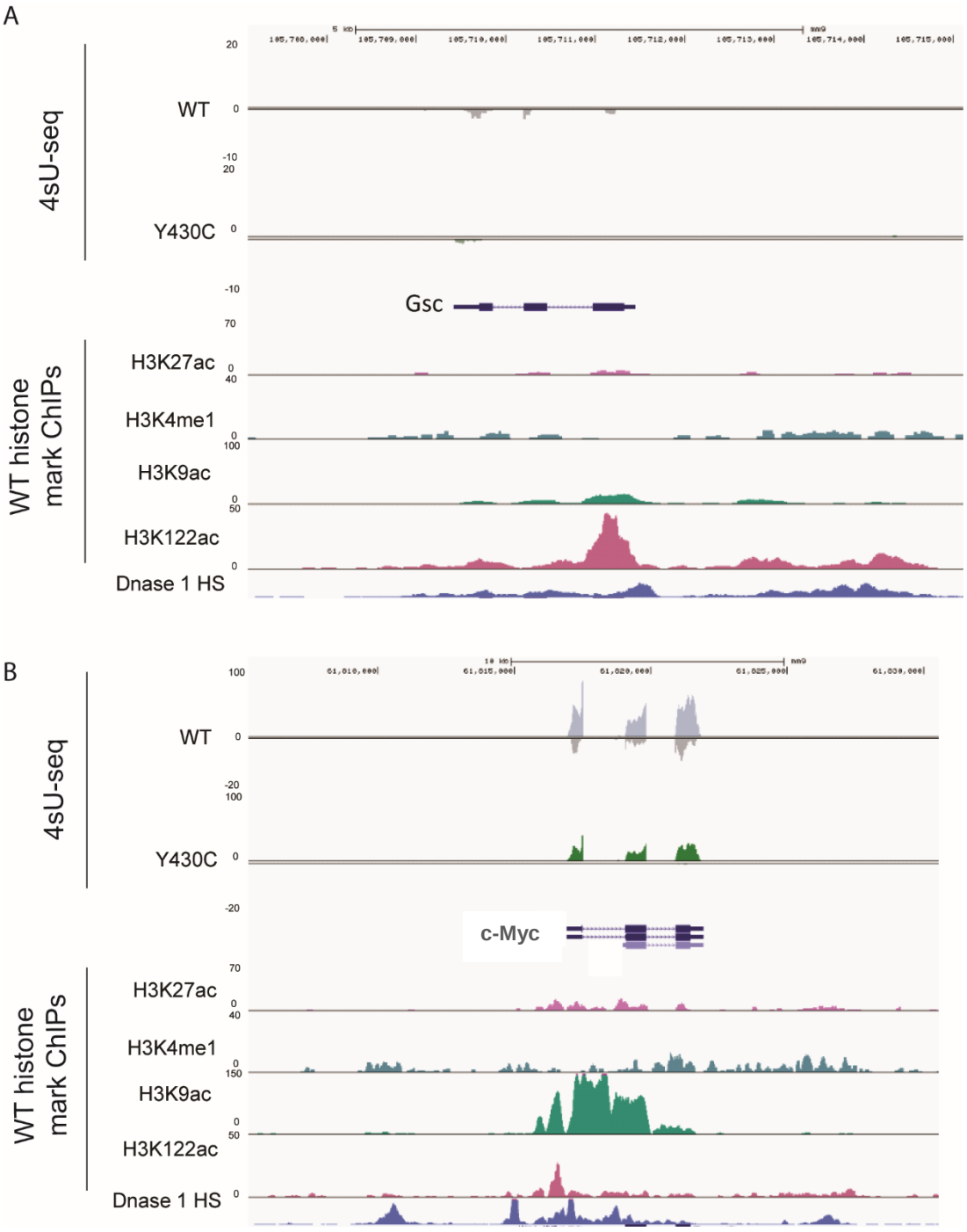


Figure Ch4-20. Genome browser view of transcription at *Gsc* and *c-Myc* after normalisation to spike-in. UCSC browser screenshot showing RNA-seq (RP10M) over **A)** *Gsc* and **B)** *c-Myc* genes. Tracks 1 and 2 = 4sU-seq in WT and Y430C mESCs, tracks 3-6 = previously published ChIP-seq data for various histone modifications, track 7 = DNase I hypersensitivity.

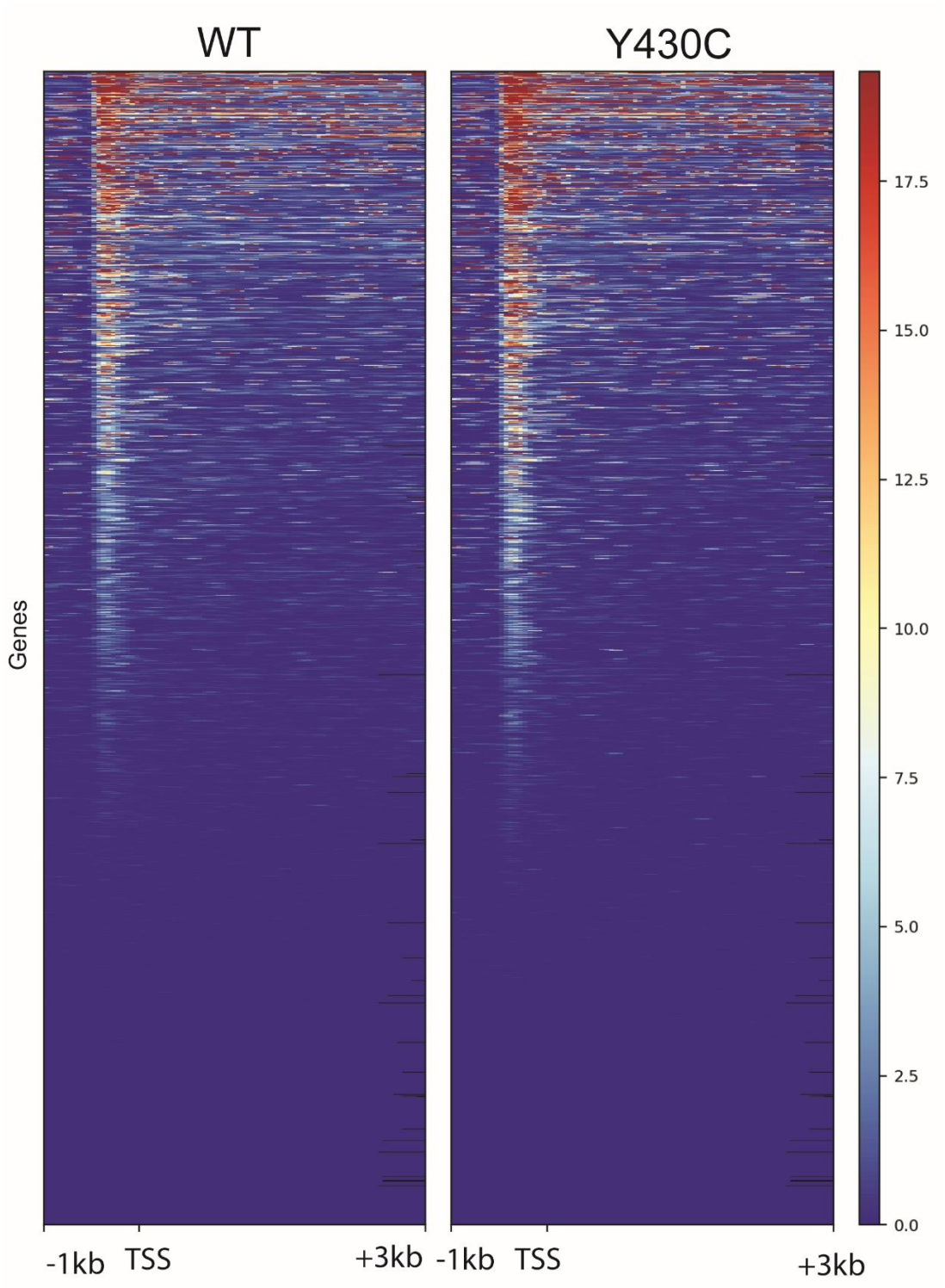


Figure Ch4-21. Transcription over gene bodies, with spike-in control. Heatmaps show distribution of reads across transcribed regions in WT and Y430C RNA-seq data, after normalisation to spike-in. Colour scale represents log2 read coverage normalised by reads per kilobase per million. Regions 1 Kb upstream and 3 Kb downstream of the transcription start sites are plotted.

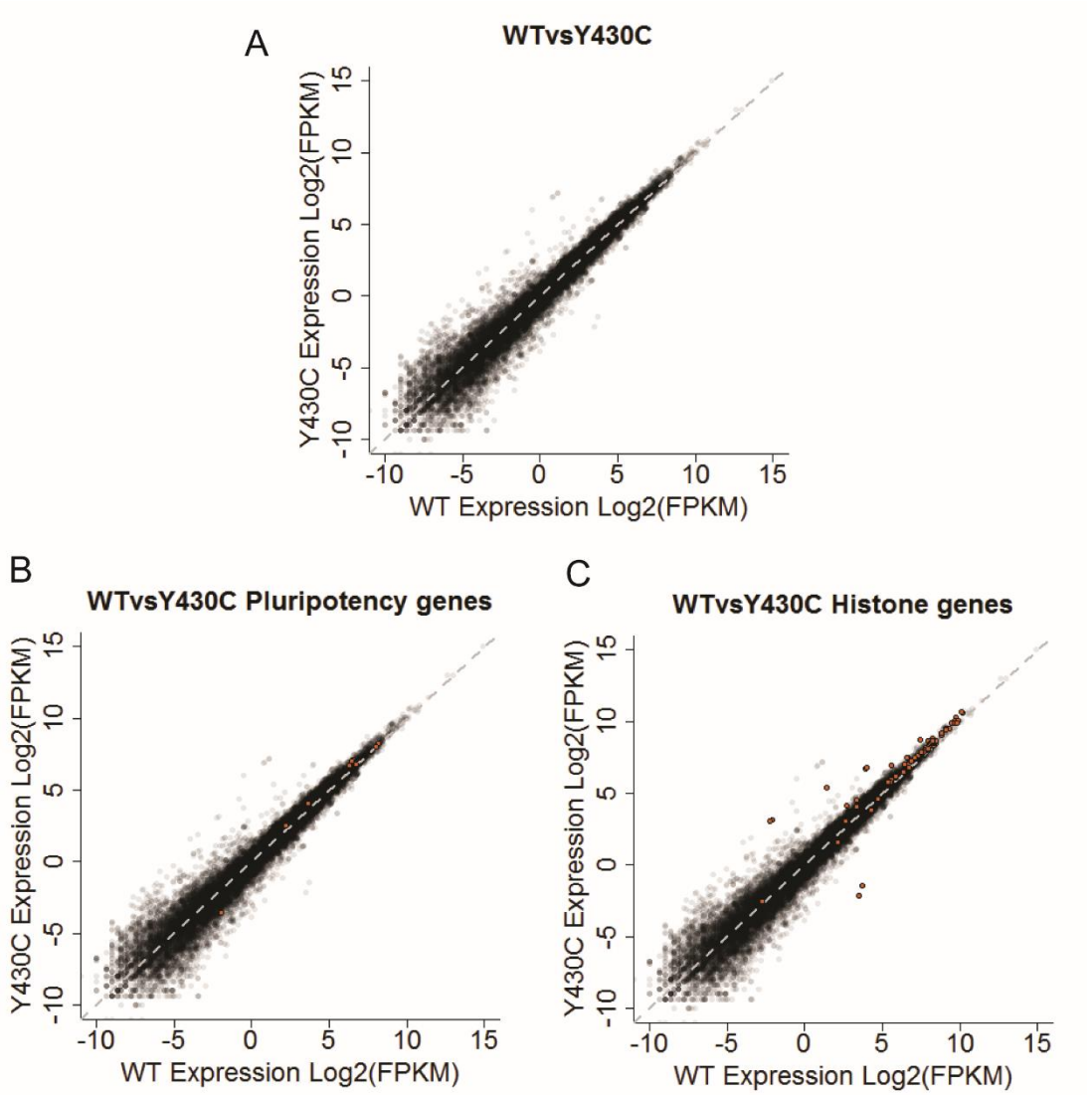


Figure Ch4-22. Expression of exons in WT vs Y430C mESCs, normalised to spike-in control. A) Log2 (FPKM) for exons in WT data (x axis) plotted against Log2 (FPKM) for Y430C data (y-axis). **B&C)** Same plots as in **A)**, with a set of pluripotency genes and histone cluster genes respectively, highlighted in orange. Values are the average of 3 biological replicates. Grey dotted line = regression line

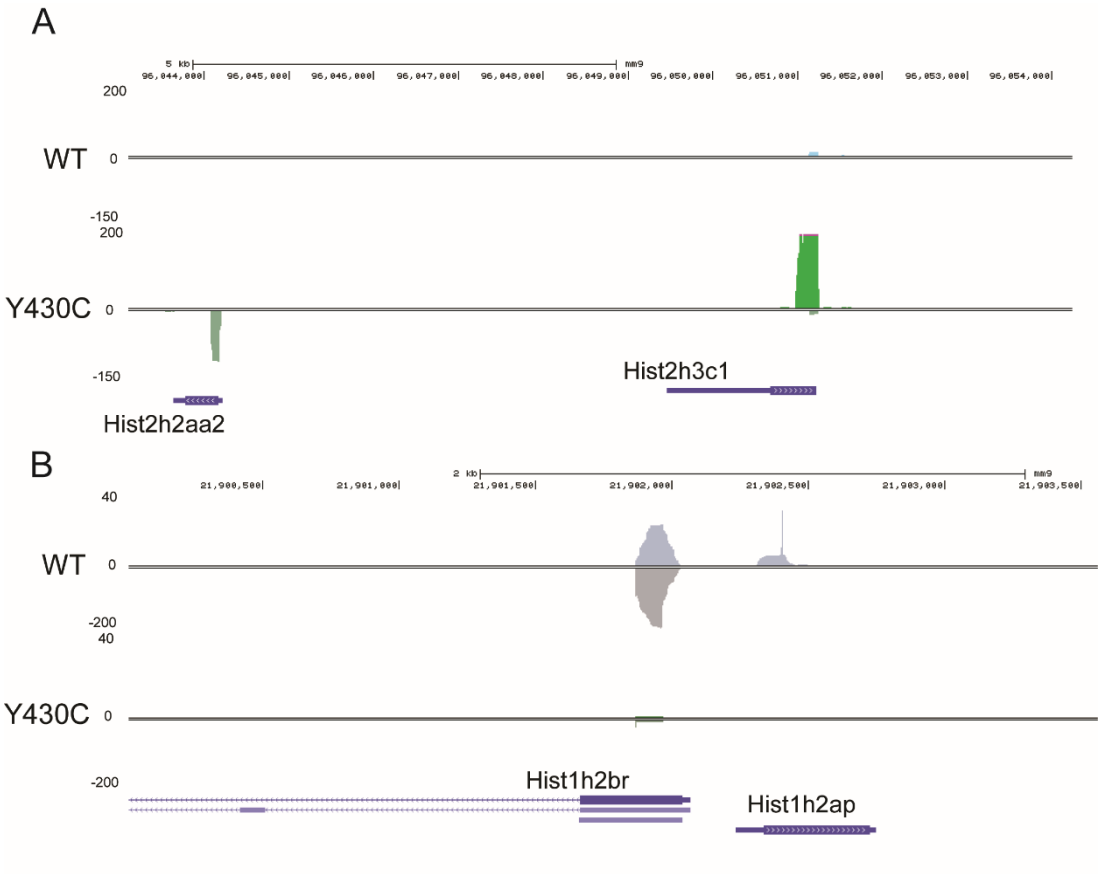


Figure Ch4-23. Genome browser view of transcription after normalisation to spike-in. UCSC browser screenshot showing RNA-seq (RP10M) over **A)** *Hist2h2aa2* and *Hist2h3c1* and **B)** *Hist1h2br* and *Hist1h2ap* genes. Tracks 1 and 2 = 4sU-seq in WT and Y430C mESCs

4.4 Assessing the effect of reduced BRD4 CRE-occupancy on enhancer transcription

Like promoters, active enhancers themselves recruit RNAPII, resulting in the bi-directional transcription of a novel class of transcripts, termed eRNAs (Kim *et al.*, 2010). These eRNAs are short (<2 Kb) (Kim *et al.*, 2010) and they are rapidly degraded by the exosome (Ogami, Chen and Manley, 2018). It has been suggested that the transcription of eRNAs may have a functional role in the regulation of gene expression (Meng and Blaine, 2017).

BRD4 has been implicated in the regulation of eRNAs. On the one hand BRD4 has been shown to facilitate eRNA transcription (Kanno *et al.*, 2014). Others suggest however, that BRD4 inhibits eRNA synthesis through the recruitment of a complex of 7SK snRNA and the chromatin remodelling complex BAF to enhancers. BAF remodels chromatin and decreases the binding of certain TFs. 7SK is particularly enriched at SEs and these regions were found to be most sensitive to 7SK depletion (Flynn *et al.*, 2016).

eRNAs are hard to study through steady-state RNA assays such as RNA-seq, due to their unstable nature, however their transcription can be captured through nascent transcription assays such as 4sU-seq (Azofeifa *et al.*, 2018). To address whether the decreased BRD4 occupancy at enhancers and SEs in Y430C cells causes an increase in eRNA transcription, I looked at my 4sU-seq data specifically over enhancer regions.

I found that reads were indeed generated over active TEs and SEs (Figure Ch4-24, 25&26A), but none over sonic brain enhancers 2 and 4 (SBE2 and SBE4) (Figure Ch4-26B), which are inactive in mESCs. At these specific regions I was unable to see an obvious difference between the amount of transcription in WT and Y430C cells (Figure Ch4-24&25). The ATAC-seq track shown in figures Ch4-24-26 was generated in mESCs by Yatendra Kumar (Bickmore Lab) and allows me to identify regions of open chromatin, where I would expect transcription to occur; ATAC-seq uses a hyperactive Tn5 transposase to insert sequencing adapters into chromatin. These adapters are inserted specifically into regions accessible to

the transposase, meaning that the sequencing reads can be used to identify regions of increased accessibility and to infer nucleosome positioning (Buenrostro *et al.*, 2015).

I generated heatmaps showing the distribution of 4sU-seq reads across SEs and TEs (figure Ch4-26A&B), and confirmed that there was little difference in read coverage between WT and Y430C cells. TEs showed a very small enrichment of reads compared to a 60 Kb surrounding area (Figure Ch4-27A). However when I looked specifically at those TEs that were bound by BRD4 (identified by intersecting TE regions with peaks called in WT BRD4 ChIP-seq), I saw an increase in the read coverage over TEs (Figure Ch4-27C). A similar increase was also seen for SEs that are bound by BRD4 (Figure Ch-27D). These data suggest a link between eRNAs and BRD4.

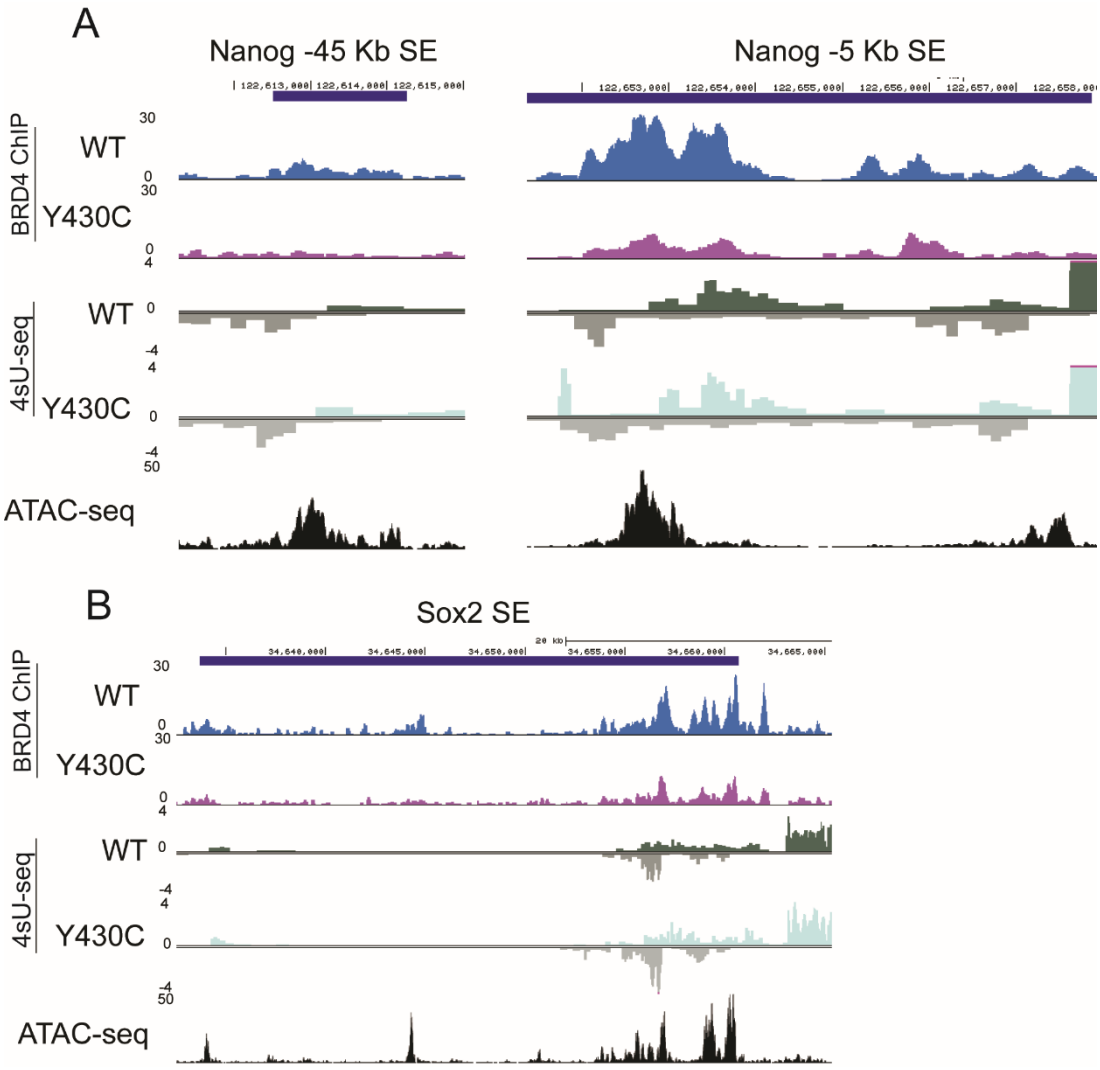


Figure Ch4-24. Nascent transcription at SEs of BRD4 target genes. UCSC browser screenshot showing 4sU-seq (RP10M) over **A)** *Nanog* SEs and **B)** *Sox2* SE. Tracks 1 and 2 = 4sU-seq in WT and Y430C mESCs, tracks 3-6 = previously published 4sU-seq data for various histone modifications, track 7 = DNase I hypersensitivity

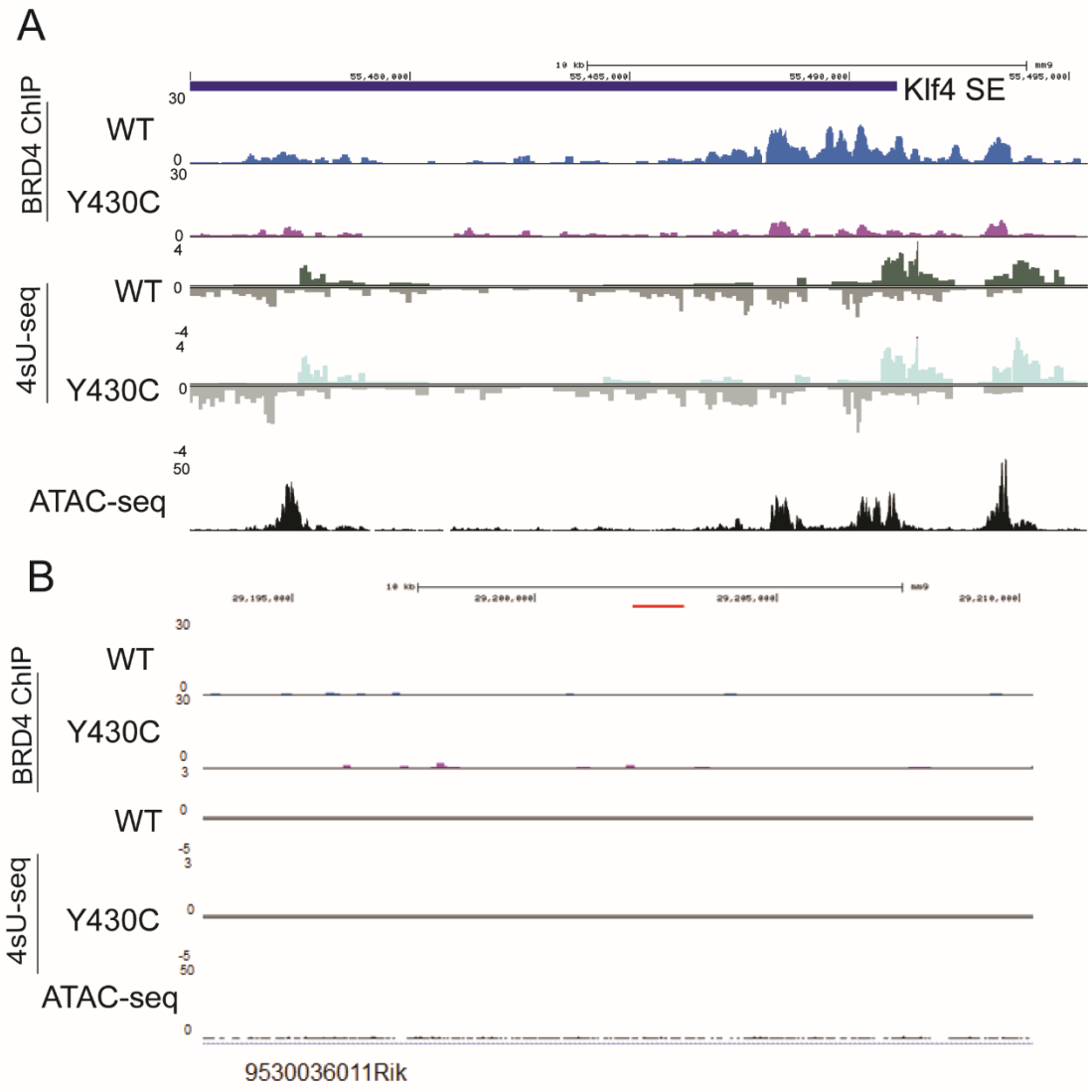


Figure Ch4-25. Nascent transcription at *Klf4* SE and the inactive SBE2 enhancer. UCSC browser screenshot showing 4sU-seq (RP10M) over **A)** *Klf4* SE and **B)** a region of Chr5 including SBE2 (SBE2 = Chr5: 29202018-29203152, red bar). Tracks 1 and 2 = 4sU-seq in WT and Y430C mESCs, tracks 3-6 = previously published 4sU-seq data for various histone modifications, track 7 = DNase I hypersensitivity

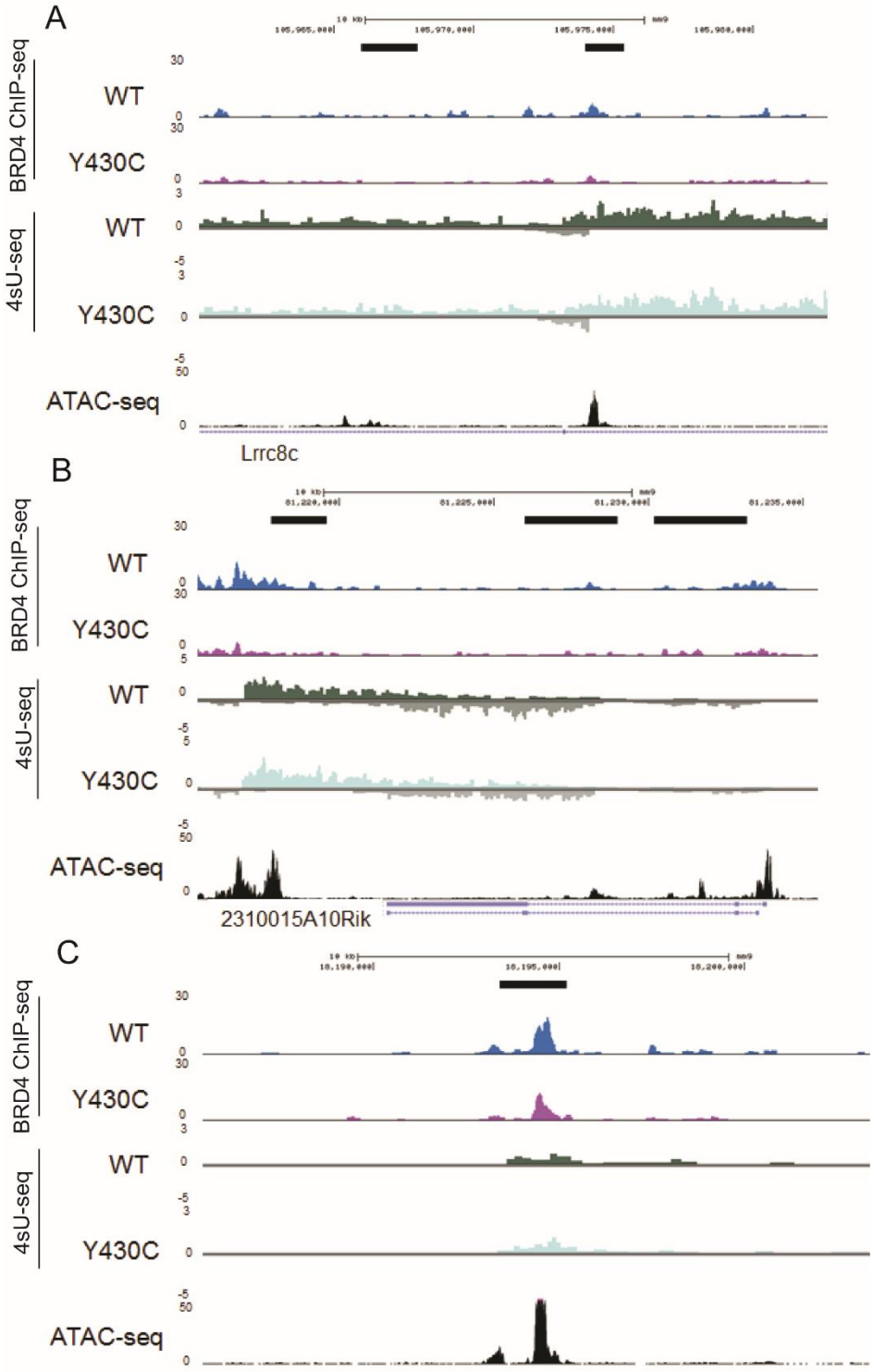


Figure Ch4-26. Nascent transcription at TEs. UCSC browser screenshot showing 4sU-seq (RP10M) over TEs associated with **A)** *Lrrc8c,* **B)** *2310015A10Rik* and **C)** *Gnai1*. Tracks 1 and 2 = 4sU-seq in WT and Y430C mESCs, tracks 3-6 = previously published 4sU-seq data for various histone modifications, track 7 = DNase I hypersensitivity. Black bars indicates strong enhancers as defined by ChromHMM

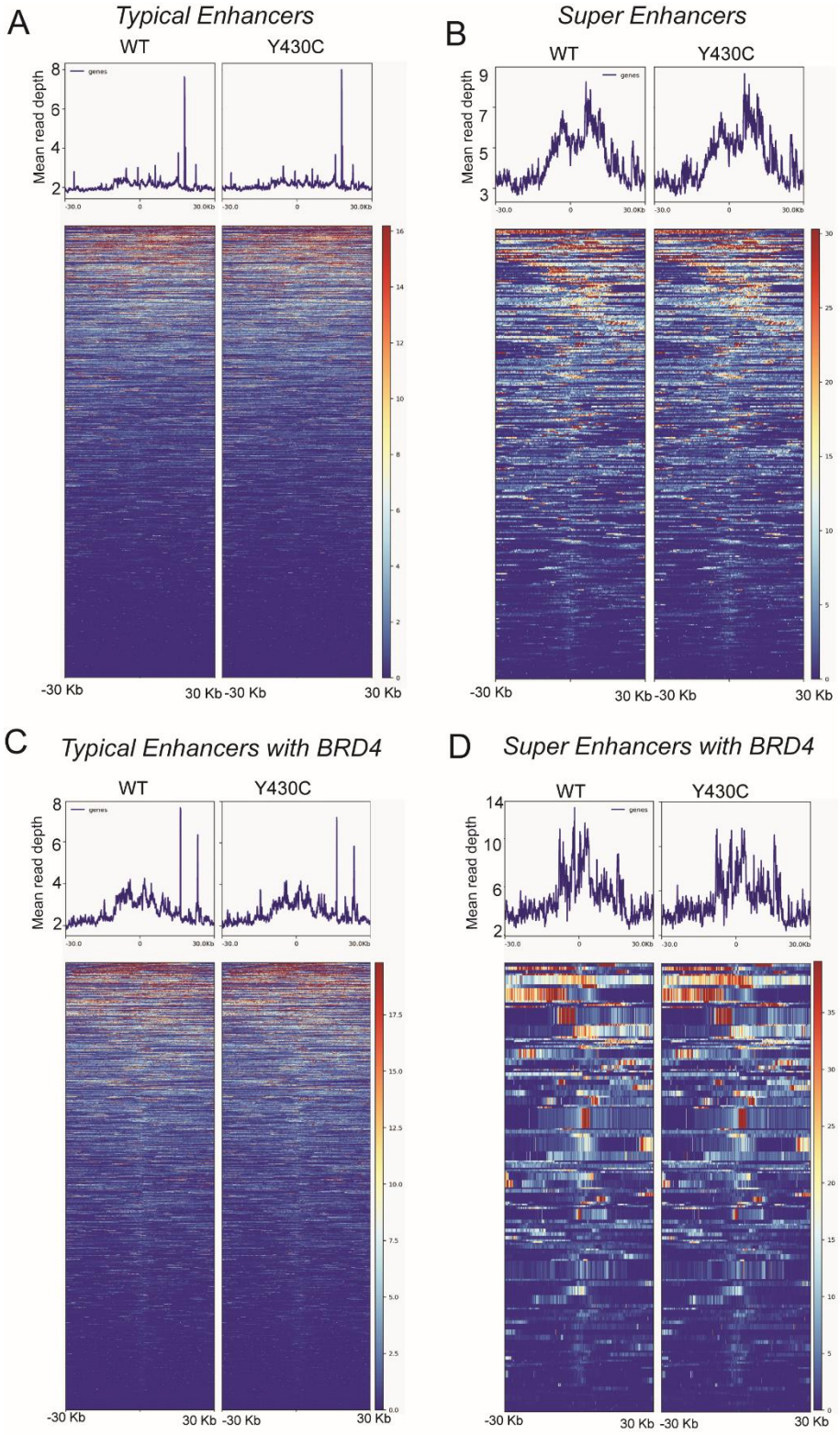


Figure Ch4-27. Nascent transcription over enhancers. Heatmaps show distribution of reads across **A)** TEs, **B)** SEs, **C)** TEs with BRD4 peaks and **D)** SEs with BRD4 peaks in WT and Y430C 4sU-seq data. Colour scale represents log2 read coverage normalised by RPKM. Plots show regions +/- 30Kb.

4.5 Effect on pluripotency

BRD4 binding to CREs has been shown to maintain self-renewal and pluripotency in mESCs (Di Micco *et al.*, 2014). I therefore wondered whether the decrease in BRD4 at these sites would cause a loss of the pluripotent stem-cell state of Y430C mESCs.

To address this I used IF to compare the levels of NANOG, ESRRB, SOX2 and OCT4 in WT and Y430C mESCs. These TFs are all involved in maintaining the ES cell genetic programme, with NANOG, SOX2 and OCT4 being considered 'master regulators' (X. Zhang *et al.*, 2008). Furthermore expression of *Sox2, Nanog* and *Esrrb* has been shown to be downregulated upon differentiation (Nair *et al.*, 2015), making their expression a good marker of ESCs.

Both WT and Y430C mESCs expressed all four stem-cell markers as expected (Figure Ch4-28). Expression of these proteins, particularly NANOG and ESRRB, was somewhat heterogeneous. This is not surprising however, as pluripotency-related genes are known to show heterogeneous expression and temporal fluctuations in mESCs cultured in standard LIF containing serum (Marucci and Lucia, 2017). These fluctuations are thought to correlate with potency of the cells; mESCs are pluripotent at the population level, but subpopulations may show varying propensities towards differentiation (Marucci and Lucia, 2017). Growing the cells in 2I media – which uses a combination of LIF and two inhibitors to maintain ESCs in a naive ground state (Sim *et al.*, 2017) – might overcome this heterogeneity.

I observed very little difference between the intensity of stem-cell markers in WT and Y430C mESCs, suggesting that there is no difference in the expression of these genes at the protein level, as well as at the transcript level. I would therefore expect the Y430C mESCs to behave similarly to WT cells in terms of their stem-cell properties. IF is not a quantitative measure of protein expression however, so a quantification of these results is needed for confirmation.

