

**Rapid proteolysis of RAD21 in primary cells
defines the role of cohesin in cell type-specific
gene expression**

by

Felix Daniel Weiss

A thesis submitted to Imperial College London for the degree of
Doctor of Philosophy

MRC London Institute of Medical Sciences

Imperial College School of Medicine

I, Felix Daniel Weiss, declare that the work presented in this thesis is my own, and that any work carried out by other has been acknowledged and appropriately referenced in the text.

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.'

Abstract

Cohesin contributes to the organisation of the mammalian genome in 3D space, but its role in gene regulation remains poorly understood. To address this role, I have implemented a system that enables temporal control over cohesin function in primary mammalian cells by the inducible cleavage of the essential cohesin subunit RAD21, using tobacco etch virus protease (RAD21-TEV). Building on previous work in the lab that had linked cohesin to the expression of inflammatory genes, I used RAD21-TEV to show that genes encoding key mediators of the macrophage inflammatory response are directly reliant on cohesin for their correct expression. The deregulation of these genes led to the collapse of the macrophage response to microbial signals. I next investigated the role of cohesin in neuronal gene expression, because patients with mutations in cohesin and cohesin regulators show intellectual disability as a feature of Cornelia de Lange Syndrome (CdLS). Using RAD21-TEV in mature mouse neurons, I was able to show that acute and chronic cohesin depletion resulted in similar deregulation of neuronal genes. This indicated that cohesin was continuously required for neuronal gene expression. Interestingly, cohesin-dependent neuronal gene expression could be rescued by restoring cohesin after a period of cohesin depletion. To test the relevance of these findings to human disease, I isolated neuronal nuclei from post-mortem CdLS and control brains. RNA-sequencing showed similar gene expression changes in human as in mouse cohesin-deficient neurons. These studies advance our understanding of cohesin's role in gene regulation, and suggest that the impact of cohesin deficiency on neuronal gene expression may in principle be reversible.

Acknowledgments

The past four years have been an absolute joy. The challenging academic while simultaneously happy, friendly, fun and supportive atmosphere in Lymphocyte Development is exactly what makes academic labs so special.

Thank you Matthias for your ever-present guidance and support. Learning from you and working together has allowed me to grow as a scientist and foster my independent thinking. I hope we continue to talk science long after my time here is done. Thank you also to Mandy for your sage advice, and especially the lifts home when I'm too lazy to cycle.

Thank you to all the members of Lymphocyte Development who have come and gone in the past four years. Sergi, you made my PhD experience infinitely better. I thank you so much for your training and support. I am glad to have made a friend for life, and look forward to future collaborations. Ludovica, your undying support and hilarious comments have meant life was never boring or lonely in the lab. Stephan, thank you for being a constant conveyor belt of fascinating, although sometimes dubious, scientific anecdotes. Thank you Gopu for the bioinformatics help and free pints. Thank you so much to James, Bhavik, Tom and Roberta, for your dedicated cell sorting. Chiara, Irene, Bea, Ediem, Tianyi and Hakan, I hope my next office will have as much character. Thank you Teresa for being such a clean and courteous bench mate. To all the rest of you, it's been a pleasure.

Finally, thank you so much to my friends, family and Luke. Thank you to my mum for spending hours on the phone with me learning all about cohesin, and guiding me through the ups and downs. Thank you Sasha for challenging my concepts. Thank you Luke for being so patient and loving with me, especially during the long weekends and late nights in the lab. I promise I'll take you to Kew Gardens one day. See you all on the other side...

Abstract	4
Acknowledgments	5
Figures and Tables	10
Abbreviations	12

6

1.3 Aims of this study	63
Chapter 2: Materials and Methods	65
2.1 Materials	65
2.1.1 Antibodies	65
2.1.2 Cell lines	66
2.1.3 Primers	66
2.2 Methods	69
2.2.1 Genomic DNA extraction and genomic PCR	69
2.2.2 Primary neuron cell culture	70
2.2.3 Primary macrophage cell culture	72
2.2.4 L929 conditioned media production	73
2.2.5 HEK 293 cell culture	73
2.2.6 Lentivirus production	73
2.2.7 Retrovirus production	74
2.2.8 Retrovirus transduction of foetal liver cells	75
2.2.9 Lentivirus transduction of primary neurons	76
2.2.10 Isolation of nuclei from post-mortem human tissue	77
2.2.11 Fluorescent activated cell sorting – FACS	78
2.2.12 Western blot	79
2.2.13 RNA extraction and sequencing	79
2.2.14 DNA delivery into primary neurons	80
2.2.15 Immunofluorescence (IF)	81
2.2.16 Cell cycle analysis – Propidium Iodide (PI) staining	81
2.2.17 Chromosome Conformation Capture – 4C	82
2.2.18 Cloning strategy	86
2.2.19 Plasmid construction and bacteria transformation	89

Chapter 3: The role of cohesin in macrophage inflammatory gene

expression	91
3.1 Introduction	91
3.2 Inducible genes are sensitive to cohesin dosage	93
3.3 Generation of RAD21-TEV macrophages from foetal liver	96
3.4 Characterisation of RAD21-TEV proteolytic cleavage in macrophages	98
3.5 Rapid RAD21-TEV cleavage leads to transcriptional deregulation	100
3.6 Mechanisms of cohesin regulation of the inflammatory response	101
3.7 How do direct targets of cohesin regulation contribute to the loss of LPS response gene regulatory network	104
3.8 Discussion	106

Chapter 4: The role of cohesin in neuronal gene regulation and function __ 108

4.1 Introduction	108
4.2 NIPBL ^{+/-} neurons display abnormal gene regulation	110
4.3 Conditional cohesin depletion in post-mitotic neurons	112
4.4 Role of cohesin in neuronal maturation	115
4.5 Generation of RAD21-TEV neurons from foetal brain	117
4.6 Characterisation of RAD21-TEV depletion in neurons	118
4.7 Cohesin is required for maintaining 3D interactions in mature neurons	120
4.8 Cohesin is required in mature neurons to maintain neuronal gene expression	121
4.9 Cohesin dependent gene regulation in neurons	123
4.10 The role of cohesin in activity dependent gene expression	126
4.11 Discussion	129

Chapter 5: Can changes in gene expression caused by the loss of cohesin

Function be rescued?	132
5.1 Introduction	132
5.2 Establishing a proteolytic based cohesin depletion and rescue system	133

5.3	24 hours of doxycycline controlled RAD21-TEV cleavage leads to changes in gene expression	135
5.4	Changes in gene expression following the removal and rescue of cohesin expression	137
5.5	Discussion	140

Chapter 6: Transcriptional changes in the neurons of Cornelia de Lange Syndrome (CdLS) patients _____ 142

6.1	Introduction	142
6.2	Identification of CdLS patients and samples	143
6.3	Development of flow cytometry based neuronal nuclei purification	145
6.4	CdLS patients display abnormal gene regulation	147
6.5	Cohesin dependent gene regulation in CdLS neurons	149
6.6	Discussion	151

Chapter 7: Discussion and future perspectives _____ 154

Bibliography _____ 160

Figures and Tables

List of Figures

1.1	Visualisation of the genome in 3D space	16
1.2	Chromosome conformation capture based techniques	19
1.3	The structure of cohesin	22
1.4	Organisation of the sonic hedgehog (<i>Shh</i>) locus in 3D space and enhancer based regulation of limb development	39
1.5	AID and TEV based systems for targeted proteolytic cleavage	50
1.6	Macrophages respond to surrounding danger signals and generate a transcriptional response	54
1.7	CdLS is a spectrum disease and can be caused by mutations to multiple genes	61
2.1	Vectors used and generated for inducible control of TEV protease	88
3.1	Cohesin is required for inflammatory gene expression	95
3.2	RAD21-TEV foetal liver produces macrophages	97
3.3	Depletion of RAD21-TEV in mature macrophages	99
3.4	Rapid cohesin depletion leads to transcriptional deregulation, affecting inducible genes and inflammatory response	101
3.5	Features of cohesin dependent genes and 4C analysis of chromatin contacts	103
3.6	Hierarchical nature of inducible gene expression and the effect of longer term cohesin depletion	105
4.1	Nipbl ^{+/-} neurons display abnormal gene regulation	111
4.2	Cohesin deletion in post-mitotic neurons and changes in gene expression	114
4.3	Cohesin is required for correct neuronal maturation	116
4.4	Neuronal cultures generated from E14.5 embryos	118
4.5	Characterisation of RAD21-TEV cleavage in neurons	119
4.6	Interactions at the <i>Pcdhb</i> locus following RAD21-TEV cleavage	121
4.7	Cohesin sustains neuronal gene expression in mature neurons	122
4.8	Cohesin dependent neuronal gene regulation	125
4.9	IEG regulation in cohesin deficient neurons	128

5.1 Doxycycline based control of cohesin expression	134
5.2 Doxycycline mediated loss of cohesin leads to change in gene expression	136
5.3 Rescue of changes in gene expression following cohesin depletion	139
6.1 Purification and validation of neuronal nuclei isolation by flow cytometry	146
6.2 Transcriptional deregulation in CdLS patient neurons	148
6.3 Changes in gene expression are consistent between cohesin depletion model systems and human CdLS patients	151

List of Tables

1. Primers for genotyping	66
2. Primers for cloning	67
3. Primers for protocadherin B 4C	67
4. Primers for <i>Ifnar1</i> 4C	68
5. Primers for <i>Egr2</i> 4C	68
6. Primers for <i>Cebpb</i> 4C	69
7. CdLS patients for neuronal analysis	142

Abbreviations

4'OHT	4-hydroxytamoxifen
3C	Chromosome conformation capture
AID	Auxin inducible degron
AML	Acute myeloid leukaemia
AraC	Cytosine β -D-arabinofuranoside
BME	Basal Medium Eagle
bp	Base pair
CdLS	Cornelia de Lange Syndrome
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation sequencing
CTCF	CCCTC-binding factor
D	Day
DAPI	4', 6-diamidino-2-phenylindole
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Deoxyribonucleic acid
E	Embryonic day
EDTA	Ethylene diamine tetraacetic acid
EGTA	Ethylene glycol tetraacetic acid
ERT2	Modified Oestrogen receptor 2
EtOH	Ethanol
FACS	Fluorescence activated cell sorting
FCS	Foetal calf serum
FISH	Fluorescent in situ hybridisation
g	Gram
GFP	green fluorescent protein
h	hour
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HP1	Heterochromatin protein 1
HSPC	Hematopoietic stem and progenitor cells
IEG	Immediate early gene
IF	Immunofluorescence
IMDM	Iscove's Modified Dulbecco's Medium
IRES	Internal ribosome entry site
Kb	Kilobase pair
KCl	Potassium chloride
kDa	kilodalton
KO	Knockout
L	Litre
LCR	Locus control region
LPS	Lipopolysaccharide
m	Milli
M	Molar
min	Minute
mRNA	Messenger RNA
NBD	Nucleotide binding domain