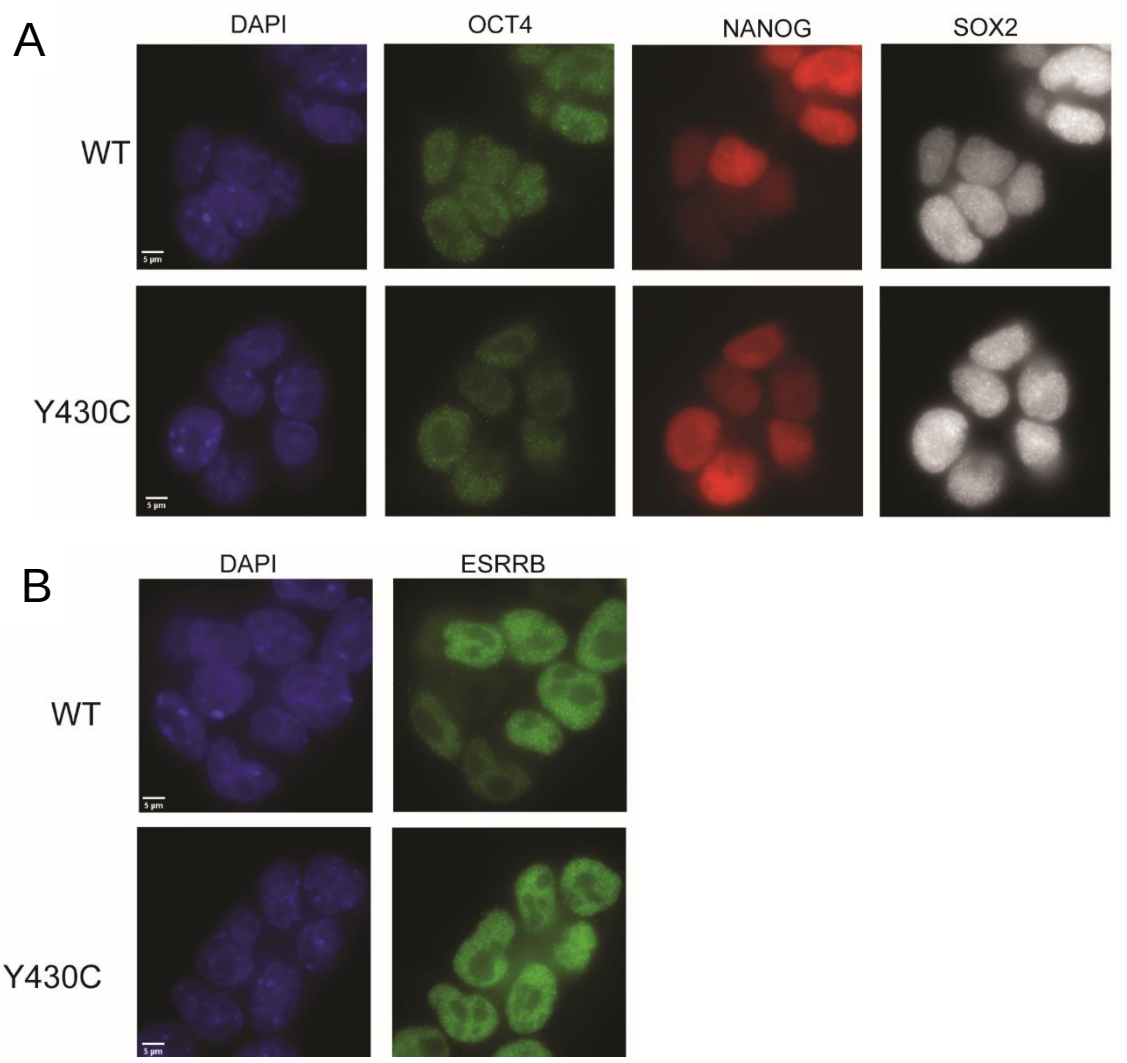


Figure Ch4-28. Expression of stem cell marker proteins in WT and Y430C mESCs. Representative images of WT and Y430C mESCs upon **A)** DAPI staining, and OCT4, NANOG, and SOX2 IF, and **B)** DAPI staining and ESRRB IF. Scale bar = 5 μm

4.6 Conclusions

The aim of this chapter was to assess how the decreased affinity of BRD4-Y430C for acetyl-lysines affects genome wide binding of BRD4, and what consequences this has for gene expression.

I found that BRD4 is enriched at CREs compared to other genomic regions, but that this enrichment is seen in both WT and Y430C mESCs. However I did find decreased binding the Y430C-BRD4 compared to WT. The decreased affinity of BRD4-Y430C to acetyl-lysines appears to affect binding at TEs, SEs and promoters more than other regions. Given the known regulatory roles of BRD4 at these regions, it might be assumed that the loss of BRD4 at these sites would influence gene expression.

I was able to show that the inhibition of BRD4 binding by (+)-JQ1 did decrease the expression of pluripotency related genes – known targets of BRD4 in mESCs. However, despite decreased BRD4 binding at CREs in Y430C cells, there was no change in the nascent transcription of BRD4 target genes, nor on a more global level. These discrepancies could be explained by the lack of JQ1 selectivity for BET proteins; JQ1 has been shown to effectively inhibit BRD2, BRD3 and BRD4 (Filippakopoulos *et al.*, 2010). This lack of selectivity is common amongst many of the available BETi and is likely due to the high degree of sequence conservation, especially at those residues important for acetyl-lysine binding (Z. Liu *et al.*, 2017). Some more specific BETi are available and could be used to test this. For instance, RVX-208 is specific to BRD4-BD2 (Ferri, Petosa and McKenna, 2016), whilst BAY1238097 shows some selectivity for BRD4 - binding to BRD4 BDs 10-fold and 39-fold more than BRD3 and BRD2 respectively – but is no longer in development (reasons undisclosed)(Z. Liu *et al.*, 2017).

I observed no differences at the protein level for BRD4 target genes, with IF for stem cell markers showing similar expression in WT and Y430C cells. These results were surprising, and somewhat contradictory to recent publications that report transcriptional misregulation in CdLS patients. Nevertheless, I conclude that the diminished level of BRD4 present at CREs in Y430C mESCs is sufficient for its roles in transcriptional regulation and stem-cell maintenance, at least in the context of mESCs. Moreover, the results in this chapter are compatible with the absence of any major developmental abnormalities in the Y430C mouse embryos (Figure Ch3-4).

BRD4 has been implicated in the stimulation of eRNA elongation, in a manner dependent on BD-mediated acetyl-lysine interactions (Kanno *et al.*, 2014), but also in the suppression of eRNA transcription, through 7SK-BAF recruitment (Flynn *et al.*, 2016). Whilst I was able to detect transcription at TEs and SEs in the mESCs, I found that there was no difference in the amounts of transcription at these regions between WT and Y430C cells. I found that eRNA transcription is more common at enhancers occupied by BRD4, which could support either of the proposed roles of BRD4 in eRNA regulation, or a combination of them both - BRD4 at enhancers may promote the transcriptional elongation of eRNAs, and at the same time recruit 7SK-BAF to limit pervasive transcription.

Chapter 5: An altered DNA damage response in Y430C mESCs

5.1 Comparing the cell cycles of WT and Y430C mESCs

5.1.1 Y430C mESCs proliferate more slowly than WT

During routine culturing of the Y430C-BRD4 mESCs and their WT counterparts, I noticed that the mutant cells were doubling in number more slowly. To quantify this difference, I conducted some simple cell growth assays, plating the cells at a known density and counting them every 24 hours for the following 4 days (Figure Ch5-1). This confirmed my observation – by day 3 the WT cultures had reached saturation, whilst the number of cells in the Y430C cultures were less than half those of the WT and continued to double over the next 24 hours. This suggests that the mutant cells may be dividing twice as slowly as the WT cells, and therefore have a slowed cell cycle. Alternatively, the Y430C cells could be doubling at the same rate as the WT cells but undergoing apoptosis, resulting in a slower increase in cell number. The lack of a G1/S phase cell cycle checkpoint renders ESCs prone to apoptosis in response to DNA damage (Wang *et al.*, 2009).

5.1.2 Flow cytometry profiles of WT and Y430C mESCs

Flow cytometry with propidium iodide stained WT and Y430C mESCs, showed that there was indeed a difference in the cell cycle profiles of these cells (Figure Ch5-2A&B). A larger proportion of Y430C cells were in the G2/M and S phases of the cell cycle compared to WT (Figure Ch5-2). Figure Ch5-2 shows the results for one biological replicate, but the second replicate showed a similar trend. I did flow cytometry for the biological replicate cell lines and found similar trends.

As ES cells lack a G1 checkpoint (Aladjem *et al.*, 1998), the differences in cell cycle between WT and Y430C mESCs cannot be caused by impaired activation of the G1/S checkpoint in the mutant cells. Instead, the Y430C profile could be the result of a problem with DNA damage during replication, activating the intra-S checkpoint and slowing the synthesis of new DNA (Willis and Rhind, 2009) or activation of the G2/M checkpoint, delaying the cells from transitioning into mitosis (Liu *et al.*, 2000).

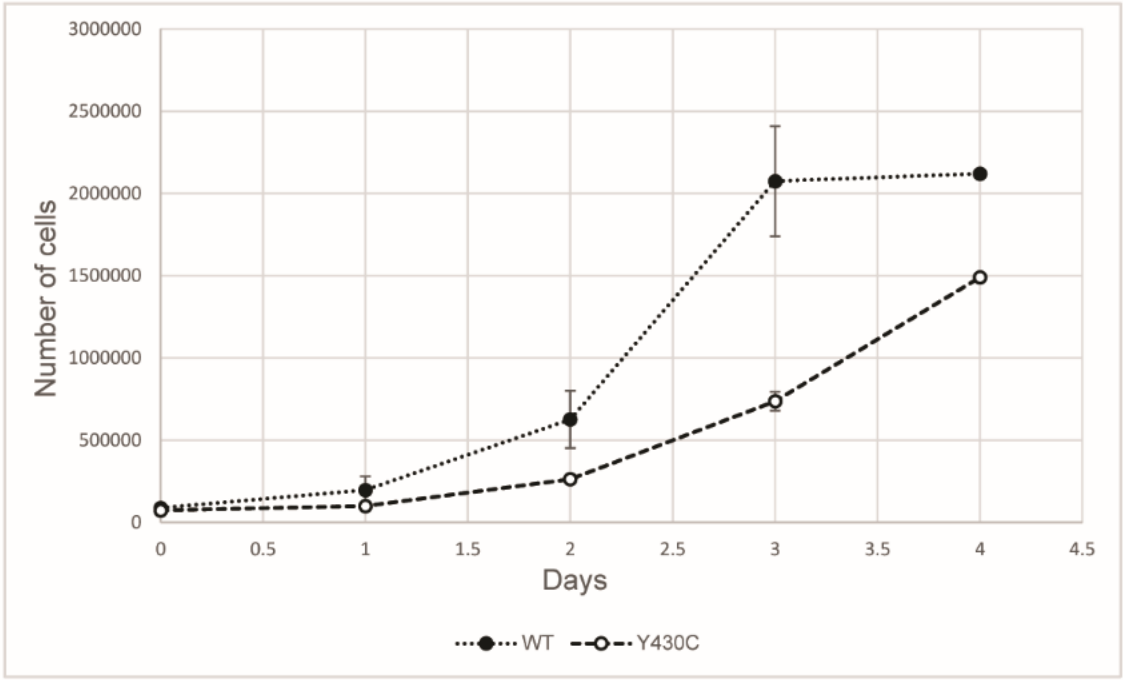


Figure Ch5-1. Y430C mESCs proliferate more slowly than WT mESCs. Graph shows average number of WT and Y430C cells per well at 1, 2, 3 and 4 days post seeding. Data is represented as mean +/- SEM from 3 technical replicates.

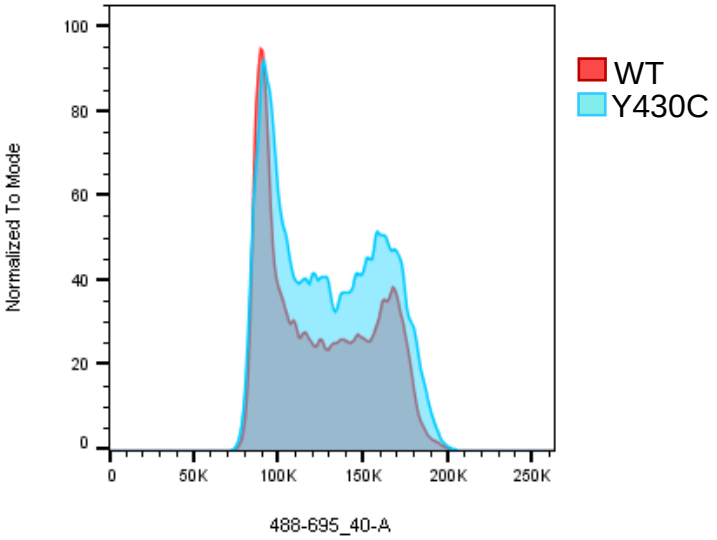


Figure Ch5-2. WT and Y430C cell cycle profiles determined by flow cytometry. Overlaid flow cytometry profiles for WT and Y430C cells. Graphs illustrate the cell count normalised to the mode, which correlates to propidium iodide intensity.

5.1.3 Comparing CHK1-P levels in WT and Y430C mESCs in response to DNA damage

Cell cycle checkpoints can be activated as a result of the DDR. Recognition of DNA damage by sensor proteins initiates a signalling cascade that results in the phosphorylation and activation of the checkpoint kinases CHK1 and CHK2. These kinases activate cell cycle checkpoints to delay progression through the cell cycle, preventing propagation of the damage to the next generation of cells (Bartek and Lukas, 2003).

CHK1 is the main checkpoint kinase required for activation of the G2/M (Zhao, Watkins and Piwnica-Worms, 2002), and intra-S cell cycle checkpoints, including in mESCs (Liu *et al.*, 2000). To determine whether the altered cell cycle in the Y430C mESCs was caused by increased G2/M checkpoint activation I used immunoblotting to analyse the phosphorylation of CHK1 in response to DNA damage in WT and Y430C mESCs.

Cells were treated with NCS, a radiomimetic drug which mainly induces DSBs (Povirk, 2012), and allowed to 'recover' for various lengths of time before nuclear protein extraction. Immunoblotting showed an increase in phosphorylated CHK1 (CHK1-P) in both WT and mutant ESCs 1 hour post NCS treatment, which is resolved by 16 hours. However, the levels

of CHK1-P are greater in the Y430C mESCs, suggesting an increase in checkpoint activation in the mutant cells (Figure Ch5-3A). Lamin B was used as a loading control; the similar band intensities across the different time-points indicates that a similar amount of protein has been loaded in each lane, meaning the differences in CHK1-P intensity are not a result of variable loading.

Increased CHK1-P supports the theory that the altered cell cycle in the Y430C mESCs is a result of a G2/M delay and/or prolonged S phase, indicating that there may be increased DNA damage signalling in the BRD4 mutant cells.

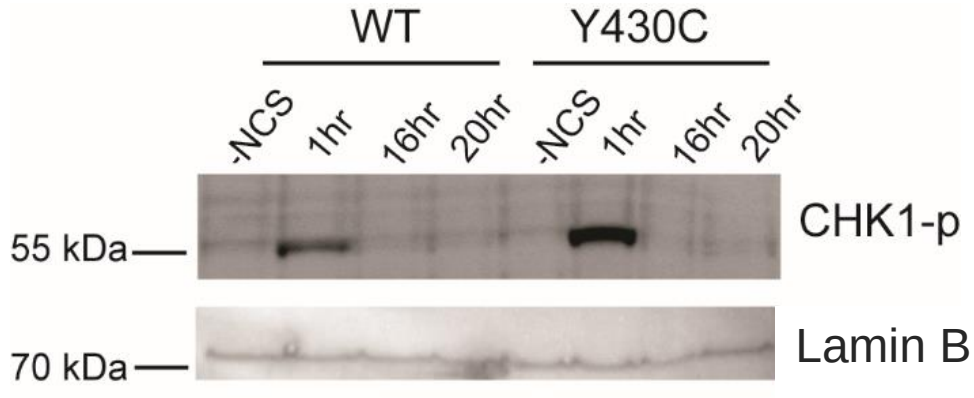


Figure Ch5-3. CHK1 phosphorylation in response to DSBs. Immunoblot for CHK1-P and Lamin B (control) in WT and Y430C mESCs after time-course of NCS treatment. Cells were treated with NCS and allowed to recover for 1, 16 and 20 hours. –NCS = untreated cells.

5.2 Analysing the DDR in WT and Y430C mESCs

5.2.1 The effect of NCS treatment on DDR signalling in WT and Y430C mESCs

The increased G2/M cell cycle checkpoint activation made me question whether Y430C mESCs were more prone to DNA damage than WT cells. BRD4 is known to regulate DNA damage signalling and repair through cell cycle checkpoint activation, recruitment of DDR proteins and chromatin remodelling (Floyd *et al.*, 2013; Pongas *et al.*, 2017; Li *et al.*, 2018; Zhang *et al.*, 2018). An increased sensitivity to DNA damage has been observed in some cell lines derived from CdLS patients, including those with NIPBL mutations. Increased sensitivity occurs especially in G2 and this has been suggested to be due to defective repair of DSBs by HR (Vrouwe *et al.*, 2007).

I therefore set out to compare the responses of both cell lines to DNA damage. Phosphorylation of the histone H2A variant H2AX (γH2AX) occurs early on in the DDR (Figure Ch1-22) and is often the preferred marker of DNA damage. However, mESCs are known to have high basal levels of γH2AX, even in the absence of genotoxic stress (Turinetto *et al.*, 2012). Indeed, through some preliminary experiments, I found that a large proportion of untreated WT mESCs contained >10 γH2AX foci. These γH2AX foci did not colocalise with 53BP1, a marker of DNA damage that functions downstream of γH2AX (Zimmermann and De Lange, 2014) (data not shown). This indicated that γH2AX was not an appropriate marker for this experiment and I instead chose to use 53BP1 to mark sites of DNA damage.

53BP1 forms large foci at DSBs (Zimmermann and De Lange, 2014) (Figure Ch1-23), and is important for the regulation of DNA repair pathway choice (Ochs *et al.*, 2016) and promotion of cell cycle checkpoint signalling (Ljungman, 2010). 53BP1 can be recruited to DSBs by BET proteins including BRD4 (Stanlie *et al.*, 2014; Li *et al.*, 2018).

I treated both WT and Y430C cells with NCS, allowed them to recover for various lengths of time, and then carried out IF for 53BP1 (Figure Ch5-4). Nuclei contained a small number of foci prior to induction of damage, and this increased dramatically 1 hour after NCS treatment, indicating the induction of a large number of DSBs. By 16 and 20 hours these foci

were reduced in number, likely due to the repair of these breaks and a decrease in DNA damage signalling.

Interestingly, there were some notable differences in 53BP1 foci size and kinetics between the WT and Y430C cells. Firstly, the 53BP1 foci in the Y430C cells appeared to be larger than those in WT cells (Figure Ch5-4). To determine whether this observation was quantifiably significant, I used a macro script in ImageJ to measure the area of the foci in each cell. This confirmed that the foci were larger in Y430C cells at all time points and was highly significant after NCS treatment (Figure Ch5-5A). Secondly, whilst the number of foci decreased from 1 hour to 16 hours in both cell lines, this decrease was smaller in the Y430C cells (Figure Ch5-5B); between 1 and 16 hours after NCS treatment the number of cells with > 5 foci decreased around 46-fold in WT cells, compared to just 4.6-fold in Y430C cells.

The greater size and number of persistent 53BP1 foci in the Y430C cells could be a result of increased DNA damage signalling; as the DDR progresses the associated factors and modifications spread from the original site of damage, and this is reflected in the size of the protein foci (Polo and Jackson, 2011). BRD4 is known to play a role in limiting this spread and to shield chromatin from ATM kinase signalling. After depletion of *Brd4* (RNAi knockdown), there is an increase in the number and size of 53BP1 (and γH2AX) irradiation induced foci (IRIF) (Floyd *et al.*, 2013). This is very similar to the observation I have made in Y430C-BRD4 mutant cells, and suggests that Y430 cells have a perturbed DDR.

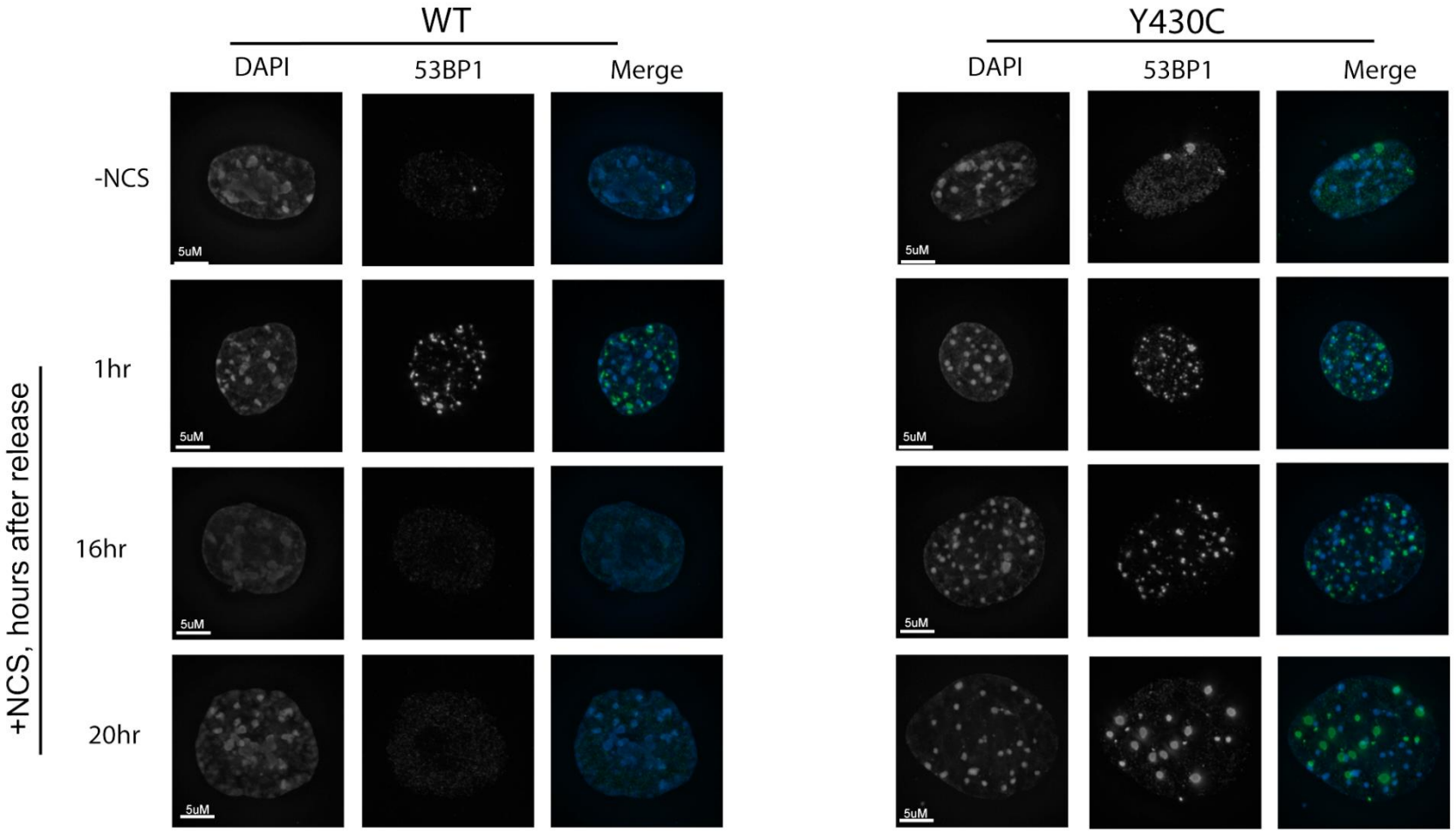


Figure Ch5-4. Imaging 53BP1 foci in WT and Y430C cells in response to DSB induction. Representative images of WT and Y430C mESCs upon 53BP1 IF and DAPI staining without treatment (-NCS) and at 1, 16 and 20 hours post-NCS treatment (25 ng/ml). Scale bar = 5 μm

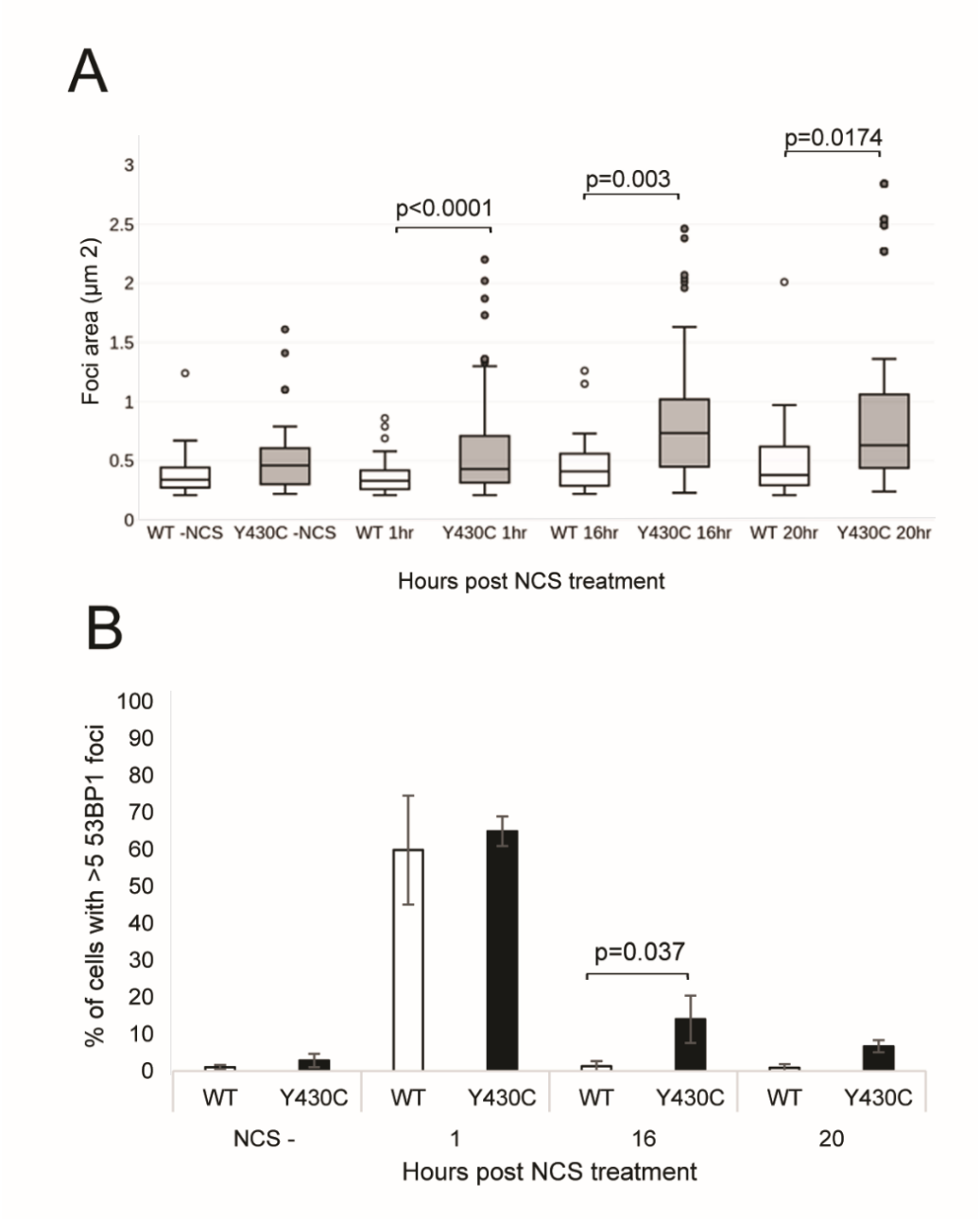


Figure Ch5-5. Counting and measuring 53BP1 foci in WT and Y430C cells in response to DSB induction. A) Box-plot shows area of 53BP1 foci (μm^2) in WT and Y430C cells after treatment with NCS and release for the indicated times. –NCS = untreated cells. Horizontal lines within boxes show medians, boxes are inter-quartile ranges and whiskers are range. Outliers are shown as circles. P-values were calculated with t-test. **B)** Graph shows percentage of cells that contain > 5 53BP1 foci after release from NCS treatment for given time points. NCS- = untreated cells. Data are represented as mean +/- SEM from 3 biological replicates. P-values calculated with t-test.

5.2.2 The effect of BRD4 inhibition on DNA damage response signalling in WT and Y430C mESCs

If the increased size and number of 53BP1 foci in Y430C cells after treatment with DNA damage inducing agents is indeed attributable to the reduced binding of the mutant BRD4 to chromatin, then inhibition of BRD4 binding by JQ1 should cause a similar phenotype. To test this, I treated WT and Y430C cells with (+)-JQ1 or (-)-JQ1 before repeating the NCS treatments, followed by 53BP1 IF.

Treatment with (+)-JQ1 caused increased numbers of 53BP1 foci in both WT and Y430C cells compared to (-)-JQ1, both before and after treatment with NCS (Figure Ch5-6). This is in contrast to the previous experiment where, in the absence of any JQ1, significant differences between the number of foci in WT and Y430C cells were only seen after NCS treatment (Figure Ch5-5). This is not surprising as I have shown by ChIP that binding of BRD4 is more strongly perturbed by (+)-JQ1 than by the reduced acetylated histone affinity of the Y430C mutation (Figure Ch4-10). Furthermore, RT-qPCR showed that (+)-JQ1 treatment causes dysregulation of BRD4 target gene transcription (Figure Ch4-11A) which I have shown does not occur in the Y430C mESCs (Figure Ch411B). The significant difference in the number of cells with >5 53BP1 foci between the mutant and WT cells was lost with (+)-JQ1 treatment.

I conclude that the increased DDR phenotype observed in the Y430C cells could indeed be explained by the impaired binding of BRD4 to chromatin.

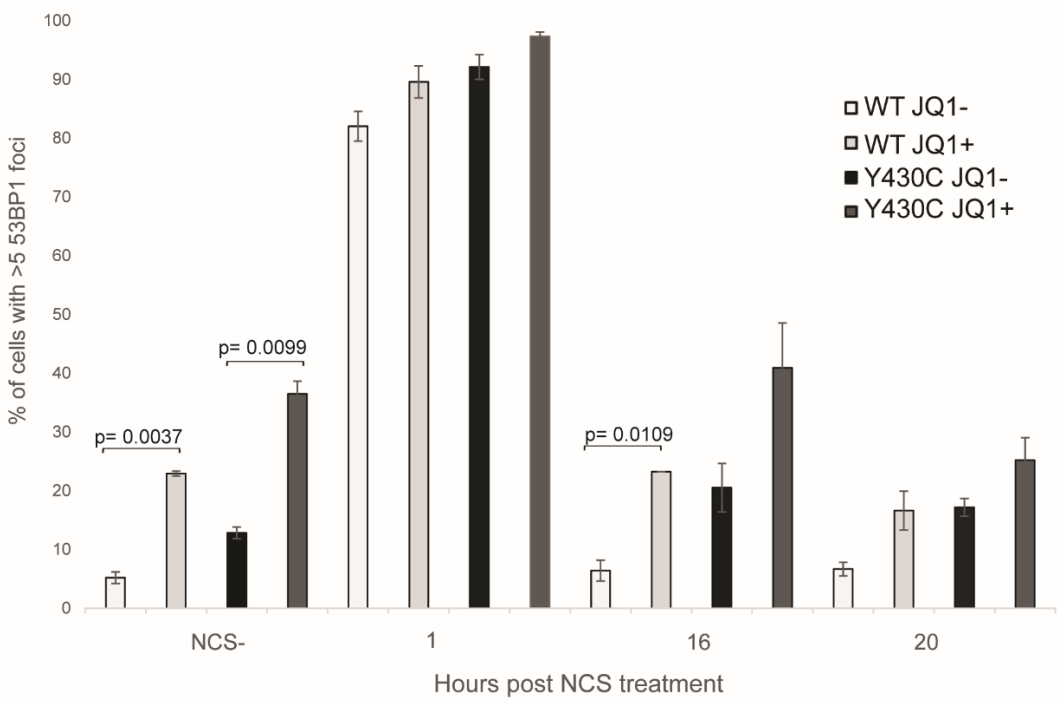


Figure Ch5-6. Percentage of cells with >5 53BP1 foci after NCS treatment, with or with JQ1+/-. Graph shows percentage of cells that contain > 5 53BP1 foci after release from NCS treatment for given time points. NCS- = untreated cells. Data is represented as mean +/- SEM from 3 biological replicates. P-values calculated with t-test. JQ1- = (-)-JQ1, JQ1+ = (+)-JQ1.

5.3 Investigating potential causes of increased 53BP1 foci in Y430C mESCs

The increase in the number and size of persistent 53BP1 foci after DNA damage induction, associated with the reduced BRD4 binding to chromatin in Y430C cells, suggests that the role of BRD4 in inhibiting the spread of DNA damage signalling (Li *et al.*, 2018) could be perturbed in the Y430C cells. I hypothesise that the decrease in affinity of BRD4 to acetyl-lysines causes decreased recruitment of BRD4 to DSBs, resulting in the increased spread of DNA damage signalling. However, there are other potential explanations for this phenotype, which I will discuss in more detail.

5.3.1 Sensitivity to replication stress?

Activation of the G2/M and/or intra-S checkpoints can also occur in response to replication stress; the presence of stalled replication forks induces the phosphorylation of CHK1, leading to the phosphorylation and subsequent degradation of CDC25A (Zhao, Watkins and Piwnica-Worms, 2002), which is required for the progression through the cell cycle (Timofeev *et al.*, 2010). I therefore thought it was possible that the increased DDR phenotype in the Y430C mESCs could be caused by an increased sensitivity to replication stress.

5.3.1.1 The effect of HU treatment on DDR signalling in WT and Y430C mESCs

I tested the response of WT and Y430C mESCs to DNA damage induced by HU, to determine if damage induced by replication stress would result in a greater increase in 53BP1 foci number than damage induced by NCS.

HU is an inhibitor of ribonucleotide reductase, the enzyme responsible for the reduction of ribonucleoside diphosphates to deoxyribonucleotides, required for DNA replication and repair. Consequently, HU treatment reversibly inhibits DNA synthesis, resulting in replication stress – the stalling of replication forks and DNA damage (Singh and Xu, 2016). I treated WT and Y430C cells with 50 mM HU and left them to recover for 1, 16 and 20 hours, before carrying out IF for 53BP1. The overall trend was similar to that seen with NCS treatment (Figure Ch5-7).

If the Y430C cells were more sensitive to DNA damage caused by replication stress, I would expect to see a greater difference between the WT and Y430C cells when treated with HU than with NCS. However, the difference in the size of 53BP1 foci between WT and Y430C cells wasn't as large after HU treatment as it was for NCS. Whilst a slight increase in foci area was observed in Y430C cells at all time points, this was only significantly higher than the WT cells at 1 hour after HU treatment (Figure Ch5-7A).

As for NCS, I found that there were significantly more cells with >5 and >10 foci 20 hours after HU treatment in the mutant Y430C cells compared to WT (Figure Ch5-7B). Furthermore, the Y430C cells were still slower to recover from the treatment (return towards the levels of 53BP1 foci seen before HU treatment). The magnitude of the difference between WT and Y430C cells in response to HU however, was not strikingly larger than that seen with NCS treatment, suggesting that the mutant cells are not specifically sensitive to replication related DNA damage.

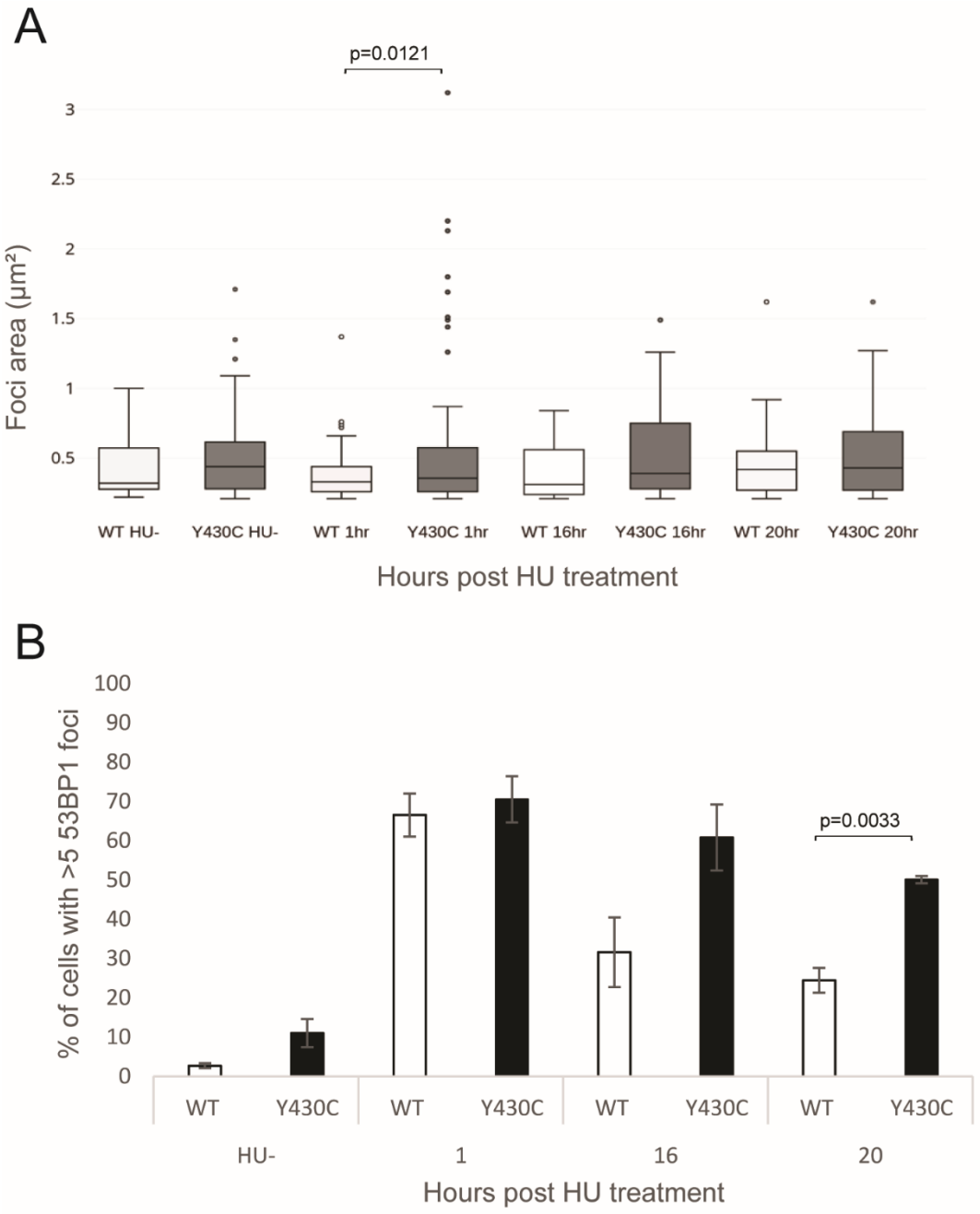


Figure Ch5-7. Counting and measuring 53BP1 foci in WT and Y430C cells in response to replication stress. A) Box-plot show area of 53BP1 foci (µm^2) in WT and Y430C cells after treatment with HU and release for the indicated times. –HU = untreated cells. Horizontal lines within boxes show medians, boxes are inter-quartile ranges and whiskers are range. Outliers are shown as circles. P-values were calculated with t-test. **B)** Graph shows percentage of cells that contain greater than 5 53BP1 foci after release from HU treatment for given time points. HU- = untreated cells. Data is represented as mean +/- SEM from 3 biological replicates. P-values calculated with t-test.

5.3.1.2 Do 53BP1 foci represent OPT domains?

Another explanation for the presence of persistent 53BP1 foci, is that they represent OPT (OCT1/PTF/transcription) domains – nuclear foci that arise in G1 but disappear in S phase, which were originally thought to represent sites of active transcription (Pombo *et al.*, 1998). 53BP1, along with γH2AX and MDC1, co-localise with these OPT domains (Harrigan *et al.*, 2011). These sites are enriched at common fragile sites, which often arise after incomplete DNA replication, and their incidence increases in response to replication stress. This suggests that these domains form in G1 as a consequence of incomplete DNA synthesis in the previous S phase and the induction of the DDR (Harrigan *et al.*, 2011).

If the 53BP1 foci that persist in Y430C mESCs, but not in WT cells, are in fact OPT domains, we would expect to see a large increase in the number of foci in the Y430C cells in response to replication stress. However, I found that incubating the cells with a low dose of aphidicolin (APH) actually had a larger effect on the WT cells than the Y430C cells. In WT cells the number of cells with >1, 5 or 10 foci increased significantly upon APH treatment. In Y430C cells, there was an increase in the number of foci but this was not statistically significant (Figure Ch5-8A). Furthermore, IF for OCT1 in the WT and Y430C mESCs showed that the persistent 53BP1 foci do not co-localise with OCT1 (Figure Ch5-8B).

Together these data suggest that the persistent 53BP1 foci I see in the Y430C cells by IF are not caused by an increased sensitivity to replication stress.

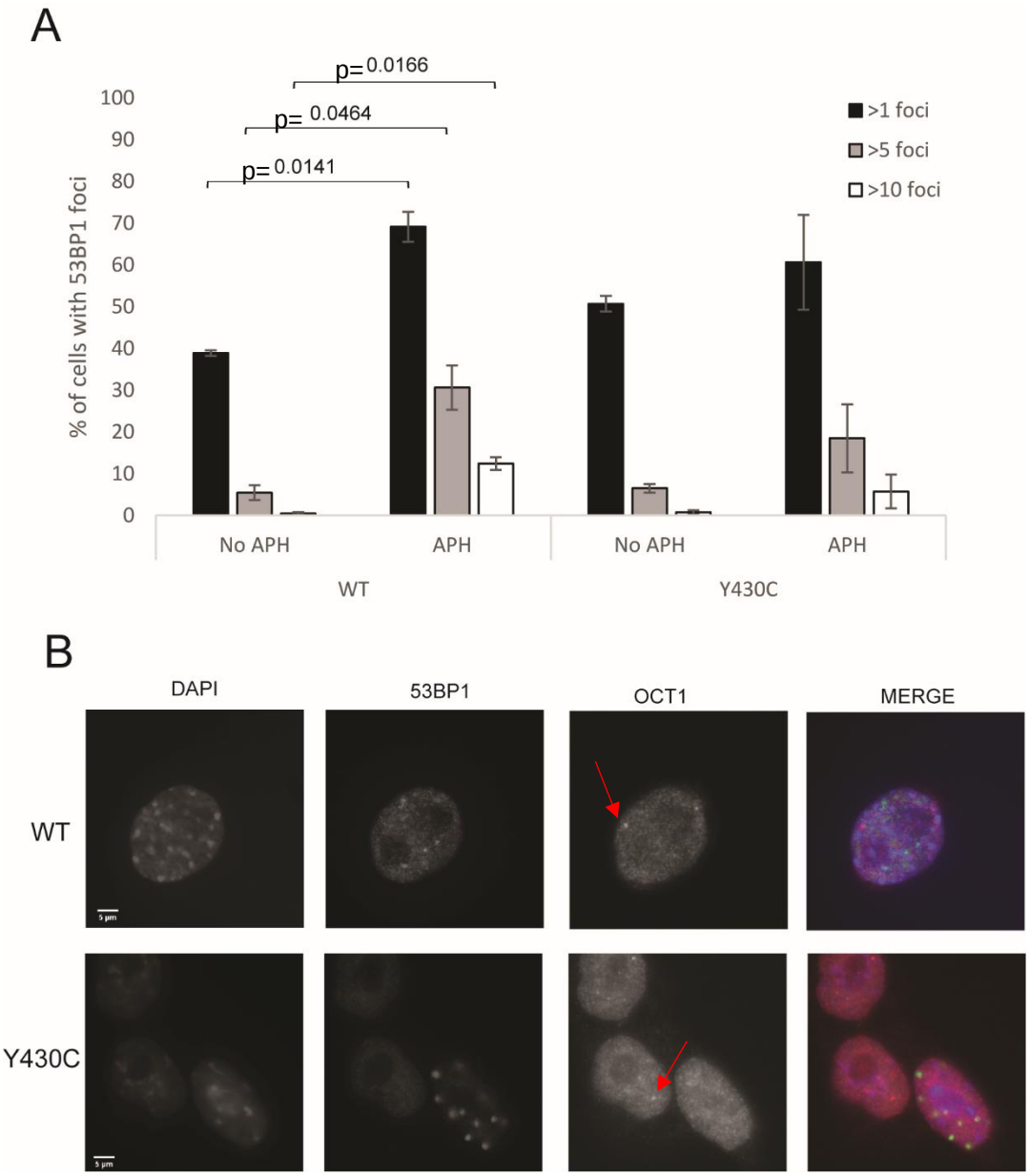


Figure Ch5-8. Response of WT and Y430C mESCs to replication stress. A) Graph shows percentage of cells that contain greater than 1, 5 or 10 53BP1 foci with or without APH treatment. Data is represented as mean +/- SEM from 3 biological replicates. P-values calculated with t-test. **B)** Representative images of WT and Y430C mESCs upon DAPI staining and 53BP1 and OCT1 IF 16 hours post NCS treatment. Red arrows point to OCT1 foci. Scale bar = 5 μm

5.3.2 Impaired DNA repair?

The increase in foci number in the Y430C cells compared to WT is most significant at 16 and 20 hours post NCS treatment (Figure Ch5-5B). This could indicate that there is a problem with DNA repair in the *Brd4* mutant cells, causing 53BP1 to remain at the sites of damage for longer.

5.3.2.1 Regulation of transcription of DNA repair genes

BRD4 is directly involved in DNA repair through the transcriptional regulation of genes encoding DNA repair proteins (C. Sun *et al.*, 2018; Li *et al.*, 2018; Wilson *et al.*, 2018). If the Y430C mutation were to affect the transcription of these genes, it could cause the increased 53BP1 foci that I see in these cells. I therefore returned to my 4sU-seq and RNA-seq results to check for any differences in the transcription of DDR genes.

Others have defined sets of genes that are mis-regulated when DNA damage repair pathways are perturbed; Peng *et al.*, 2014 used genome-wide expression profiling to measure the cellular transcriptome reprogramming in HR-deficient (HRD) cells (deficient in *BRCA1, RAD51* or *MCPH1*). Through microarray analysis of these cells they defined a HRD gene signature – a set of 230 genes whose expression differed by a factor of 2 or more. Li *et al.*, 2018 found a set of NHEJ related genes to be differentially expressed after treatment of cells with JQ1.

I compared the WT and Y430C profiles for these gene sets. Visualising the data on the UCSC genome browser, I did not observe any changes in transcription for both the 4sU-seq and RNA-seq datasets (Figure Ch5-9&10**)** (subset of genes shown). Plotting expression of exons in WT against Y430C also shows that these genes are expressed at a similar level in both cell lines (Figure Ch5-11). This suggests that transcriptional dysregulation of DNA repair genes is not the cause of the aberrant DDR signalling in the Y430C mESCs.

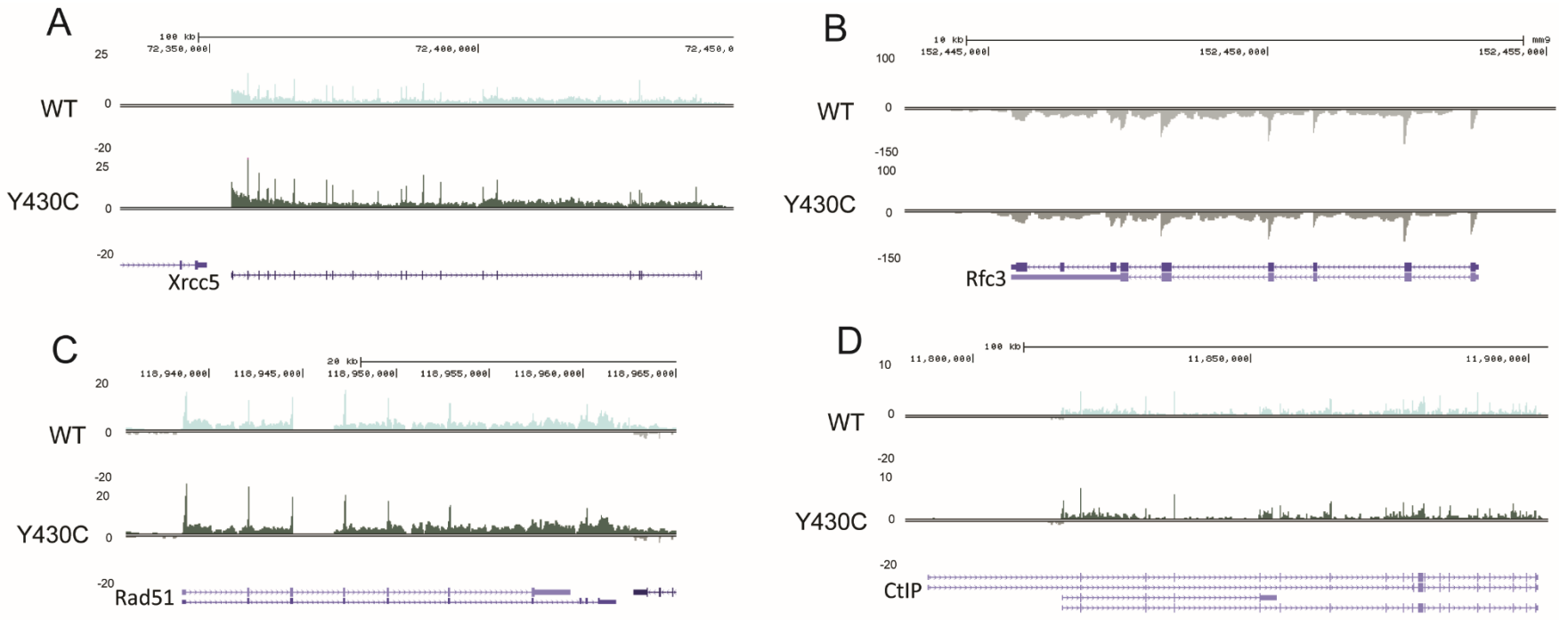


Figure Ch5-9. Transcription of DNA damage related genes is similar in WT and Y430C mESCs. UCSC browser screenshot showing RP10M of WT and Y430C 4sU-seq over the transcribed regions of **A)** *Xrcc5*, **B)** *Rfc3*, **C)** *Rad51* and **D)** *CtIP*.

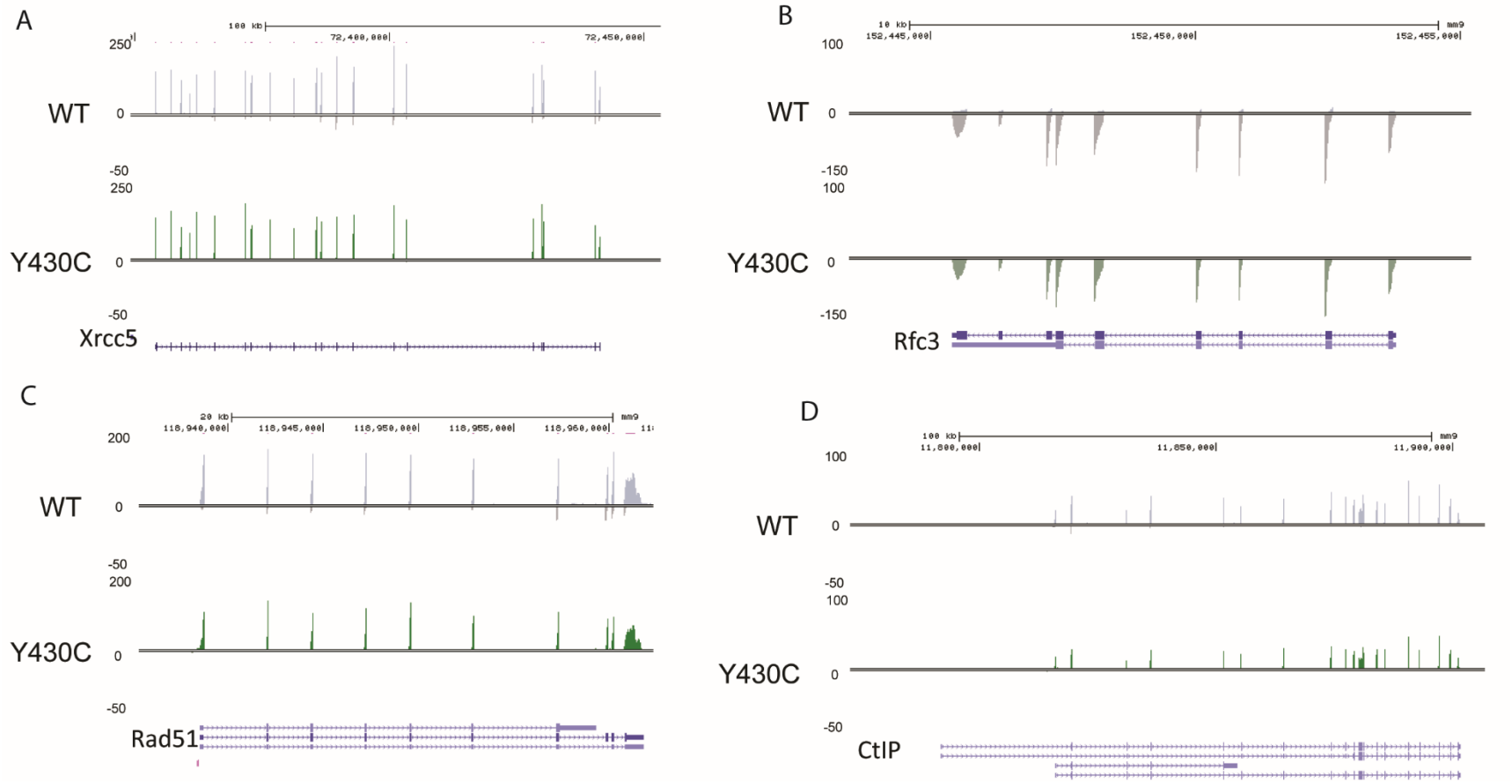


Figure Ch5-10. Transcription of DNA damage related genes is similar in WT and Y430C mESCs. UCSC browser screenshot showing RP10M of WT and Y430C RNA-seq normalised to spike-in control over the transcribed regions of **A)** *Xrcc5,* **B)** *Rfc3,* **C)** *Rad51* and **D)** *CtIP.*

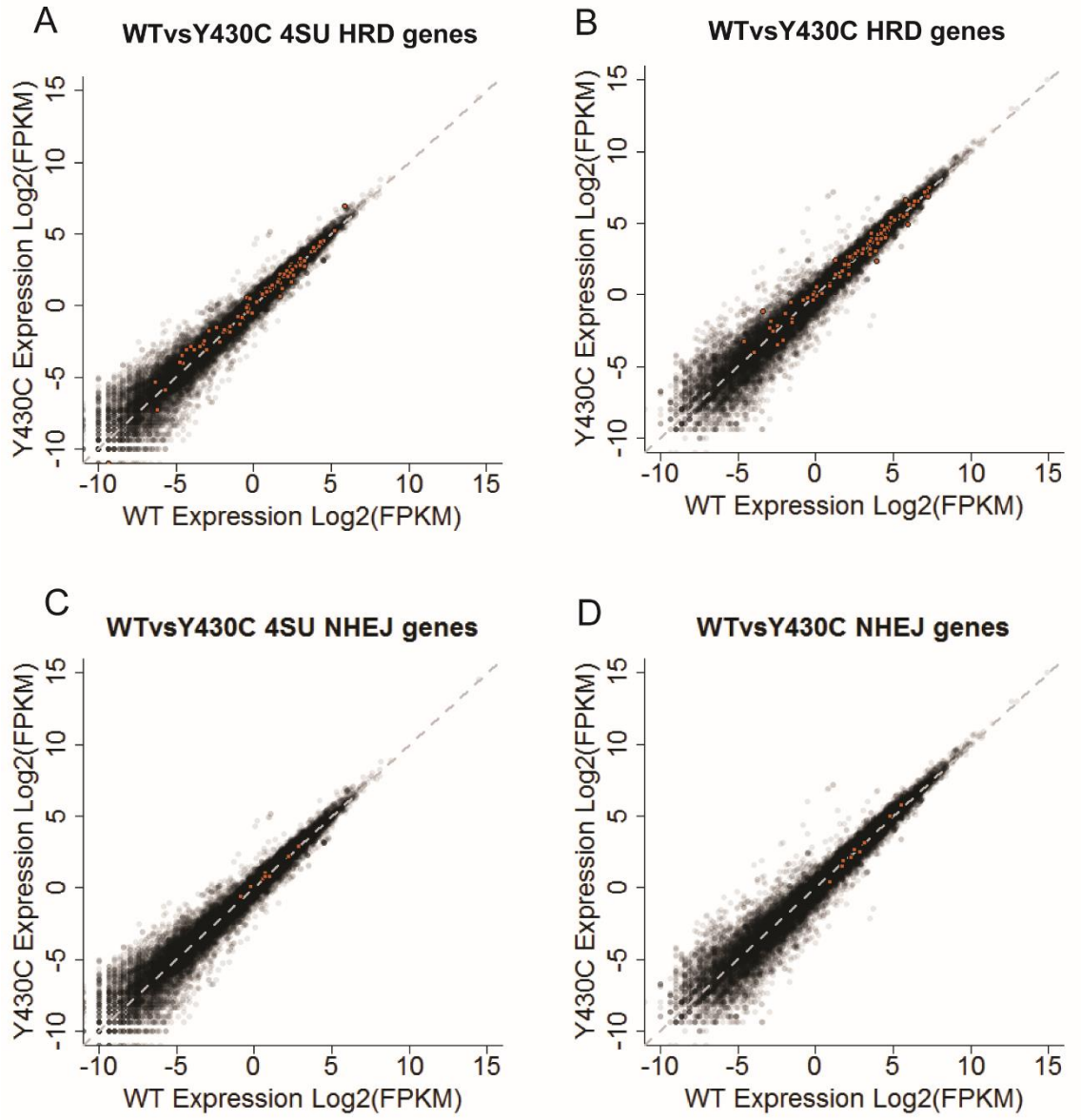


Figure Ch5-11. Expression of DDR genes in WT vs Y430C mESCs. Log2 (FPKM) of WT exons against Y430C exons in **A&C)** 4sU data and **B&D)** RNA-seq data. HDR genes (**A&B)** and NHEJ genes **(C&D)** are highlighted in orange. Values are the average of 2 (4sU) or 3 (RNA) biological replicates. Grey dotted line = regression line

5.3.2.2 RAD51 binding at DSBs

The absence of differences between DDR gene transcription in WT and Y430C cells does not rule out a problem with DNA repair itself; BRD4 may also regulate the chromatin environment around a DSB, altering accessibility to DDR proteins (Floyd *et al.*, 2013; Pongas *et al.*, 2017), and recruit and stabilise proteins of the DNA damage repair complex (Li *et al.*, 2018).

RAD51 is a DDR protein that binds ssDNA after end-resection and is important for the regulation of HR at many steps (section 1.5.4.2)(Wright, Shah and Heyer, 2018b). To compare HR in the WT and Y430C mESCs, I induced damage with NCS and used IF to visualise RAD51 foci (Figure Ch5-12). Compared to WT, significantly more Y430C mESCs had RAD51 foci prior to NCS treatment and at 16 and 20 hours post treatment (Figure Ch5-12&13). This corresponds with the results of the 53BP1 foci experiments (Figure Ch5-4&5) and supports the idea that Y430C cells have higher levels of endogenous breaks, and more persistent breaks, in response to induced damage. In contrast to 53BP1 focus formation however, considerably more WT cells had RAD51 foci than the Y430C cells at 1 hour post NCS treatment. This may indicate a slower rate of RAD51 recruitment to DSBs in the mutant cells, which could explain the increase in persistent foci after damage - if RAD51 is recruited more slowly, this could result in slower repair of the breaks. These results are consistent with the findings of Wilson *et al.*, 2018, who found that BET protein inhibition caused a reduction in the number of cells expressing RAD51 foci and a reduction in HR repair efficiency.

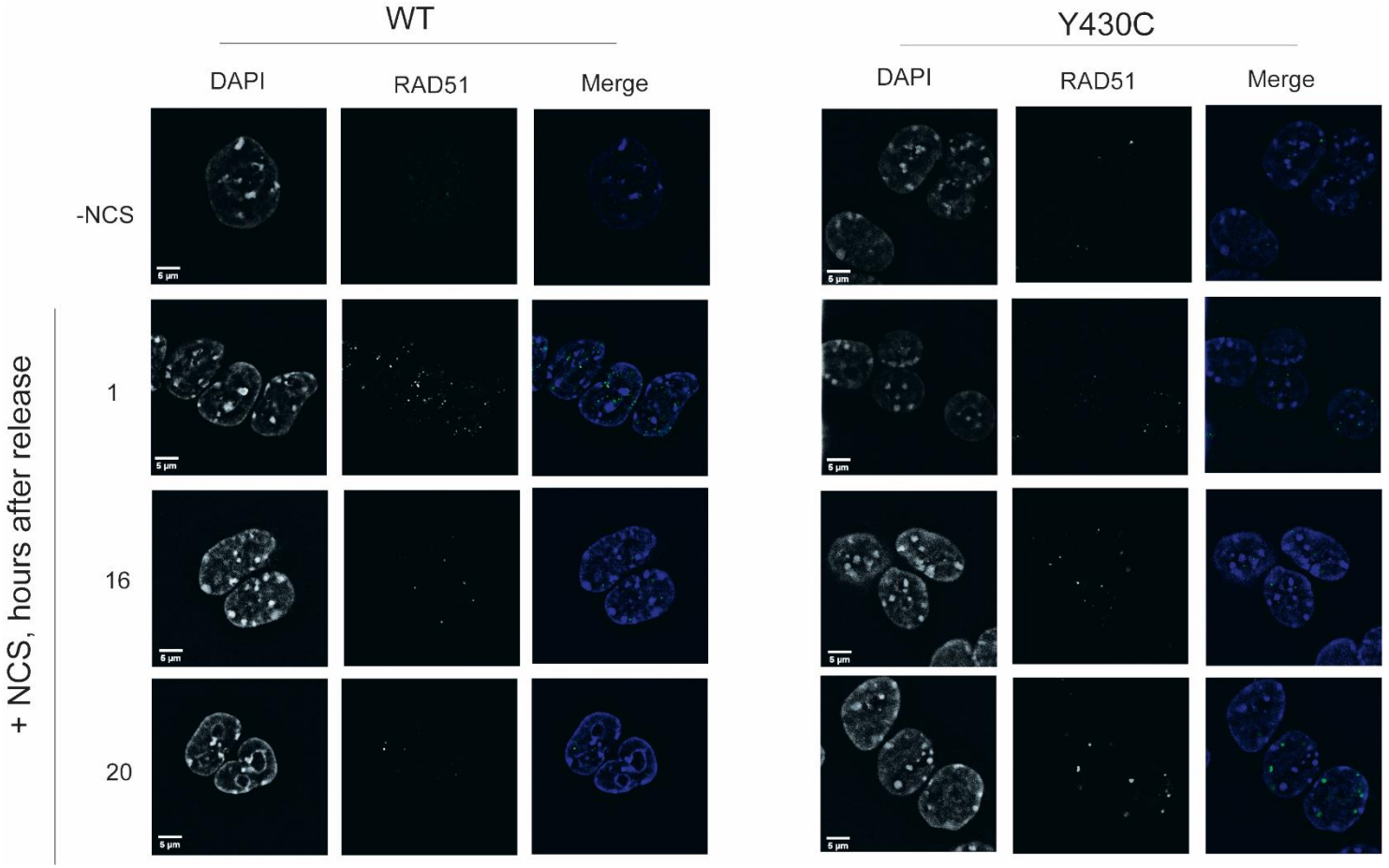


Figure Ch5-12. Imaging RAD51 foci in WT and Y430C cells in response to DSB induction. Representative images of WT and Y430C mESCs upon RAD51 IF and DAPI staining without treatment (-NCS) and at 1, 16 and 20 hours post NCS treatment. Scale bar = 5 µm

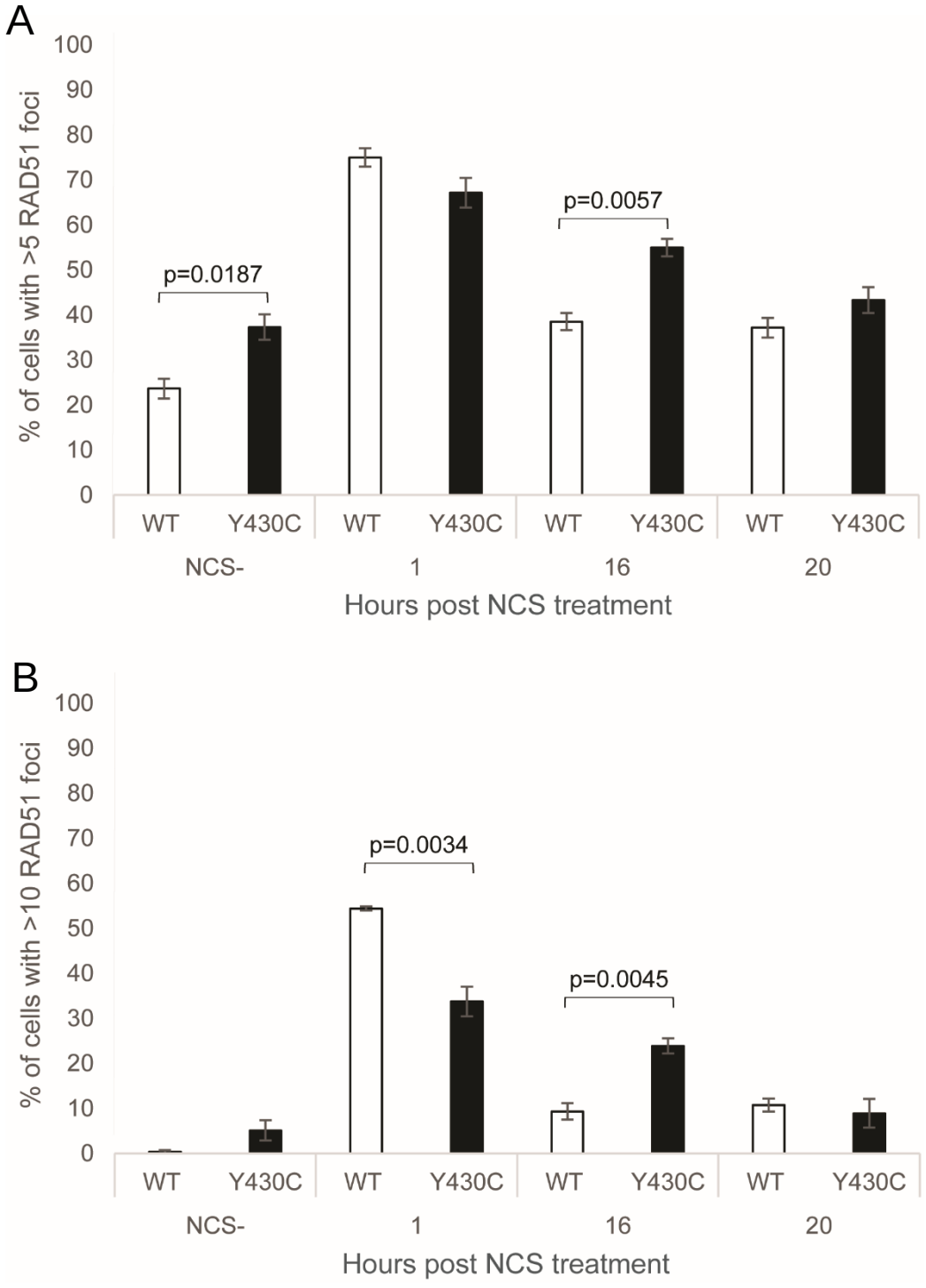


Figure Ch5-13. Percentages of cells with RAD51 foci in WT and Y430C cells in response to DSB induction. Graphs show percentage of cells that contain **A)** >5 and **B)** >10 RAD51 foci after release from NCS treatment for given time points. NCS- = untreated cells. Data is represented as mean +/- SEM from 3 biological replicates. P-

5.3.3 eRNA transcription?

It has been suggested that BRD4 inhibits bidirectional transcription from enhancers (section 4.4) (Flynn *et al.*, 2016). Such bidirectional transcription can result in convT – simultaneous transcription of a region on the sense and anti-sense strands -, particularly at SEs, where transcription start sites are highly clustered. This can lead to DNA damage through RNAPII collisions and activation-induced cytidine deamination (AID) (Meng *et al.*, 2014). Therefore the loss of BRD4 from enhancers could result in increased damage mediated by convT.

Through 4sU-seq, I have identified eRNAs at TEs and SEs. There are instances where these enhancers appear to be transcribed on both strands at the same time (Figure Ch4-24&25), suggesting convT. However, without carrying out single-cell analysis, there is no way of knowing whether this is occurring in one cell – ie. convT – or in different cells. Furthermore, whilst these eRNAs were present, I was unable to detect a difference in levels between WT and Y430C cells.

To test whether transcription was contributing to the increased number of 53BP1 foci in Y430C mESCs, I treated WT and Y430C cells with 500 nM Triptolide (Trp) for 2 hours, and then carried out IF for 53BP1. Trp inhibits TFIIH to prevent transcription initiation (Chen *et al.*, 2015). Trp treatment caused a significant decrease in the number of Y430C cells with >1 and >5 foci compared to the DMSO treated control cells (Figure Ch5-14). A slight decrease was observed in the WT cells after Trp treatment, but this was not significant. This suggests that transcription may be a cause of endogenous DNA damage in the Y430C mESCs.

Trp is not specific to eRNAs however - it causes a global inhibition of transcription. Therefore, it is impossible to say from this data whether the decrease in 53BP1 foci in Y430C cells upon Trp treatment is caused by decreased eRNA transcription, or from decreased transcription in general. Furthermore, since Trp inhibits transcription initiation, it may not be the best inhibitor to use. An inhibitor of transcriptional elongation, such as DRB or flavopiridol (Bensaude, 2011), may be more appropriate, as it is during the elongation of eRNAs that RNAPII collisions would occur.

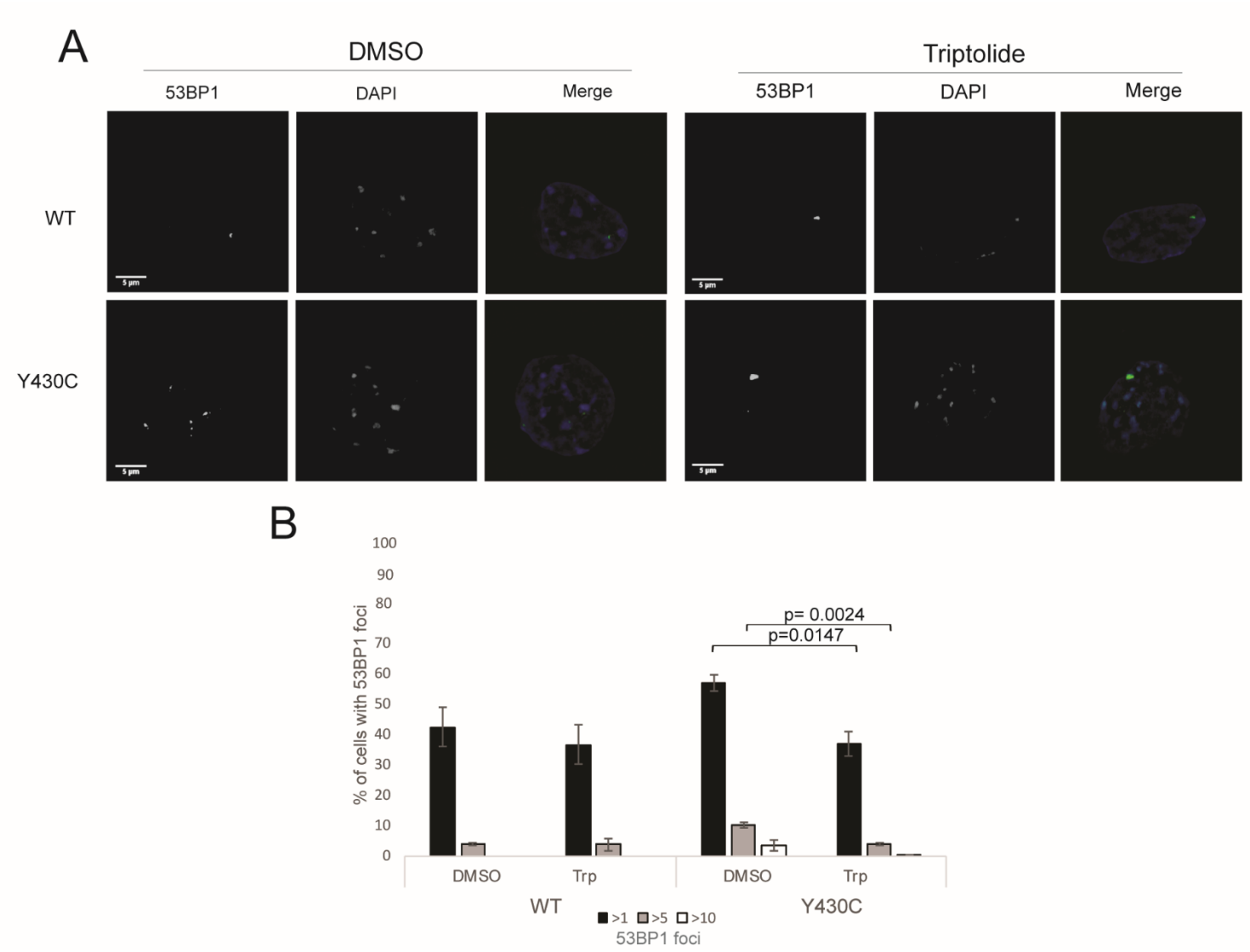

A
DMSO
Triptolide
53BP1
DAPI
Merge
53BP1
DAPI
Merge
WT
Y430C
5 μm
B
% of cells with 53BP1 foci
100
90
80
70
60
50
40
30
20
10
0
p= 0.0024
p=0.0147
DMSO
Trp
DMSO
Trp
WT
Y430C
>1
>5
>10
53BP1 foci

Figure Ch5-14. Decreased number of 53BP1 foci after transcriptional inhibition in Y430C mESCs. A) Representative images of WT and Y430C mESCs upon 53BP1IF and DAPI staining, after treatment with DMSO control or 500 nM triptolide (Trp) for 2 hours. **B)** Graph shows percentage of cells that contain >1, 5 or 10 53BP1 foci in WT and Y430C mESCs after treatment with DMSO control or 500 nM Trp for 2 hours. Data is represented as mean +/- SEM from 3 biological replicates. P-values were calculated with t-test.

5.4 Conclusions

I have shown, through flow cytometry, that there is a difference between the cell cycles of WT and Y430C mESCs. The current gating suggests that a greater proportion of Y430C cells are in the G2/M phase of the cell cycle than in WT cells, which could be due to increased activation of the G2/M checkpoint. The categorisation of cells into cell cycle stages may need to be altered however, and it is possible that there are actually more Y430C cells in S phase than in WT cells, suggesting increased activation of the intra-S phase checkpoint. Either way, the flow cytometry results are supported by increased CHK1-P upon DNA damage in the Y430C cells compared to WT, and delay at either checkpoint would result in slower progression through the cell cycle.

Cell cycle checkpoints are activated in response to replication stress and/or DNA damage (Bartek and Lukas, 2003). Induction of DSBs with NCS, and through replication fork stalling with HU, caused DDR signalling in WT and Y430C mESCs. After removal of the drugs the signalling decreased over time, but this decrease was less pronounced in the mutant cells. This shows that Y430C cells exhibit increased DDR signalling in response to ionising radiation and replication induced damage. Inhibition of BRD4 binding by JQ1 had a similar effect, increasing the number of foci in WT cells to greater than that seen in Y430C cells. Therefore, I conclude that the DDR phenotype in the Y430C mutant cells is a result of the decreased affinity of BRD4-Y430C to chromatin.

Stalling of the cell cycle during S phase in the Y430C cells would suggest a problem with DNA replication. However, whilst APH treatment causes a significant increase in the number of WT cells with >5 53BP1 foci, there is little change in Y430C cells. Furthermore the 53BP1 foci in these cells do not co-localise with replication stress-induced OPT domains, and it has been reported that the loss of BRD4 binding (through BET inhibition by (+)-JQ1)

causes decreased CHK1 phosphorylation and intra-S phase checkpoint signalling in cells undergoing replication stress (Zhang *et al.*, 2018), which does not fit with the increased CHK1-P or the slower cell cycle that I have found in the Y430C cells. Together these results suggest that the increased DDR signalling in Y430C cells is not caused by an increased sensitivity to replication stress.

The Y430C mutation does not appear to affect the transcription of DDR proteins, despite it being shown by others that BRD4 regulates the expression of these genes (C. Sun *et al.*, 2018; Li *et al.*, 2018; Wilson *et al.*, 2018). However, RAD51 IF showed a decrease in foci formation shortly after DSB induction, and an increase in persistent RAD51 foci at later time points, suggesting delayed RAD51 recruitment to DSBs in Y430C mESCs and impaired HR.