NES	Normalise enrichment score
NP-40	Nonylphenyl Polyethylene Glycol
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
PI	Propidium Iodide
RFP	Red fluorescent protein
RNA	Ribonucleic acid
RNAi	RNA interference
rpm	Rotations per minute
SDS	Sodium Dodecylsulfate
SMC	Structural maintenance of chromosomes
TCR	T cell receptor
TEV	Tobacco etch virus
Tris	Tris(hydroxymethyl)aminomethane
WT	Wildtype
YY1	Yin yang 1

Chapter 1: Introduction

Our understanding of how chromatin is organised in the 3D space of the nucleus has developed dramatically over the past two decades. The development of chromosome conformation capture (3C) techniques, and the constantly decreasing cost of sequencing has allowed for the mapping of 3D genome organisation in ever greater detail. These have further been complemented by developments in super-resolution microscopy. Mechanisms of the drivers and facilitators of 3D genome organisation have started to be identified through genome editing and targeted proteolysis of architectural proteins and their co-factors. Organisation of the 3D genome is highly conserved across evolution and is therefore thought to be critical for genome function including recombination and repair, replication, and transcription. As the fundamental principles of 3D genome organisation have emerged, so too have its roles in these essential processes.

1.1 The 3D Genome

1.1.1 The different levels of 3D genome organisation

The genome is organised in 3D space into structures of various sizes. At the largest scale one has chromosome territories, where loci preferentially interact with other loci on the same

chromosome as opposed to forming cross chromosome interactions (Lieberman-Aiden et al., 2009).

The genome is then further subdivided into regions known as compartments. Compartments are regions of interacting DNA that are defined by their epigenetic status. Compartments containing active/euchromatin are known as A compartments, while compartments that contain inactive/heterochromatin are known as B compartments (Lieberman-Aiden et al., 2009). A and B compartments segregate in 3D space and preferentially interact with other compartments of the same type (Figure 1.1A).

Within compartments reside domains. Domains are regions of preferentially interacting chromatin that span between 40kB and 3Mb (Figure 1.1A, (Dixon et al., 2012, Sexton et al., 2012, Nora et al., 2012, Rao et al., 2014). DNA within the same domain is more likely to interact with itself than DNA in neighbouring domains. Domains can contain multiple genes and multiple regulatory elements (Rao et al., 2017, Nora et al., 2017, Lupianez et al., 2015).

While compartments and domains cover large regions of DNA, within which all DNA is more likely to interact with itself, chromatin loops represent distinct point to point interactions (Figure 1.1A). Chromatin loops can be formed between specific enhancers and promoters (Palstra et al., 2003), but have also been reported to be formed between co-regulated genes (Schoenfelder et al., 2010). The number and positioning of chromatin loops varies between cell types, with between 3000 and 10,000 having been identified (Rao et al., 2014). Some chromatin loops are essential for the expression of certain genes in specific cell types, the function of others remains unclear.

The different structures of 3D genome organisation will be explained in further detail later in the introduction.

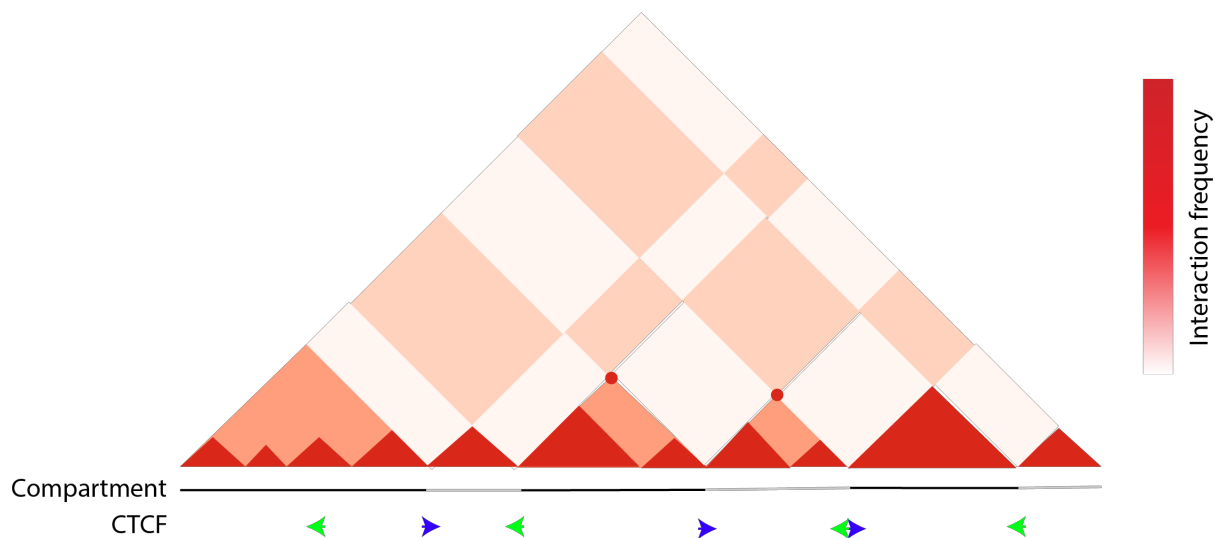


Figure 1.1 Visualisation of the genome in 3D space. This cartoon representing a Hi-C map reveals structures of the genome in 3D space. Structures are different sizes, compartments spanning several megabases are visualised as the plaid pattern, while domains, the dark red triangles, are normally under 1Mb. Chromatin loops are identified by the dark red points. Interactions often, although not exclusively, occur between convergent CTCF sites (represented by the arrows). Compartment interaction is indicated by the black or grey line, as A and B compartments preferentially self-interact.

1.1.2 Visualising the genome – from 3C to Hi-C

Molecular biology techniques to study the organisation of the genome in 3D space are based on the chromosome conformation capture (3C) method and further 3C-based techniques (Figure 1.2, (Lieberman-Aiden et al., 2009, Rao et al., 2017, Schmitt et al., 2016).

3C relies on the combination of protein crosslinking and the ligation of proximal DNA fragments to observe long range chromatin interactions (Dekker et al., 2002). Cells are treated with formaldehyde, to stabilise chromatin structure. The crosslinked chromatin is digested with restriction enzymes, and the fragments are then religated. Fragments that are closest to each other will ligate. DNA sequencing or PCR is used to establish the frequency of ligation between two known genomic loci (Figure 1.2). 3C therefore allows for the quantification of interactions between two known regions in different conditions. While 3C led to discoveries highlighting the importance of genome interactions in gene regulation, it produces a biased view. A key deficit in 3C is that interactions are not compared with local background interactions so their relative strength compared to other interactions cannot be properly determined. In addition, the determining of what regions to assess interactions between rests with the experimentalist. This non-genome wide approach means more important 3D genome interactions may be missed, or certain interactions may be given more importance in gene regulation events, when in fact it is one not targeted that has therefore never been observed.

In order to develop a more genome wide approach, circular chromosome conformation capture (4C) was developed (Figure 1.2). 4C allows for the study of interactions between one known location, and the rest of the genome (Zhao et al., 2006). Much of 4C is the same as 3C, cells are fixed, digested and ligated. However, 4C allows for the identification of interactions between a chosen locus and the rest of the genome. This relies on the formation of ligated rings and primers designed to amplify in both directions from your target genomic area. These PCR products are then sequenced and used to map interactions.

The transition to fully genome wide mapping of 3D genome interactions came with the development of Hi-C (Figure 1.2, (Lieberman-Aiden et al., 2009). This relies on the fixation, digestion, ligation and massive sequencing of DNA reads to generate whole genome interaction maps. Other Hi-C related methods have been developed such as chromosome conformation capture carbon copy (5C) (Dostie et al., 2006) which relies on generating 3C libraries and then using many primers to tile regions of interest and map interaction. This generates maps of pre-selected regions. Capture-HiC involves generating Hi-C libraries, but using biotinylated RNA or DNA oligonucleotide probes, enriching for target sequences such as promoters (Dryden et al., 2014). This massively reduces the amount of DNA fragments and sequencing required to identify sequence specific genome wide interactions.

The development of 3C techniques has been critical to furthering our understanding of the 3D genome and its important in gene regulation, development and disease.