Together these results lead me to hypothesise that the Y430C mutation reduces BRD4 recruitment to DSBs (through reduced affinity for acetyl-lysine), and this affects the DDR in two ways: 1. Reduced inhibition of DDR protein spreading, leading to larger 53BP1 foci. 2. Delayed repair through HR. This slower rate of HR could be a result of a reduction in RAD51 recruitment by BRD4, or it could be that the loss of BRD4 affects HR upstream of RAD51, decreasing ssDNA for RAD51 to bind.

BRD4 is thought to supress eRNA transcription to protect against convT and the subsequent DNA damage that this can cause (Flynn *et al.*, 2016). I have found no difference in levels of eRNA between WT and Y430C mESCs, suggesting that the increased number and size of 53BP1 foci in Y430C cells is not caused by an increase in convT at enhancers. However, inhibition of transcription with Trp did cause a significant decrease in endogenous 53BP1 foci in Y430C cells. As eRNAs are seen in both WT and Y430C cells, it is possible that their transcription is a contributing factor to DSBs in both cell lines but that the increase in DDR signalling and reduced repair in Y430C cells renders these cells more sensitive to eRNA-mediated damage. This would explain why the inhibition of transcription had a larger effect on the Y430C cells than on the WT cells.

Chapter 6: Genomic occupancy of 53BP1 and NIPBL in mESCs

6.1 Exploring 53BP1 occupancy by ChIP-seq

In Chapter 4, I showed that transcription occurs at enhancer elements (Figure Ch4-24&25), producing eRNAs, and that this is enriched at enhancers occupied by BRD4 (Figure Ch4-26). BRD4 has been proposed to protect against DNA damage at these regions, by recruiting a complex of 7SK-BAF that inhibits excessive eRNA transcription (Flynn *et al.*, 2016).

In Chapter 5, I showed that Y430C mESCs have larger and more persistent 53BP1 foci in response to DNA damage, suggesting an increase in DNA damage signalling. If these foci are caused by convT at enhancers, I would expect to see 53BP1 localised at these regions of the genome. To determine the binding pattern of 53BP1 across the genome, I carried out ChIP-seq in the WT and Y430C mESCs using an antibody against 53BP1.

This data is based on only one ChIP-seq replicate and so needs to be replicated and validated by ChIP-qPCR before any final conclusions can be drawn.

6.1.1 Quality control

The 53BP1 antibody that I used for IF (Novus; 100-304) in Chapter 5 has not been validated for ChIP-seq, and so I chose to use an alternative from the same manufacturer (Novus; 100-904). Novus; 100-904 is a combination of Novus; 100-304 and another antibody (Novus; 100-504) and has been previously used for ChIP by others.

I performed 53BP1 ChIP using chromatin from WT and Y430C mESCs and made libraries for paired-end sequencing from the ChIP DNA, using their corresponding input samples as controls.

FASTQC analysis showed that the sequencing results were suitable for further downstream applications. The per base sequence quality for this ChIP-seq was found to be acceptable for all samples (Figure Ch6-1), with scores generally above 28, suggesting a high accuracy of base calling. However, I did see a dip in the quality of the last base for all of the 53BP1 ChIP-seq samples. This is a common problem, attributable to the way the Illumina sequencing platform works. Illumina sequencing uses sequencing by synthesis. In this method all the sequences in a run are sequenced simultaneously in ensembles of identical molecules called

clusters. Sequencing is carried out by running a chemistry cycle, in which a tagged base is added to the end of each cluster, generating a single output sequence. The assumption is that every molecule is extended by one base. However, in reality, a small number of molecules will remain on the previous base, meaning that in the next cycle the signal coming from one cluster is actually a mix of signals from the current and previous base. The heterogeneity of the signal increases with each cycle, resulting in a decrease of per base sequence quality with increasing read length (Cliften, 2014).

The samples showed no evidence of adaptor content or overrepresented sequences, and whilst there was low level sequence duplication, this was not enough to cause the samples to fail the QC.

6.1.2 Mapping

I mapped the sequenced DNA to the mm9 assembly of the mouse genome using Bowtie2 (Langmead and Salzberg, 2012), and used Samtools (Li *et al.*, 2009) flagstat to calculate the total number of reads, the number of reads correctly paired, and the number of reads successfully mapped (Table Ch6-1). This showed that all samples had over 60 million reads, 100% of which had corresponding pairs for all samples. The percentage of mapped reads was high for all samples, ranging from 87.37-97.87%.

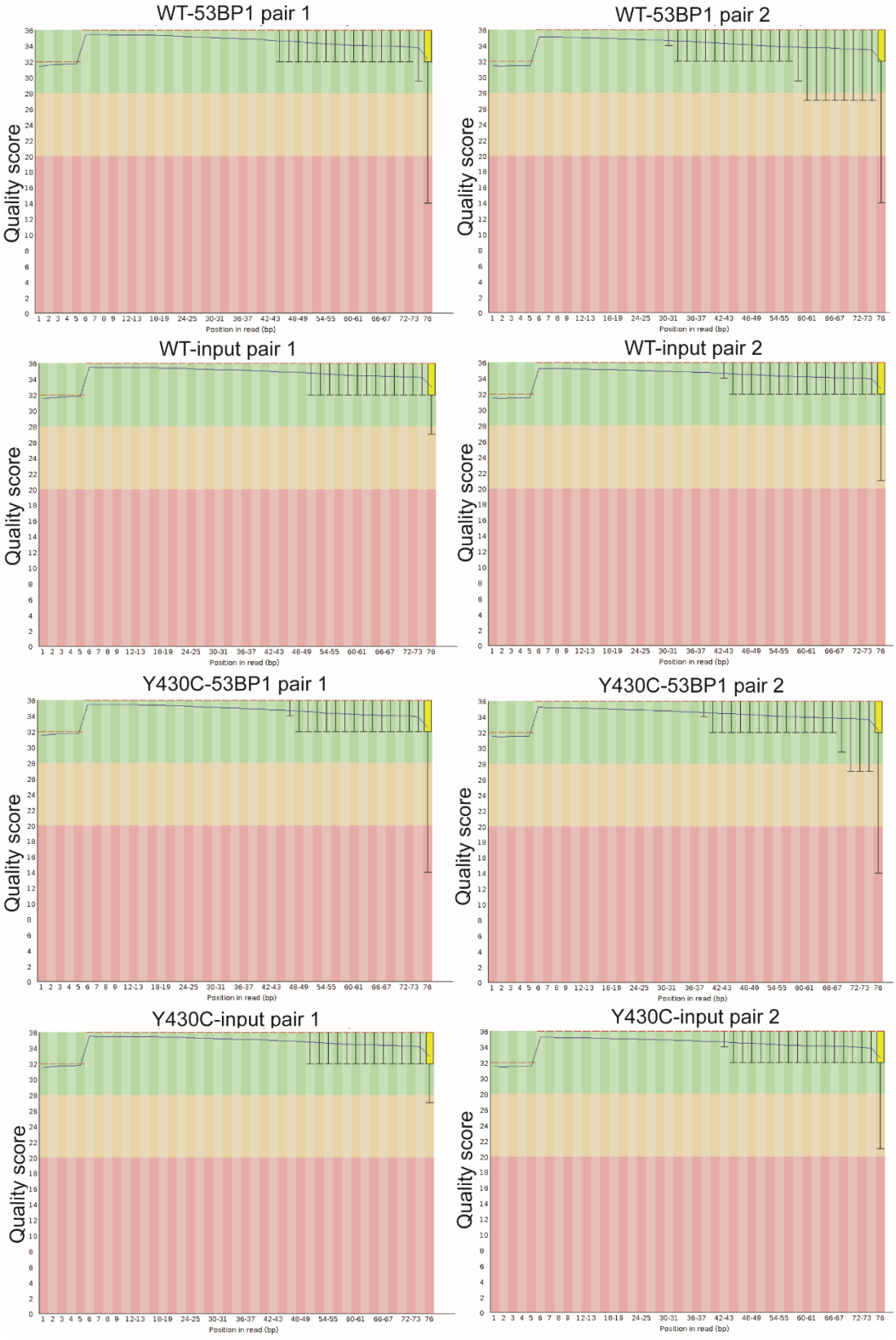


Figure Ch6-1. 53BP1 ChIP-seq per base sequence quality. Graphs show quality (phred) scores at each base position for the ChIP-seq FASTQ files, as calculated by FASTQC

	ChIP-seq mapping statistics		
	Total reads	**% reads correctly paired**	**% reads aligned**
WT 53BP1	68,296,055	100	90.11
Y430C 53BP1	82,185,592	100	87.38
WT input	96,017,791	100	97.81
Y430C input	66,892,195	100	97.87

Table Ch6-1. Mapping statistics for WT and Y430C ChIP-seq FASTQ files. Files were mapped with Bowtie2 and mapping statistics calculated with Samtools Flagstat

6.1.3 Visualisation

I converted the mapped reads to bigwig format and visualised them on the UCSC genome browser. What was instantly apparent was the similarity in the 53BP1 binding pattern to that of BRD4 at enhancers; there are large peaks in the 53BP1 ChIP-seq at the SEs and TEs where I found BRD4 to be enriched (Figures Ch6-2 and Ch6-3A). This is interesting, as a role for 53BP1 at enhancers has not been well studied.

The BRD4 ChIP-seq showed the number of reads across these BRD4 target regions to be higher in WT cells compared to Y430C cells (Figure Ch4-5). However, there is no obvious difference between WT and Y430C over the same regions for 53BP1 ChIP-seq (Figure Ch6-2). Surprisingly there does appear to be more reads in the Y430C sample across the KLF4 SE (Figure Ch6-2A), but this decrease is not apparent at the KLF4 promoter or consistent over other SEs (Figure Ch6-2B&C).

There is very little evidence of 53BP1 binding at the *Gsc* promoter (Figure Ch6-3A) where BRD4 binding is also negligible (Figure Ch4-6). However, there are large, sharp, 53BP1 peaks at the TSS of many genes, such as *Klf4* (Figure Ch6-2A), *Nanog* (Figure Ch6-2B), *Sox2* (Ch6-2C) and *c-Myc* (Figure Ch6-3B), which are not present in the BRD4 ChIP-seq. As with the 53BP1 peaks at enhancers of these genes, the 53BP1 promoter peaks are of a similar size in WT and Y430C cells.

BRD4 has been shown to recruit 53BP1 (Li *et al.*, 2018), and so it is surprising that 53BP1 occupancy doesn't decrease in Y430C mESCs at regions where BRD4 binding is decreased. To compare occupancy at BRD4 target sites on a genome wide level, I generated heatmaps of the read coverage across all of the SEs, TEs and promoters in WT and Y430C mESCs

(Figure Ch6-4). These show an enrichment of reads over the centre of these regions versus the surrounding area, and a slight decrease in the number of reads at all three regions in Y430C cells compared to WT - although this decrease is not as strong as is seen in the BRD4 ChIP.

The enrichment of 53BP1 signal around the centre of promoters (Figure Ch6-4A) is much tighter than that of BRD4 (Figure Ch4-7A), and this is consistent with the sharp peaks seen at TSSs (Figure Ch6-2&3). The heatmaps of 53BP1 at TEs and SEs (Figure Ch6-4B&C) also show a different pattern to those of BRD4 (Figure Ch4-7A) – 53BP1 heat maps show a dip in 53BP1 occupancy in the centre, which may correlate with a nucleosome free region.

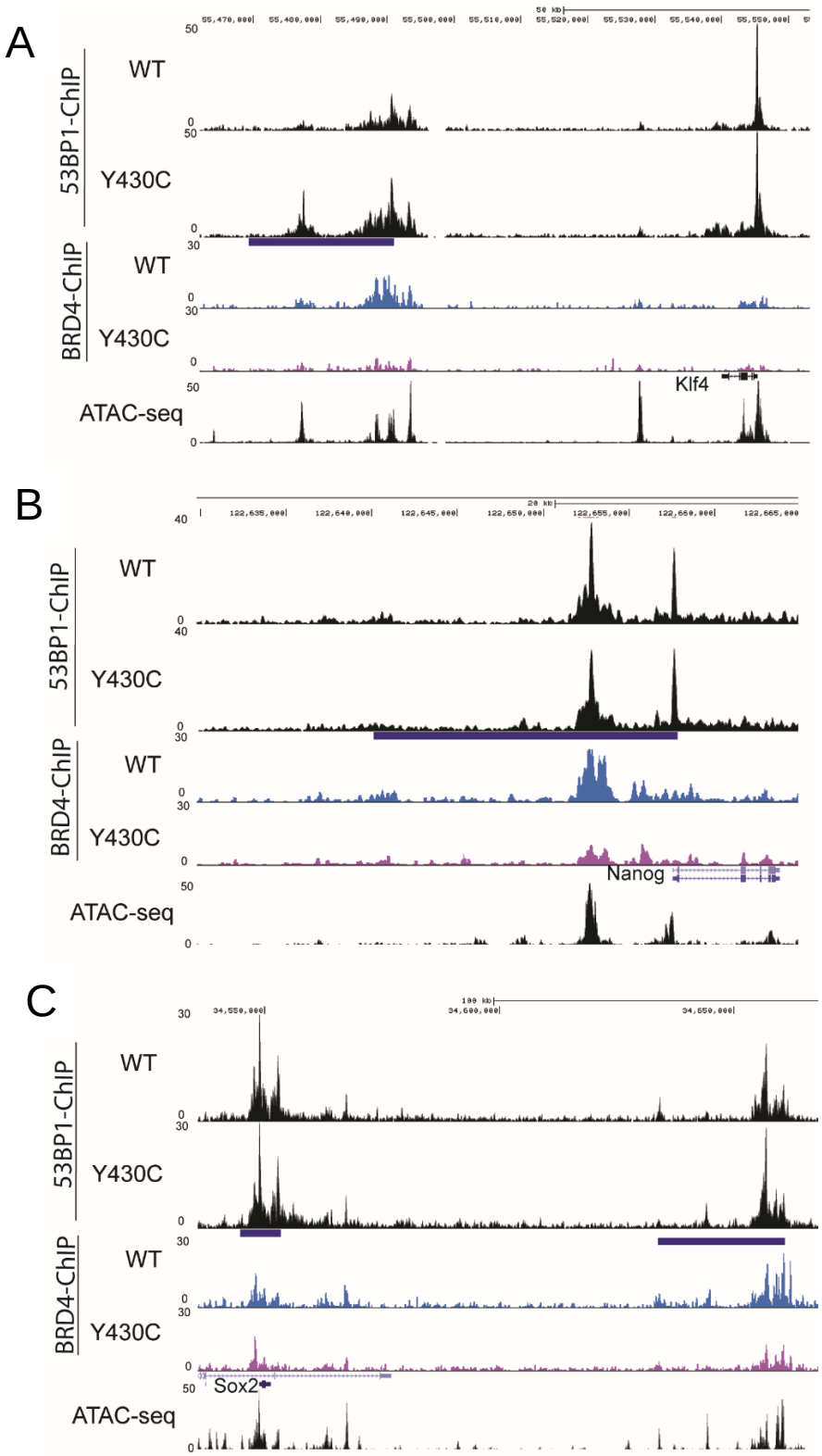


Figure Ch6-2. Genome browser view of 53BP1 binding. UCSC browser screenshot showing reads per 10 million (RP10M) over the extended A) *Klf4* locus, B) *Nanog* locus and C) *Sox2* locus for various datasets. Tracks 1 and 2 = 53BP1 ChIP-seq in WT and Y430C mESCs, tracks 3 and 4 = BRD4 ChIP-seq in WT and Y430C mESCs. Track 5 = ATAC-seq from mESCs (provided by Yatendra Kumar). Blue bars = SE

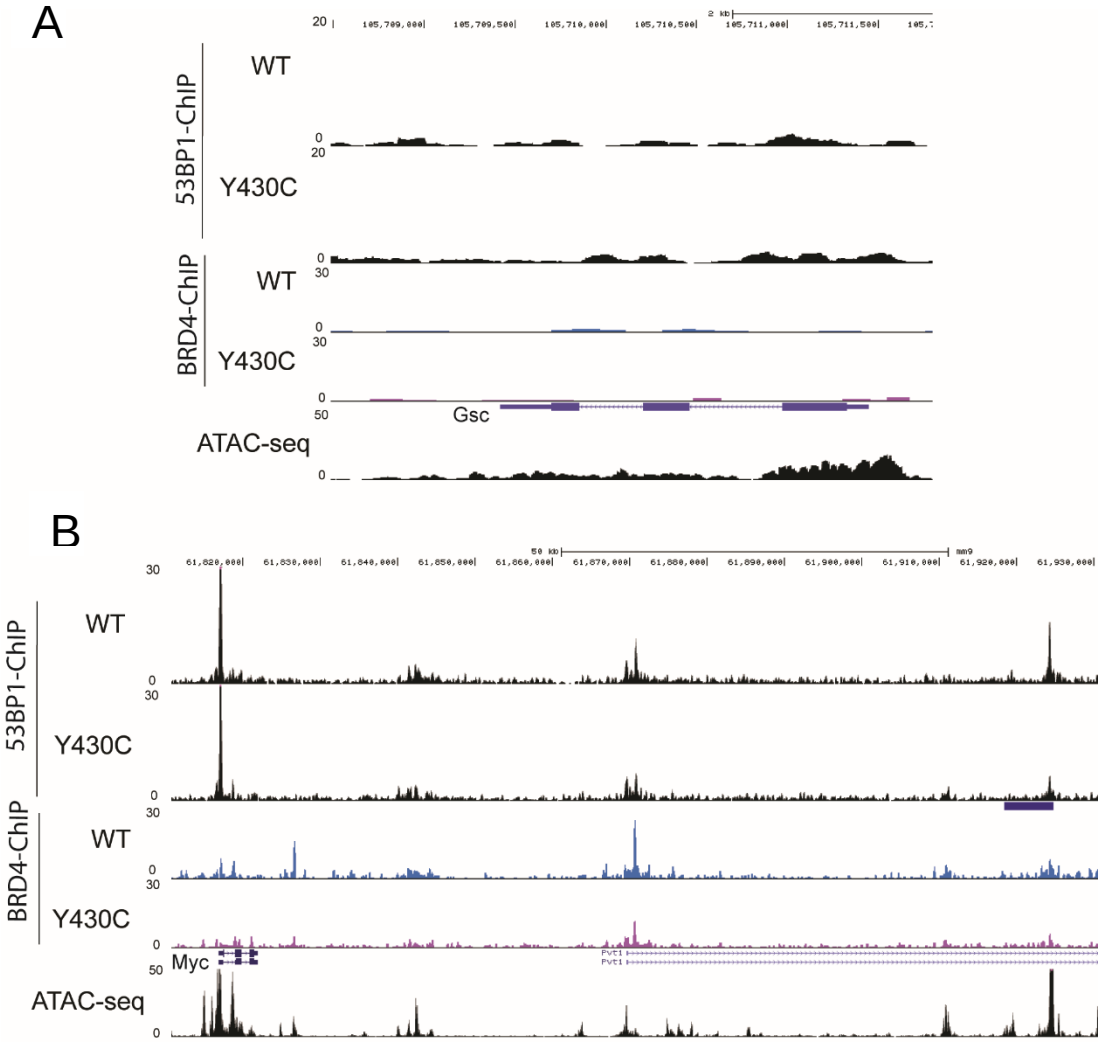


Figure Ch6-3. Genome browser view of 53BP1 binding. UCSC browser screenshot showing RP10M over the extended **A**) *Gsc* locus and **B)** *Myc* locus for various datasets. Tracks 1 and 2 = 53BP1 ChIP-seq in WT and Y430C mESCs, tracks 3 and 4 = BRD4 ChIP-seq in WT and Y430C mESCs. Track 5 = ATAC-seq from mESCs (provided by Yatendra Kumar). Blue bars = SE

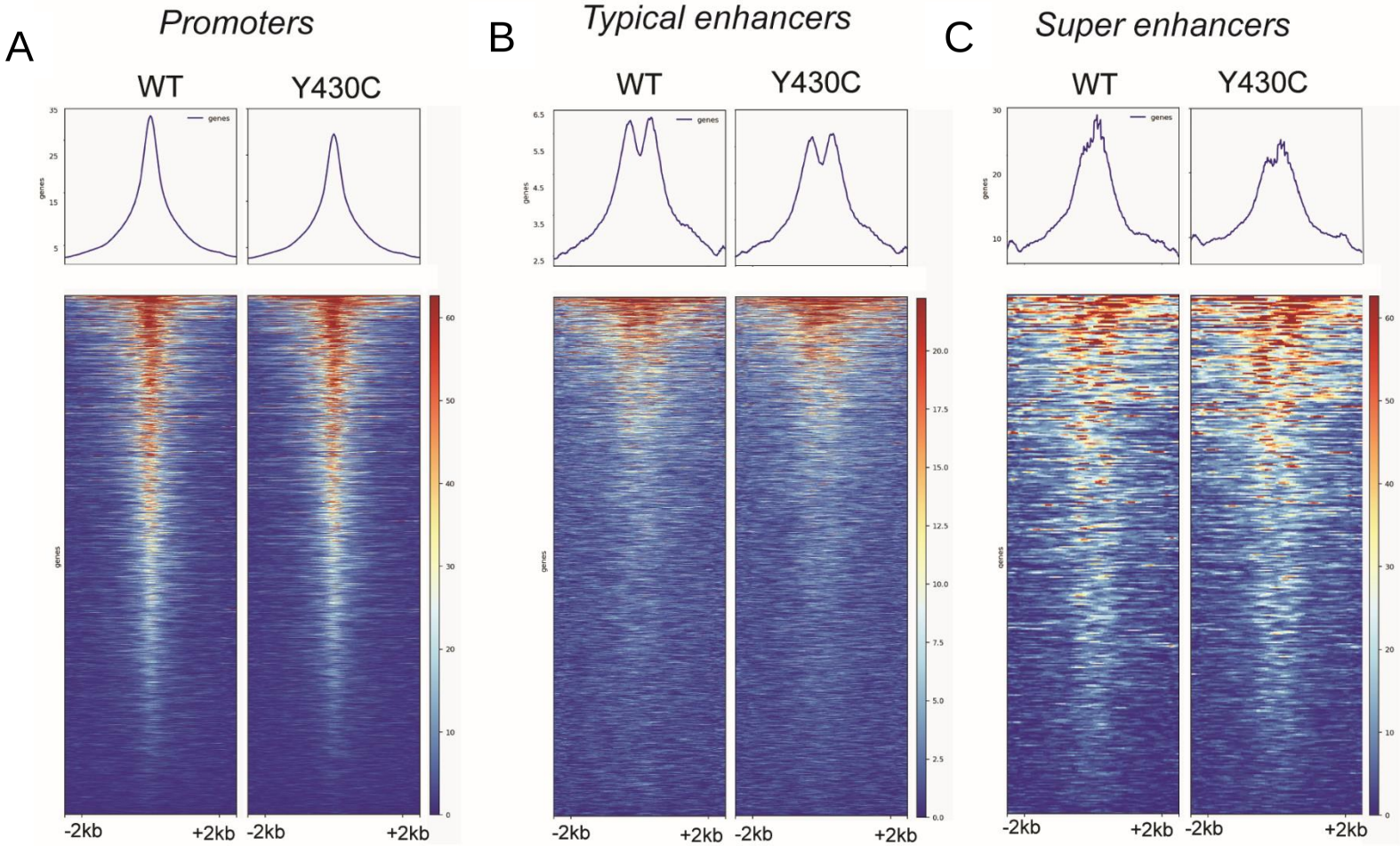


Figure Ch6-4. Binding of 53BP1 at CREs. Heatmaps show distribution of 53BP1 ChIP-seq reads across **A)** super enhancers, **B)** typical enhancers and **C)** promoters in WT and Y430C cells. Colour scale represents log2 read coverage normalised by RPKM. Regions are scaled to 4 Kb

6.1.4 Peak-calling

I called significant peaks in the WT and Y430C datasets as before, using the MACS2 software. Overall there were more peaks in the 53BP1 ChIP-seq in both WT and Y430C mESCs than in the BRD4 ChIP-seq, which may be attributable to the peaks of 53BP1 at TSSs. There were also more peaks in the WT 53BP1 ChIP-seq compared to Y430C 53BP1 ChIP-seq (31205 and 29871 respectively) (Table Ch6-2). The decrease in 53BP1 peaks in the Y430C cells is less than the decrease seen in the BRD4 ChIP-seq (1.04 fold versus 1.76 fold), which fits with the smaller difference seen in the heatmaps.

regioneR analysis showed 53BP1 to be most significantly enriched over promoters, TEs and SEs (Figure Ch6-5), and least enriched at heterochromatin, repressed regions and weakly transcribed regions. That 53BP1 shows similar genomic patterns of enrichment to BRD4 is in concordance with the idea that BRD4 might recruit 53BP1 to chromatin.

6.1.4.1 Differential peak calling

I next used MACS2 bdgdiff to call differential binding events between WT and Y430C 53BP1 ChIP-seq (Table Ch6-2). 15,752 regions were found to have significantly more reads in WT than in Y430C, whilst no regions were significantly increased in Y430C. These results suggest that 53BP1 binds at a similar number of regions in both cell lines, but that less 53BP1 binds at each region in the Y430C cells. Coupled with the heatmap data (Figure Ch6-4), which shows only a small difference in read coverage at CREs between the two cell lines, this indicates that the decreased 53BP1 occupancy may be occurring at regions other than those defined in the TE, SE and promoter files.

I visualised the 53BP1 ChIP-seq across the regions where the peak size was found to be most different between WT and Y430C, confirming the difference (Figure Ch6-6). These genomic regions vary in their function, but include strong enhancers (defined by ChromHMM) (Figure Ch6-6A). regioneR analysis of these regions suggests that they are enriched in promoters, TEs and SEs (Figure Ch6-5).

The BRD4 ChIP-seq data (Figure Ch6-6, tracks 3&4) shows that the regions where 53BP1 binding is differential between WT and Y430C cells are also occupied by BRD4, with a decrease in BRD4 occupancy in the Y430C cells. These BRD4 peaks tend to be smaller and narrower than those present at the SEs (Figure Ch6-3&4). Whilst the decrease in BRD4 occupancy in Y430C cells doesn't appear to affect 53BP1 recruitment to SEs (Figure Ch6-

3&4), it may be that at the non-SE regions, where it is already limited, a required threshold of BRD4 fails to be reached, leading to a decrease in 53BP1 recruitment.

	Peak calling results		
	Total peaks	**Specific peaks**	**% specific peaks**
WT	31,205	15,752	8.2
Y430C	29,871	0	0.8×10^{-5}

Table Ch6-2. Mapping statistics for WT and Y430C ChIP-seq FASTQ files. Files were mapped with Bowtie2 and mapping statistics calculated with Samtools Flagstat

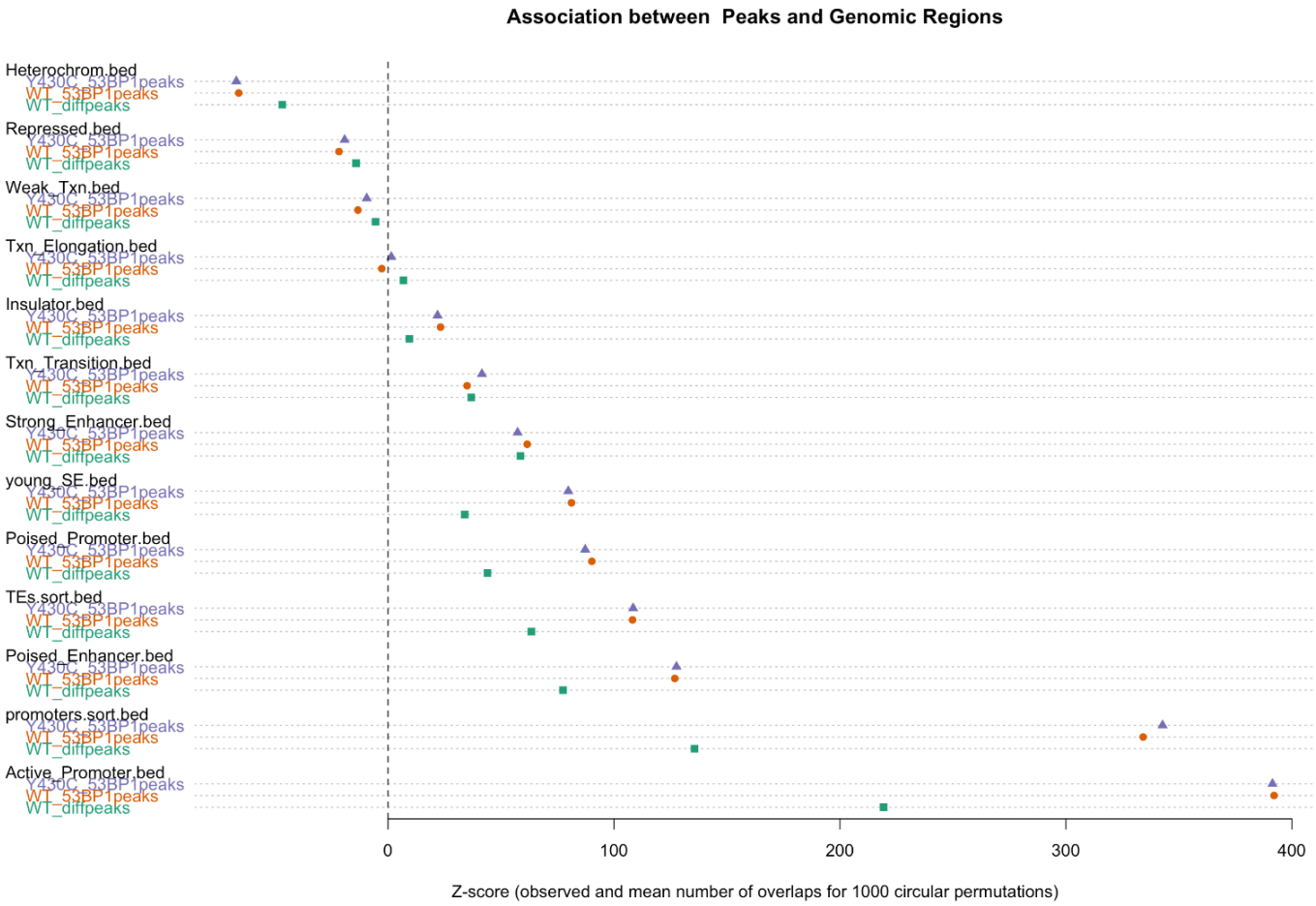


Figure Ch6-5. Enrichment of functional genomic regions in 53BP1 peaks. Plots show Z-score calculated by regioneR. These represent the likelihood of 53BP1 peaks overlapping with specific genomic regions. Peaks were called by MACS2 callpeaks ($p<0.00001$). Diffpeaks are those specific to WT, called by MACS2 bdgdiff. Genomic regions were defined by ChromHMM in mESCs, excluding Promoter, TEs and SEs which were defined by Flynn *et al.,* 2016.

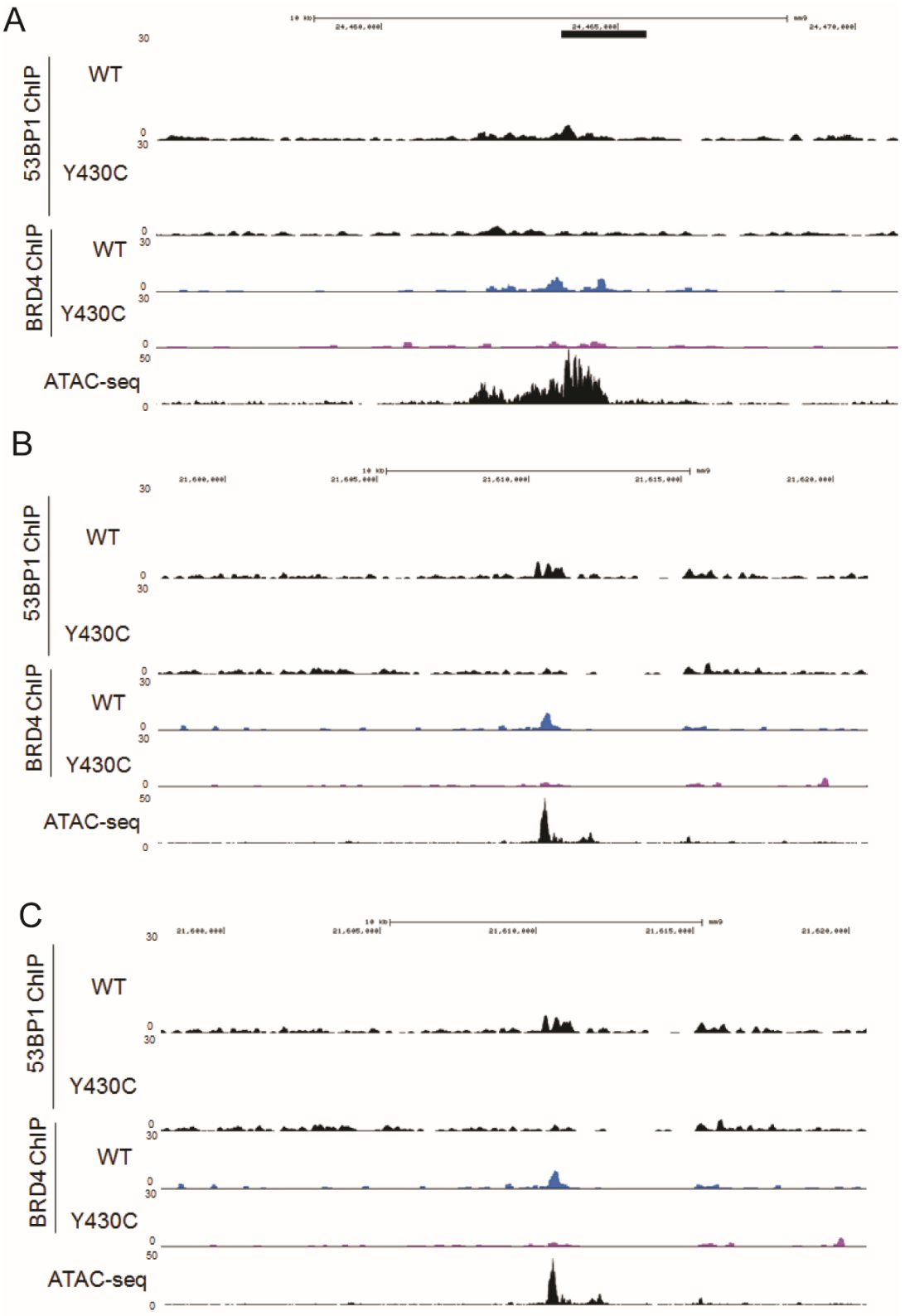


Figure Ch6-6. Genome browser view of 53BP1 binding. UCSC browser screenshot showing RP10M over regions of differential WT and Y430C BRD4 binding. Tracks 1 and 2 = 53BP1 ChIP-seq in WT and Y430C mESCs, tracks 3 and 4 = BRD4 ChIP-seq in WT and Y430C mESCs. Track 5 = ATAC-seq from mESCs (provided by Yatendra Kumar). Black bars = strong enhancers defined by ChromHMM.

6.2 Investigating a link between Y430C positive CdLS and classical CdLS.

Heterozygous LOF mutations in NIPBL are responsible for most cases of CdLS (Ansari *et al.*, 2014). Interestingly, NIPBL-deficient CdLS patient cells were found to have impaired early recruitment of 53BP1 to DSBs (Vrouwe *et al.*, 2007; Enervald *et al.*, 2013), highlighting a potential link between the mechanisms underlying CdLS caused by BRD4 and NIPBL mutations.

6.2.1 Comparing NIPBL, BRD4 and 53BP1 occupancy in mESCs

I took publicly available NIPBL ChIP-seq data from V6.5 (C57BL/6-129) mESCs (GSM560350) and mapped this to the mm9 genome, as for the BRD4 and 53BP1 ChIP-seq datasets. I then uploaded this to the UCSC genome browser to allow a visual comparison of the binding of NIPBL, BRD4 and 53BP1 (Figure Ch6-7).

The binding pattern of NIPBL was similar to that of BRD4, with large peaks of NIPBL at CREs. This is not surprising as both BRD4 and NIPBL are known to bind promoters, TEs and SEs of pluripotency related genes in mESCs (Kagey *et al.*, 2010; Di Micco *et al.*, 2014). However, NIPBL also binds at promoters, in a pattern similar to that of 53BP1. The striking resemblance of the 53BP1 ChIP-seq peaks to the NIPBL peaks at promoters, and its co-localisation with both BRD4 and NIPBL at SEs, suggests that recruitment of 53BP1 could be the step that links CdLS caused by the Y430C mutations to classical CdLS.