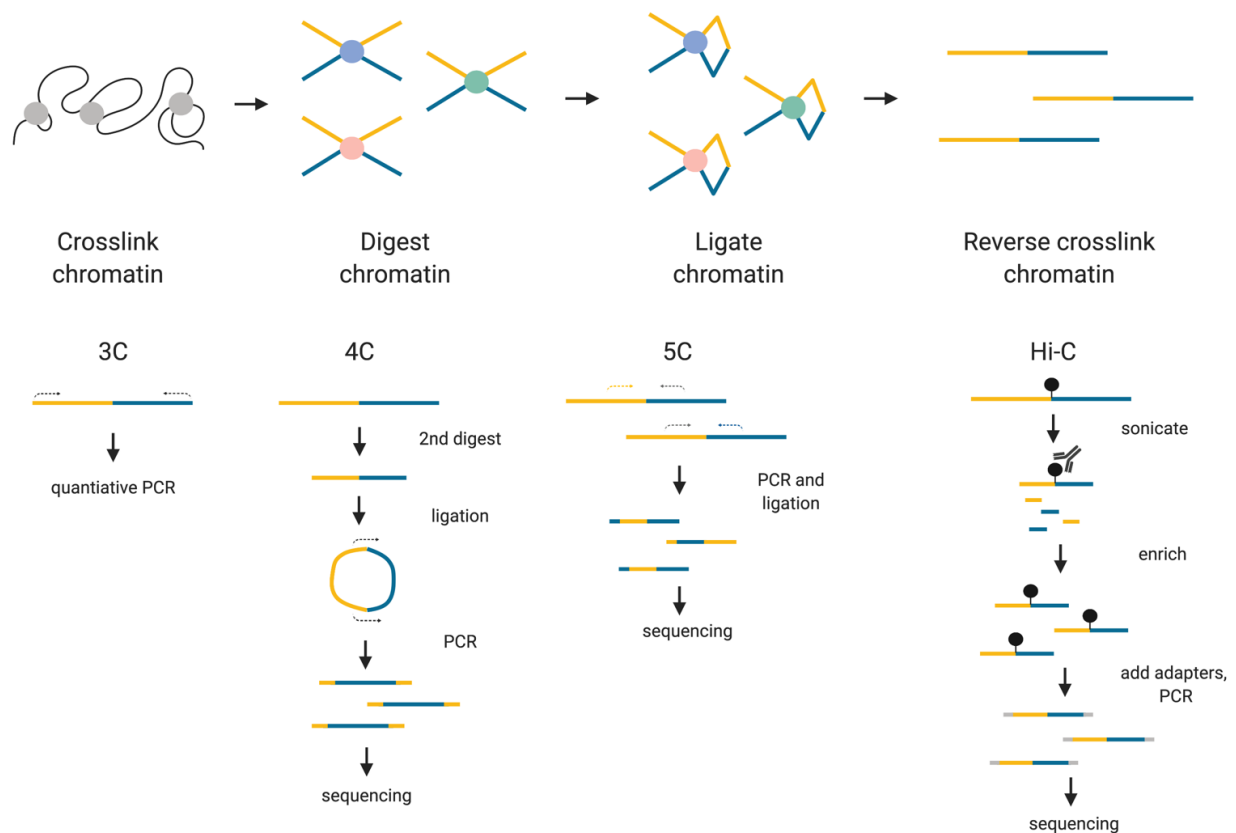


Figure 1.2 Chromosome conformation capture based techniques. 3C based techniques are used to study the 3D organisation of the genome. They all involve common early steps, but differ further downstream depending on whether local or genome wide interactions are studied. For 5C, known primers (represented in blue and yellow) are used with universal primers (grey) to detect interactions. Hi-C involves biotinylating DNA following reverse crosslinking in order to enrich for ligated DNA fragments.

1.1.3 Architectural proteins

The different levels of 3D genome organisation are controlled by a group of proteins and protein complexes known as architectural proteins. They work together to define boundaries of interactions and maintain organisation. Disruption to architectural proteins leads to

disruption of 3D genome organisation. While the regulators of compartments remain unknown. The sequence specific DNA binding protein CCCTC-binding factor (CTCF) and the protein complex, cohesin, are known to be important for determining and maintaining most of domain and loop structure (Rao et al., 2017, Schwarzer et al., 2017, Nora et al., 2017).

Recently, other proteins such as the transcription factor Yin Yang 1 (YY1) have emerged as being important for mediating 3D interactions (Weintraub et al., 2017, Beagan et al., 2017). Enhancers and promoters are enriched for YY1 and cohesin, and the deletion of YY1 binding sites or the depletion of YY1 protein leads to reduced enhancer-promoter interactions (Weintraub et al., 2017). This raises further questions on the roles of transcription factors in mediating 3D genome organisation.

1.1.4 Cohesin

Initially identified as being essential for cell cycle progression, cohesin is now also known to be important for the organisation of the genome in 3D space during interphase (Rao et al., 2017). The cohesin protein complex is a highly evolutionary conserved structure made up of four key subunits (Figure 1.3). In mammals cohesin contains two structural maintenance of chromosome (SMC) proteins, SMC1 and SMC3, and two non-SMC proteins, RAD21 (Scc1 in *S. cerevisiae*) and SA1/2. SMC1 and SMC3 form rod-shaped proteins with a hinge domain at one end and an ATP dependent nucleotide binding domain (NBD) at the other (Figure 1.3). The SMC NBD is made up of both N- and C-terminal amino acids, while the hinge domain contains

amino acids from the centre of the polypeptide (Nasmyth and Haering, 2009). This is due to the SMC proteins having antiparallel coiled coils that cause the polypeptide to double back onto itself. A heterodimer of SMC1 and SMC3 forms the central structure of the cohesin complex, where the hinge domains contact at the top and the SMC NBDs are at the bottom, forming a “V” shape. Bridging the gap between the SMC1 and SMC3 NBDs is the kleisin protein RAD21/Scc1. This results in a ring-shaped structure (Gruber et al., 2003). Less is known about SA1 and SA2. They bind to RAD21, and are thought to mediate interactions between other proteins and the cohesin complex (Remeseiro et al., 2012, Xiao et al., 2011).

In mammalian cells, cohesin relies on the activity of the cohesin loading complex made up of proteins NIPBL and MAU2 for it to be loaded onto DNA (Nasmyth and Haering, 2009). The loss of NIPBL is akin to a loss of cohesin function (Schwarzer et al., 2017), and leads to a disorganisation of the 3D genome. It is thought NIPBL loads cohesin to DNA by recruiting it to DNA and stimulating its ATPase activity. The SMC1 and SMC3 proteins in cohesin are both able to hydrolyse ATP, and in the absence of this form rings but are unable to be loaded onto DNA (Gruber et al., 2006, Weitzer et al., 2003). Furthermore, work has shown that chromatin enters the cohesin ring through the SMC1 and SMC3 hinge domain (Buheitel and Stemmann, 2013). Therefore, it is thought that ATP hydrolysis causes opening of the SMC1/3 hinge domain and the entrapment of DNA into the cohesin ring, leading to cohesin’s association with chromatin.

Despite cohesin’s ability to bind DNA, it is unable to either load or remove itself from chromatin. These processes are carried out by a series of co-factors.

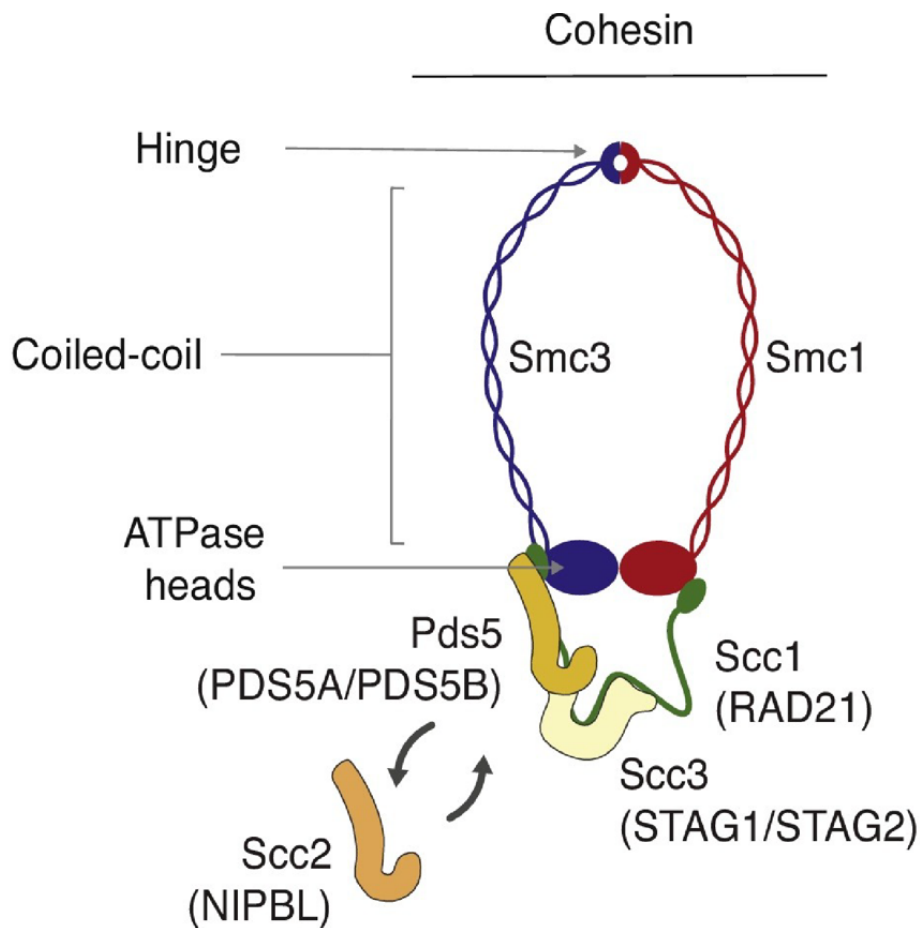


Figure 1.3 The structure of cohesin. A schematic representation of the cohesin complex and NIPBL. The subunits are labelled with their budding yeast and mammalian (in brackets) terms. SMC3 and SMC1 fold back on themselves to form the coiled-coil domains, with the hinge at the top and ATPase heads at the bottom. The heads associated with the alpha-kleisin RAD21 and PDS5. STAG1 and STAG2 are associated with RAD21, and NIPBL determines cohesin loading. Adapted from Sedeno Cacciatore and Rowland, 2019.

1.1.5 Cohesin loading, maintenance and removal with DNA

Much like cohesin being unable to load itself onto DNA, it also relies on other proteins for it to dissociate from DNA. Cohesin-chromatin binding is maintained by the acetylation of SMC3 by ECO1. Acetylation of SMC3 NBD prevents dissociation of RAD21 with SMC3, and the

opening of the SMC3-RAD21 gate by which cohesin is removed from chromatin (Buheitel and Stemmann, 2013). In order for cohesin to be removed from chromatin, SMC3 is deacetylated by HDAC8. SMC3 deacetylation alone is not sufficient to cause cohesin removal from chromatin. The protein WAPL is further required for cohesin removal presumably by opening the SMC3-RAD21 gate. The loss of WAPL leads to increased levels of chromatin bound cohesin, as well as abnormal 3D genome organisation due to increased cohesin residency (Kueng et al., 2006, Gandhi et al., 2006, Wutz et al., 2017).

Cohesin removal also occurs through an alternative mechanism based on cleavage of the RAD21 kleisin protein by separase. This method is essential for progression of the cell cycle as cleavage of cohesin allows for the metaphase to anaphase transition, and the separation of chromosomes to the future daughter cell and parent cell. Separase is inhibited through the rest of mitosis by phosphorylation by CDK1. Once chromosomes have bioriented, separase can be activated leading to cleavage of RAD21. Separase mediated cleavage of RAD21 leads to rapid and efficient removal of the cohesin complex from DNA.

1.1.6 CTCF

CCCTC-binding factor (CTCF) is a sequence specific DNA binding protein that was initially identified as an insulator protein, displaying an ability to interfere with enhancer mediated gene regulation (Gaszner and Felsenfeld, 2006). It is now known to be important for 3D

genome organisation where it is found preferentially bound at domain boundaries (Rao et al., 2014) and its disruption leads to a reduction in domain insulation (Nora et al., 2017).

CTCF is highly conserved in eukaryotes and has a highly conserved DNA binding domain up of 11 zinc-fingers (Ohlsson et al., 2001). CTCF binds between 55,000 and 65,000 sites in the mammalian genome (Chen et al., 2012a), many are conserved across cell-types however 30-60% are cell-type specific (Ong and Corces, 2014). CTCF binding is thought to be at least partly regulated by the methylation of the CpGs in the consensus binding sequence (Engel et al., 2004), contributing to changes in gene expression (Lai et al., 2010, Chang et al., 2010). However not all CTCF sites contain CpG, and the methylation of nearby CpGs may play a role in CTCF binding (Kaaij et al., 2019). CTCF binding and its effect on gene expression may therefore have an effect on cell-type specific gene expression. Differential CTCF binding also occurs in the absence of changes to DNA methylation, as a result there must be alternative regulators that determine CTCF occupancy.

While proteins such as YY1, the transcription factor Kaiso, PARP1, JUND and zinc-finger protein ZNF143 have been implicated in CTCF function at specific loci (Zlatanova and Caiafa, 2009), cohesin has been identified as having the greatest role and interaction with CTCF genome wide (Wendt et al., 2008, Parelho et al., 2008, Rubio et al., 2008). Interactions between CTCF and cohesin take place between the C-terminal region of CTCF and the SA1 or SA2 subunits of cohesin (Kojic et al., 2018, Xiao et al., 2011). A majority of CTCF-binding sites are co-occupied by cohesin (Hadjur et al., 2009, Parelho et al., 2008). However, the loss of cohesin does not affect CTCF binding (Rao et al., 2017). Therefore, while cohesin interacts with CTCF, the stability and loading of CTCF on the genome is not regulated by cohesin but by

the methylation status of DNA and interactions with other DNA binding proteins such as chromatin modifiers and transcription factors.

1.1.7 Chromatin loops

Chromatin loops appear as single spots on HiC maps that show a greater interaction frequency than surrounding DNA (Figure 1.1A). Loops identified in an immortalised B-cell line were shown to frequently have a known promoter at one end, and a known enhancer at the other. (Enhancers are DNA sequences that exert regulatory effects on target genes. They often exist at sites distant from their target genes. They are discussed in greater detail later in the introduction.) This happened far more likely than one would expect by chance (Rao et al., 2014). These loops included classic known promoter-enhancer loops, including *Myc* and alpha-globin (Rao et al., 2014). Secondly, genes that have an enhancer-promoter loop are much more highly expressed than those that don't (Rao et al., 2014). This suggests an important role for enhancer-promoter interactions in gene regulation.

Loops can also be cell-type specific. As discussed previously, enhancers activation can be highly cell-type specific in order to regulate cell-type specific gene expression. Unique loops have been identified across multiple cell lines, with a corresponding increase in expression of the promoter targeted by its enhancer (Rao et al., 2014).

A classic example of loop formation and its role in gene regulation is that of the locus control region (LCR) in globin gene regulation. Globin genes transition through expression during

development. Initially embryonic globin is expressed, followed by foetal, and finally adult-type (β -globin). The expression of each globin subtype is controlled in developmentally appropriate manner through the generation of looped contacts between the LCR and the appropriate globin region during development (Tolhuis et al., 2002, Palstra et al., 2003). Deletion of the LCR leads to deregulated transcription of β -globin in an erythroblast cell line (Bartman et al., 2016), whereas forcing interaction between the LCR and β -globin genes leads to increased transcription (Bartman et al., 2016). Interestingly, the study showed that enhancer mediated regulation of genes seems to act through determining the number of active transcription sites in each cell, as opposed to the number of transcripts produced at each site. Forced looping between β -globin and the LCR led to an increased number of active β -globin alleles, but the number of transcripts generated at each transcription site remained the same (Bartman et al., 2016).

Further studies using forced chromatin looping have demonstrated the importance of enhancer-promoter interactions in gene regulation. Light-activated-dynamic-looping (LADL) is a technique whereby inactivated Cas9 tethered to a truncated version of the CIB1 protein (dCas9-CBN) from *Arabidopsis thaliana* is forced to heterodimerise with cryptochrome2 (CRY2) upon stimulation with blue light. By targeting two genomic regions with dCas9-CBN, introducing CRY2 into the nucleus, and treating cells with blue light, the two genomic regions are forced to interact forming a loop (Kim et al., 2019). Targeting an enhancer normally associated with *Klf4*, as well as targeting the promoter of *Zfp462*, both with dCas9-CBN, activation with blue light forced the enhancer away from its *Klf4* target and generated an interaction instead with the *Zfp462* promoter. This led to a corresponding increase in

expression of *Zfp462*. However, interestingly there was no effect on *Klf4* expression despite the loss of its enhancer contact (Kim et al., 2019).

Chromatin looping therefore appears important in regulating gene expression through enhancer activity. It appears that chromatin looping confers the ability for enhancers to exert their regulatory activity as discussed earlier by allowing enhancers to physically contact target promoters, leading to the regulation of cell-type specific genes and fine tuning the levels of expression.

Cohesin and CTCF were initially identified as being potentially important for chromatin looping based on overlaps of chromatin immunoprecipitation sequencing (ChIP-seq) data (Rao et al., 2014, Hadjur et al., 2009, Parelho et al., 2008, Seitan et al., 2011). Previous work has shown that cohesin was preferentially bound to DNase I hypersensitive sites, many of which contained the consensus binding motif for CTCF, and that CTCF binding was required for the localisation of cohesin at these sites (Parelho et al., 2008). It was then found that a majority of loops are bound by CTCF and cohesin at their anchor sites (Rao et al., 2014). Furthermore, the consensus DNA binding sequence (5'-CCACNAGGTGGCAG-3') for CTCF is not palindromic, and as such can exist in two orientations, either forward or backward. Overwhelming, loops are formed between two convergent CTCF sites, that is, CTCF motifs that point towards each other as opposed to away or in the same direction (Rao et al., 2014). It was shown that inverting CTCF sites disrupts loop formation without affecting CTCF binding, and that genes targeted by enhancers in these altered loop formations are deregulated (de Wit et al., 2015, Guo et al., 2015). Furthermore, the depletion of CTCF leads to the loss of

chromatin loops that are anchored between convergent CTCF sites, and they could be restored with the rescue of CTCF expression (Nora et al., 2017).

While CTCF is important for demarcating the boundaries of loop formation and laying down the blueprint of interactions, cohesin has been shown to be important for generating the physical interactions and maintaining organisation. The loss of cohesin leads to the loss of organisation of the genome in 3D space below compartment level (Rao et al., 2017), including domains and chromatin loops. Mice heterozygous for the cohesin loading protein NIPBL display reduced interactions between the β -globin LCR and β -globin genes, and reduced cohesin binding at the same sites, as well as a decrease in expression of β -globin genes (Chien et al., 2011). Cohesin has also been shown to be essential for long-range chromatin interactions and gene regulation at the *Ifng* cytokine locus in mouse and human (Hadjur et al., 2009) as well as the T-cell receptor (TCR) α locus (*Tcra*) (Seitan et al., 2011).

Chromatin loops appear to be important for the formation of enhancer-promoter interactions, which are through to allow enhancers to exert their regulatory effect. Examples have shown that the loss of chromatin looping can lead to deregulated gene expression, and the forcing of loops can activate genes. Additionally, high correlation between loop formation and expression levels also suggests their important role in enhancer mediated gene regulation. However, while the loss of specific chromatin loops can lead to changes in gene expression, the same rules have not been identified genome wide. The depletion of cohesin and subsequent loss of all chromatin loops leads to minimal changes in gene regulation (Rao et al., 2017). Therefore, while chromatin looping is certainly important in some situations, its role in gene regulation genome wide remains unclear.

What has been shown is that architectural proteins cohesin and CTCF play fundamental but different roles in chromatin loop formation and maintenance. CTCF is essential for laying the blueprint and defining regions of interactions, while cohesin is responsible for actually generating them. In the absence of CTCF or cohesin, 3D genome organisation is compromised.

1.1.8 Domains

The advent of high powered 3C techniques such as 4C, 5C and HiC have allowed for the study of 3D genome organisation at larger scales and greater resolution. One of the most striking discoveries was that chromosomes are subdivided into domains (Figure 1.1A, 1.4A, (Dixon et al., 2012, Nora et al., 2012, Sexton et al., 2012). Domains represent regions of the genome which are more likely to interact with themselves than surrounding regions. They can range in size from 40kB to 3Mb and can contain multiple genes and multiple regulatory elements such as enhancers (Figure 1.4A, (Rao et al., 2014). While domains can vary in structure across organisms and cell types, in mammals they are largely invariant across all tissues (Dekker and Heard, 2015). How domains interact with each other in larger organisation units known as compartments does however change dramatically across cell-types, and as such it has been suggested that domains are the building blocks of the genome in 3D space which are reconfigured in their associations with one another during development, but themselves do not change (Dekker and Mirny, 2016).

As discussed previously, loci within domains are more likely to interact with other loci within the same domain, and are relatively insulated from the loci outside of their domain. Work recently has shown that a function of the insulated domain is to provide a framework in which distal regulatory elements are more likely to interact with their target genes than with surrounding loci. This allows enhancers for example, to exert their regulatory effect through as they come into 3D proximity with their target genes. While domains do not facilitate point to point interactions, they merely provide a space in which certain interactions are more likely to occur.

How domain structure is determined is still being uncovered. A majority of domains are bordered by convergent CTCF sites, where they are also enriched for cohesin. Domain formation and loop formation are therefore rooted in the same active process. Domains are often formed as a result of loop formation between two anchors, and the active process of bringing the anchors together drives formation of a preferentially interacting region in between them, known as a domain. Some domains are not bordered by CTCF and therefore their blueprint is dependent on other architectural proteins (Rao et al., 2014). The loss of cohesin function leads to the loss of all domains (Rao et al., 2017, Schwarzer et al., 2017). Together these studies provide strong evidence that cohesin is essential for the organisation of the genome into domains, and for facilitating interactions where the limitations of which have been determined by CTCF and other architectural proteins.