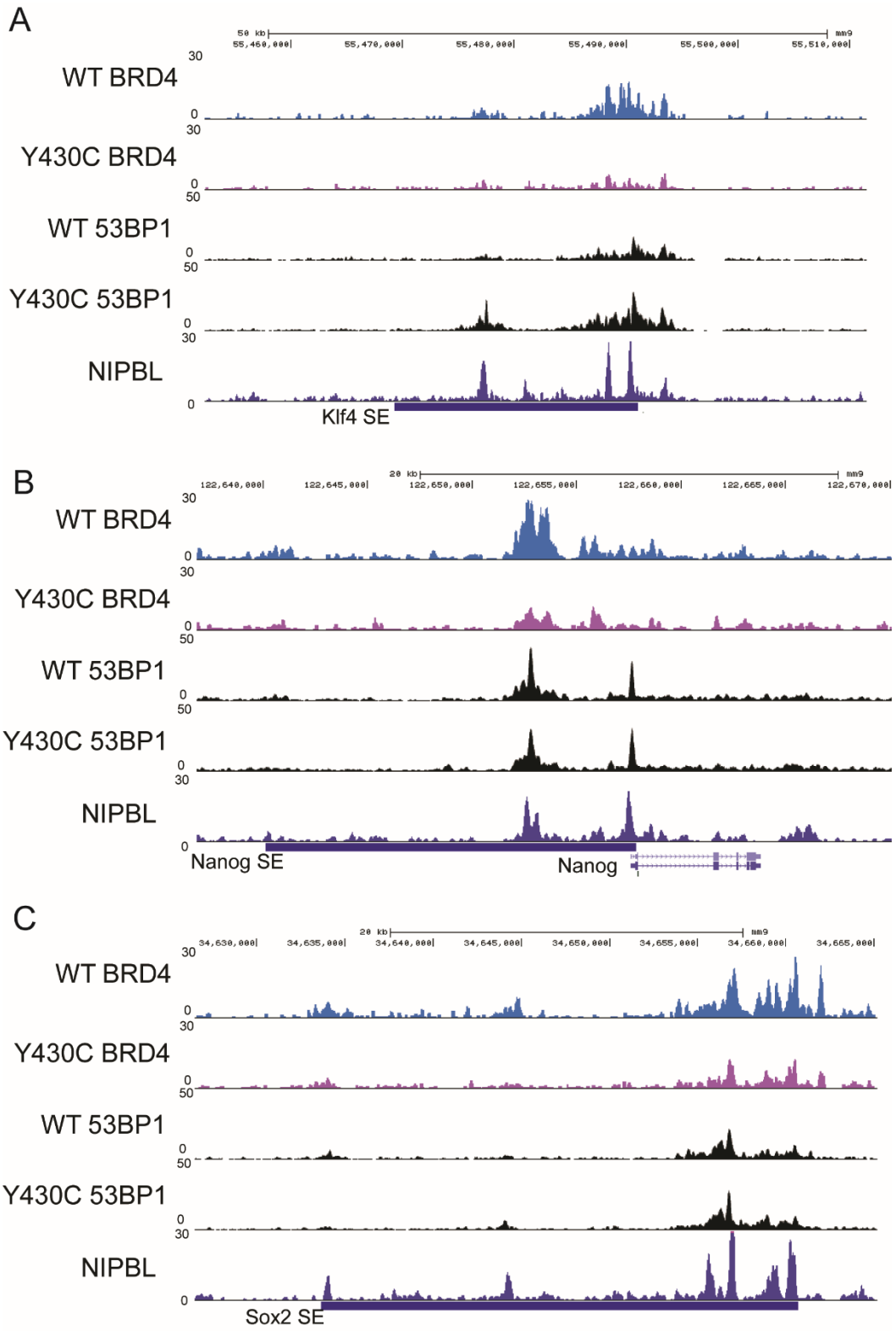


Figure Ch6-7. Comparison on BRD4, 53BP1 and NIPBL binding at SEs. UCSC browser screenshot showing RP10M over **A)** *Klf4*, **B)** *Nanog* and **C)** *Sox2* SEs. Tracks 1 and 2 = BRD4 ChIP-seq in WT and Y430C mESCs, tracks 3 and 4 = 53BP1 ChIP-seq in WT and Y430C mESCs, track 5 and 6 = NIPBL ChIP-seq from V6.5 (C57BL/6-129) mESCs (GSM560350)

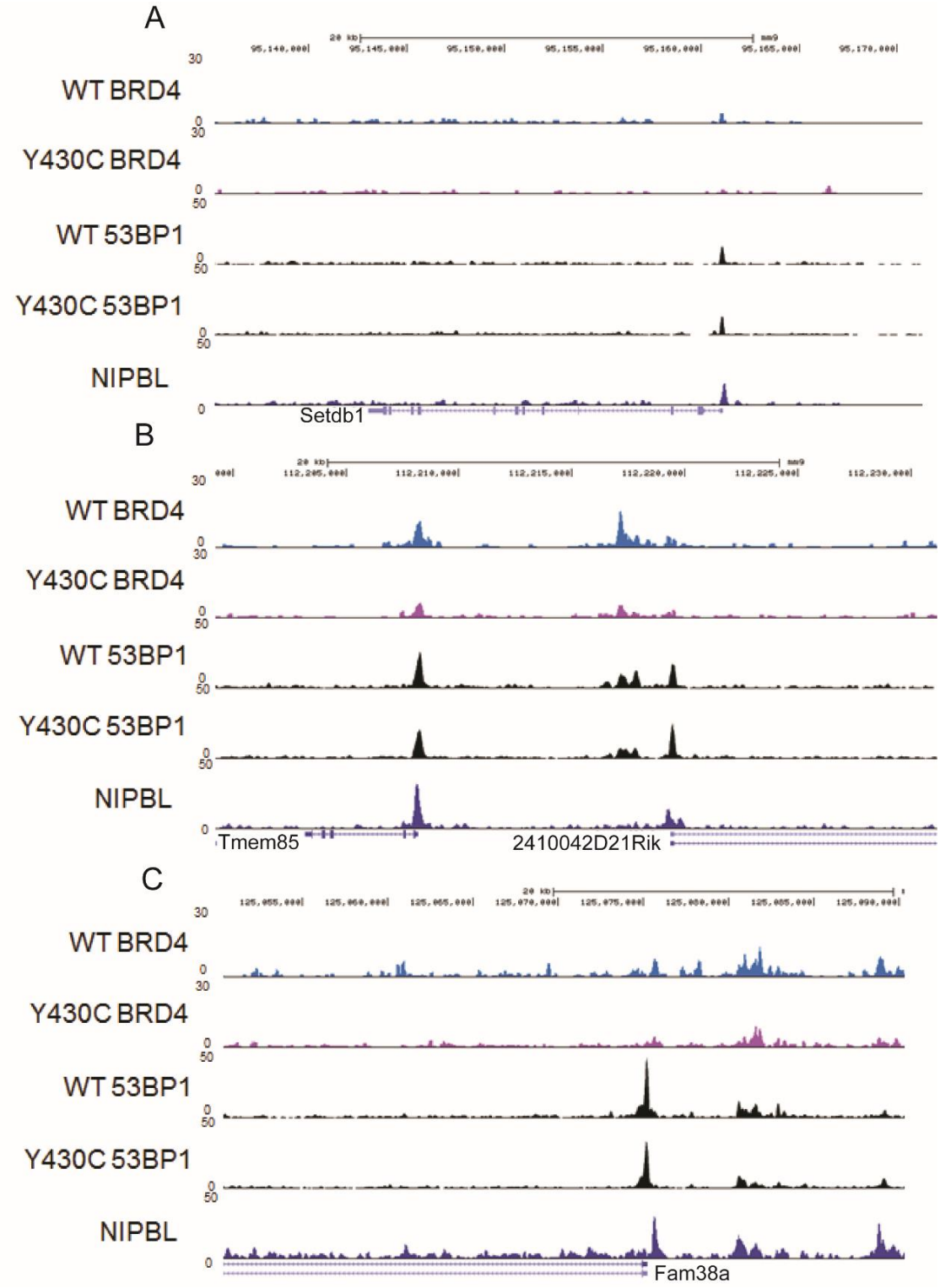


Figure Ch6-8. Comparison on BRD4, 53BP1 and NIPBL binding at promoters. UCSC browser screenshot showing RP10M over **A)** *Setdb1*, **B)** *Tmem85* and *2410042D21Rik* and **C)** *Fam38a* promoters. Tracks 1 and 2 = BRD4 ChIP-seq in WT and Y430C mESCs, tracks 3 and 4 = 53BP1 ChIP-seq in WT and Y430C mESCs, track 5 and 6 = NIPBL ChIP-seq from V6.5 (C57BL/6-129) mESCs (GSM560350)

6.2.2 Analysing endogenous DNA damage in NIPBL deficient LCLs

The correlation between NIPBL binding, 53BP1 binding, and regions of differential BRD4 binding in the mESCs, suggested that the *NIPBL* CdLS patient mutations could also perturb the DDR. So far, I have focussed on the mESCs carrying the Y430C mutation in *BRD4* identified in one patient with CdLS. However, to analyse the effects of *NIPBL* mutations on DNA damage signalling, I used two LCLs previously derived from CdLS patients with heterozygous mutations in *NIPBL* - I1206del (Tonkin *et al.*, 2004) and R2298H (Gillis *et al.*, 2004).

53BP1 IF after treatment with NCS, shows that these patient LCLs have an increased number of 53BP1 foci compared to the WT LCLs in the absence of any exogenous damaging agent (Figure Ch6-8A&B), suggesting that the NIBPL mutations in these patients may affect the DDR. 1 hour after NCS treatment, nearly all WT, R2298H and I1206del cells had >5 53BP1 foci. At 16 and 20 hours post NCS treatment, these foci had begun to be resolved - although this was much less drastic than in the mESCs (Figure Ch4-5B), with ~70-80% of cells still containing >5 foci. In contrast to the mESCs, there was no significant difference in the number of cells containing >5 53BP1 foci at 16 or 20 hours post NCS treatment between WT and mutant LCLs. As all three LCLs still had high numbers of foci containing cells, it is possible that repair in these cells is slower than in mESCs and that a difference would become more apparent at a later time point. Before NCS treatment, the foci present in R2298H and I1206del LCLs are larger than in WT (Figure Ch6-8C), as well as more plentiful. This is consistent with the larger foci in the mESC mutant cells, and could mean that the spread of the DDR is greater upon mutation of *BRD4* or *NIPBL*.

6.2.3 Cell cycle analysis in NIBPL deficient LCLs

As the NIPBL deficient LCLs showed an increased number of 53BP1 foci, I wondered if they might have an altered cell cycle; the BRD4 mutant ES cells showed slower cell cycle with an apparent arrest in the G2/M phase of the cell cycle. Cell cycle analysis by flow cytometry showed very similar profiles for all three LCLs (Figure Ch6-9) (Table Ch6-3). The lack of cell cycle changes observed with the NIPBL mutation may be due to the use of LCLs, which have different cell cycle timings to the mESCs in which the *Brd4* mutation is modelled. Furthermore, LCLs are immortalised cell lines established by the transformation of B-cells with the Epstein-Barr Virus (Hui-Yuen *et al.*, 2011)

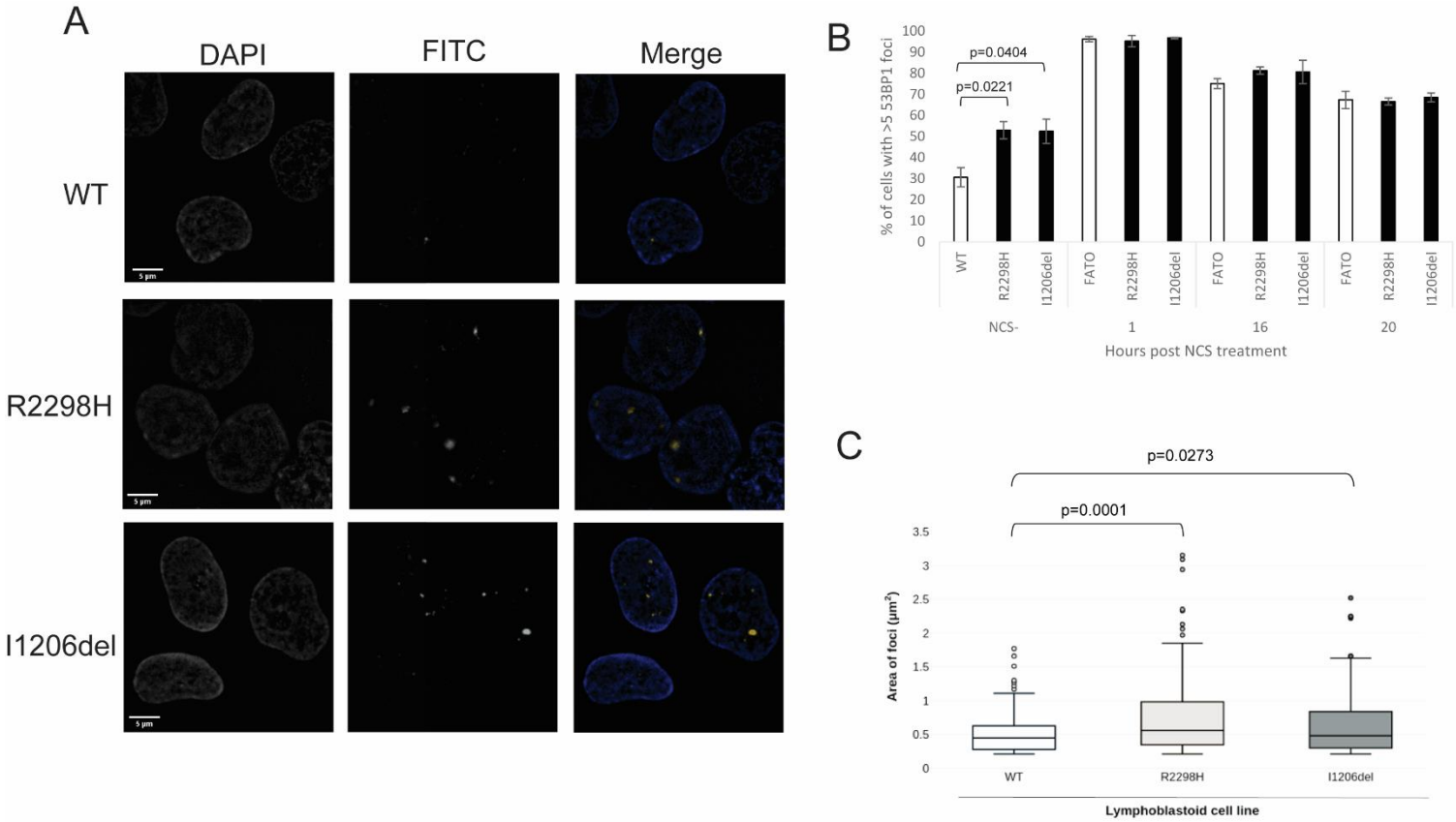


Figure Ch6-8. 53BP1 foci in WT and NIPBL-deficient LCLs. A) Representative images of WT, R2298H and I1206del LCLs upon 53BP1 IF and DAPI staining in untreated cells. Scale bar = 5μm. **B)** Graph shows percentage of cells that contain >5 53BP1 foci in untreated cells (NCS-) and 1, 16 and 20 hours post NCS treatment. Data are represented as mean +/- SEM from 3 biological replicates. **C)** Box-plot shows area of 53BP1 foci (μm^2) in WT, R2298H and I1206del LCLs without treatment with NCS. Horizontal lines within boxes show medians, boxes are inter-quartile ranges and whiskers are range. Outliers are shown as circles. P-values were calculated with t-test.

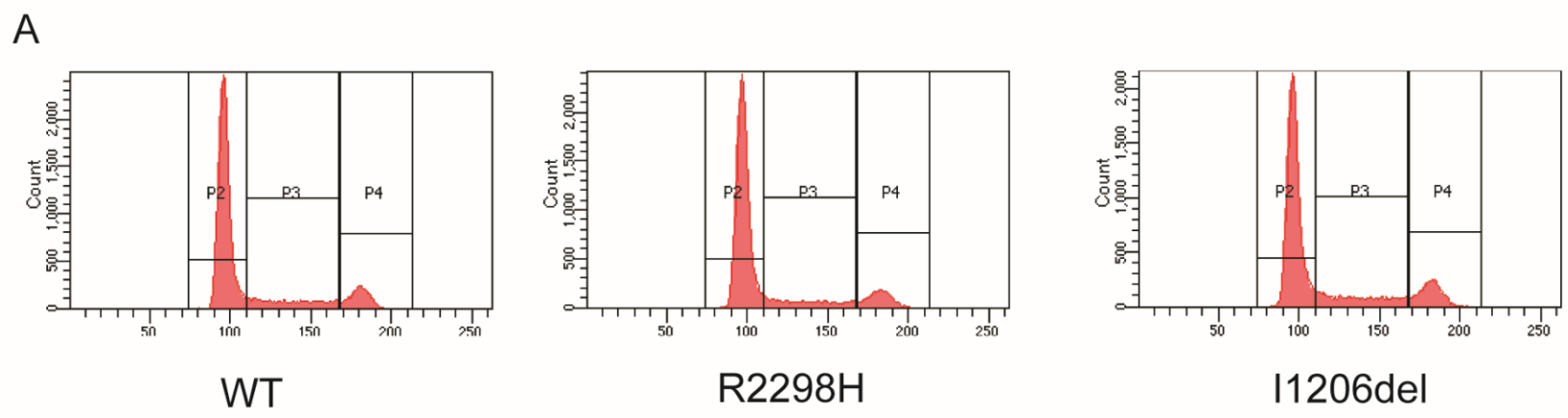


Figure Ch6-9. WT, R2998H and I1206del cell cycle profiles determined by flow cytometry. Graphs illustrate the cell count, which correlates to propidium iodide intensity. Gated cells were manually categorized into the different phases of the cell cycle – P2=G0/G1, P3=S and P4=G2/M.

	% of cells		
	G0/G1	**S**	**G2/M**
WT	68.8	18.3	13.3
R2298H	71.5	15.5	13.2
I1206del	65	19.6	15.7

Table Ch6-3. % of LCLs in each cell cycle stage

6.3 Conclusions

My preliminary 53BP1 ChIP-seq results show that 53BP1 is enriched at enhancers, similarly to BRD4. This fits with the results of others, who have shown BRD4 to recruit 53BP1 to chromatin (Li *et al.*, 2018). Unlike BRD4, 53BP1 is not decreased in the Y430C cells at SEs. However, there are a set of regions at which BRD4 occupancy is relatively low, where 53BP1 occupancy is higher in the WT cells compared to the Y430C cells. I hypothesise that at these regions, the decrease in BRD4 affinity for acetyl-lysines means that there is too low a level of BRD4 binding in Y430C cells for the full recruitment of 53BP1.

DNA-damage-induced histone H4 acetylation is thought to recruit BRD4 to chromatin (Li *et al.*, 2018). If the Y430C mutation decreases BRD4 binding in response to damage, this result could suggest a subsequent decrease in 53BP1 recruitment to DSBs, as I saw for RAD51 (Figure Ch5-12&13), which may perturb DSB repair. I did not observe any difference between 53BP1 foci levels in WT and Y430C cells at 1 hour post NCS, however it may be that the large number of foci present in all cells at this time masks a more subtle difference.

There is less 53BP1 binding in the Y430C cells compared to WT, as assayed by ChIP-seq. This is contradictory to my previous results, where I saw more 53BP1 foci by IF. However, this may be because the ChIP-seq peaks do not correspond to the foci that are present in IF after DSB induction. The 53BP1 peaks in ChIP-seq are unlikely to be DSBs as they are small, sharp peaks expanding <1 Kb (Figure Ch6-2,3,5&6), whereas the observed spreading of DDR factors around a DSB by ChIP is usually of several Mb (Iacovoni *et al.*, 2010). The lack of 53BP1 peaks at DSBs in the ChIP-seq may be explained by the random manner of these breaks. If they are occurring in a different locus in each cell, this would not be picked up by ChIP, which combines the signals from millions of cells.

53BP1 detection at enhancers by ChIP therefore implies that the binding of 53BP1 at these elements is consistent across the population of cells. This suggests a functional role for 53BP1 at enhancers, which will be discussed further in section 7.3.

53BP1 ChIP-seq also shows large, sharp peaks at promoters, which do not resemble those of my BRD4 ChIP-seq, but are strikingly similar to those seen in NIPBL ChIP-seq. NIPBL may be responsible for recruiting 53BP1 to promoters, and a loss of NIPBL (as in CdLS

patients with *NIPBL* mutations) could result in decreased 53BP1 recruitment. If this is the case, it would suggest that the loss of 53BP1 recruitment links the *BRD4*-Y430C mutation with the *NIPBL* mutations that cause CdLS. Furthermore LCLs with mutations in NIBPL show signs of a perturbed DDR (Figure Ch6-8&9). Therefore a misregulation of the DDR may be a feature common to CdLS mutations.

Chapter 7: Discussion

7.1 Overview of results

Recently David FitzPatrick's lab identified de novo mutations in the BET protein BRD4 in three patients with CdLS-like syndromes, leading us to investigate how the disruption of BRD4's normal function may result in CdLS. I focused on a Tyr to Cys point mutation located in BD2 of BRD4, termed Y430C, identified in a patient with a classical CdLS phenotype. *In silico* modelling of the mutation suggested that it would decrease the affinity of the BD to acetyl-lysines, and I confirmed this through experiments in a mESC line that I engineered to be homozygous for the Y430C mutation. I found that Y430C-BRD4 interacts less with acetylated H3K9 and H3K27 than WT-BRD4, and ChIP-seq showed that this decreased affinity impairs the binding of BRD4 to CREs, where it is usually enriched.

BRD4 is involved in the regulation of transcription (Jang *et al.*, 2005; Kanno *et al.*, 2014), and in mESCs is known to bind the SEs of pluripotency associated genes and promote their transcription (Di Micco *et al.*, 2014). However, I found the transcription of pluripotency related genes and global gene transcription to be similar between the WT and Y430C mESCs, indicating that the decreased occupancy of mutant BRD4 at CREs does not affect transcription in these cells.

I observed a slower doubling rate of the Y430C mESCs compared to WT and propose that this is caused by increased G2/M cell cycle checkpoint activation. The G2/M checkpoint is activated by the DDR (Schmitt *et al.*, 2007), suggesting that the Y430C mESCs may have a problem with DNA damage. I found that the mutant *Brd4* cells exhibit increased DDR signalling in response to DSB induction, as evidenced by the increased size and number of 53BP1 foci, and may have defective DSB repair through impaired HR.

ChIP-seq of 53BP1 showed that it is enriched at CREs. At enhancers, 53BP1 binds in a pattern similar to that of BRD4. This is in agreement with recent studies that have found BRD4 to recruit 53BP1 to chromatin (Li *et al.*, 2018), and suggests a functional role of 53BP1 at enhancers that is yet to be elucidated. Furthermore, the decreased affinity of BRD4-Y430C for acetyl-lysines appears to affect the binding of 53BP1 at a subset of binding sites which could have further implications for DNA repair.

At promoters, large, sharp 53BP1 ChIP-seq peaks correlate with those of NIPBL, suggesting NIPBL could be involved in recruitment or regulation of 53BP1 to these regions.

Furthermore, NIPBL co-localises with both BRD4 and 53BP1 at enhancers. Coupled with NIPBL's known roles in DNA damage signalling and repair (Bot *et al.*, 2017), and my finding that BRD4 interacts with NIPBL and cohesin ring components, this suggested that an impaired DDR could be a common mechanism of CdLS. My preliminary results in NIPBL-deficient LCLs from CdLS patients corroborate this, with evidence of increased DNA damage signalling in the absence of exogenous damaging agents.

7.2 Transcriptional dysregulation in CdLS

Currently, dysregulated transcription is the preferred mechanism for how de novo mutations cause CdLS. Cohesin interacts with a number of transcriptional regulators (Kline *et al.*, 2017) and many groups have reported altered gene expression profiles in CdLS patients (compared to healthy controls) (Liu *et al.*, 2009; Mannini *et al.*, 2015; Yuan *et al.*, 2015; Boudaoud *et al.*, 2017; Mills *et al.*, 2018). Furthermore, there is growing evidence for overlapping phenotypes between CdLS and developmental syndromes caused by mutations in transcriptional regulators – leading to the idea that CdLS may be best classified as a 'transcriptomopathy' (Kline *et al.*, 2017).

Given BRD4's characterised function as a transcriptional regulator (Di Micco *et al.*, 2014; Bhagwat *et al.*, 2016; Hajmirza *et al.*, 2018; Rahnamoun *et al.*, 2018), I fully expected the BRD4-Y430C mutation to support the 'transcriptome disruption model' of CdLS – my hypothesis being that perturbation of BD2 would decrease BRD4 binding to acetyl-lysines at CREs and therefore inhibit transcriptional initiation and elongation of BRD4 target genes. This was not the case; I saw few significant changes in the transcriptomes of Y430C mESCs compared to WT, even at the pluripotency related genes – known targets of BRD4 in mESCs (Di Micco *et al.*, 2014).

Does this mean that I can discount transcriptional dysregulation as a contributing factor to CdLS in the patients with BRD4 mutations? No. I think it likely that the loss of BRD4 binding observed in the Y430C cells causes changes in transcription that I am unable to detect. One explanation for this may be the use of ES cells.

ES cells display unique regulation of transcription. In contrast to differentiated cells, ES cell genomes are transcriptionally hyperactive - they express large regions at low levels and have elevated levels of chromatin remodelling proteins and transcription machinery. Global reduction of the active portion of the genome, coupled with increased activation of specific genes, promotes the differentiation of ES cells down specific lineages. Furthermore, chromatin proteins display hyperdynamic and looser binding to chromatin in ES cells than in undifferentiated cells (Efroni *et al.*, 2008). It may be that decreased BRD4 binding would affect transcription in differentiated cells but this is compensated for by the high levels of transcriptional machinery in ES cells, or that transcription in ES cells does not require such 'tight' BRD4 chromatin binding.

It is also possible that the way that we culture ES cells – supplementing with the exogenous cytokine LIF to prevent spontaneous differentiation (Ohtsuka, Nakai-Futatsugi and Niwa, 2015) – masks the effects of decreased BRD4 binding. This method would select against differentiating cells, whilst selecting for pluripotent and self-renewing cells. Since transcriptional dysregulation of ES cells might cause differentiation, cells that are resistant to this may be expected to outcompete the other cells – resulting in a population that shows few transcriptional differences.

To investigate this further, I could repeat my experiments with differentiated cell lines, as discussed in section 7.8.4.

7.3 A functional interaction between BRD4 and 53BP1 at enhancers?

BRD4 has been shown to recruit a complex of the snRNA 7SK and the SWI/SNF chromatin remodelling complex BAF to TEs and SEs (Flynn *et al.*, 2016). In mESCs, BAF maintains nucleosome depleted regions and promotes elevated nucleosome occupancy adjacent to these regions, suppressing ncRNA transcription (Hainer *et al.*, 2015). At TEs and SEs, 7SK bridges interactions important for BAF's activity. Loss of 7SK results in global reduction of BAF155 occupancy at enhancers and increased eRNA transcription. Unregulated eRNA

transcription can result in convT, and this can cause DNA damage through both RNAPII collisions and AID (Flynn *et al.*, 2016).

Through 4sU-seq I detected eRNA transcripts at TEs and SEs in WT mESCs, showing that some transcription is permitted at enhancers even in the presence of BRD4. This suggests that convT could be a feature of WT cells and may cause DSBs that require repair.

My 53BP1 ChIP-seq showed that 53BP1 occupies TEs and SEs similarly to BRD4, however a role of 53BP1 at enhancers has not been reported. 53BP1 can be recruited by BRD4 (Li *et al.*, 2018) and I found that decreased 53BP1 occupancy correlated with decreased BRD4 occupancy at a subset of regions. I propose that 53BP1 could be recruited by BRD4 at TEs and SEs to aid the repair of DSBs that arise through convT (Figure Ch7-1). 53BP1 is extremely important for accurate DNA repair. It supresses the nucleolytic resection of DNA ends at breaks, inhibiting repair by HR and thus regulating the choice between HR and NHEJ (Noordermeer, Adam, Setiaputra, Barazas, Pettitt, Ling, Olivieri, Álvarez-Quilón, Moatti, Zimmermann, Annunziato, Krastev, Song, Brandsma, Frankum, Brough, Sherker, Landry, Szilard, Munro, McEwan, de Rugy, *et al.*, 2018). In the absence of 53BP1, repair foci are slow to resolve and 53BP1 knockout mice exhibit growth retardation, radiation sensitivity and genomic instability (Fernandez-Capetillo *et al.*, 2002).

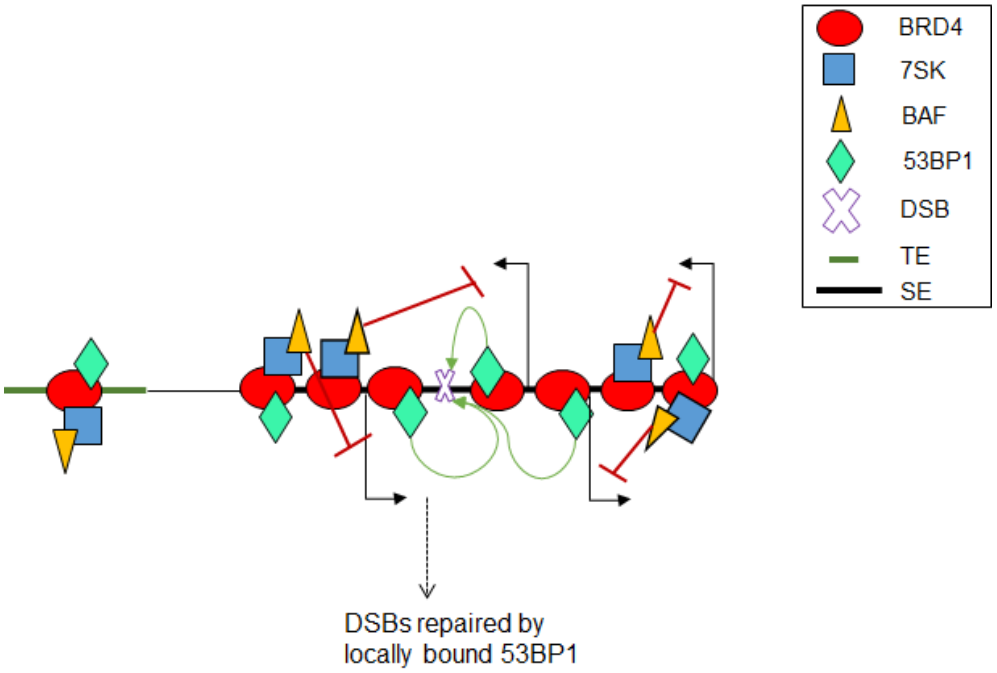


Figure Ch7-1. Model for the roles of BRD4 and 53BP1 at enhancer elements. BRD4 binds to acetyl-lysines in TEs and SEs and recruits the 7SK-BAF complex and 53BP1. 7SK-BAF surpresses pervasive eRNA transcription, whilst 53BP1 promotes the repair of DSBs that do occur.

7.4 BRD4 in the DDR

7.4.1 Current literature

BRD4 is overexpressed in a number of different cancers (Zhang *et al.*, 2015; Pongas *et al.*, 2017; Dong *et al.*, 2018) and this has recently been exploited through the design of small molecule inhibitors, such as JQ1, that decrease BRD4's binding to chromatin and have shown efficacy in both solid tumours and hematologic malignancies (Ocaña, Nieto-Jiménez and Pandiella, 2017). The success of these therapies relies in part on the downregulation of BRD4 target proto-oncogenes such as *MYC* (Ba *et al.*, 2018; Muhar *et al.*, 2018), but also by sensitising cancer cells to DNA damaging agents (Zhang *et al.*, 2018).

This has led to a wealth of recent literature regarding the role of BRD4 in the DDR. BRD4 localises to DSBs through recognition of H4 tail acetyl-lysine residues (Dhar *et al.*, 2017; Li *et al.*, 2018), where it is thought to promote DNA damage signalling through the regulation of DNA replication checkpoint signalling (Zhang *et al.*, 2018), prevention of heterochromatin formation (Pongas *et al.*, 2017), and recruitment of DDR proteins (Li *et al.*, 2018). Previously however, BRD4 isoform b had been shown to inhibit DDR signalling through the recruitment of the condensin II complex, which causes chromatin compaction and decreased accessibility of the DSB to DDR proteins (Floyd *et al.*, 2013). Finally, the binding of BRD4 at enhancers is thought to inhibit DSB formation through the suppression of eRNA transcription (Flynn *et al.*, 2016).

7.4.2 How my results fit with the current literature

The increase in the size and number of 53BP1 foci that I have shown in the Y430C mESCs, supports the idea that BRD4 acts to inhibit the spread of DNA damage signalling. Y430C-BRD4's decreased affinity for acetyl-lysines may decrease the recruitment of BRD4 to DSBs, leading to a loss of inhibition of the DDR signal. Analysis of chromatin compaction in WT and Y430C mESCs by FISH, would help confirm whether this increase in signalling is in fact caused by enhanced accessibility to DDR proteins upon loss of BRD4 binding – I could look specifically at loci where BRD4 is differentially bound between WT and Y430C cells, and/or at an induced DSB.

Through 4sU-seq I showed that eRNA transcript levels were similar in both WT and Y430C mESCs, indicating that the decrease in BRD4 binding at TEs and SEs does not affect the proposed role of BRD4 in supressing eRNA transcription. It may be that the diminished levels of BRD4 at these regions in Y430C cells is sufficient for inhibition of eRNA transcription. ChIP-seq or Chromatin Isolation by RNA Purification (ChIRP)-seq could be used to map the genome occupancy of BAF or 7SK respectively, in order to determine whether the decreased BRD4 levels affect 7SK-BAF binding at TEs and SEs. Furthermore, it would be interesting to map the location of DSBs in the Y430C mESCs, to see whether the additional 53BP1 foci observed in these cells are localised at enhancer elements. There are many methods available for this, which I will discuss in section 7.8.1.

The persistent 53BP1 foci, delayed recruitment of RAD51 in response to DSBs, and decrease in 53BP1 binding at certain sites that I have identified in the Y430C mESCs, is in agreement with the proposed role of BRD4 in recruiting DNA repair proteins and promoting DSB repair. The multiple functions of BRD4, suggested by my own results and put forth by others, suggests that it plays a complex role in the DDR, and may fine-tune the response to ensure the most appropriate outcome.

7.5 How does increased DDR signalling and decreased repair cause CdLS?

I have shown evidence for a defective DDR in the Y430C mESCs, suggesting that BRD4 mutations could cause CdLS via this mechanism. Increased sensitivity of CdLS patient cells to DNA damage has been reported previously (Vrouwe *et al.*, 2007; Enervald *et al.*, 2013), indicating that this may also occur in CdLS caused by mutations in proteins other than BRD4 – e.g. NIPBL. However, the question remains of how a dysregulated DDR might result in the CdLS phenotype.

The most obvious outcome of an impaired DDR is an increased risk of cancer, and DDR gene alterations are prevalent in many human cancer types (Knijnenburg *et al.*, 2018). These mutations may affect break repair, leading to increased genomic instability, and allow the

proliferation of cells harbouring deleterious mutations (Karanika *et al.*, 2015). CdLS patients do not, however, have an increased incidence of cancer (OMIM 122470).

The phosphorylation and activation of the checkpoint kinases, CHK1 and CHK2, is the end result of the DDR signalling cascade (Maréchal and Zou, 2013) and can lead to a delay in the cell cycle until DNA damage is repaired (Bartek and Lukas, 2003). The loss of BRD4's inhibition of DDR signalling in the Y430C-BRD4 cells, resulted in increased CHK1 phosphorylation and a slower cell cycle. This delay in the cell cycle, when occurring during development, could cause some aspects of the CdLS phenotype. The cell cycle is tightly regulated to ensure proper development and tissue homeostasis (Barnum and O'Connell, 2014), and loss of certain cell cycle regulators have been associated with reduced body size (Rane *et al.*, 1999; Sicinski *et al.*, 1995; Ciemerych *et al.*, 2002) neurological abnormalities (Sicinski *et al.*, 1995; Ciemerych *et al.*, 2002), craniofacial abnormalities (Humbert *et al.*, 2000) and heart defects (Cloud *et al.*, 2002) - all prominent features of CdLS.

Recurrent infections have also been reported at a high frequency in CdLS patients and are a significant cause of mortality. The increased risk of infection in CdLS patients is thought to be a result of decreased percentages of T regulatory cells and T follicular helper cells (Jyonouchi *et al.*, 2013). Cohesin binds sites flanking the enhancer and promoter of the T cell receptor alpha locus (*Tcra*) and is required for long-range promoter-enhancer interactions and H3K4 trimethylation that facilitates *Tcra* rearrangement. *Tcra* rearrangement drives the differentiation of thymocytes into T-cells, and thymocytes deficient in cohesin differentiate with reduced efficiency (Seitan *et al.*, 2011). BRD4 is also critical for normal T-cell development – BRD4 regulates the differentiation of T-cells by promoting the transcription of certain genes (Kagoya *et al.*, 2016; K. Cheung *et al.*, 2017; Gegonne *et al.*, 2018).

Immunological defects are also present in patients with developmental disorders such as LiDS and RS-SCID, which show some overlap in phenotype with CdLS (Davis and Chen, 2013a). In LiDS and RS-SCID the immune defects are the result of impaired NHEJ, which is essential for maturation of immune cells (Chaudhuri and Alt, 2004). The proposed impairment of DNA repair in the Y430C mESCs suggests that mutations in BRD4 could cause immunological defects in a similar way to those seen in patients with LiDS and RS-SCID. Whilst it is plausible that impaired BRD4 function could cause immune defects, these have not been reported in the patients with *BRD4* mutations.

7.6 A model for the role of the Y430C-BRD4 mutation in CdLS

Combining the proposed roles of BRD4 (in the DDR and enhancer transcription) with my own results comparing WT and Y430C mESCs, I can begin to build a model for how the Y430C mutation might cause a CdLS phenotype (Figure Ch7-2). I propose:

- The Y430C mutation causes decreased affinity of BRD4 to acetyl-lysines, affecting the binding of BRD4 to DSBs and to CREs.
- At DSBs the reduced BRD4 occupancy allows an increase in the spread of DDR signalling - potentially through increased accessibility of other DDR proteins to DSBs via chromatin decompaction – and decreases the efficiency of repair, through a delay in the recruitment of the RAD51 and 53BP1.
- In the absence of DSBs the effects of decreased BRD4 binding in the Y430C mESCs may be focussed at enhancers, where BRD4 is usually enriched. The loss of BRD4 at enhancers decreases 53BP1 recruitment, perturbing its 'protective' role against breaks that may occur from convT of eRNAs.
- The resulting increase in DDR signalling leads to increased cell cycle checkpoint activation and a slower cell cycle that, when occurring during development, could cause the CdLS phenotype.