It has also been shown that domain structure is variable but constrained at a single cell level. While domains differ between single-cells, their aggregate structure resembles those identified in bulk Hi-C. Following the depletion of cohesin, while domain like structures

remain, they are not represented in bulk Hi-C as the constraining activity of cohesin has been lost (Bintu et al., 2018), and they are randomly generated blocks of interacting chromatin. Cohesin is therefore not necessarily important for domain formation, but for the maintenance of domains where interactions are more likely to occur compared to background interactions.

While the loss of cohesin leads to the loss of all domains and smaller 3D structures, the depletion of CTCF has a different effect on 3D organisation. The loss of CTCF leads to the reduction of insulation at over 80% of boundaries. Of the remaining 18% of boundaries, these were identified as not being bordered by CTCF and where insulation is not affected by the loss of CTCF (Nora et al., 2017).

Inversion of CTCF sites at domain boundaries also affects their structure (Guo et al., 2015). Protocadherins (*Pcdh*) are a family of membrane bound proteins which a large number of variable structures. This is due to the presence of *Pcdh* clusters which contain “variable regions”. These regions contain a large number of variable exons that are alternately expressed in different PCDH proteins. Each variable exon is preceded by a promoter which is bound by CTCF. CTCF is also found at the *Pcdh* enhancers further downstream. These CTCF binding sites are positioned in a convergent orientation. Using CRISPR/Cas9, inversion of these CTCF sites led to the loss of interaction between the *Pcdh* enhancer and the variable exon promoters, and decreased expression of the variable exons. The same approach was used in the β -Globin locus with similar effect (Guo et al., 2015).

This work, amongst others, has shown that CTCF orientation in mammals, as well as other as of yet uncovered architectural proteins, determine the direction of 3D DNA interactions, and

together with cohesin are responsible for determining domain organisation. In the absence of cohesin there are serious genome wide consequences on the 3D organisation of the genome including the loss of all domains.

1.1.9 Compartments

Compartments are long range self-organising regions of preferentially interacting chromatin. Compartments appear as a plaid pattern in Hi-C maps, and are spread over great distances. They are formed of genomic regions with similar chromatin modifications. Genomic regions containing predominantly active chromatin modifications will associated with similarly modified active chromatin, and form “A” compartments. On the other hand, genomic regions containing inactive, or silenced chromatin, display an interaction preference for each other and come together to form “B” compartments (Figure 1.1A, (Lieberman-Aiden et al., 2009).

Since their initial discovery, compartments have been further subclassified into distinct groups (Rao et al., 2014). Two of them, A1 and A2, are associated with the originally identified A compartments. They are both gene dense, contain highly expressed genes, and are labelled with active chromatin marks such as H3K4me1, H3K27ac, H3K79me2 and H3K36me3. They are depleted at the nuclear lamina and at nucleolus-associated domains (NADs). A1 and A2 are separated from each other by replication timing, A1 compartments finish replicating at the beginning of S phase, while A2 compartment DNA is replicated into the middle of S phase. A2 compartments are also enriched for H3K9me3, have a lower GC content and contain longer genes (Rao et al., 2014).

The other subfamilies of compartments are known as B1, B2, B3 and B4 compartments. These all stem from the initially more broadly defined B compartments. Subcompartment B1 is positively associated with H3K27me₃, while depleted in H3K36me₃, implying it is predominantly made up of heterochromatin. However, this is counterintuitive to the observed B1 overlap with A compartments. B2 is present at the nuclear lamina and NADs, and contains a majority of the pericentromeric heterochromatin. B3 is also enriched at the nuclear lamina, but depleted at NADs. Finally, B4, the smallest of all subcompartments as it only spans a total of 11Mb of DNA, is enriched for many KRAB-ZNF superfamily genes, and display both activating and repressive chromatin marks (Rao et al., 2014).

Unlike domains, compartments are highly variable between cell types, and change during development and differentiation (Dixon et al., 2015, Lieberman-Aiden et al., 2009). This is unsurprising given they are determined by chromatin marks, and as cells develop they differentially activate and repress thousands of genes.

The mechanisms that govern compartment formation remain largely unknown, however recent work has implicated phase separation as being a driver behind compartment formation. Heterochromatin domains have been reported to form through phase-separated HP1 condensates that enrich for heterochromatin factors and exclude those of euchromatin (Strom et al., 2017). It is appealing to think that compartments are therefore generally formed through homotypic interactions between chromatin modifications.

Cohesin does seem to play a role in some aspects of compartment organisation. Following the loss of cohesin, compartments display increased strength, and the number of compartments increases. Compartment size is also reduced (Schwarzer et al., 2017, Rao et al., 2017). Compartments generated following the loss of cohesin correlate better with transcriptional activity and chromatin states than those in control cells, suggesting that cohesin plays an almost antagonistic role (Schwarzer et al., 2017). Compartment formation is therefore not reliant on cohesin, and the formation of domains by cohesin seems in fact to antagonise compartment formation and localisation (Rao et al., 2017).

1.1.10 Loop extrusion – how cohesin generates a 3D framework

While compartments are formed through self-organisation, domains and loops are reliant on the activity of proteins and protein complexes to generate their interactions. Cohesin is thought to generate 3D interactions by a process known as loop extrusion, where the cohesin ring physically pulls chromatin through in an active process.

During mitosis, all interphase 3D genome organisation is lost as chromosomes condense. Therefore, when cells exit mitosis they need to reconstruct the 3D organisation of their genomes. Compartments, due to being largely defined by chromatin state, are most likely regenerated due to retained histone modifications and other proteins that remain bound to DNA during mitosis and that re-establish the epigenetic landscape.

These mechanisms however cannot dictate the re-establishment of CTCF and cohesin mediated interactions such as loops and domains. For these structures to be re-established a complex needs to be present that binds DNA, can translocate along DNA, and be blocked by other complexes (such as CTCF) in a directional manner. Proteins have been identified that have translocase ability, and helicases can translocate along DNA without unwinding it, and these proteins share similarities with cohesin at their ATPase domains. In addition, cohesin shares structural similarities with kinesins and myosins, motor protein complexes responsible for cytoplasmic transport. Cohesin therefore is an attractive candidate to be the extruder that establishes and maintains 3D genome organisation.

Following the loss of cohesin function by proteolytic degradation, and the resulting loss of all domains and loops, its reintroduction leads to the rapid reestablishment of these structures (Rao et al., 2017). This supports the model that cohesin forms domains by dynamically binding and extruding DNA (Fudenberg et al., 2016). In this model, cohesin extrudes chromatin continuously as it is bound, until it reaches a “roadblock”, such as CTCF, where it stalls. This would explain the overlap of CTCF and cohesin binding at domain boundaries, but also explain the lower levels of cohesin often found in domains. The directionality of CTCF binding sites dictating domain formation – that domains are formed between convergent CTCF sites – also agrees with this model.

How exactly cohesin extrudes remains unknown. Different models have been suggested including diffusion of cohesin by Brownian motion along chromatin, the ability for cohesin to be its own ATP driven motor, and the “pushing” of cohesin along chromatin by other chromatin bound complexes such as Pol II.

The diffusion of cohesin along chromatin to explain its movement does not fit with models of the speed at which 3D genome organisation is re-established following rescue of cohesin function. In addition, the “pushing” idea does not explain how loops form in inactive regions of the genome, and is also too slow to explain reestablishment (Rowley and Corces, 2018).

SMC complexes do have ATPase activity and as such could act as their own motor. Cohesin’s close cousin, condensin (another SMC complex important for chromosome compaction in the cell cycle), has been shown to move along chromatin in vitro (Terakawa et al., 2017). In addition, the energy requirement for loop extrusion is in line with other cellular processes, and the removal of ATP followed by the loss and reintroduction of cohesin inhibited the reestablishment of cohesin-CTCF mediated 3D genome organisation, showing it is an ATP dependent process (Vian et al., 2018). Mutations that inactivate the ATPase ability of cohesin led to the same defects, strongly suggesting cohesin ATPase activity and loop extrusion drives 3D genome organisation (Vian et al., 2018).

Loop extrusion through cohesin motor activity seems to be the best current fit for models of loops extrusion. While work has yet to directly show that cohesin is able to extrude chromatin in eukaryotic cells, supporting evidence for the model continues to grow.

1.1.11 Enhancers

Regulation of genes is not just determined by transcription factors and transcriptional machinery binding to promoter regions but also due to distal regions of DNA known as enhancers. They are defined as sequences of non-coding DNA that induce spatiotemporally precise gene regulation, however the mechanisms by which this occurs remains unknown (Furlong and Levine, 2018). What is known is that enhancers exert regulatory effect through the recruitment of combinations of sequence-specific transcription factors, which in turn recruit RNA polymerase II and other transcriptional coactivators to target genes (Haberle and Stark, 2018, Long et al., 2016). Enhancers display gene specificity and not all enhancers can activate all genes.

Enhancers are identified based on a distinct chromatin signature. They are accessible, marked by H3K4me1 or H3K4ac27 and bind transcribing Pol II (Catarino and Stark, 2018). However, sequences that contain these features may not necessarily display “enhancer” functions, and need to be experimentally validated through reporter assays. Such experiments have highlighted the importance of enhancers in determining the diversity of gene expression observed across cell types and throughout development, and the resulting complexity of metazoa.

While enhancers have been shown to be important for the dynamic regulation of genes. Not all genes rely on enhancers for expression. Those that do tend to be genes with precise spatiotemporal expression patterns, such as during development (Levine, 2010, Long et al.,

2016), cell-type specific genes (Heinz et al., 2015), or genes which need to rapidly regulated in response to external stimuli, such as inflammatory response genes in macrophages (Glass and Natoli, 2016), and activity dependent genes in neurons (Kim et al., 2010).

Multiple enhancers can control the activity of multiple genes and promoters with spatiotemporal precision. The gene *Sonic hedgehog* (*Shh*) relies on the differential activity of 11 distinct enhancer elements to drive correct expression in different tissues such as the central nervous system, epithelia and limbs during mouse embryogenesis (Anderson et al., 2014). Loss of specific *Shh* targeting enhancer elements leads to disruption of *Shh* regulation in specific tissues at specific times (Figure 1.4 A, B). Removal of the limb specific ZRS enhancer leads to the loss of expression of *Shh* only in the developing limb, leading to reduced limb growth, while other tissues remain unaffected (Anderson et al., 2014).