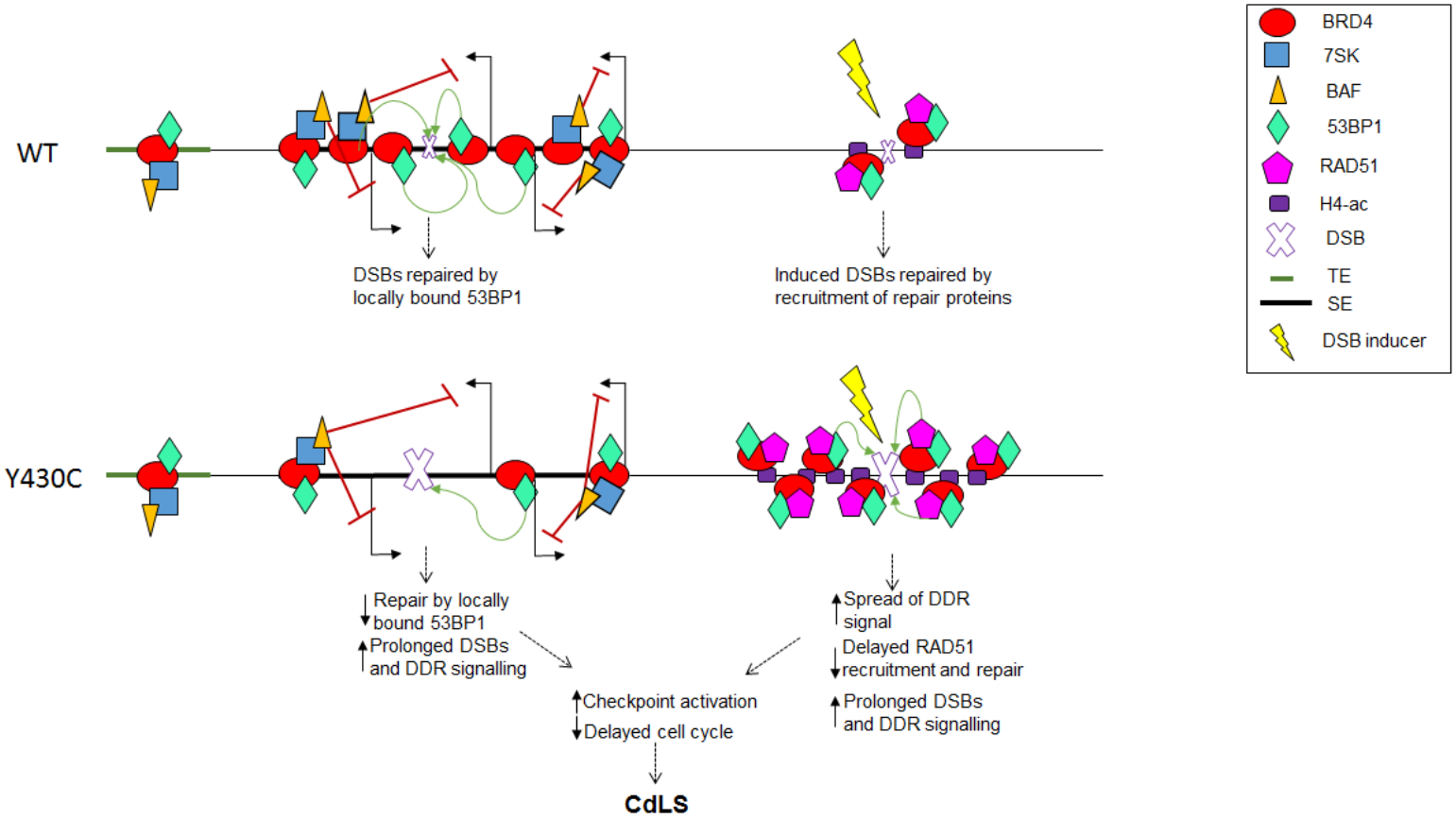


Figure Ch7-2. Model for how BRD4-Y430C may cause CdLS. BRD4 binds to acetyl-lysines in TEs and SEs and recruits the 7SK-BAF complex and 53BP1. 7SK-BAF suppresses pervasive eRNA transcription, whilst 53BP1 promotes the repair of DSBs that do occur.

7.7 A common mechanism for CdLS?

CdLS is a heterogeneous syndrome. Cases caused by mutations in different genes share the main complement of symptoms, although the severity of the phenotype does vary (Boyle *et al.*, 2015). This suggests that the underlying mutations all perturb the function of the same pathway. Originally, due to mutations in the cohesin components, *SMC1*, *SMC3* and *RAD21*, and the cohesin regulators, *NIPBL* and *HDAC8,* CdLS was classified as a cohesinopathy – one of a group of disorders caused by mutations in the cohesin pathway (Skibbens *et al.*, 2013). However, reports associating non-cohesin genes with CdLS (Ansari *et al.*, 2014; Gallagher *et al.*, 2015; Yuan *et al.*, 2015) broadened the aetiology to more than that of a simple cohesinopathy. It is now thought that it is perturbation of cohesin's role in a larger pathway – that of transcription – that underlies the phenotype.

I have suggested that a Y430C mutation in *BRD4* causes CdLS via mis-regulation of the DDR pathway. The patient with this mutation has a classical phenotype similar to those caused by *NIPBL* mutations, with a CdLS-like facial appearance, microcephaly, intellectual disability and cardiac malformation. Limb defects – which are usually associated with the more severe cases caused by *NIPBL* splice-site and nonsense mutations (Boyle *et al.*, 2015) – are absent in the Y430C patient. However, it may be that a splice-site or nonsense *BRD4* mutation would result in a more severe phenotype, similarly to *NIPBL* mutations. Given the strong overlap in phenotypes caused by *NIPBL* and *BRD4* mutations, I would expect the 'classical' cohesin mutations of CdLS to also affect the DDR pathway.

There is a lot of evidence that cohesin is involved in the DDR – it is recruited to DSBs and required for cell cycle checkpoint activation and DNA repair (Frattini *et al.*, 2017; Countryman *et al.*, 2018; Villa-Hernández and Bermejo, 2018) – and there have been reports of increased sensitivity to DNA damage in CdLS patients with *NIPBL* mutations (Vrouwe *et al.*, 2007; Enervald *et al.*, 2013), supporting the idea of misregulated DDR as a common mechanism. Furthermore, my own results in NIPBL deficient LCLs from CdLS patients suggest that there may be an increase in endogenous DDR signalling.

Interestingly NIPBL has also been implicated in the suppression of intergenic or antisense intragenic transcription. *Nipbl* deficient cells showed an increase in ncRNA transcription, which was mainly bidirectional and occurred at poised promoters and active enhancers.

Therefore, the loss of NIPBL binding at enhancers could cause increased eRNA transcription and convT similarly to what I have proposed for BRD4 (Schwarzer *et al.*, 2017).

Recently, mutations in a number of transcriptional regulators were reported in patients with phenotypes overlapping those of CdLS (Ansari *et al.*, 2014; Woods *et al.*, 2014; Izumi *et al.*, 2015; Yuan *et al.*, 2015), promoting the reclassification of CdLS as a transcriptomopathy. These genes do not appear to be involved in the DDR, indicating that these may represent a separate subgroup of CdLS to the BRD4 mutations. The absence of transcriptional dysregulation in the Y430C mESCs could suggest that, whilst the majority of CdLS cases (caused by mutations in *NIPBL*) feature both transcription and DDR misregulation, the less classical cases fall into two distinct categories – those caused by transcriptional mis-regulation and those caused by DDR defects. However, as discussed above, it is also possible that BRD4 mutations also cause transcriptional differences that I am unable to detect.

7.8 Future directions

The 53BP1 ChIP-seq presented in Chapter 6 is based on only 1 replicate and so must be repeated to ensure its reproducibility. The flow cytometry experiments also need to be repeated, to ensure assignation of the cells to their correct cell cycle stage. There are also many other experiments that I would like to have done, that could help substantiate my model or take it further.

7.8.1 Mapping DSBs

The model I have proposed for how the Y430C mutation causes CdLS posits that, in the absence of exogenous damaging agents, DSBs would be enriched at enhancer elements, particularly SEs. Mapping the location of DSBs in the WT and Y430C cells would therefore be a logical next step.

Many protocols have been recently developed to map DSBs, each with its own limitations. Iacovoni *et al.*, used ChIP-seq to map γH2AX at breaks by the *Asi*SI restriction enzyme, but it may not be sensitive enough to identify DSBs at non-sequence specific sites – where they occur at different locations in each cell – and does not detect DSBs directly (Iacovoni *et al.*, 2010).

Methods that map the DSB directly include Genome-wide, Unbiased Identification of DSBs Enabled by Sequencing (GUIDE-seq) (Wang *et al.*, 2015) and integrase-defective lentiviral vector (IDLV)-mediated DNA break capture (Tsai *et al.*, 2015), both of which are technically challenging and detect DSBs through the quantification of NHEJ products, meaning that they may overlook DSBs repaired through alternative pathways (Yan *et al.*, 2017).

Methods such as direct in situ Breaks Labelling, Enrichment on Streptavidin and next-generation Sequencing (BLESS) (Crosetto *et al.*, 2013), and its derivatives END-seq (Canela *et al.*, 2016), DSB capture (Lensing *et al.*, 2016) and most recently Breaks Labeling In Situ and Sequencing (BLISS), map DSBs by ligation of sequencing adaptors to DNA ends. These methods are suitable for studying endogenous DSBs and, with each reiteration, are decreasing in required input and labour intensiveness, and increasing in sensitivity (Yan *et al.*, 2017).

7.8.2 Analysing DSB repair

I have also suggested that DSB repair is perturbed in Y430C mESCs. This was concluded from the reduced number of RAD51 foci 1 hour post DSB induction, and the greater number of 53BP1 and RAD51 foci at 16 and 20 hours post DSB induction, in Y430C mESCs compared to WT, but it would be interesting to investigate this more thoroughly.

Cas9 could be used to introduce a site specific DSB, followed by targeted ChIP-seq to analyse the binding of repair proteins to this region (as in Iacovoni *et al.*, 2010). ChIP-seq would allow the analysis of the spread and kinetics of protein binding. By carrying out ChIP-seq for multiple proteins involved in the different repair pathways, I could determine whether recruitment of proteins from a specific pathway is perturbed.

Cas9 could also be utilised to measure the efficiency of HR versus NHEJ. If WT and Y430C cells are transfected with Cas9-gRNAs and a repair template, the frequencies of different outcomes can be used as a proxy for the efficiency of different repair pathways: incorporation of the repair template represents HR, the presence of small indels represents NHEJ and no change suggests no repair.

7.8.3 BRD4 recruitment to breaks

I have shown that Y430C-BRD4 has reduced affinity for acetyl-lysine residues and that this affects BRD4 occupancy at CREs. It has been shown by others that lysine residues in the

H4 tail are acetylated at sites of DNA damage and that this recruits BRD4 to DSBs (Li *et al.*, 2018). I have extrapolated this and suggested that the Y430C mutation decreases BRD4's recruitment to DSBs, however it would be good to confirm this experimentally. The induction of a site-specific break by Cas9 followed by BRD4 ChIP-qPCR in the WT and Y430C mESCs could be used to test this.

Additionally, ChIP for BAF components, DDR proteins such as 53BP1 and RAD51, and ChIRP for 7SK RNA, would be useful to determine if decreased BRD4 recruitment does affect the further recruitment of these proteins.

7.8.4 Alternative cell lines

My experiments were mainly carried out using mESCs. There is a wealth of datasets, both publicly available and produced within our lab, for the epigenetic profiles of these cells, and their enhancers are well defined (Rao, 2012). Furthermore, they are easy to grow, treat, and manipulate. However, as discussed above, their culture conditions and unique gene expression programme may mask transcriptional differences between the WT and Y430C cells. The use of a differentiated cell line may illuminate any differences. LCLs are often used for studies of patient mutations as they can be derived from patients themselves, rather than engineered to carry the mutation. However, the CdLS patient with the Y430C mutation did not agree to this.

It would be interesting to use cell lines that could inform us about a specific phenotypic feature of CdLS, or a particular perturbation of the DDR. For instance, adipocytes may be used to focus on the low percentage of body fat seen in CdLS patients, or immune lymphocytes to test whether processes such as V(D)J recombination and CSR are affected.

References

Van der Aa, N., Vandeweyer, G. and Kooy, R. F. (2010) 'A boy with mental retardation, obesity and hypertrichosis caused by a microdeletion of 19p13.12', *European Journal of Medical Genetics*, 53(5), pp. 291–293. doi: 10.1016/j.ejmg.2010.05.006.

Acs, K. *et al.* (2011) 'The AAA-ATPase VCP/p97 promotes 53BP1 recruitment by removing L3MBTL1 from DNA double-strand breaks', *Nature Structural and Molecular Biology*. doi: 10.1038/nsmb.2188.

Aladjem, M. I. *et al.* (1998) 'ES cells do not activate p53-dependent stress responses and undergo p53-independent apoptosis in response to DNA damage', *Current Biology*, 8(3), pp. 145–155. doi: 10.1016/S0960-9822(98)70061-2.

Albulescu, L. O. *et al.* (2012) 'A quantitative, high-throughput reverse genetic screen reveals novel connections between pre-mRNA splicing and 5' and 3' end transcript determinants', *PLoS Genetics*, 8(3). doi: 10.1371/journal.pgen.1002530.

Alekseyenko, A. A. *et al.* (2015) 'The oncogenic BRD4-NUT chromatin regulator drives aberrant transcription within large topological domains', *Genes and Development*, 29(14), pp. 1507–1523. doi: 10.1101/gad.267583.115.

Alexander, J. L. and Orr-Weaver, T. L. (2016) 'Replication fork instability and the consequences of fork collisions from rereplication', *Genes and Development*, 30(20), pp. 2241–2252. doi: 10.1101/gad.288142.116.

Ali, M. *et al.* (2012) 'Tandem PHD aingers of MORF/MOZ acetyltransferases display selectivity for acetylated histone H3 and are required for the association with chromatin', *Journal of Molecular Biology*, 424(5), pp. 328–338. doi: 10.1016/j.jmb.2012.10.004.

Allen, B. L. and Taatjes, D. J. (2015) 'The Mediator complex: A central integrator of transcription', *Nature Reviews Molecular Cell Biology*, pp. 155–166. doi: 10.1038/nrm3951.

Altmann, T. and Gennery, A. R. (2016) 'DNA ligase IV syndrome; a review', *Orphanet Journal of Rare Diseases*, pp. 1–7. doi: 10.1186/s13023-016-0520-1.

Andros, C. C. *et al.* (2015) 'A novel application of radiomimetic compounds as antibiotic drugs', *Journal of Pharmacy and Pharmacology*, 67(10), pp. 1371–1379. doi: 10.1111/jphp.12432.

Ansari, M. *et al.* (2014) 'Genetic heterogeneity in Cornelia de Lange syndrome (CdLS) and CdLS-like phenotypes with observed and predicted levels of mosaicism', *Journal of Medical Genetics*, 51(10), pp. 659–668. doi: 10.1136/jmedgenet-2014-102573.

Aparicio, T., Baer, R. and Gautier, J. (2014) 'DNA double-strand break repair pathway choice and cancer', *DNA Repair*, 19, pp. 169–175. doi: 10.1016/j.dnarep.2014.03.014.

Ashley, A. K. *et al.* (2014) 'DNA-PK phosphorylation of RPA32 Ser4/Ser8 regulates replication stress checkpoint activation, fork restart, homologous recombination and mitotic catastrophe', *DNA Repair*, 21, pp. 131–139. doi: 10.1016/j.dnarep.2014.04.008.

Awasthi, P., Foiani, M. and Kumar, A. (2016) 'ATM and ATR signaling at a glance', *Journal of Cell Science*, 129(6), pp. 1285–1285. doi: 10.1242/jcs.188631.

Azofeifa, J. G. *et al.* (2018) 'Enhancer RNA profiling predicts transcription factor activity', *Genome Research*, 28(3), pp. 334–344. doi: 10.1101/gr.225755.117.

Ba, M. *et al.* (2018) 'BRD4 promotes gastric cancer progression through the transcriptional and epigenetic regulation of c-MYC', *Journal of Cellular Biochemistry*, 119(1), pp. 973–982. doi: 10.1002/jcb.26264.

Badoe, E. (2006) 'Classical cornelia de lange syndrome.', *Ghana medical journal*, 40(4), pp. 148–150.

Bailey, T. *et al.* (2013) 'Practical Guidelines for the Comprehensive Analysis of ChIP-seq Data', *PLoS Computational Biology*, 9(11). doi: 10.1371/journal.pcbi.1003326.

Bakhoum, S. F. *et al.* (2017) 'Mitotic DNA Damage Response: At the Crossroads of Structural and Numerical Cancer Chromosome Instabilities', *Trends in Cancer*, pp. 225–234. doi: 10.1016/j.trecan.2017.02.001.

Baldock, R. A. A. *et al.* (2015) 'ATM Localization and Heterochromatin Repair Depend on Direct Interaction of the 53BP1-BRCT2 Domain with γH2AX', *Cell Reports*, 13(10), pp. 2081–2089. doi: 10.1016/j.celrep.2015.10.074.

Ball, A. R., Chen, Y. Y. and Yokomori, K. (2014) 'Mechanisms of cohesin-mediated gene regulation and lessons learned from cohesinopathies', *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, pp. 191–202. doi: 10.1016/j.bbagrm.2013.11.002.

Baranello, L. *et al.* (2016) 'RNA Polymerase II Regulates Topoisomerase 1 Activity to Favor Efficient Transcription', *Cell*, 165(2), pp. 357–371. doi: 10.1016/j.cell.2016.02.036.

Baranovskiy, A. G. *et al.* (2014) 'Structural basis for inhibition of DNA replication by aphidicolin', *Nucleic Acids Research*, 42(22), pp. 14013–14021. doi: 10.1093/nar/gku1209.

Barnum, K. J. and O'Connell, M. J. (2014) 'Cell cycle regulation by checkpoints', *Methods in Molecular Biology*, 1170, pp. 29–40. doi: 10.1007/978-1-4939-0888-2_2.

Barrington, C., Finn, R. and Hadjur, S. (2017) 'Cohesin biology meets the loop extrusion model', *Chromosome Research*, pp. 51–60. doi: 10.1007/s10577-017-9550-3.

Bartek, J. and Lukas, J. (2003) 'Chk1 and Chk2 kinases in checkpoint control and cancer', *Cancer Cell*, pp. 421–429. doi: 10.1016/S1535-6108(03)00110-7.

Beck, B. and Fenger, K. (1985) 'Mortality, Pathological Findings and Causes of Death in the de Lange Syndrome', *Acta Pædiatrica*, 74(5), pp. 765–769. doi: 10.1111/j.1651-2227.1985.tb10028.x.

Beckouët, F. *et al.* (2016) 'Releasing Activity Disengages Cohesin's Smc3/Scc1 Interface in a Process Blocked by Acetylation', *Molecular Cell*, 61(4), pp. 563–574. doi: 10.1016/j.molcel.2016.01.026.

Bensaude, O. (2011) 'Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity?', *Transcription*. doi: 10.4161/trns.2.3.16172.

Berti, M. and Vindigni, A. (2016) 'Replication stress: Getting back on track', *Nature Structural and Molecular Biology*, pp. 103–109. doi: 10.1038/nsmb.3163.

Bettini, L. R. *et al.* (2018) 'Rings and bricks: Expression of cohesin components is dynamic during development and adult life', *International Journal of Molecular Sciences*, 19(2). doi:

10.3390/ijms19020438.

Bhagwat, A. S. *et al.* (2016) 'BET Bromodomain Inhibition Releases the Mediator Complex from Select cis-Regulatory Elements', *Cell Reports*, 15(3), pp. 519–530. doi: 10.1016/j.celrep.2016.03.054.

Bibbo, G. and Piotto, L. (2014) 'Background ionising radiation: A pictorial perspective', *Australasian Physical and Engineering Sciences in Medicine*, 37(3), pp. 575–581. doi: 10.1007/s13246-014-0286-5.

Biswas, U. *et al.* (2016) 'Distinct Roles of Meiosis-Specific Cohesin Complexes in Mammalian Spermatogenesis', *PLoS Genetics*, 12(10). doi: 10.1371/journal.pgen.1006389.

Blackford, A. N. and Jackson, S. P. (2017) 'ATM, ATR, and DNA-PK: The Trinity at the Heart of the DNA Damage Response', *Molecular Cell*, pp. 801–817. doi: 10.1016/j.molcel.2017.05.015.

Bloom, M. S., Koshland, D. and Guacci, V. (2018) 'Cohesin function in cohesion, condensation, and DNA repair is regulated by wpl1p via a common mechanism in Saccharomyces cerevisiae', *Genetics*, 208(1), pp. 111–124. doi: 10.1534/genetics.117.300537.

Bogu, G. K. *et al.* (2016) 'Chromatin and RNA Maps Reveal Regulatory Long Noncoding RNAs in Mouse', *Molecular and Cellular Biology*. doi: 10.1128/MCB.00955-15.

Bohgaki, M. *et al.* (2013) 'RNF168 ubiquitylates 53BP1 and controls its response to DNA double-strand breaks', *Proceedings of the National Academy of Sciences*, 110(52), pp. 20982–20987. doi: 10.1073/pnas.1320302111.

Bonaglia, M. C. *et al.* (2010) 'Genotype-phenotype relationship in three cases with overlapping 19p13.12 microdeletions', *European Journal of Human Genetics*, 18(12), pp. 1302–1309. doi: 10.1038/ejhg.2010.115.

Borde, V. and de Massy, B. (2013) 'Programmed induction of DNA double strand breaks during meiosis: Setting up communication between DNA and the chromosome structure', *Current Opinion in Genetics and Development*, pp. 147–155. doi: 10.1016/j.gde.2012.12.002.

Borrie, M. S. *et al.* (2017) 'Binding, sliding, and function of cohesin during transcriptional activation', *Proceedings of the National Academy of Sciences*, 114(7), pp. E1062–E1071. doi: 10.1073/pnas.1617309114.

Bose, T. and Gerton, J. L. (2010) 'Cohesinopathies, gene expression, and chromatin organization', *Journal of Cell Biology*, pp. 201–210. doi: 10.1083/jcb.200912129.

Bot, C. *et al.* (2017) 'Independent mechanisms recruit the cohesin loader protein NIPBL to sites of DNA damage', *Journal of Cell Science*, p. jcs.197236. doi: 10.1242/jcs.197236.

Boudaoud, I. *et al.* (2017) 'Connected gene communities underlie transcriptional changes in cornelia de lange syndrome', *Genetics*, 207(1), pp. 139–151. doi: 10.1534/genetics.117.202291.

Bouwman, B. A. M. and de Laat, W. (2015) 'Getting the genome in shape: The formation of loops, domains and compartments', *Genome Biology*. doi: 10.1186/s13059-015-0730-1.

Boyle, M. I. *et al.* (2015) 'Cornelia de Lange syndrome', *Clin Genet*, 88(1), pp. 1–12. doi: 10.1111/cge.12499.

Broustas, C. G. and Lieberman, H. B. (2014) 'DNA Damage Response Genes and the

Development of Cancer Metastasis', *Radiation Research*, 181(2), pp. 111–130. doi: 10.1667/RR13515.1.

Brown, E. J. and Baltimore, D. (2000) 'ATR disruption leads to chromosomal fragmentation and early embryonic lethality', *Genes and Development*, 14(4), pp. 397–402. doi: 10.1101/gad.14.4.397.

Brown, J. D. *et al.* (2018) 'BET bromodomain proteins regulate enhancer function during adipogenesis', *Proceedings of the National Academy of Sciences*, p. 201711155. doi: 10.1073/pnas.1711155115.

Bryant, P. E. and Johnston, P. J. (1993) 'Restriction-endonuclease-induced DNA double-strand breaks and chromosomal aberrations in mammalian cells', *Mutation research*, 299(3–4), pp. 289–296.

Buenrostro, J. D. *et al.* (2015) 'ATAC-seq: A method for assaying chromatin accessibility genome-wide', *Current Protocols in Molecular Biology*, 2015, p. 21.29.1-21.29.9. doi: 10.1002/0471142727.mb2129s109.

Bunch, H. *et al.* (2015) 'Transcriptional elongation requires DNA break-induced signalling', *Nature Communications*, 6. doi: 10.1038/ncomms10191.

Bunting, S. F. *et al.* (2010) '53BP1 inhibits homologous recombination in brca1-deficient cells by blocking resection of DNA breaks', *Cell*, 141(2), pp. 243–254. doi: 10.1016/j.cell.2010.03.012.

Busslinger, G. A. *et al.* (2017) 'Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl', *Nature*, 544(7651), pp. 503–507. doi: 10.1038/nature22063.

Calderwood, S. K. (2016) 'A critical role for topoisomerase IIb and DNA double strand breaks in transcription', *Transcription*, 7(3), pp. 75–83. doi: 10.1080/21541264.2016.1181142.

Callén, E. *et al.* (2009) 'Essential Role for DNA-PKcs in DNA Double-Strand Break Repair and Apoptosis in ATM-Deficient Lymphocytes', *Molecular Cell*, 34(3), pp. 285–297. doi: 10.1016/j.molcel.2009.04.025.

Canela, A. *et al.* (2016) 'DNA Breaks and End Resection Measured Genome-wide by End Sequencing', *Molecular Cell*, 63(5), pp. 898–911. doi: 10.1016/j.molcel.2016.06.034.

Caron, P. *et al.* (2015) 'Non-redundant Functions of ATM and DNA-PKcs in Response to DNA Double-Strand Breaks', *Cell Reports*, 13(8), pp. 1598–1609. doi: 10.1016/j.celrep.2015.10.024.

Carroll, T. S. *et al.* (2014) 'Impact of artifact removal on ChIP quality metrics in ChIP-seq and ChIP-exo data', *Frontiers in Genetics*, 5(APR). doi: 10.3389/fgene.2014.00075.

Casamassimi, A. *et al.* (2017) 'Transcriptome profiling in human diseases: New advances and perspectives', *International Journal of Molecular Sciences*. doi: 10.3390/ijms18081652.

Castronovo, P. *et al.* (2009) 'Premature chromatid separation is not a useful diagnostic marker for Cornelia de Lange syndrome', *Chromosome Research*, 17(6), pp. 763–771. doi: 10.1007/s10577-009-9066-6.

Chang, E. Y. C. and Stirling, P. C. (2017) 'Replication fork protection factors controlling R-loop bypass and suppression', *Genes*. doi: 10.3390/genes8010033.

Chang, H. H. Y. *et al.* (2017a) 'Non-homologous DNA end joining and alternative pathways to double-strand break repair', *Nature Reviews Molecular Cell Biology*, 18(8), pp. 495–506. doi: 10.1038/nrm.2017.48.

Chang, H. H. Y. *et al.* (2017b) 'Non-homologous DNA end joining and alternative pathways to double-strand break repair', *Nature Reviews Molecular Cell Biology*, 18(8), pp. 495–506. doi: 10.1038/nrm.2017.48.

Chapuy, B. *et al.* (2013) 'Discovery and Characterization of Super-Enhancer-Associated Dependencies in Diffuse Large B Cell Lymphoma', *Cancer Cell*, 24(6), pp. 777–790. doi: 10.1016/j.ccr.2013.11.003.

Chaudhuri, J. and Alt, F. W. (2004) 'Class-switch recombination: Interplay of transcription, DNA deamination and DNA repair', *Nature Reviews Immunology*, pp. 541–552. doi: 10.1038/nri1395.

Chen, F. *et al.* (2015) 'Stably paused genes revealed through inhibition of transcription initiation by the TFIIH inhibitor triptolide', *Genes and Development*. doi: 10.1101/gad.246173.114.

Chen, H., Lisby, M. and Symington, L. (2013) 'RPA Coordinates DNA End Resection and Prevents Formation of DNA Hairpins', *Molecular Cell*, 50(4), pp. 589–600. doi: 10.1016/j.molcel.2013.04.032.

Chen, K. *et al.* (2016) 'The Overlooked Fact: Fundamental Need for Spike-In Control for Virtually All Genome-Wide Analyses', *Molecular and Cellular Biology*, 36(5), pp. 662–667. doi: 10.1128/MCB.00970-14.

Chen, Y. *et al.* (2012) 'Systematic evaluation of factors influencing ChIP-seq fidelity', *Nature Methods*, 9(6), pp. 609–614. doi: 10.1038/nmeth.1985.

Cheung, K. *et al.* (2017) 'BET N-terminal bromodomain inhibition selectively blocks Th17 cell differentiation and ameliorates colitis in mice', *Proceedings of the National Academy of Sciences*, 114(11), pp. 2952–2957. doi: 10.1073/pnas.1615601114.

Cheung, K. L. *et al.* (2017) 'Distinct Roles of Brd2 and Brd4 in Potentiating the Transcriptional Program for Th17 Cell Differentiation', *Molecular Cell*, 65(6), p. 1068–1080.e5. doi: 10.1016/j.molcel.2016.12.022.

Ciemerych, M. A. *et al.* (2002) 'Development of mice expressing a single D-type cyclin', *Genes and Development*, 16(24), pp. 3277–3289. doi: 10.1101/gad.1023602.

Citterio, E. (2015) 'Fine-tuning the ubiquitin code at DNA double-strand breaks: Deubiquitinating enzymes at work', *Frontiers in Genetics*. doi: 10.3389/fgene.2015.00282.

Clapier, C. R. and Cairns, B. R. (2009a) 'The Biology of Chromatin Remodeling Complexes', *Annual Review of Biochemistry*, 78(1), pp. 273–304. doi: 10.1146/annurev.biochem.77.062706.153223.

Clapier, C. R. and Cairns, B. R. (2009b) 'The Biology of Chromatin Remodeling Complexes', *Annual Review of Biochemistry*, 78(1), pp. 273–304. doi: 10.1146/annurev.biochem.77.062706.153223.

Clark, S. L. *et al.* (2012) 'STRUCTURE-FUNCTION OF THE TUMOR SUPPRESSOR BRCA1', *Computational and Structural Biotechnology Journal*, 1(1), p. e201204005. doi:

10.5936/csbj.201204005.

Cliften, P. (2014) 'Base Calling, Read Mapping, and Coverage Analysis', in *Clinical Genomics*, pp. 91–107. doi: 10.1016/B978-0-12-404748-8.00007-1.

Clouaire, T. and Legube, G. (2015) 'DNA double strand break repair pathway choice: a chromatin based decision?', *Nucleus (Austin, Tex.)*, 6(2), pp. 107–13. doi: 10.1080/19491034.2015.1010946.

Cloud, J. E. *et al.* (2002) 'Mutant Mouse Models Reveal the Relative Roles of E2F1 and E2F3 In Vivo', *Molecular and Cellular Biology*, 22(8), pp. 2663–2672. doi: 10.1128/MCB.22.8.2663-2672.2002.

Countryman, P. *et al.* (2018) 'Cohesin SA2 is a sequence-independent DNA-binding protein that recognizes DNA replication and repair intermediates', *Journal of Biological Chemistry*, 293(3), pp. 1054–1069. doi: 10.1074/jbc.M117.806406.

Cox, J. *et al.* (2014) 'Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ', *Molecular & Cellular Proteomics*, 13(9), pp. 2513–2526. doi: 10.1074/mcp.M113.031591.

Crosetto, N. *et al.* (2013) 'Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing', *Nature Methods*, 10(4), pp. 361–365. doi: 10.1038/nmeth.2408.

Cucinotta, C. E. and Arndt, K. M. (2016) 'SnapShot: Transcription Elongation', *Cell*, p. 1058–1058.e1. doi: 10.1016/j.cell.2016.07.039.

Cuella-Martin, R. *et al.* (2016) '53BP1 Integrates DNA Repair and p53-Dependent Cell Fate Decisions via Distinct Mechanisms', *Molecular Cell*, 64(1), pp. 51–64. doi: 10.1016/j.molcel.2016.08.002.

D'Adda Di Fagagna, F. (2008) 'Living on a break: Cellular senescence as a DNA-damage response', *Nature Reviews Cancer*, pp. 512–522. doi: 10.1038/nrc2440.

D'Alessandro, G. and d'Adda di Fagagna, F. (2017) 'Transcription and DNA Damage: Holding Hands or Crossing Swords?', *Journal of Molecular Biology*, pp. 3215–3229. doi: 10.1016/j.jmb.2016.11.002.

Daniel, B. *et al.* (2014) 'Mapping the genomic binding sites of the activated retinoid X receptor in murine bone marrow-derived macrophages using chromatin immunoprecipitation sequencing', *Methods in Molecular Biology*, 1204, pp. 15–24. doi: 10.1007/978-1-4939-1346-6_2.

Dasgupta, T. *et al.* (2016) 'HDAC8 Inhibition Blocks SMC3 Deacetylation and Delays Cell Cycle Progression without Affecting Cohesin-dependent Transcription in MCF7 Cancer Cells', *Journal of Biological Chemistry*, 291(24), pp. 12761–12770. doi: 10.1074/jbc.M115.704627.

Davis, A. J., Chen, B. P. C. and Chen, D. J. (2014) 'DNA-PK: A dynamic enzyme in a versatile DSB repair pathway', *DNA Repair*, 17, pp. 21–29. doi: 10.1016/j.dnarep.2014.02.020.

Davis, A. J. and Chen, D. J. (2013a) 'DNA double strand break repair via non-homologous end-joining.', *Translational cancer research*, 2(3), pp. 130–143. doi: 10.3978/j.issn.2218-676X.2013.04.02.

Davis, A. J. and Chen, D. J. (2013b) 'DNA double strand break repair via non-homologous

end-joining.', *Translational cancer research*, 2(3), pp. 130–143. doi: 10.3978/j.issn.2218-676X.2013.04.02.

Deardorff, M. A. *et al.* (2007) 'Mutations in Cohesin Complex Members SMC3 and SMC1A Cause a Mild Variant of Cornelia de Lange Syndrome with Predominant Mental Retardation', *The American Journal of Human Genetics*, 80(3), pp. 485–494. doi: 10.1086/511888.

Deardorff, M. A., Bando, M., *et al.* (2012) 'HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle', *Nature*, 489(7415), pp. 313–317. doi: 10.1038/nature11316.

Deardorff, M. A., Wilde, J. J., *et al.* (2012) 'RAD21 mutations cause a human cohesinopathy', *American Journal of Human Genetics*, 90(6), pp. 1014–1027. doi: 10.1016/j.ajhg.2012.04.019.

Decker, T. M. *et al.* (2017) 'Transcriptome analysis of dominant-negative Brd4 mutants identifies Brd4-specific target genes of small molecule inhibitor JQ1', *Scientific Reports*, 7(1). doi: 10.1038/s41598-017-01943-6.

Densham, R. M. and Morris, J. R. (2017) 'The BRCA1 Ubiquitin ligase function sets a new trend for remodelling in DNA repair', *Nucleus*, 8(2), pp. 116–125. doi: 10.1080/19491034.2016.1267092.

Devaiah, B. N. *et al.* (2012) 'BRD4 is an atypical kinase that phosphorylates Serine2 of the RNA Polymerase II carboxy-terminal domain', *Proceedings of the National Academy of Sciences*, 109(18), pp. 6927–6932. doi: 10.1073/pnas.1120422109.

Dhar, S. *et al.* (2017) 'The tale of a tail: Histone H4 acetylation and the repair of DNA breaks', *Philosophical Transactions of the Royal Society B: Biological Sciences*. doi: 10.1098/rstb.2016.0284.

Dong, X. *et al.* (2018) 'BRD4 regulates cellular senescence in gastric cancer cells via E2F/miR-106b/p21 axis', *Cell Death and Disease*, 9(2). doi: 10.1038/s41419-017-0181-6.

Donzelli, M. and Draetta, G. F. (2003) 'Regulating mammalian checkpoints through Cdc25 inactivation', *EMBO Reports*. doi: 10.1038/sj.embor.embor887.

Dorsett, D. and Krantz, I. D. (2009) 'On the molecular etiology of Cornelia de Lange syndrome', *Annals of the New York Academy of Sciences*, pp. 22–37. doi: 10.1111/j.1749-6632.2008.03450.x.

Drabik, A., Ciborowski, P. and Silberring, J. (2013) 'Quantitative Measurements in Proteomics', in *Proteomic Profiling and Analytical Chemistry*, pp. 135–150. doi: 10.1016/B978-0-444-59378-8.00007-4.

Drané, P. *et al.* (2017) 'TIRR regulates 53BP1 by masking its histone methyl-lysine binding function', *Nature*. doi: 10.1038/nature21358.

Eden, E. *et al.* (2007) 'Discovering motifs in ranked lists of DNA sequences', *PLoS Computational Biology*, 3(3), pp. 0508–0522. doi: 10.1371/journal.pcbi.0030039.

Eden, E. *et al.* (2009) 'GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists', *BMC Bioinformatics*, 10. doi: 10.1186/1471-2105-10-48.

Efroni, S. *et al.* (2008) 'Global Transcription in Pluripotent Embryonic Stem Cells', *Cell Stem Cell*. doi: 10.1016/j.stem.2008.03.021.

Elbatsh, A. M. O. *et al.* (2016) 'Cohesin Releases DNA through Asymmetric ATPase-Driven

Ring Opening', *Molecular Cell*, 61(4), pp. 575–588. doi: 10.1016/j.molcel.2016.01.025.

Enervald, E. *et al.* (2013) 'A regulatory role for the cohesin loader NIPBL in nonhomologous end joining during immunoglobulin class switch recombination', *J Exp Med*, 210(12), pp. 2503–2513. doi: 10.1084/jem.20130168.

Engels, H. *et al.* (2007) 'DNA microarray analysis identifies candidate regions and genes in unexplained mental retardation', *Neurology*, 68(10), pp. 743–750. doi: 10.1212/01.wnl.0000256367.70365.e0.

Ernst, J. and Kellis, M. (2012) 'ChromHMM: Automating chromatin-state discovery and characterization', *Nature Methods*, pp. 215–216. doi: 10.1038/nmeth.1906.

Falck, J., Coates, J. and Jackson, S. P. (2005) 'Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage', *Nature*, 434(7033), pp. 605–611. doi: 10.1038/nature03442.

Fernandez-Capetillo, O. *et al.* (2002) 'DNA damage-induced G2-M checkpoint activation by histone H2AX and 53BP1', *Nature Cell Biology*, 4(12), pp. 993–997. doi: 10.1038/ncb884.

Ferri, E., Petosa, C. and McKenna, C. E. (2016) 'Bromodomains: Structure, function and pharmacology of inhibition', *Biochemical Pharmacology*, pp. 1–18. doi: 10.1016/j.bcp.2015.12.005.

Filippakopoulos, P. *et al.* (2010) 'Selective inhibition of BET bromodomains', *Nature*, 468(7327), pp. 1067–1073. doi: 10.1038/nature09504.

Filippakopoulos, P. *et al.* (2012) 'Histone recognition and large-scale structural analysis of the human bromodomain family', *Cell*, 149(1), pp. 214–231. doi: 10.1016/j.cell.2012.02.013.

Floyd, S. R. *et al.* (2013) 'The bromodomain protein Brd4 insulates chromatin from DNA damage signalling', *Nature*, 498(7453), pp. 246–250. doi: 10.1038/nature12147.

Flynn, R. A. *et al.* (2016) '7SK-BAF axis controls pervasive transcription at enhancers', *Nature Structural and Molecular Biology*, 23(3), pp. 231–238. doi: 10.1038/nsmb.3176.

Frattini, C. *et al.* (2017) 'Cohesin Ubiquitylation and Mobilization Facilitate Stalled Replication Fork Dynamics', *Molecular Cell*, 68(4), p. 758–772.e4. doi: 10.1016/j.molcel.2017.10.012.

Fujisawa, T. and Filippakopoulos, P. (2017) 'Functions of bromodomain-containing proteins and their roles in homeostasis and cancer', *Nature Reviews Molecular Cell Biology*, pp. 246–262. doi: 10.1038/nrm.2016.143.

Gade, P. and Kalvakolanu, D. V (2012a) 'Chromatin immunoprecipitation assay as a tool for analyzing transcription factor activity.', *Methods in molecular biology (Clifton, N.J.)*, 809, pp. 85–104. doi: 10.1007/978-1-61779-376-9_6.

Gade, P. and Kalvakolanu, D. V (2012b) 'Chromatin immunoprecipitation assay as a tool for analyzing transcription factor activity.', *Methods in molecular biology (Clifton, N.J.)*, 809, pp. 85–104. doi: 10.1007/978-1-61779-376-9_6.

Galdeano, C. and Ciulli, A. (2016) 'Selectivity on-target of bromodomain chemical probes by structure-guided medicinal chemistry and chemical biology', *Future Medicinal Chemistry*, pp. 1655–1680. doi: 10.4155/fmc-2016-0059.

Gallagher, D. *et al.* (2015) 'Ankrd11 is a chromatin regulator involved in autism that is essential for neural development', *Developmental Cell*, 32(1), pp. 31–42. doi: 10.1016/j.devcel.2014.11.031.

Gallant, N. M. *et al.* (2011) 'Pontocerebellar hypoplasia in association with de novo 19p13.11p13.12 microdeletion', *American Journal of Medical Genetics, Part A*, 155(11), pp. 2871–2878. doi: 10.1002/ajmg.a.34286.

Gan, W. *et al.* (2011) 'R-loop-mediated genomic instability is caused by impairment of replication fork progression', *Genes and Development*, 25(19), pp. 2041–2056. doi: 10.1101/gad.17010011.

Gard, S. *et al.* (2009) 'Cohesinopathy mutations disrupt the subnuclear organization of chromatin (The Journal of Cell Biology (2009) 187, 4, (455-462))', *Journal of Cell Biology*, p. 749. doi: 10.1083/jcb.20090607520091116c.

Gegonne, A. *et al.* (2018) 'Immature CD8 Single-Positive Thymocytes Are a Molecularly Distinct Subpopulation, Selectively Dependent on BRD4 for Their Differentiation', *Cell Reports*, 24(1), pp. 117–129. doi: 10.1016/j.celrep.2018.06.007.

Gel, B. *et al.* (2016) 'regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests.', *Bioinformatics (Oxford, England)*, 32(2), pp. 289–91. doi: 10.1093/bioinformatics/btv562.

Gelot, C. *et al.* (2016) 'The Cohesin Complex Prevents the End Joining of Distant DNA Double-Strand Ends', *Molecular Cell*, 61(1), pp. 15–26. doi: 10.1016/j.molcel.2015.11.002.

Gillette, T. G. and Hill, J. A. (2015) 'Readers, writers, and erasers: Chromatin as the whiteboard of heart disease', *Circulation Research*, pp. 1245–1253. doi: 10.1161/CIRCRESAHA.116.303630.

Gillis, L. A. *et al.* (2004) 'NIPBL Mutational Analysis in 120 Individuals with Cornelia de Lange Syndrome and Evaluation of Genotype-Phenotype Correlations', *The American Journal of Human Genetics*, 75(4), pp. 610–623. doi: 10.1086/424698.

Goldstein, M. and Kastan, M. B. (2015) 'The DNA Damage Response: Implications for Tumor Responses to Radiation and Chemotherapy', *Annual Review of Medicine*, 66(1), pp. 129–143. doi: 10.1146/annurev-med-081313-121208.

Gong, Y. *et al.* (2018) 'Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries', *Nature Communications*, 9(1). doi: 10.1038/s41467-018-03017-1.

Gonzales-Cope, M. *et al.* (2016) 'Histone H4 acetylation and the epigenetic reader Brd4 are critical regulators of pluripotency in embryonic stem cells', *BMC Genomics*, 17(1). doi: 10.1186/s12864-016-2414-y.

Guleria, A. and Chandna, S. (2016) 'ATM kinase: Much more than a DNA damage responsive protein', *DNA Repair*, pp. 1–20. doi: 10.1016/j.dnarep.2015.12.009.

Gunjan, A., Paik, J. and Verreault, A. (2005) 'Regulation of histone synthesis and nucleosome assembly', in *Biochimie*, pp. 625–635. doi: 10.1016/j.biochi.2005.02.008.

Haarhuis, J. H. I. *et al.* (2013) 'WAPL-mediated removal of cohesin protects against segregation errors and aneuploidy', *Current Biology*, 23(20), pp. 2071–2077. doi:

10.1016/j.cub.2013.09.003.

Haarhuis, J. H. I. *et al.* (2017) 'The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension', *Cell*, 169(4), p. 693–707.e14. doi: 10.1016/j.cell.2017.04.013.

Hainer, S. J. *et al.* (2015) 'Suppression of pervasive noncoding transcription in embryonic stem cells by esBAF', *Genes and Development*, 29(4), pp. 362–378. doi: 10.1101/gad.253534.114.

Hajmirza, A. *et al.* (2018) 'BET Family Protein BRD4: An Emerging Actor in NFκB Signaling in Inflammation and Cancer', *Biomedicines*, 6(1), p. 16. doi: 10.3390/biomedicines6010016.

Hanamshet, K., Mazina, O. and Mazin, A. (2016) 'Reappearance from Obscurity: Mammalian Rad52 in Homologous Recombination', *Genes*. Multidisciplinary Digital Publishing Institute, 7(9), p. 63. doi: 10.3390/genes7090063.

Harrigan, J. A. *et al.* (2011) 'Replication stress induces 53BP1-containing OPT domains in G1 cells', *Journal of Cell Biology*, 193(1), pp. 97–108. doi: 10.1083/jcb.201011083.

Harris, M. E. *et al.* (1991) 'Regulation of Histone mRNA in the Unperturbed Cell Cycle: Evidence Suggesting Control at Two Posttranscriptional Steps', *MOLECULAR AND CELLULAR BIOLOGY*, 11(5), pp. 2416–2424. doi: 10.1128/MCB.11.5.2416.Updated.

Hartlerode, A. J. *et al.* (2012) 'Impact of Histone H4 Lysine 20 Methylation on 53BP1 Responses to Chromosomal Double Strand Breaks', *PLoS ONE*, 7(11). doi: 10.1371/journal.pone.0049211.

Hartlerode, A. J. *et al.* (2015) 'Recruitment and activation of the ATM kinase in the absence of DNA-damage sensors', *Nature Structural and Molecular Biology*, 22(9), pp. 736–743. doi: 10.1038/nsmb.3072.

Hauf, S., Waizenegger, I. C. and Peters, J. M. (2001) 'Cohesin cleavage by separase required for anaphase and cytokinesis in human cells', *Science*, 293(5533), pp. 1320–1323. doi: 10.1126/science.1061376.

Hay, D. *et al.* (2016) 'Genetic dissection of the α-globin super-enhancer in vivo', *Nature Genetics*, 48(8), pp. 895–903. doi: 10.1038/ng.3605.

He, G. *et al.* (2005) 'Induction of p21 by p53 following DNA damage inhibits both Cdk4 and Cdk2 activities', *Oncogene*. doi: 10.1038/sj.onc.1208474.

Heidinger-Pauli, J. M. *et al.* (2010) 'Systematic Reduction of Cohesin Differentially Affects Chromosome Segregation, Condensation, and DNA Repair', *Current Biology*. doi: 10.1016/j.cub.2010.04.018.

Her, J. and Bunting, S. F. (2018) 'How cells ensure correct repair of DNA double-strand breaks', *Journal of Biological Chemistry*, p. jbc.TM118.000371. doi: 10.1074/jbc.TM118.000371.

Hinshaw, S. M. *et al.* (2015) 'Structural evidence for Scc4-dependent localization of cohesin loading', *eLife*, 4(JUNE), pp. 1–15. doi: 10.7554/eLife.06057.

Hinshaw, S. M. *et al.* (2017) 'The Kinetochore Receptor for the Cohesin Loading Complex', *Cell*, 171(1), p. 72–84.e13. doi: 10.1016/j.cell.2017.08.017.

Hnisz, D. *et al.* (2013) 'Super-enhancers in the control of cell identity and disease.', *Cell*, 155(4), pp. 934–47. doi: 10.1016/j.cell.2013.09.053.

Hoeijmakers, J. H. J. (2009) 'DNA damage, aging, and cancer.', *The New England journal of medicine*, 361(15), pp. 1475–85. doi: 10.1056/NEJMra0804615.

Horne, G. A. *et al.* (2015) 'Nanog Requires BRD4 to Maintain Murine Embryonic Stem Cell Pluripotency and Is Suppressed by Bromodomain Inhibitor JQ1 Together with Lefty1', *Stem Cells and Development*, 24(7), pp. 879–891. doi: 10.1089/scd.2014.0302.

House, N. C. M., Koch, M. R. and Freudenreich, C. H. (2014) 'Chromatin modifications and DNA repair: Beyond double-strand breaks', *Frontiers in Genetics*, 5(SEP), pp. 1–18. doi: 10.3389/fgene.2014.00296.

Houzelstein, D. *et al.* (2002) 'Growth and Early Postimplantation Defects in Mice Deficient for the Bromodomain-Containing Protein Brd4 ', *Molecular and Cellular Biology*, 22(11), pp. 3794–3802. doi: 10.1128/MCB.22.11.3794-3802.2002.

Hsieh, C.-L. *et al.* (2014) 'Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation', *Proceedings of the National Academy of Sciences*, 111(20), pp. 7319–7324. doi: 10.1073/pnas.1324151111.

Hsu, C. C. *et al.* (2018) 'Recognition of histone acetylation by the GAS41 YEATS domain promotes H2A.Z deposition in non-small cell lung cancer', *Genes and Development*, 32(1), pp. 58–69. doi: 10.1101/gad.303784.117.

Huang, D. W., Sherman, B. T. and Lempicki, R. A. (2009a) 'Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists', *Nucleic Acids Research*, 37(1), pp. 1–13. doi: 10.1093/nar/gkn923.

Huang, D. W., Sherman, B. T. and Lempicki, R. A. (2009b) 'Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources', *Nature Protocols*, 4(1), pp. 44–57. doi: 10.1038/nprot.2008.211.

Hui-Yuen, J. *et al.* (2011) 'Establishment of Epstein-Barr Virus Growth-transformed Lymphoblastoid Cell Lines', *Journal of Visualized Experiments*. doi: 10.3791/3321.

Huisman, S. A. *et al.* (2013) 'High rate of mosaicism in individuals with Cornelia de Lange syndrome', *Journal of Medical Genetics*, 50(5), pp. 339–344. doi: 10.1136/jmedgenet-2012-101477.

Humbert, P. O. *et al.* (2000) 'E2F4 is essential for normal erythrocyte maturation and neonatal viability', *Molecular Cell*, 6(2), pp. 281–291. doi: 10.1016/S1097-2765(00)00029-0.

Hussong, M. *et al.* (2017) 'The bromodomain protein BRD4 regulates splicing during heat shock', *Nucleic Acids Research*, 45(1), pp. 382–394. doi: 10.1093/nar/gkw729.

Huyen, Y. *et al.* (2004) 'Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks', *Nature*, 432(7015), pp. 406–411. doi: 10.1038/nature03114.

Hwang, J. K., Alt, F. W. and Yeap, L.-S. (2015) 'Related Mechanisms of Antibody Somatic Hypermutation and Class Switch Recombination.', *Microbiology spectrum*, 3(1), pp. MDNA3-0037-2014. doi: 10.1128/microbiolspec.MDNA3-0037-2014.

Iacovoni, J. S. *et al.* (2010) 'High-resolution profiling of γh2AX around DNA double strand breaks in the mammalian genome', *EMBO Journal*, 29(8), pp. 1446–1457. doi: 10.1038/emboj.2010.38.

Ikura, T. *et al.* (2007a) 'DNA Damage-Dependent Acetylation and Ubiquitination of H2AX Enhances Chromatin Dynamics', *Molecular and Cellular Biology*, 27(20), pp. 7028–7040. doi: 10.1128/MCB.00579-07.

Ikura, T. *et al.* (2007b) 'DNA Damage-Dependent Acetylation and Ubiquitination of H2AX Enhances Chromatin Dynamics', *Molecular and Cellular Biology*, 27(20), pp. 7028–7040. doi: 10.1128/MCB.00579-07.

Illumina (2011) 'Quality Scores for Next-Generation Sequencing', *Http://Res.Illumina.Com/Documents/Products/Technotes/Technote_Q-Scores.Pdf*, pp. 1–2.

Iyer, D. R. and Rhind, N. (2017) 'The intra-S checkpoint responses to DNA damage', *Genes*. doi: 10.3390/genes8020074.

Izumi, K. *et al.* (2015) 'Germline gain-of-function mutations in AFF4 cause a developmental syndrome functionally linking the super elongation complex and cohesin', *Nature Genetics*, 47(4), pp. 338–344. doi: 10.1038/ng.3229.

Jahnke, P. *et al.* (2008) 'The Cohesin loading factor NIPBL recruits histone deacetylases to mediate local chromatin modifications', *Nucleic Acids Research*, 36(20), pp. 6450–6458. doi: 10.1093/nar/gkn688.

Jang, M. K. *et al.* (2005) 'The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription', *Molecular Cell*, 19(4), pp. 523–534. doi: 10.1016/j.molcel.2005.06.027.

Jang, M. K., Shen, K. and McBride, A. A. (2014) 'Papillomavirus Genomes Associate with BRD4 to Replicate at Fragile Sites in the Host Genome', *PLoS Pathogens*, 10(5). doi: 10.1371/journal.ppat.1004117.

Jelsig, A. M. *et al.* (2012) 'A case of microdeletion of 19p13 with intellectual disability, hypertrichosis, synophrys, and protruding front teeth', *European Journal of Medical Genetics*, 55(10), pp. 564–567. doi: 10.1016/j.ejmg.2012.06.009.

Jensen, D. R. *et al.* (2009) 'A novel chromosome 19p13.12 deletion in a child with multiple congenital anomalies', *American Journal of Medical Genetics, Part A*, 149(3), pp. 396–402. doi: 10.1002/ajmg.a.32691.

Jette, N. and Lees-Miller, S. P. (2015) 'The DNA-dependent protein kinase: A multifunctional protein kinase with roles in DNA double strand break repair and mitosis', *Progress in Biophysics and Molecular Biology*, pp. 194–205. doi: 10.1016/j.pbiomolbio.2014.12.003.

Johnson, D. S. *et al.* (2007) 'Genome-wide mapping of in vivo protein-DNA interactions', *Science*, 316(5830), pp. 1497–1502. doi: 10.1126/science.1141319.

Ju, B. G. *et al.* (2006) 'A topoisomerase IIβ-mediated dsDNA break required for regulated transcription', *Science*, 312(5781), pp. 1798–1802. doi: 10.1126/science.1127196.

Jung, M. *et al.* (2014) 'Affinity map of bromodomain protein 4 (BRD4) interactions with the histone H4 tail and the small molecule inhibitor JQ1', *Journal of Biological Chemistry*, 289(13), pp. 9304–9319. doi: 10.1074/jbc.M113.523019.

Jyonouchi, S. *et al.* (2013) 'Immunologic Features of Cornelia de Lange Syndrome', *PEDIATRICS*. doi: 10.1542/peds.2012-3815.

Kagey, M. H. *et al.* (2010) 'Mediator and cohesin connect gene expression and chromatin architecture', *Nature*, 467(7314), pp. 430–435. doi: 10.1038/nature09380.

Kagoya, Y. *et al.* (2016) 'BET bromodomain inhibition enhances T cell persistence and function in adoptive immunotherapy models', *Journal of Clinical Investigation*, 126(9), pp. 3479–3494. doi: 10.1172/JCI86437.

Kanno, T. *et al.* (2014) 'BRD4 assists elongation of both coding and enhancer RNAs by interacting with acetylated histones', *Nature Structural and Molecular Biology*, 21(12), pp. 1047–1057. doi: 10.1038/nsmb.2912.

Karanika, S. *et al.* (2015) 'DNA damage response and prostate cancer: Defects, regulation and therapeutic implications', *Oncogene*, pp. 2815–2822. doi: 10.1038/onc.2014.238.

Kari, V. *et al.* (2016) 'Loss of CHD1 causes DNA repair defects and enhances prostate cancer therapeutic responsiveness', *EMBO Reports*, 17, pp. 1609–1623. doi: 10.15252/embr.

Kawauchi, S. *et al.* (2009) 'Multiple organ system defects and transcriptional dysregulation in the Nipbl(+/-) mouse, a model of Cornelia de Lange Syndrome.', *PLoS genetics*, 5(9), p. e1000650. doi: 10.1371/journal.pgen.1000650.

Kawauchi, S. *et al.* (2016) 'Using mouse and zebrafish models to understand the etiology of developmental defects in Cornelia de Lange Syndrome', *American Journal of Medical Genetics, Part C: Seminars in Medical Genetics*, 172(2), pp. 138–145. doi: 10.1002/ajmg.c.31484.

Kaygun, H. and Marzluff, W. F. (2005) 'Regulated degradation of replication-dependent histone mRNAs requires both ATR and Upf1', *Nature Structural and Molecular Biology*, 12(9), pp. 794–800. doi: 10.1038/nsmb972.

Khan, F. A. and Ali, S. O. (2017) 'Physiological Roles of DNA Double-Strand Breaks', *Journal of Nucleic Acids*. doi: 10.1155/2017/6439169.

Kim, T. K. *et al.* (2010) 'Widespread transcription at neuronal activity-regulated enhancers', *Nature*, 465(7295), pp. 182–187. doi: 10.1038/nature09033.

Kitajima, T. S., Kawashima, S. A. and Watanabe, Y. (2004) 'The conserved kinetochore protein shugoshin protects centromeric cohesion during meiosis', *Nature*. doi: 10.1038/nature02312.

Klapstein, K., Chou, T. and Bruinsma, R. (2004) 'Physics of RecA-mediated homologous recognition', *Biophysical Journal*, 87(3), pp. 1466–1477. doi: 10.1529/biophysj.104.039578.

De Klein, A. *et al.* (2000) 'Targeted disruption of the cell-cycle checkpoint gene ATR leads to early embryonic lethality in mice', *Current Biology*, 10(8), pp. 479–482. doi: 10.1016/S0960-9822(00)00447-4.

Klein, R. H. *et al.* (2017) 'Characterization of enhancers and the role of the transcription factor KLF7 in regulating corneal epithelial differentiation', *Journal of Biological Chemistry*, 292(46), pp. 18937–18950. doi: 10.1074/jbc.M117.793117.

Kline, A. D. *et al.* (2007) 'Cornelia de Lange syndrome: Clinical review, diagnostic and scoring systems, and anticipatory guidance', *American Journal of Medical Genetics, Part A*, pp. 1287–1296. doi: 10.1002/ajmg.a.31757.

Kline, A. D. *et al.* (2017) 'Cornelia de Lange syndrome and molecular implications of the

cohesin complex: Abstracts from the 7th biennial scientific and educational symposium 2016', in *American Journal of Medical Genetics, Part A*. doi: 10.1002/ajmg.a.38161.

Knijnenburg, T. A. *et al.* (2018) 'Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas', *Cell Reports*, 23(1), p. 239–254.e6. doi: 10.1016/j.celrep.2018.03.076.

Kocyłowski, M. K. *et al.* (2015) 'Ubiquitin-H2AX fusions render 53BP1 recruitment to DNA damage sites independent of RNF8 or RNF168', *Cell Cycle*, 14(11), pp. 1748–1758. doi: 10.1080/15384101.2015.1010918.

Koprinarova, M., Schnekenburger, M. and Diederich, M. (2015) 'Role of Histone Acetylation in Cell Cycle Regulation', *Current Topics in Medicinal Chemistry*, 16(7), pp. 732–744. doi: 10.2174/1568026615666150825140822.

Korb, E. *et al.* (2015) 'BET protein Brd4 activates transcription in neurons and BET inhibitor Jq1 blocks memory in mice', *Nature Neuroscience*, 18(10), pp. 1464–1473. doi: 10.1038/nn.4095.

Kosaki, K. *et al.* (2011) 'Branchial arch defects and 19p13.12 microdeletion: Defining the critical region into a 0.8M base interval', *American Journal of Medical Genetics, Part A*, 155(9), pp. 2212–2214. doi: 10.1002/ajmg.a.33908.

Krantz, I. D. *et al.* (2004) 'Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B', *Nat Genet*, 36(6), pp. 631–635. doi: 10.1038/ng1364.

Kurat, C. F. *et al.* (2011) 'Restriction of histone gene transcription to S phase by phosphorylation of a chromatin boundary protein', *Genes and Development*, 25(23), pp. 2489–2501. doi: 10.1101/gad.173427.111.

Kurimasa, A. *et al.* (1999) 'Requirement for the kinase activity of human DNA-dependent protein kinase catalytic subunit in DNA strand break rejoining.', *Molecular and cellular biology*, 19(5), pp. 3877–84. doi: 10.1128/MCB.19.5.3877.

Landt, S. G. *et al.* (2012) 'ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia', *Genome Research*, pp. 1813–1831. doi: 10.1101/gr.136184.111.

Langmead, B. and Salzberg, S. L. (2012) 'Fast gapped-read alignment with Bowtie 2', *Nature Methods*, 9(4), pp. 357–359. doi: 10.1038/nmeth.1923.

Lavin, M. F. *et al.* (2015) 'ATM-dependent phosphorylation of all three members of the MRN complex: From sensor to adaptor', *Biomolecules*, pp. 2877–2902. doi: 10.3390/biom5042877.

Lee, H.-S. *et al.* (2010a) 'A cooperative activation loop among SWI/SNF, gamma-H2AX and H3 acetylation for DNA double-strand break repair.', *The EMBO journal*, 29(8), pp. 1434–1445. doi: 10.1038/emboj.2010.27.

Lee, H.-S. *et al.* (2010b) 'A cooperative activation loop among SWI/SNF, gamma-H2AX and H3 acetylation for DNA double-strand break repair.', *The EMBO journal*, 29(8), pp. 1434–1445. doi: 10.1038/emboj.2010.27.

Lensing, S. V. *et al.* (2016) 'DSBCapture: In situ capture and sequencing of DNA breaks', *Nature Methods*, 13(10), pp. 855–857. doi: 10.1038/nmeth.3960.

Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25(16), pp. 2078–2079. doi: 10.1093/bioinformatics/btp352.

Li, X. *et al.* (2018) 'BRD4 Promotes DNA Repair and Mediates the Formation of TMPRSS2-ERG Gene Rearrangements in Prostate Cancer', *Cell Reports*, 22(3), pp. 796–808. doi: 10.1016/j.celrep.2017.12.078.

Li, Y. *et al.* (2017) 'USP13 regulates the RAP80-BRCA1 complex dependent DNA damage response', *Nature Communications*, 8. doi: 10.1038/ncomms15752.

Lindgren, E. *et al.* (2014) 'Inactivation of the budding yeast cohesin loader Scc2 alters gene expression both globally and in response to a single DNA double strand break', *Cell Cycle*, 13(23), pp. 3645–3658. doi: 10.4161/15384101.2014.964108.

Liu, E. T., Pott, S. and Huss, M. (2010) 'Q&A: ChIP-seq technologies and the study of gene regulation', *BMC Biology*, 8. doi: 10.1186/1741-7007-8-56.

Liu, J. *et al.* (2009) 'Transcriptional Dysregulation in NIPBL and Cohesin Mutant Human Cells', *PLoS Biology*, 7(5), p. e1000119. doi: 10.1371/journal.pbio.1000119.

Liu, J. and Baynam, G. (2010) 'Cornelia de Lange syndrome.', *Advances in experimental medicine and biology*, 685, pp. 111–123. doi: 10.1111/cge.12499.

Liu, L. F. and Wang, J. C. (1987) 'Supercoiling of the DNA template during transcription.', *Proceedings of the National Academy of Sciences*, 84(20), pp. 7024–7027. doi: 10.1073/pnas.84.20.7024.

Liu, Q. *et al.* (2000) 'Chk1 is an essential kinase that is regulated by Atr and required for the G2/M DNA damage checkpoint', *Genes and Development*, 14(12), pp. 1448–1459. doi: 10.1101/gad.14.12.1448.

Liu, W. *et al.* (2013) 'Brd4 and JMJD6-associated anti-pause enhancers in regulation of transcriptional pause release', *Cell*, 155(7), pp. 1581–1595. doi: 10.1016/j.cell.2013.10.056.

Liu, X. *et al.* (2017) 'PAXX promotes KU accumulation at DNA breaks and is essential for end-joining in XLF-deficient mice', *Nature Communications*, 8. doi: 10.1038/ncomms13816.

Liu, X. L. *et al.* (2016) 'Novel ATM mutations with ataxia-telangiectasia', *Neuroscience Letters*, 611, pp. 112–115. doi: 10.1016/j.neulet.2015.11.036.

Liu, Z. *et al.* (2017) 'Drug Discovery Targeting Bromodomain-Containing Protein 4', *Journal of Medicinal Chemistry*, pp. 4533–4558. doi: 10.1021/acs.jmedchem.6b01761.

Ljungman, M. (2010) 'The DNA damage response-Repair or despair?', *Environmental and Molecular Mutagenesis*, pp. 879–889. doi: 10.1002/em.20597.

Lomax, M. E., Folkes, L. K. and O'Neill, P. (2013) 'Biological consequences of radiation-induced DNA damage: Relevance to radiotherapy', *Clinical Oncology*, 25(10), pp. 578–585. doi: 10.1016/j.clon.2013.06.007.

Lopez-Burks, M. E. *et al.* (2016) 'Genetic enhancement of limb defects in a mouse model of Cornelia de Lange syndrome', *American Journal of Medical Genetics, Part C: Seminars in Medical Genetics*, 172(2), pp. 146–154. doi: 10.1002/ajmg.c.31491.

Lopez-Serra, L. *et al.* (2014) 'The Scc2-Scc4 complex acts in sister chromatid cohesion and transcriptional regulation by maintaining nucleosome-free regions', *Nature Genetics*, 46(10),

pp. 1147–1151. doi: 10.1038/ng.3080.

Lovén, J. *et al.* (2013) 'Selective inhibition of tumor oncogenes by disruption of super-enhancers', *Cell*, 153(2), pp. 320–334. doi: 10.1016/j.cell.2013.03.036.

Lowe, R. *et al.* (2017) 'Transcriptomics technologies', *PLoS Computational Biology*, 13(5). doi: 10.1371/journal.pcbi.1005457.

Lu, H. *et al.* (2017) 'Cell cycle-dependent phosphorylation regulates RECQL4 pathway choice and ubiquitination in DNA double-strand break repair', *Nature Communications*, 8(1), p. 2039. doi: 10.1038/s41467-017-02146-3.

Lu, X. *et al.* (2016) 'Multiple P-TEFbs cooperatively regulate the release of promoter-proximally paused RNA polymerase II', *Nucleic Acids Research*, 44(14), pp. 6853–6867. doi: 10.1093/nar/gkw571.

Lupiáñez, D. G. *et al.* (2015) 'Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions', *Cell*, 161(5), pp. 1012–1025. doi: 10.1016/j.cell.2015.04.004.

M., H. *et al.* (2016) 'The bromodomain BRD4 regulates heat-shock induced co-transcriptional splicing', *Oncology Research and Treatment*, p. 67. Available at: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed18a&NEWS=N&AN=72212502.

Ma, J. and Wang, M. D. (2016) 'DNA supercoiling during transcription', *Biophysical Reviews*, pp. 75–87. doi: 10.1007/s12551-016-0215-9.

Machida, S. *et al.* (2018) 'Structural Basis of Heterochromatin Formation by Human HP1', *Molecular Cell*, 69(3), p. 385–397.e8. doi: 10.1016/j.molcel.2017.12.011.

Makrantoni, V. and Marston, A. L. (2018) 'Cohesin and chromosome segregation', *Current Biology*, pp. R688–R693. doi: 10.1016/j.cub.2018.05.019.

Mallette, F. A. *et al.* (2012) 'RNF8- and RNF168-dependent degradation of KDM4A/JMJD2A triggers 53BP1 recruitment to DNA damage sites', *EMBO Journal*, 31(8), pp. 1865–1878. doi: 10.1038/emboj.2012.47.

Mandemaker, I. K. *et al.* (2017) 'DNA damage-induced histone H1 ubiquitylation is mediated by HUWE1 and stimulates the RNF8-RNF168 pathway', *Scientific Reports*. doi: 10.1038/s41598-017-15194-y.

Mannini, L. *et al.* (2013) 'Mutation Spectrum and Genotype-Phenotype Correlation in Cornelia de Lange Syndrome', *Human Mutation*, 34(12), pp. 1589–1596. doi: 10.1002/humu.22430.

Mannini, L. *et al.* (2015) 'Mutant cohesin affects RNA polymerase II regulation in Cornelia de Lange syndrome', *Scientific Reports*, 5. doi: 10.1038/srep16803.

Marcos-Alcalde, Í. *et al.* (2017) 'Two-step ATP-driven opening of cohesin head', *Scientific Reports*, 7(1). doi: 10.1038/s41598-017-03118-9.

Maréchal, A. and Zou, L. (2013) 'DNA damage sensing by the ATM and ATR kinases', *Cold Spring Harbor Perspectives in Biology*, 5(9). doi: 10.1101/cshperspect.a012716.

Marucci, L. and Lucia (2017) 'Nanog Dynamics in Mouse Embryonic Stem Cells: Results

from Systems Biology Approaches', *Stem Cells International*, 2017, pp. 1–14. doi: 10.1155/2017/7160419.

Maruyama, T. *et al.* (2002) 'A Mammalian Bromodomain Protein, Brd4, Interacts with Replication Factor C and Inhibits Progression to S Phase', *Molecular and Cellular Biology*, 22(18), pp. 6509–6520. doi: 10.1128/MCB.22.18.6509-6520.2002.

Matsuoka, S. *et al.* (2007) 'ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage', *Science*, 316(5828), pp. 1160–1166. doi: 10.1126/science.1140321.

McKeown, M. R. *et al.* (2017) 'Superenhancer analysis defi nes novel epigenomic subtypes of non-APL AML, including an RARα dependency targetable by SY-1425, a potent and selective RARα agonist', *Cancer Discovery*, 7(10), pp. 1136–1153. doi: 10.1158/2159-8290.CD-17-0399.

McRae, J. F. *et al.* (2017) 'Prevalence and architecture of de novo mutations in developmental disorders', *Nature*, 542(7642), pp. 433–438. doi: 10.1038/nature21062.

Mehta, A. and Haber, J. E. (2014) 'Sources of DNA double-strand breaks and models of recombinational DNA repair', *Cold Spring Harbor Perspectives in Biology*, 6(9). doi: 10.1101/cshperspect.a016428.

Mei, Q. *et al.* (2017) 'Regulation of DNA replication-coupled histone gene expression.', *Oncotarget*, 8(55), pp. 95005–95022. doi: 10.18632/oncotarget.21887.

Mellacheruvu, D. *et al.* (2013) 'The CRAPome: A contaminant repository for affinity purification-mass spectrometry data', *Nature Methods*, 10(8), pp. 730–736. doi: 10.1038/nmeth.2557.

Meng, F. L. *et al.* (2014) 'Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability', *Cell*, 159(7), pp. 1538–1548. doi: 10.1016/j.cell.2014.11.014.

Meng, H. and Blaine, B. (2017) 'Emerging Roles of Transcriptional Enhancers In Chromatin Looping And Promoter-Proximal Pausing Of RNA Polymerase II', *Journal of Biological Chemistry*, p. jbc.R117.813485. doi: 10.1074/jbc.R117.813485.

Mfopou, J. K. *et al.* (2014) 'Efficient definitive endoderm induction from mouse embryonic stem cell adherent cultures: A rapid screening model for differentiation studies', *Stem Cell Research*, 12(1), pp. 166–177. doi: 10.1016/j.scr.2013.10.004.

Di Micco, R. *et al.* (2014) 'Control of embryonic stem cell identity by brd4-dependent transcriptional elongation of super-enhancer-associated pluripotency genes', *Cell Reports*, 9(1), pp. 234–247. doi: 10.1016/j.celrep.2014.08.055.

Michl, J., Zimmer, J. and Tarsounas, M. (2016) 'Interplay between Fanconi anemia and homologous recombination pathways in genome integrity', *The EMBO Journal*, 35(9), pp. 909–923. doi: 10.15252/embj.201693860.

Mills, J. A. *et al.* (2018) 'NIPBL+/-haploinsufficiency reveals a constellation of transcriptome disruptions in the pluripotent and cardiac states', *Scientific Reports*, 8(1). doi: 10.1038/s41598-018-19173-9.

Misulovin, Z. *et al.* (2008) 'Association of cohesin and Nipped-B with transcriptionally active regions of the Drosophila melanogaster genome', *Chromosoma*, 117(1), pp. 89–102. doi:

10.1007/s00412-007-0129-1.

Mortazavi, A. *et al.* (2008) 'Mapping and quantifying mammalian transcriptomes by RNA-Seq', *Nature Methods*, 5(7), pp. 621–628. doi: 10.1038/nmeth.1226.

Moyal, L. *et al.* (2011) 'Requirement of ATM-Dependent Monoubiquitylation of Histone H2B for Timely Repair of DNA Double-Strand Breaks', *Molecular Cell*, 41(5), pp. 529–542. doi: 10.1016/j.molcel.2011.02.015.

Muhar, M. *et al.* (2018) 'SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis', *Science*, 360(6390), pp. 800–805. doi: 10.1126/science.aao2793.

Murayama, Y. *et al.* (2018) 'Establishment of DNA-DNA Interactions by the Cohesin Ring', *Cell*, 172(3), p. 465–477.e15. doi: 10.1016/j.cell.2017.12.021.

Murr, R. *et al.* (2006) 'Histone acetylation by Trrap-Tip60 modulates loading of repair proteins and repair of DNA double-strand breaks', *Nature Cell Biology*, 8(1), pp. 91–99. doi: 10.1038/ncb1343.

Musio, A. *et al.* (2006) 'X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations', *Nature Genetics*, 38(5), pp. 528–530. doi: 10.1038/ng1779.

Muto, A. *et al.* (2014) 'Nipbl and Mediator Cooperatively Regulate Gene Expression to Control Limb Development', *PLoS Genetics*, 10(9). doi: 10.1371/journal.pgen.1004671.

Nagy, Z. and Soutoglou, E. (2009) 'DNA repair: easy to visualize, difficult to elucidate', *Trends in Cell Biology*, pp. 617–629. doi: 10.1016/j.tcb.2009.08.010.

Nair, G. *et al.* (2015) 'Heterogeneous lineage marker expression in naive embryonic stem cells is mostly due to spontaneous differentiation', *Scientific Reports*, 5. doi: 10.1038/srep13339.

Najafova, Z. *et al.* (2017) 'BRD4 localization to lineage-specific enhancers is associated with a distinct transcription factor repertoire', *Nucleic Acids Research*, 45(1), pp. 127–141. doi: 10.1093/nar/gkw826.

Nepal, M. *et al.* (2017) 'Fanconi Anemia Signaling and Cancer', *Trends in Cancer*. doi: 10.1016/j.trecan.2017.10.005.

Newkirk, D. A. *et al.* (2017) 'The effect of Nipped-B-like (Nipbl) haploinsufficiency on genome-wide cohesin binding and target gene expression: Modeling Cornelia de Lange syndrome', *Clinical Epigenetics*, 9(1). doi: 10.1186/s13148-017-0391-x.

Nishiyama, T. *et al.* (2010) 'Sororin mediates sister chromatid cohesion by antagonizing Wapl', *Cell*, 143(5), pp. 737–749. doi: 10.1016/j.cell.2010.10.031.

Nolen, L. D. *et al.* (2013) 'Regional chromatin decompaction in cornelia de lange syndrome associated with NIPBL disruption can be uncoupled from cohesin and CTCF', *Human Molecular Genetics*, 22(20), pp. 4180–4193. doi: 10.1093/hmg/ddt265.

Noordermeer, S. M., Adam, S., Setiaputra, D., Barazas, M., Pettitt, S. J., Ling, A. K., Olivieri, M., Álvarez-Quilón, A., Moatti, N., Zimmermann, M., Annunziato, S., Krastev, D. B., Song, F., Brandsma, I., Frankum, J., Brough, R., Sherker, A., Landry, S., Szilard, R. K., Munro, M. M., McEwan, A., Goullet de Rugy, T., *et al.* (2018) 'The shieldin complex mediates 53BP1-dependent DNA repair', *Nature*, 560(7716), pp. 117–121. doi: 10.1038/s41586-018-0340-7.

Noordermeer, S. M., Adam, S., Setiaputra, D., Barazas, M., Pettitt, S. J., Ling, A. K., Olivieri,

M., Álvarez-Quilón, A., Moatti, N., Zimmermann, M., Annunziato, S., Krastev, D. B., Song, F., Brandsma, I., Frankum, J., Brough, R., Sherker, A., Landry, S., Szilard, R. K., Munro, M. M., McEwan, A., de Rugy, T. G., *et al.* (2018) 'The shieldin complex mediates 53BP1-dependent DNA repair', *Nature*. doi: 10.1038/s41586-018-0340-7.

Nora, E. P. *et al.* (2017) 'Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization', *Cell*, 169(5), p. 930–944.e22. doi: 10.1016/j.cell.2017.05.004.