The physical interaction between enhancers and promoters (Figure 1.4) are thought to be important for allowing enhancers to exert their regulatory effect on target genes. While interactions have been shown to be mediated by proteins discussed earlier such as cohesin and CTCF, transcription factors are increasingly being implicated as being important for mediating enhancer-promoter interactions. Enhancers as well as promoters contain transcription factor binding sites, often for the same transcription factors, or those known to interact with each other. There are several mechanisms proposed by which transcription factors can mediate 3D genome interactions. Direct oligomerisation of transcription factors or oligomerisation with cofactors has been proposed to form interactions between distal chromatin sites (Kim and Shendure, 2019). In mouse erythroid cells LIM domain binding

protein (Ldb1) is recruited by the cofactor Lmo2 and bridges interactions with the transcription factors Gata1, Tal1 and E2A (Love et al., 2014b).

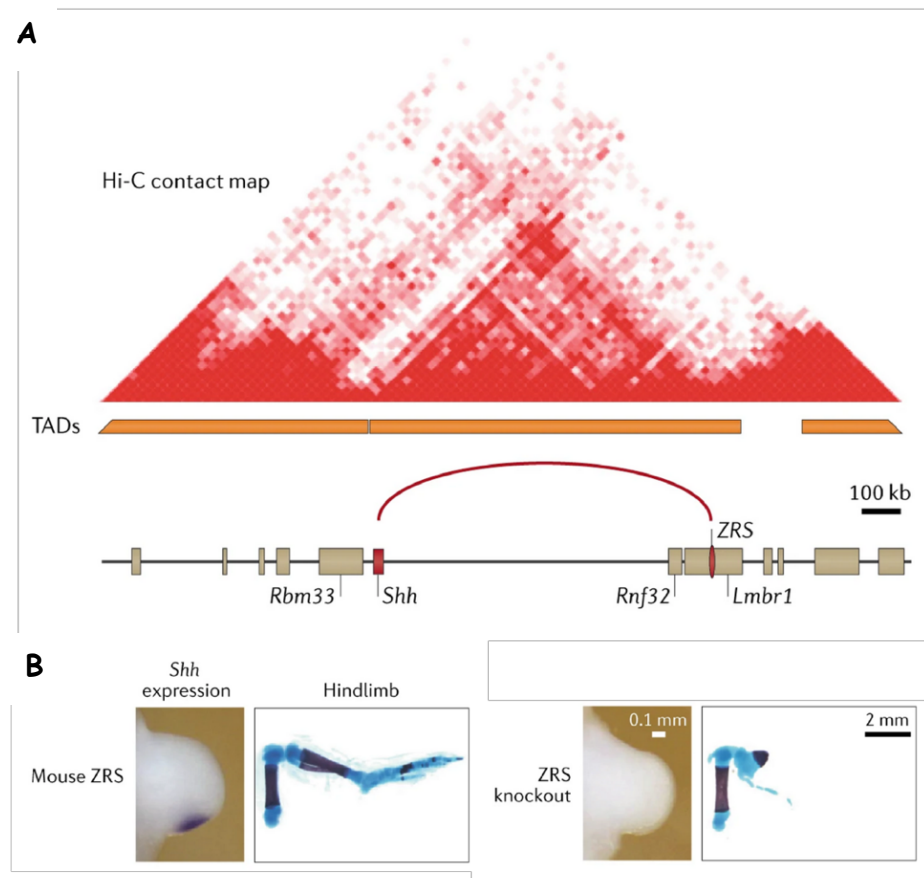


Figure 1.4 Organisation of the sonic hedgehog (*Shh*) locus in 3D space and enhancer based regulation of limb development. (A) A section of a Hi-C map showing the organisation of the *Shh* locus, and the interaction between the ZRS and *Shh* gene. (B) The loss of the ZRS leads to the loss of *Shh* expression and abnormal hindlimb development. Adapted from (Schoenfelder and Fraser, 2019).

Recently stabilisation of interactions between distal genomic regions based on liquid-liquid phase separation has been proposed. Phase separated droplets have a composition distinct from the surround solution. Transcription factors are enriched for intrinsically disordered regions (IDRs) which have been shown to be important for condensate formation *in vitro* and *in vivo* (Banani et al., 2017, Boija et al., 2018, Cho et al., 2018). Enhancers that share

transcription factor binding sites have been shown to colocalise in a manner that appears to be phase separated, where their transcription factors are not known to oligomerise (Denholtz et al., 2013, Nair et al., 2019, de Wit et al., 2013). This is consistent with the possibility of condensate formation mediating their interactions. Phase separation however remains contentious and the distinction between phase separation and smaller protein-protein assemblies remains beyond our abilities.

Metazoa have thousands of distinct cell subtypes. These unique cells must maintain their unique function in order for the organism to survive optimally. This requires to the precise regulation of cell-type specific genes. While some genes are active in all cell-types, known as housekeeping genes, others are highly cell-type specific. Housekeeping genes are associated with universal cellular functions such as homeostasis, growth, translation and transportation. Cell-type specific genes are genes that confer unique function onto cells, for example, making neurons able to form memories, and macrophages able to fight infection. Cell-type specific gene regulation is thought to be at least partly regulated by cell-type specific enhancer activity (Heinz et al., 2015, Nord et al., 2013). While the mammalian genome contains millions of potential enhancers, only several thousand are active in a given cell type (Consortium, 2012, Nord et al., 2013).

Lineage determining factors are essential for the establishment and maintenance of cell-type specific function (Heinz et al., 2015). They are expressed during development and cell differentiation, and are able to prime and activate cell-type specific enhancers in a cell-type specific manner, conferring unique cellular function (Ghisletti et al., 2010, Heinz et al., 2010, Heinz et al., 2015). A classic example is that of PU.1 which is a lineage determining

transcription factor required for both the development of macrophages and B cells. PU.1 binding sites are highly variable between B cells and macrophages, and PU.1 therefore occupies largely unique sites between the two cell types (Heinz et al., 2010). Macrophage-specific binding sites of PU.1 overlap with the binding sites of other macrophage specific lineage determining transcription factors, and the same is observed in B cells. The loss of PU.1 or associated lineage determining transcription factors revealed an interdependence on each other, and they require each other to bind to DNA and exert their regulatory function determining cell differentiation. The binding of PU.1 and associated factors is required for the monomethylation of H3K4, and inducing chromatin accessibility, both known enhancer marks (Heinz et al., 2010, Ghisletti et al., 2010). Therefore, lineage determining transcription factors are thought to activate cell-type specific enhancers, which in turn regulate cell-type specific genes, allowing for unique cell function.

Enhancers also play an important role in the dynamic regulation of genes in response to stimuli. For example, the regulation of inflammatory genes in macrophages relies on the activation and repression of many enhancer elements (Ostuni et al., 2013, Glass and Natoli, 2016). Stimulation of macrophages with lipopolysaccharide (LPS) – a component of the bacterial cell wall – is used to mimic pathogen detection by macrophages. LPS activates downstream signalling pathways through toll-like receptor 4 (TLR4) eventually leading to the induction of an inflammatory gene expression program of hundreds of genes and thousands of regulatory elements (Bhatt et al., 2012). Work has shown that following LPS stimulation, regions of DNA become accessible, gain H3K27ac and H3K4me1, and begin transcribing, the hallmarks of enhancer activity (Ostuni et al., 2013, Kaikkonen et al., 2013). Activated enhancers in macrophages contain binding sites for known transcriptional regulators of the

inflammatory gene expression program such as PU.1, STAT and JUNB, highlighting their importance in regulating this process (Ostuni et al., 2013). Loss of enhancer activation in macrophages in response to inflammatory stimuli leads to a loss of inducible gene regulation (Cuartero et al., 2018, Ostuni et al., 2013).

Much like macrophages, neurons have an inducible gene expression program. Neurons upregulate genes in response to depolarisation of their membrane caused either by treatment with potassium chloride (KCl), or with proteins that bind to activating receptors on the synapse such as brain derived neurotrophic factor (BDNF). Much like in macrophages, following stimulation of neurons with KCl regions of DNA gain hallmark signs of enhancer activation, including H3K27ac and transcription (Kim et al., 2010, Malik et al., 2014). Activated enhancers contain binding sites for transcription factors known to be essential for the neuronal inducible gene expression such as FOS and ARC (Malik et al., 2014, Kim et al., 2010). These inducible genes are essential for the synaptic development and function, and for brain specific processes such as memory formation (Frey et al., 1996, Nguyen et al., 1994). Targeting the enhancers of the inducible gene *Fos* using an inactivated Cas9 fused with a Krüppel-associated box (KRAB) domain to induce heterochromatin formation, led to the loss of correct *Fos* regulation in response to KCl and BDNF stimulation (Joo et al., 2016). This highlights the importance of enhancers in the inducible gene expression program in neurons.

As well as displaying high specificity such as the example above, enhancers can also display redundancy. More commonly, genes are controlled through the activity of multiple redundant enhancers that allows for resistance to genetic variation, often multiple enhancers need to be deleted for there to be an observable effect on target gene expression

(Osterwalder et al., 2018). The presence of multiple enhancer elements also allows for the generation of new regulatory activities through evolution, as they provide a template for mutation and the gaining of new spatiotemporal activation and subsequent changes to gene expression.

It is also possible that enhancer redundancy exists in order to guarantee expression of target genes across cells. Should an enhancer fail in one cell, others can take over and regulate their target genes. By having so many redundant enhancers, different ones can fail in different cells across a population, but ultimately have no effect on gene regulation (Perry et al., 2011).

Overall it is clear that enhancers play a crucial role in gene regulation. They allow for dynamic spatiotemporal transcriptional regulation, which gives rise to key processes during development and allows different cell types to carry out their specific functions.

While the precise mechanisms that underlie enhancer function remain unknown, it is now known that 3D genome organisation plays a crucial role in allowing enhancers to exert their regulatory function on target genes (Cuartero et al., 2018, Schoenfelder and Fraser, 2019, Merckenschlager and Nora, 2016). In order for enhancers to exert their regulatory effect, they must be brought into close proximity to their target promoter to allow for recruitment of the transcriptional machinery. The 3D organisation of the genome provides a framework in which enhancers come into contact with their target promoters and through this facilitates dynamic enhancer driven gene regulation.

1.1.12 The role of 3D genome organisation in transcriptional regulation

The role of 3D genome organisation in gene regulation still remains relatively unknown. Initial work focused on specific loci where expression was known to be regulated by a key enhancer, such as the T-cell-receptor and β -Globin loci (Seitan et al., 2011, Palstra et al., 2003). Here it was shown that the loss of cohesin or loss of interactions had detrimental effects on gene expression. With the advent of Hi-C, questions related to how genome wide changes to genome organisation affect changes in gene expression have started to be addressed.