Noshir Mehta, D. and Bhatia, R. (2013) 'Cornelia De-Lange Syndrome: A Case Report', *International Journal of Clinical Pediatric Dentistry*, 6, pp. 115–118. doi: 10.5005/jp-journals-10005-1201.

Nowsheen, S. *et al.* (2018) 'ZNF506-dependent positive feedback loop regulates H2AX signaling after DNA damage', *Nature Communications*, 9(1). doi: 10.1038/s41467-018-05161-0.

Nuebler, J. *et al.* (2018) 'Chromatin organization by an interplay of loop extrusion and compartmental segregation', *Proceedings of the National Academy of Sciences*, 115(29), pp. E6697–E6706. doi: 10.1073/pnas.1717730115.

O'Driscoll, M. (2012) 'Diseases associated with defective responses to DNA damage', *Cold Spring Harbor Perspectives in Biology*, 4(12). doi: 10.1101/cshperspect.a012773.

O'Driscoll, M. and Jeggo, P. A. (2008) 'The role of the DNA damage response pathways in brain development and microcephaly: Insight from human disorders', *DNA Repair*, 7(7), pp. 1039–1050. doi: 10.1016/j.dnarep.2008.03.018.

Ocaña, A., Nieto-Jiménez, C. and Pandiella, A. (2017) 'BET inhibitors as novel therapeutic agents in breast cancer', *Oncotarget*, 8(41). doi: 10.18632/oncotarget.19744.

Ochs, F. *et al.* (2016) '53BP1 fosters fidelity of homology-directed DNA repair', *Nature Structural and Molecular Biology*, 23(8), pp. 714–721. doi: 10.1038/nsmb.3251.

Ogami, K., Chen, Y. and Manley, J. (2018) 'RNA Surveillance by the Nuclear RNA Exosome: Mechanisms and Significance', *Non-Coding RNA*. doi: 10.3390/ncrna4010008.

Ogiwara, H. *et al.* (2011) 'Histone acetylation by CBP and p300 at double-strand break sites facilitates SWI/SNF chromatin remodeling and the recruitment of non-homologous end joining factors', *Oncogene*, 30(18), pp. 2135–2146. doi: 10.1038/onc.2010.592.

Ohtsuka, S., Nakai-Futatsugi, Y. and Niwa, H. (2015) 'LIF signal in mouse embryonic stem cells', *JAK-STAT*. doi: 10.1080/21623996.2015.1086520.

Oka, Y. *et al.* (2011) 'Recruitment of the cohesin loading factor NIPBL to DNA double-strand breaks depends on MDC1, RNF168 and HP1γ in human cells', *Biochemical and Biophysical Research Communications*, 411(4), pp. 762–767. doi: 10.1016/j.bbrc.2011.07.021.

Olley, G. *et al.* (2018) 'BRD4 interacts with NIPBL and BRD4 is mutated in a Cornelia de Lange–like syndrome', *Nature Genetics*. doi: 10.1038/s41588-018-0042-y.

Oristo, S., Lee, H. J. and Maunula, L. (2018) 'Performance of pre-RT-qPCR treatments to discriminate infectious human rotaviruses and noroviruses from heat-inactivated viruses: applications of PMA/PMAxx, benzonase and RNase', *Journal of Applied Microbiology*, 124(4), pp. 1008–1016. doi: 10.1111/jam.13737.

Orlando, D. A. *et al.* (2014) 'Quantitative ChIP-Seq normalization reveals global modulation of the epigenome', *Cell Reports*, 9(3), pp. 1163–1170. doi: 10.1016/j.celrep.2014.10.018.

Paix, A. *et al.* (2017) 'Precision genome editing using synthesis-dependent repair of Cas9-induced DNA breaks', *Proceedings of the National Academy of Sciences*, p. 201711979. doi: 10.1073/pnas.1711979114.

Panamarova, M. *et al.* (2016) 'The BAF chromatin remodelling complex is an epigenetic regulator of lineage specification in the early mouse embryo', *Development*, 143(8), pp. 1271–1283. doi: 10.1242/dev.131961.

Panier, S. and Boulton, S. J. (2014) 'Double-strand break repair: 53BP1 comes into focus', *Nature Reviews Molecular Cell Biology*, pp. 7–18. doi: 10.1038/nrm3719.

Pannunzio, N. R., Watanabe, G. and Lieber, M. R. (2017) 'Nonhomologous DNA End Joining for Repair of DNA Double-Strand Breaks', *Journal of Biological Chemistry*, p. jbc.TM117.000374. doi: 10.1074/jbc.TM117.000374.

Parenti, I. *et al.* (2017) 'Mutations in chromatin regulators functionally link Cornelia de Lange syndrome and clinically overlapping phenotypes', *Human Genetics*, 136(3), pp. 307–320. doi: 10.1007/s00439-017-1758-y.

Park, P. J. (2009) 'ChIP-seq: Advantages and challenges of a maturing technology', *Nature Reviews Genetics*, pp. 669–680. doi: 10.1038/nrg2641.

Patel, M. C. *et al.* (2013) 'BRD4 Coordinates Recruitment of Pause Release Factor P-TEFb and the Pausing Complex NELF/DSIF To Regulate Transcription Elongation of Interferon-Stimulated Genes', *Molecular and Cellular Biology*, 33(12), pp. 2497–2507. doi: 10.1128/MCB.01180-12.

Peng, G. *et al.* (2014) 'Genome-wide transcriptome profiling of homologous recombination DNA repair', *Nature Communications*. doi: 10.1038/ncomms4361.

Petela, N. J. *et al.* (2018) 'Scc2 Is a Potent Activator of Cohesin's ATPase that Promotes Loading by Binding Scc1 without Pds5', *Molecular Cell*, 70(6), p. 1134–1148.e7. doi: 10.1016/j.molcel.2018.05.022.

Petrenko, N. *et al.* (2016) 'Mediator Undergoes a Compositional Change during Transcriptional Activation', *Molecular Cell*, 64(3), pp. 443–454. doi: 10.1016/j.molcel.2016.09.015.

Pillai, S., Dasgupta, P. and Chellappan, S. P. (2015) 'Chromatin immunoprecipitation assays: Analyzing transcription factor binding and histone modifications in vivo', in *Chromatin Protocols: Third Edition*, pp. 429–446. doi: 10.1007/978-1-4939-2474-5_25.

Pintacuda, G. *et al.* (2017) 'hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing', *Molecular Cell*, 68(5), p. 955–969.e10. doi: 10.1016/j.molcel.2017.11.013.

Polo, S. E. and Jackson, S. P. (2011) 'Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications.', *Genes & development*, 25(5), pp. 409–33. doi: 10.1101/gad.2021311.

Pombo, A. *et al.* (1998) 'Regional and temporal specialization in the nucleus: A transcriptionally-active nuclear domain rich in PTF, Oct1 and PIKA antigens associates with

specific chromosomes early in the cell cycle', *EMBO Journal*, 17(6), pp. 1768–1778. doi: 10.1093/emboj/17.6.1768.

Pommier, Y. *et al.* (2016) 'Roles of eukaryotic topoisomerases in transcription, replication and genomic stability', *Nature Reviews Molecular Cell Biology*, pp. 703–721. doi: 10.1038/nrm.2016.111.

Pongas, G. *et al.* (2017) 'BRD4 facilitates DNA damage response and represses CBX5/ Heterochromatin protein 1 (HP1)', *Oncotarget*, 8(31), pp. 51402–51415. doi: 10.18632/oncotarget.17572.

Potts, P. R., Porteus, M. H. and Yu, H. (2006) 'Human SMC5/6 complex promotes sister chromatid homologous recombination by recruiting the SMC1/3 cohesin complex to double-strand breaks', *EMBO Journal*, 25(14), pp. 3377–3388. doi: 10.1038/sj.emboj.7601218.

Povirk, L. F. (2012) 'Processing of Damaged DNA Ends for Double-Strand Break Repair in Mammalian Cells', *ISRN Molecular Biology*, 2012, pp. 1–16. doi: 10.5402/2012/345805.

Pradeepa, M. M. *et al.* (2016) 'Histone H3 globular domain acetylation identifies a new class of enhancers', *Nature Genetics*, 48(6). doi: 10.1038/ng.3550.

Price, B. D. and D'Andrea, A. D. (2013) 'Chromatin remodeling at DNA double-strand breaks', *Cell*, pp. 1344–1354. doi: 10.1016/j.cell.2013.02.011.

Prochazkova, J. and Loizou, J. I. (2016) 'Programmed DNA breaks in lymphoid cells: Repair mechanisms and consequences in human disease', *Immunology*, 147(1), pp. 11–20. doi: 10.1111/imm.12547.

Qiao, Q. *et al.* (2017) 'AID Recognizes Structured DNA for Class Switch Recombination', *Molecular Cell*, 67(3), p. 361–373.e4. doi: 10.1016/j.molcel.2017.06.034.

Rahman, S. *et al.* (2011) 'The Brd4 Extraterminal Domain Confers Transcription Activation Independent of pTEFb by Recruiting Multiple Proteins, Including NSD3', *Molecular and Cellular Biology*, 31(13), pp. 2641–2652. doi: 10.1128/MCB.01341-10.

Rahnamoun, H. *et al.* (2018) 'RNAs interact with BRD4 to promote enhanced chromatin engagement and transcription activation.', *Nature Structural & Molecular Biology*, 25, pp. 687–697.

Ramalingam, K. *et al.* (2012) 'Seckel syndrome: a report of a case.', *Journal of the Indian Society of Pedodontics and Preventive Dentistry*, 30(3), pp. 258–261. doi: 10.4103/0970-4388.105021.

Rane, S. G. *et al.* (1999) 'Loss of Cdk4 expression causes insulin-deficient diabetes and Cdk4 activation results in β-islet cell hyperplasia', *Nature Genetics*, 22(1), pp. 44–54. doi: 10.1038/8751.

Rao, S. (2012) 'Embryonic Stem Cells: A Perfect Tool for Studying Mammalian Transcriptional Enhancers', *Journal of Stem Cell Research & Therapy*, 01(S10). doi: 10.4172/2157-7633.S10-007.

Rao, S. S. P. *et al.* (2017) 'Cohesin Loss Eliminates All Loop Domains', *Cell*, 171(2), p. 305–320.e24. doi: 10.1016/j.cell.2017.09.026.

Remeseiro, S. *et al.* (2013) 'Reduction of Nipbl impairs cohesin loading locally and affects

transcription but not cohesion-dependent functions in a mouse model of Cornelia de Lange Syndrome', *Biochimica et Biophysica Acta - Molecular Basis of Disease*, 1832(12), pp. 2097–2102. doi: 10.1016/j.bbadis.2013.07.020.

Ren, B. *et al.* (2000) 'Genome-wide location and function of DNA binding proteins', *Science*, 290(5500), pp. 2306–2309. doi: 10.1126/science.290.5500.2306.

Revenkova, E. *et al.* (2009) 'Cornelia de Lange syndrome mutations in SMC1A or SMC3 affect binding to DNA', *Human Molecular Genetics*, 18(3), pp. 418–427. doi: 10.1093/hmg/ddn369.

Rocha, P. P. *et al.* (2015) 'Breaking TADs: Insights into hierarchical genome organization', *Epigenomics*, pp. 523–526. doi: 10.2217/epi.15.25.

Rodgers, K. and Mcvey, M. (2016) 'Error-Prone Repair of DNA Double-Strand Breaks', *Journal of Cellular Physiology*, pp. 15–24. doi: 10.1002/jcp.25053.

Rogakou, E. P. *et al.* (1999) 'Megabase chromatin domains involved in DNA double-strand breaks in vivo.', *The Journal of cell biology*, 146(5), pp. 905–16. doi: 10.1083/jcb.146.5.905.

Rohatgi, S. *et al.* (2010) 'Facial diagnosis of mild and variant CdLS: Insights from a dysmorphologist survey', *American Journal of Medical Genetics, Part A*, 152(7), pp. 1641–1653. doi: 10.1002/ajmg.a.33441.

Rollins, R. A., Morcillo, P. and Dorsett, D. (1999) 'Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes', *Genetics*, 152(2), pp. 577–593.

Rother, M. B. and van Attikum, H. (2017) 'DNA repair goes hip-hop: SMARCA and CHD chromatin remodellers join the break dance', *Philosophical Transactions of the Royal Society B: Biological Sciences*. doi: 10.1098/rstb.2016.0285.

van Ruiten, M. S. and Rowland, B. D. (2018) 'SMC Complexes: Universal DNA Looping Machines with Distinct Regulators', *Trends in Genetics*, pp. 477–487. doi: 10.1016/j.tig.2018.03.003.

Russell, K. L. *et al.* (2001) 'Dominant paternal transmission of Cornelia de Lange syndrome: A new case and review of 25 previously reported familial recurrences', *American Journal of Medical Genetics*, 104(4), pp. 267–276. doi: 10.1002/ajmg.10066.

Rutkowski, A. J. and Dölken, L. (2017) 'High-resolution gene expression profiling of RNA synthesis, processing, and decay by metabolic labeling of newly transcribed RNA using 4-thiouridine', in *Methods in Molecular Biology*, pp. 129–140. doi: 10.1007/978-1-4939-6518-2_10.

Sabari, B. R. *et al.* (2018) 'Coactivator condensation at super-enhancers links phase separation and gene control', *Science*, pp. 1–16. doi: 10.1126/science.aar3958.

Saldivar, J. C., Cortez, D. and Cimprich, K. A. (2017) 'The essential kinase ATR: Ensuring faithful duplication of a challenging genome', *Nature Reviews Molecular Cell Biology*, pp. 622–636. doi: 10.1038/nrm.2017.67.

Santos, R. *et al.* (2016) 'Conditional Creation and Rescue of Nipbl-Deficiency in Mice Reveals Multiple Determinants of Risk for Congenital Heart Defects', *PLoS Biology*, 14(9). doi: 10.1371/journal.pbio.2000197.

Sanyal, S. *et al.* (2018) 'Mutations that prevent methylation of cohesin render sensitivity to DNA damage in *S. pombe*', *Journal of Cell Science*, 131(13), p. jcs214924. doi: 10.1242/jcs.214924.

Schaukowitch, K. *et al.* (2014) 'Enhancer RNA facilitates NELF release from immediate early genes', *Molecular Cell*, 56(1), pp. 29–42. doi: 10.1016/j.molcel.2014.08.023.

Schiller, C. B. *et al.* (2014) 'Structural Studies of DNA End Detection and Resection in Homologous Recombination.', *Cold Spring Harbor Perspectives in Biology*, p. a017962. doi: 10.1101/cshperspect.a017962.

Schipler, A. and Iliakis, G. (2013) 'DNA double-strand-break complexity levels and their possible contributions to the probability for error-prone processing and repair pathway choice', *Nucleic Acids Research*, 41(16), pp. 7589–7605. doi: 10.1093/nar/gkt556.

Schmitt, E. *et al.* (2007) 'DNA-damage response network at the crossroads of cell-cycle checkpoints, cellular senescence and apoptosis', *Journal of Zhejiang University SCIENCE B*, 8(6), pp. 377–397. doi: 10.1631/jzus.2007.B0377.

Schuhwerk, H. *et al.* (2017) 'Kinetics of poly(ADP-ribosyl)ation, but not PARP1 itself, determines the cell fate in response to DNA damage in vitro and in vivo', *Nucleic Acids Research*, 45(19), pp. 11174–11192. doi: 10.1093/nar/gkx717.

Schwarzer, W. *et al.* (2017) 'Two independent modes of chromatin organization revealed by cohesin removal', *Nature*, 551(7678), pp. 51–56. doi: 10.1038/nature24281.

Schymkowitz, J. *et al.* (2005) 'The FoldX web server: An online force field', *Nucleic Acids Research*, 33(SUPPL. 2). doi: 10.1093/nar/gki387.

Seeber, A., Hauer, M. and Gasser, S. M. (2013) 'Nucleosome remodelers in double-strand break repair', *Current Opinion in Genetics and Development*, pp. 174–184. doi: 10.1016/j.gde.2012.12.008.

Seitan, V. C. *et al.* (2011) 'A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation', *Nature*. doi: 10.1038/nature10312.

Sharif, H. *et al.* (2017) 'Cryo-EM structure of the DNA-PK holoenzyme', *Proceedings of the National Academy of Sciences*, 114(28), pp. 7367–7372. doi: 10.1073/pnas.1707386114.

Shen, D. and Skibbens, R. V. (2017) 'Chl1 DNA helicase and Scc2 function in chromosome condensation through cohesin deposition', *PLoS ONE*, 12(11). doi: 10.1371/journal.pone.0188739.

Sherr, C. J. and Bartek, J. (2017) 'Cell Cycle–Targeted Cancer Therapies', *Annual Review of Cancer Biology*, 1(1), pp. 41–57. doi: 10.1146/annurev-cancerbio-040716-075628.

Shi, J. and Vakoc, C. R. (2014) 'The Mechanisms behind the Therapeutic Activity of BET Bromodomain Inhibition', *Molecular Cell*, pp. 72–736. doi: 10.1016/j.molcel.2014.05.016.

Shiloh, Y. and Ziv, Y. (2013) 'The ATM protein kinase: Regulating the cellular response to genotoxic stress, and more', *Nature Reviews Molecular Cell Biology*, pp. 197–210. doi: 10.1038/nrm3546.

Shin, H. Y. *et al.* (2016) 'Hierarchy within the mammary STAT5-driven Wap super-enhancer', *Nature Genetics*, 48(8), pp. 904–911. doi: 10.1038/ng.3606.

Sicinski, P. *et al.* (1995) 'Cyclin D1 provides a link between development and oncogenesis in the retina and breast', *Cell*, 82(4), pp. 621–630. doi: 10.1016/0092-8674(95)90034-9.

Siddik, Z. (2002) 'Mechanisms of action of cancer chemotherapeutic agents: DNA-interactive alkylating agents and antitumour platinum-based drugs', *Cancer Journal*, 78, pp. 156–162. doi: 10.1002/0470025077.

Sim, Y. J. *et al.* (2017) '2i Maintains a Naive Ground State in ESCs through Two Distinct Epigenetic Mechanisms', *Stem Cell Reports*. doi: 10.1016/j.stemcr.2017.04.001.

Singh, A. and Xu, Y. J. (2016) 'The cell killing mechanisms of hydroxyurea', *Genes*. doi: 10.3390/genes7110099.

Singh, V. P. and Gerton, J. L. (2015) 'Cohesin and human disease: Lessons from mouse models', *Current Opinion in Cell Biology*, pp. 9–17. doi: 10.1016/j.ceb.2015.08.003.

Sirbu, B. M. and Cortez, D. (2013) 'DNA damage response: Three levels of DNA repair regulation', *Cold Spring Harbor Perspectives in Biology*, 5(8). doi: 10.1101/cshperspect.a012724.

Sirmaci, A. *et al.* (2011) 'Mutations in ANKRD11 cause KBG syndrome, characterized by intellectual disability, skeletal malformations, and macrodontia', *American Journal of Human Genetics*, pp. 289–294. doi: 10.1016/j.ajhg.2011.06.007.

Skibbens, R. V. *et al.* (2013) 'Cohesinopathies of a Feather Flock Together', *PLoS Genetics*. doi: 10.1371/journal.pgen.1004036.

Solomon, D. A., Kim, J. S. and Waldman, T. (2014) 'Cohesin gene mutations in tumorigenesis: From discovery to clinical significance', *BMB Reports*, pp. 299–310. doi: 10.5483/BMBRep.2014.47.6.092.

Stanlie, A. *et al.* (2014) 'Chromatin reader Brd4 functions in Ig class switching as a repair complex adaptor of nonhomologous end-joining', *Molecular Cell*, 55(1), pp. 97–110. doi: 10.1016/j.molcel.2014.05.018.

Stark, G. R. and Taylor, W. R. (2006) 'Control of the G2/M transition', *Molecular Biotechnology*. doi: 10.1385/MB:32:3:227.

Stewart, G. S. *et al.* (2007) 'RIDDLE immunodeficiency syndrome is linked to defects in 53BP1-mediated DNA damage signaling.', *Proceedings of the National Academy of Sciences of the United States of America*, 104(43), pp. 16910–5. doi: 10.1073/pnas.0708408104.

Stewart, G. S. *et al.* (2009) 'The RIDDLE Syndrome Protein Mediates a Ubiquitin-Dependent Signaling Cascade at Sites of DNA Damage', *Cell*, 136(3), pp. 420–434. doi: 10.1016/j.cell.2008.12.042.

Struhl, K. (1998) 'Histone acetylation and transcriptional regulatory mechanisms.', *Genes & development*, 12(5), pp. 599–606. doi: 10.1101/gad.12.5.599.

Su, Z. and Denu, J. M. (2016) 'Reading the Combinatorial Histone Language', *ACS Chemical Biology*, pp. 564–574. doi: 10.1021/acschembio.5b00864.

Sun, C. *et al.* (2018) 'BRD4 Inhibition Is Synthetic Lethal with PARP Inhibitors through the Induction of Homologous Recombination Deficiency', *Cancer Cell*, 33(3), p. 401–416.e8. doi: 10.1016/j.ccell.2018.01.019.

Sun, W. *et al.* (2018) 'Integrative analysis of super enhancer SNPs for type 2 diabetes', *PLoS*

ONE, 13(1). doi: 10.1371/journal.pone.0192105.

Symington, L. S. (2014) 'End resection at double-strand breaks: mechanism and regulation.', *Cold Spring Harbor perspectives in biology*. Cold Spring Harbor Laboratory Press, 6(8). doi: 10.1101/cshperspect.a016436.

Tanaka, T. *et al.* (2000) 'Cohesin ensures bipolar attachment of microtubules to sister centromeres and resists their precocious separation', *Nature Cell Biology*, 2(8), pp. 492–499. doi: 10.1038/35019529.

Tang, J. *et al.* (2013) 'Acetylation limits 53BP1 association with damaged chromatin to promote homologous recombination', *Nature Structural and Molecular Biology*. doi: 10.1038/nsmb.2499.

Taniguchi, Y. (2016) 'The bromodomain and extra-terminal domain (BET) family: Functional anatomy of BET paralogous proteins', *International Journal of Molecular Sciences*. doi: 10.3390/ijms17111849.

Teloni, F. and Altmeyer, M. (2016) 'Survey and summary readers of poly(ADP-ribose): Designed to be fit for purpose', *Nucleic Acids Research*, 44(3), pp. 993–1006. doi: 10.1093/nar/gkv1383.

Tessadori, F. *et al.* (2017) 'Germline mutations affecting the histone H4 core cause a developmental syndrome by altering DNA damage response and cell cycle control', *Nature Genetics*, 49(11), pp. 1642–1646. doi: 10.1038/ng.3956.

Thorne, A. W., Myers, F. A. and Hebbes, T. R. (2004) 'Native Chromatin Immunoprecipitation', *Epigenetics Protocols. Methods in Molecular Biology*, 287, pp. 21–44.

Tian, P. *et al.* (2017) 'Fundamental CRISPR-Cas9 tools and current applications in microbial systems', *Synthetic and Systems Biotechnology*, pp. 219–225. doi: 10.1016/j.synbio.2017.08.006.

Timofeev, O. *et al.* (2010) 'Cdc25 phosphatases are required for timely assembly of CDK1-cyclin B at the G2/M transition', *Journal of Biological Chemistry*, 285(22), pp. 16978–16990. doi: 10.1074/jbc.M109.096552.

Toledo, L., Neelsen, K. J. and Lukas, J. (2017) 'Replication Catastrophe: When a Checkpoint Fails because of Exhaustion', *Molecular Cell*, pp. 735–749. doi: 10.1016/j.molcel.2017.05.001.

Tonkin, E. T. *et al.* (2004) 'NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome', *Nature Genetics*, 36(6), pp. 636–641. doi: 10.1038/ng1363.

Trapnell, C. *et al.* (2010) 'Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation', *Nature Biotechnology*, 28(5), pp. 511–515. doi: 10.1038/nbt.1621.

Trapnell, C. *et al.* (2013) 'Differential analysis of gene regulation at transcript resolution with RNA-seq', *Nature Biotechnology*, 31(1), pp. 46–53. doi: 10.1038/nbt.2450.

Trapnell, C., Pachter, L. and Salzberg, S. L. (2009) 'TopHat: Discovering splice junctions with RNA-Seq', *Bioinformatics*, 25(9), pp. 1105–1111. doi: 10.1093/bioinformatics/btp120.

Tsai, S. Q. *et al.* (2015) 'GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases', *Nature Biotechnology*, 33(2), pp. 187–198. doi: 10.1038/nbt.3117.

Turinetto, V. *et al.* (2012) 'High basal ??H2AX levels sustain self-renewal of mouse embryonic and induced pluripotent stem cells', *Stem Cells*, 30(7), pp. 1414–1423. doi: 10.1002/stem.1133.

Turriziani, B. *et al.* (2014) 'On-Beads Digestion in Conjunction with Data-Dependent Mass Spectrometry: A Shortcut to Quantitative and Dynamic Interaction Proteomics', *Biology*, 3(2), pp. 320–332. doi: 10.3390/biology3020320.

Tyanova, S. *et al.* (2016) 'The Perseus computational platform for comprehensive analysis of (prote)omics data', *Nature Methods*, pp. 731–740. doi: 10.1038/nmeth.3901.

Uhlmann, F. (2004) 'The mechanism of sister chromatid cohesion', *Experimental Cell Research*, pp. 80–85. doi: 10.1016/j.yexcr.2004.03.005.

Vahedi, G. *et al.* (2015) 'Super-enhancers delineate disease-associated regulatory nodes in T cells', *Nature*, 520(7548), pp. 558–562. doi: 10.1038/nature14154.

Vega, H. *et al.* (2005) 'Roberts syndrome is caused by mutations in ESCO2, a human homolog of yeast ECO1 that is essential for the establishment of sister chromatid cohesion', *Nature Genetics*, 37(5), pp. 468–470. doi: 10.1038/ng1548.

Vignard, J., Mirey, G. and Salles, B. (2013) 'Ionizing-radiation induced DNA double-strand breaks: A direct and indirect lighting up', *Radiotherapy and Oncology*, pp. 362–369. doi: 10.1016/j.radonc.2013.06.013.

Villa-Hernández, S. and Bermejo, R. (2018) 'Cohesin dynamic association to chromatin and interfacing with replication forks in genome integrity maintenance', *Current Genetics*, pp. 1–9. doi: 10.1007/s00294-018-0824-x.

Vincenten, N. *et al.* (2015) 'The kinetochore prevents centromere-proximal crossover recombination during meiosis', *eLife*. doi: 10.7554/eLife.10850.

Visnes, T. *et al.* (2014) 'Localisation of the SMC loading complex Nipbl/Mau2 during mammalian meiotic prophase I', *Chromosoma*, 123(3), pp. 239–252. doi: 10.1007/s00412-013-0444-7.

Vollmuth, F., Blankenfeldt, W. and Geyer, M. (2009) 'Structures of the dual bromodomains of the P-TEFb-activating protein Brd4 at atomic resolution', *Journal of Biological Chemistry*, 284(52), pp. 36547–36556. doi: 10.1074/jbc.M109.033712.

Vrouwe, M. G. *et al.* (2007) 'Increased DNA damage sensitivity of Cornelia de Lange syndrome cells: Evidence for impaired recombinational repair', *Human Molecular Genetics*, 16(12), pp. 1478–1487. doi: 10.1093/hmg/ddm098.

Walker, J. R., Corpina, R. A. and Goldberg, J. (2001) 'Structure of the Ku heterodimer bound to dna and its implications for double-strand break repair', *Nature*, 412(6847), pp. 607–614. doi: 10.1038/35088000.

Wang, B. *et al.* (2013) 'Abraxas and Rap80 form a novel BRCA1 protein complex required for the DNA damage response', *Science*, 316(5828), pp. 1194–1198. doi: 10.1126/science.1139476.Abraxas.

Wang, J. C. (2002) 'CELLULAR ROLES OF DNA TOPOISOMERASES:A MOLECULAR PERSPECTIVE', *Nature Publishing Group*, 3, pp. 430–440. doi: 10.1038/nrm831.

Wang, R. *et al.* (2012) 'Bromodomain protein Brd4 associated with acetylated chromatin is important for maintenance of higher-order chromatin structure', *Journal of Biological Chemistry*, 287(14), pp. 10738–10752. doi: 10.1074/jbc.M111.323493.

Wang, X. *et al.* (2009) 'DNA damage mediated s and g(2) checkpoints in human embryonal carcinoma cells', *Stem cells (Dayton, Ohio)*, 27(3), pp. 568–576. doi: 10.1634/stemcells.2008-0690.

Wang, X. *et al.* (2015) 'Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors', *Nature Biotechnology*, 33(2), pp. 175–179. doi: 10.1038/nbt.3127.

Wang, Z., Gerstein, M. and Snyder, M. (2009) 'RNA-Seq: A revolutionary tool for transcriptomics', *Nature Reviews Genetics*, pp. 57–63. doi: 10.1038/nrg2484.

Watrin, E. and Peters, J. M. (2009) 'The cohesin complex is required for the DNA damage-induced G2/M checkpoint in mammalian cells', *EMBO Journal*, 28(17), pp. 2625–2635. doi: 10.1038/emboj.2009.202.

Wendt, K. S. (2017) 'Resolving the genomic localization of the kollerin cohesin-loader complex', in *Methods in Molecular Biology*, pp. 115–123. doi: 10.1007/978-1-4939-6545-8_7.

Weterings, E. and Van Gent, D. C. (2004) 'The mechanism of non-homologous end-joining: A synopsis of synapsis', *DNA Repair*, pp. 1425–1435. doi: 10.1016/j.dnarep.2004.06.003.

Whyte, W. A. *et al.* (2013) 'Master transcription factors and mediator establish super-enhancers at key cell identity genes', *Cell*, 153(2), pp. 307–319. doi: 10.1016/j.cell.2013.03.035.

Willis, N. and Rhind, N. (2009) 'Regulation of DNA replication by the S-phase DNA damage checkpoint', *Cell Division*. doi: 10.1186/1747-1028-4-13.

Wilson, A. J. *et al.* (2018) 'The BET inhibitor INCB054329 reduces homologous recombination efficiency and augments PARP inhibitor activity in ovarian cancer', *Gynecologic Oncology*, 149(3), pp. 575–584. doi: 10.1016/j.ygyno.2018.03.049.

Winter, G. E. *et al.* (2017) 'BET Bromodomain Proteins Function as Master Transcription Elongation Factors Independent of CDK9 Recruitment', *Molecular Cell*, 67(1), p. 5–18.e19. doi: 10.1016/j.molcel.2017.06.004.

Woods, S. A. *et al.* (2014) 'Exome sequencing identifies a novel EP300 frame shift mutation in a patient with features that overlap cornelia de lange syndrome', *American Journal of Medical Genetics, Part A*, 164(1), pp. 251–258. doi: 10.1002/ajmg.a.36237.

Wright, W. D., Shah, S. S. and Heyer, W.-D. (2018a) 'Homologous recombination and the repair of DNA Double-Strand Breaks.', *The Journal of biological chemistry*, p. jbc.TM118.000372. doi: 10.1074/jbc.TM118.000372.

Wright, W. D., Shah, S. S. and Heyer, W.-D. (2018b) 'Homologous recombination and the repair of DNA Double-Strand Breaks.', *The Journal of biological chemistry*, p. jbc.TM118.000372. doi: 10.1074/jbc.TM118.000372.

Wu, J., Lu, L. Y. and Yu, X. (2010) 'The role of BRCA1 in DNA damage response', *Protein and Cell*, pp. 117–123. doi: 10.1007/s13238-010-0010-5.

Wu, S. Y. and Chiang, C. M. (2007) 'The double bromodomain-containing chromatin adaptor Brd4 and transcriptional regulation', *Journal of Biological Chemistry*, pp. 13141–13145. doi: 10.1074/jbc.R700001200.

Wu, T. *et al.* (2015) 'The BET family member BRD4 interacts with OCT4 and regulates pluripotency gene expression', *Stem Cell Reports*, 4(3), pp. 390–403. doi: 10.1016/j.stemcr.2015.01.012.

Wu, X. and Brewer, G. (2012) 'The regulation of mRNA stability in mammalian cells: 2.0', *Gene*, pp. 10–21. doi: 10.1016/j.gene.2012.03.021.

Wu, Y. *et al.* (2015) 'Drosophila Nipped-B Mutants Model Cornelia de Lange Syndrome in Growth and Behavior', *PLoS Genetics*, 11(11). doi: 10.1371/journal.pgen.1005655.

Xu, B. *et al.* (2001) 'Involvement of Brca1 in S-Phase and G 2 -Phase Checkpoints after Ionizing Irradiation', *Molecular and cellular biology*. doi: 10.1128/MCB.21.10.3445.

Xu, Y. *et al.* (2012) 'Histone H2A.Z Controls a Critical Chromatin Remodeling Step Required for DNA Double-Strand Break Repair', *Molecular Cell*, 48(5), pp. 723–733. doi: 10.1016/j.molcel.2012.09.026.

Yamamoto, T. and Shimojima, K. (2015) 'A novel MED12 mutation associated with non-specific X-linked intellectual disability', *Human Genome Variation*, 2, p. 15018. doi: 10.1038/hgv.2015.18.

Yan, W. X. *et al.* (2017) 'BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks', *Nature Communications*, 8. doi: 10.1038/ncomms15058.

Yang, G. *et al.* (2018a) 'Super-resolution imaging identifies PARP1 and the Ku complex acting as DNA double-strand break sensors', *Nucleic Acids Research*. doi: 10.1093/nar/gky088.

Yang, G. *et al.* (2018b) 'Super-resolution imaging identifies PARP1 and the Ku complex acting as DNA double-strand break sensors', *Nucleic Acids Research*. doi: 10.1093/nar/gky088.

Yang, Y. *et al.* (2016) 'Enhancer RNA-driven looping enhances the transcription of the long noncoding RNA DHRS4-AS1, a controller of the DHRS4 gene cluster', *Scientific Reports*, 6. doi: 10.1038/srep20961.

Ying, Q. L. *et al.* (2003) 'Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture', *Nature Biotechnology*, 21(2), pp. 183–186. doi: 10.1038/nbt780.

Yoshimura, S. H. and Hirano, T. (2016) 'HEAT repeats – versatile arrays of amphiphilic helices working in crowded environments?', *Journal of Cell Science*, p. jcs.185710. doi: 10.1242/jcs.185710.

Yuan, B. *et al.* (2015) 'Global transcriptional disturbances underlie Cornelia de Lange syndrome and related phenotypes', *Journal of Clinical Investigation*, 125(2), pp. 636–651. doi: 10.1172/JCI77435.

Zabidi, M. A. and Stark, A. (2016) 'Regulatory Enhancer–Core-Promoter Communication via Transcription Factors and Cofactors', *Trends in Genetics*, pp. 801–814. doi: 10.1016/j.tig.2016.10.003.

Zaboikin, M. *et al.* (2017) 'Non-homologous end joining and homology directed DNA repair frequency of double-stranded breaks introduced by genome editing reagents', *PLoS ONE*, 12(1). doi: 10.1371/journal.pone.0169931.

Zakari, M., Yuen, K. and Gerton, J. L. (2015) 'Etiology and pathogenesis of the cohesinopathies', *Wiley Interdisciplinary Reviews: Developmental Biology*, pp. 489–504. doi: 10.1002/wdev.190.

Zámborszky, J. *et al.* (2017) 'Loss of BRCA1 or BRCA2 markedly increases the rate of base substitution mutagenesis and has distinct effects on genomic deletions', *Oncogene*, 36(6), pp. 746–755. doi: 10.1038/onc.2016.243.

Zeng, L. and Zhou, M. M. (2002) 'Bromodomain: An acetyl-lysine binding domain', *FEBS Letters*, pp. 124–128. doi: 10.1016/S0014-5793(01)03309-9.

Zhang, J. *et al.* (2018) 'BRD4 facilitates replication stress-induced DNA damage response', *Oncogene*, pp. 1–15. doi: 10.1038/s41388-018-0194-3.

Zhang, N. *et al.* (2008) 'A handcuff model for the cohesin complex', *Journal of Cell Biology*, 183(6), pp. 1019–1031. doi: 10.1083/jcb.200801157.

Zhang, P. *et al.* (2015) 'BRD4 promotes tumor growth and epithelial-mesenchymal transition in hepatocellular carcinoma', *International Journal of Immunopathology and Pharmacology*, 28(1), pp. 36–44. doi: 10.1177/0394632015572070.

Zhang, W. *et al.* (2012) 'Bromodomain-containing protein 4 (BRD4) regulates RNA polymerase II serine 2 phosphorylation in human CD4+ T cells', *Journal of Biological Chemistry*, 287(51), pp. 43137–43155. doi: 10.1074/jbc.M112.413047.

Zhang, X. *et al.* (2008) 'Esrrb activates Oct4 transcription and sustains self-renewal and pluripotency in embryonic stem cells', *Journal of Biological Chemistry*, 283(51), pp. 35825–35833. doi: 10.1074/jbc.M803481200.

Zhang, Y. *et al.* (2008) 'Model-based analysis of ChIP-Seq (MACS)', *Genome Biology*, 9(9). doi: 10.1186/gb-2008-9-9-r137.

Zhao, H., Watkins, J. L. and Piwnica-Worms, H. (2002) 'Disruption of the checkpoint kinase 1/cell division cycle 25A pathway abrogates ionizing radiation-induced S and G2 checkpoints', *Proceedings of the National Academy of Sciences*, 99(23), pp. 14795–14800. doi: 10.1073/pnas.182557299.

Zheng, G. and Yu, H. (2015) 'Regulation of sister chromatid cohesion during the mitotic cell cycle', *Science China Life Sciences*, 58(11), pp. 1089–1098. doi: 10.1007/s11427-015-4956-7.

Zimmermann, M. and De Lange, T. (2014) '53BP1: Pro choice in DNA repair', *Trends in Cell Biology*, pp. 108–117. doi: 10.1016/j.tcb.2013.09.003.

Zuin, J. *et al.* (2014) 'A Cohesin-Independent Role for NIPBL at Promoters Provides Insights in CdLS', *PLoS Genetics*, 10(2). doi: 10.1371/journal.pgen.1004153.

Zuin, J. *et al.* (2017) 'Regulation of the cohesin-loading factor NIPBL: Role of the lncRNA NIPBL-AS1 and identification of a distal enhancer element', *PLoS Genetics*, 13(12). doi: 10.1371/journal.pgen.1007137.