The importance of chromatin loops in gene regulation has been studied relatively extensively. Their importance in gene regulation are easier to define and understand than that of larger 3D genome structures such as domains. Chromatin loops bring together genes with highly specific enhancers for correct expression, often in complex systems where a significant degree of variable transcripts are produced. Chromatin loops have been shown to be essential for the regulation of globin genes during development, and manipulation of chromatin looping can induce changes in gene expression (Deng et al., 2014). The identification of thousands of chromatin loops and their subsequent disappearance following cohesin, led to minimal effect on gene expression (Rao et al., 2017). This suggests that there are some chromatin loops that are essential for gene regulation in specific cells and at certain stages, the majority of chromatin loops are non-essential for correct gene regulation.

Understanding the functional relevance of A and B compartments in regulating gene expression is less clear and remains a challenge as we lack the knowledge and tools to freely

manipulate them. Histone modifications and chromatin compaction are sufficient to trigger A – B compartment switching, suggesting transcription factor and chromatin modifier DNA binding events can drive conformation changes at the compartment level (Stadhouders et al., 2019).

Much like chromatin loops, the role of domains in gene expression has not been fully understood. While specific examples have been identified where domain formation and function is essential, the loss of domains by cohesin depletion does not lead to much transcriptional deregulation (Rao et al., 2017, Schwarzer et al., 2017). The importance of domains in constraining interactions and facilitating preferable enhancer-promoter interactions suggests they are essential for gene regulation. It also implies that correct 3D genome organisation does not just facilitate interactions, but drives them, allowing for enhancer control of gene regulation by insulation. This would lend itself to the idea that enhancers are able to activate genes in their domain relatively indiscriminately, or at the very least gene expression within domains would be relatively coordinated. Work has shown that during the differentiation of embryonic stem cells (ESCs) into neural progenitor cells, genes in the same domain tend to be more correlated in expression than those in neighbouring domains (Nora et al., 2012).

Work from Lupianez and colleagues showed that structural variations due to genomic deletions, duplications and inversions at the *EPHA4* domain in humans causes abnormal limb development such as brachydactyly and polydactyly. This domain, and the surrounding domains, contain genes essential for the regulation of limb development, such as *WNT6*, *IHH*, *EPHA4* and *PAX3*. Mouse models replicating the human genomic variations showed changes

in interactions at these domains, and the mimicking of the human limb abnormalities. Inversion of one domain boundary led to repositioning of limb enhancers into the same domain as the *Wnt6* gene. This led to increased interactions with the gene and its upregulation in limb tissues where the enhancers were active (Lupianez et al., 2015).

Work focused on understanding the role of cohesin in gene expression has been ongoing, well before the current understanding of the scale of 3D genome organisation, and the importance cohesin plays in it.

During T-cell-receptor rearrangement, gene segments distributed over several Mb of DNA are transcribed and recombined in a development specific fashion (Stanhope-Baker et al., 1996, Jhunjhunwala et al., 2009), and it is essential for determining cells for differentiation. Cohesin binding at the T-cell-receptor α (*Tcra*) locus is found at key regions such as the *Tcra* enhancer and promoter, and forms a boundary that excludes neighbouring genes as well as the neighbouring T-cell-receptor segments that are expressed at different developmental stages (Seitan et al., 2011). When cohesin was depleted through the genetic removal of *Rad21*, there was a decrease in expression of *Tcra* and an increase in expression of surrounding genes. The loss of cohesin led to a loss of interactions between the *Tcra* promoter and enhancer, as well as the in-between *Tcra* segments. As a result, thymocytes displayed abnormal T-cell-receptor rearrangement, and abnormal differentiation (Seitan et al., 2011).

Work discussed later in this thesis shows that cohesin, and by extension 3D genome organisation is essential for the inflammatory response in macrophages (Cuartero et al., 2018). In the absence of correct domain formation and 3D organisation, inducible genes in

macrophages in response to the mimicking of a pathogen are unable to be correctly regulated, and their enhancers fail to induce (Cuartero et al., 2018). Additionally, work will show that cohesin, and by extension 3D genome organisation, is required for the expression of cell-type specific genes in neurons (unpublished).

These examples, suggest that 3D organisation is highly important for gene regulation. However, the field remains somewhat murky. Others have shown that total cohesin loss has a very limited effect on gene regulation with only 146 genes changing expression level by more than 1.75-fold (Rao et al., 2017). While the loss of cohesin loading protein NIPBL in hepatocytes led to the significant deregulation of 1124 genes, but gene ontology showed no functional enrichment (Schwarzer et al., 2017).

The discrepancies in studies may be explained by several factors. Many studies are done in cell lines in a steady-state, unchallenged in an artificial environment. Therefore, these cells do not undergo many changes in gene regulation. It is in these settings where the most minimal changes in gene expression are observed (Rao et al., 2017). Cells which are exposed to external stimuli and need to respond with a transcriptional program, suffer from greater deregulation following the loss of 3D organisation (Cuartero et al., 2018). Macrophages need cohesin, and therefore 3D organisation to generate a transcriptional response to microbial signals, and it is emerging that 3D genome organisation also plays a key role in cell development and differentiation (Zheng and Xie, 2019). Perhaps in order to really understand the importance of the 3D genome we need to use primary cells, or *in vivo* studies, that challenge systems to be dynamic and undergo changes in transcription. Then we might start to understand the importance of 3D genome. Finally, discrepancies in studies arise from each

author's own definition of "significant" gene regulation changes. The use of expression fold-change as a measure of significance, and by extension the importance of 3D genome organisation can be misleading. Often small fluctuations in lowly expressed genes can have dramatic consequences on the ability for cells to function. Low level expression of cytokines in macrophages before encountering a pathogen is essential for mounting an immune response (Bhatt et al., 2012). Disruption of lowly expressed genes may therefore have dramatic consequences on organismal function.

More work needs to be done in a variety of systems in order to understand when and where correct 3D genome organisation is so essential in gene regulation and cell function, and to potentially develop a set of rules and an understanding of which genes and processes rely on it the most.

1.2 Models for understanding 3D genome organisation in gene regulation

1.2.1 Using proteolytic cleavage to investigate protein function – the RAD21-TEV system

In order to address questions related to 3D genome organisation in gene expression, systems based on the proteolytic cleavage of architectural proteins such as CTCF or RAD21 have been

developed (Nora et al., 2017, Natsume et al., 2016, Rao et al., 2017, Tachibana-Konwalski et al., 2010). These systems offer distinct advantages over more conventional genetic knock-out and RNAi based systems. They can be more precisely temporally controlled, and the loss of protein is rapid. Genetic depletion of RAD21 takes up to three days, however similar levels of cohesin depletion can be achieved through proteolytic cleavage in a matter of hours. Genetic knock-outs and RNAi are not always efficient, and the degree of loss of protein is often greater using proteolysis-based systems. The timing and degree of protein depletion can be adjusted in proteolysis systems, and is often reversible, allowing for dosage dependent as well as rescue experiments (Natsume et al., 2016, Nora et al., 2017, Rao et al., 2017, Tachibana-Konwalski et al., 2010, Cuartero et al., 2018). Genetic and RNAi based cohesin depletion systems also lead to the presence of confounding effects, Firstly, the key requirement for cohesin in cell-cycle means that data taken from genetic depletions and RNAi experiments often have serious confounding effects on gene regulation and organisation, as cells can undergo aberrant cell cycle (Wendt et al., 2008, Kagey et al., 2010, Nitzsche et al., 2011). Secondly, knock-on effects of cohesin depletion on gene regulation can also lead to changes in gene regulation not directly caused by cohesin depletion. The rapid nature of proteolytic cleavage means that proteins can be depleted quickly to minimise secondary effects caused by aberrant cell cycle and knock-on changes from defective gene regulation.

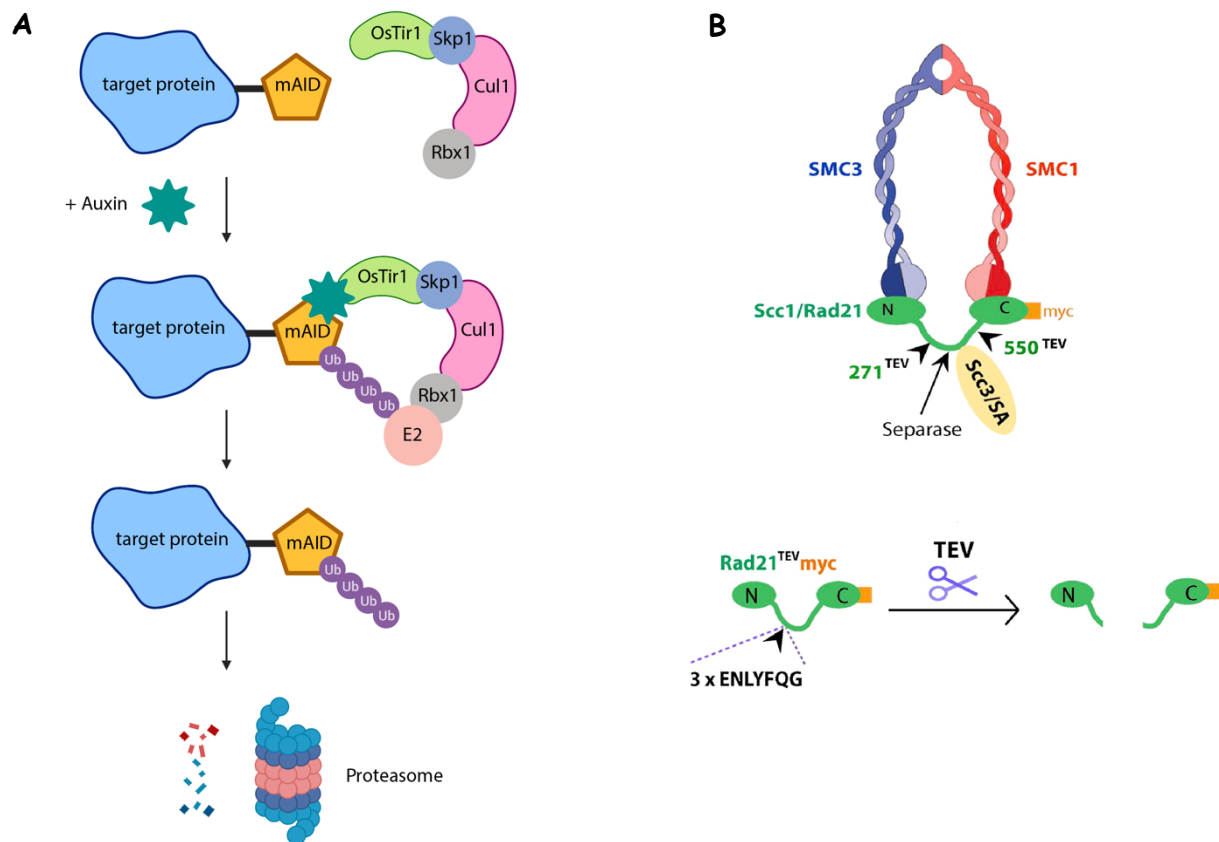


Figure 1.5 AID and TEV based systems for targeted proteolytic cleavage. (A) A schematic representation of the AID degron system. Forced expression of OsTIR1 and addition of auxin leads to recruitment of the ligase complex and the degradation of an AID tagged protein following polyubiquitinylation. (B) Schematic representation of modified cohesin complex where RAD21 contains two TEV cleavage sites, which upon TEV protease expression is degraded. The endogenous separase site is marked by the arrow. Adapted from (Pauli et al., 2008).

A common proteolytic system currently being used is the auxin-inducible degron (AID) (Figure 1.5A). In AID, a plant-specific degradation pathway, controlled by the phytohormone, auxin, is introduced into non-plant cells. Endogenous proteins are fused with an AID tag, and cells are forced to express the auxin responsive F-box protein TIR1. Treatment of the cells with auxin leads to the rapid degradation of the AID fused protein. Removal of auxin from cells leads to the termination of auxin activated degradation and the rescue of protein expression (Natsume et al., 2016). The system has been widely employed to study the role of

architectural proteins in 3D genome organisation and gene regulation with success (Rao et al., 2017, Nora et al., 2017). Its use however remains limited to cell lines, and no mouse models yet exist. This means primary cells cannot be studied. As discussed previously, the lack of clarity on the importance of 3D genome organisation on gene expression and cell function may arise from the lack of investigation into the role of 3D genome organisation in primary cells and *in vivo* studies. The AID system has yet to break this barrier. Secondly, the system is leaky. AID fused proteins in cells expressing TIR1 experience significant levels of depletion even before the addition of auxin (Nora et al., 2017), although this seems dependent on what protein is tagged in what cell type, and which promoter is used to drive TIR1 expression (Natsume et al., 2016). If there is leakiness however, changes in gene expression as well as cell function may be less dramatic than expected due to the already low baseline.

An alternative proteolytic system is that based on the tobacco etch virus (TEV) protease. The TEV protease is a viral protease that is not found in animals. It identifies and cleaves a specific sequence, the TEV sequence. Introduction of a TEV cleavage site into cohesin proteins and its cleavage by the TEV protease has been successfully used in yeast, flies and mouse oocytes (Uhlmann et al., 2000, Tachibana-Konwalski et al., 2010, Pauli et al., 2008) (Figure 1.5B).

Kim Nasmyth (University of Oxford, Department of Biochemistry) kindly provided our group with mice where RAD21 has been modified to be targetable by TEV protease. Two TEV cleavage sites were inserted in the flexible polypeptide region between the N- and C-terminal domains of RAD21 (RAD21-TEV). Cleavage of RAD21-TEV by TEV protease leads to cohesin dissociation from chromatin, much like the during the cell cycle when separase cleaves RAD21, leading to cohesin dissociation from chromatin and triggering sister-chromatid

disjunction. A Myc epitope tag was fused to the C-terminal of RAD21, allowing for its identification using molecular biology techniques.

Mice which are homozygous for RAD21-TEV, that is where both alleles of *Rad21* have been modified with the TEV cleavage site, are embryonically lethal, dying shortly before birth. However, many primary cells can be taken from RAD21-TEV homozygous embryos before death and grown in culture, including neurons, macrophages and fibroblasts. These cells express only cleavage RAD21, allowing for degradation of cohesin in primary cells. This overcomes an obstacle posed by the AID system where primary cell containing AID tagged architectural proteins have yet to be developed. This is especially important as using primary cells is key to furthering our knowledge of the importance of 3D genome organisation in gene regulation.

The development, adaptation and utilisation of the RAD21-TEV system in primary cells forms the bulk of this thesis. It is now an established powerful tool to study the role of cohesin in gene regulation in primary cells across a variety of conditions.

1.2.2 Macrophage inflammatory gene expression

Macrophages are immune cells present in all tissues of the body. They monitor and respond to the tissue environment and play key roles in homeostasis, development and immune response.

Macrophages respond to a variety of danger signals in response to invading pathogens or injury (Figure 1.6A). Different danger signals are detected in different regions in the cell via different mechanisms. The cell surface associated transmembrane Toll-like receptor 4 (TLR4) is responsible for the sensing of LPS (Chow et al., 1999). TLR3 on endosomes detects double-stranded RNA following phagosomal viral uptake. In the cytoplasm, RIG-1 senses double-stranded RNA as well, and cGAS sense DNA. Endogenous danger signals such as extracellular ATP and HMGB1 (a chromatin associated protein important for transcription) are both released following injury and trigger a macrophage response (Figure 1.6A). The goal of a macrophage in response to the wide variety of danger signals is ultimately to clear tissue of invaders, repair damage and restore homeostasis (Medzhitov and Horng, 2009).

Each signal triggers a distinct inflammatory gene expression program and modification of the regulatory landscape in order to elicit the appropriate response (Figure 1.6B). The inflammatory gene expression program has a time dependent and hierarchical structure. Genes are expressed in waves following LPS stimulation. The first genes to be expressed are primary targets, which in turn regulate the expression of secondary targets, and so on, giving the program its hierarchical nature (Bhatt et al., 2012).

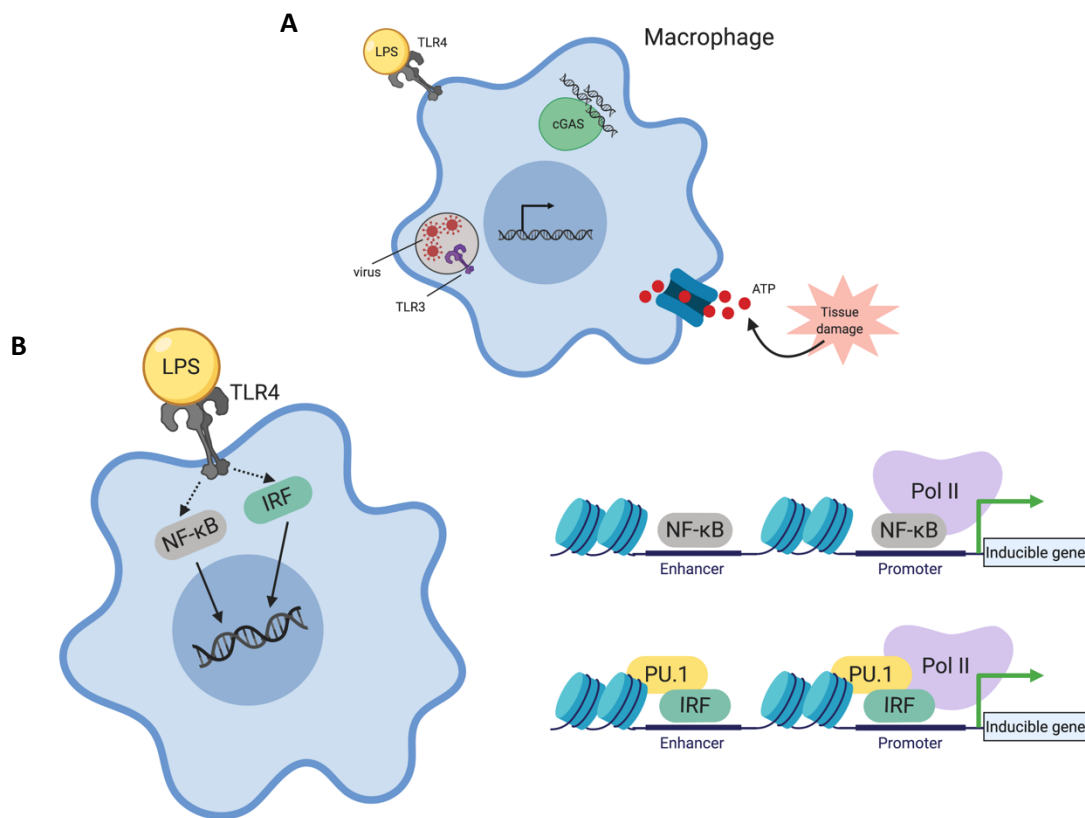


Figure 1.6 Macrophages respond to surrounding danger signals and generate a transcriptional response. (A) A schematic illustrating some of the stimuli that trigger changes in gene expression in macrophages. Stimuli can be both from exogenous and endogenous sources. (B) LPS activation of the TLR4 receptor leads to downstream signalling and activation of transcription factors, including NF-κB and IRF. These then translocate to the nucleus where they bind to regulatory elements and promoters. Along with other transcription factors they recruit RNA Pol II leading to inducible gene expression.

Activation of the LPS-dependent gene expression program in macrophages is controlled by a large number of transcription factors. Several key transcription factors that regulate the inflammatory response are already expressed before LPS stimulation. These transcription factors are generally located in the cytoplasm and their signal-dependent activation leads to

translocation to the nucleus (Figure 1.6B). This group includes proteins such as interferon (IFN) regulatory factors (IRFs), cAMP responsive-element-binding protein 1 (CREB1) and NF- κ B. These transcription factors are responsible for the regulation of primary response genes (Medzhitov and Horng, 2009). More transcription factors are newly generated following LPS stimulation. They regulate the expression of inducible genes following the primary response genes and the continuing waves of gene expression following LPS stimulation. They can maintain inflammatory gene activation over a prolonged period of time often through positive feedback loops. Finally, lineage-specific transcription factors also play an important role in regulation inflammatory gene expression in response to LPS. These transcription factors, including PU.1, IRF8 and RUNX1 are not exclusively expressed in macrophages, they can also be found in other immune cells such as B cells. These transcription factors determine the chromatin state of inducible genes, determining cell and stimulus specific responses (Medzhitov and Horng, 2009).

Enhancers also play an important role in the stimulus specific regulation of the inflammatory gene expression program in macrophages. They can be separated into three main categories, constitutive, poised and latent. Constitutive and poised enhancers are marked by accessible chromatin and active chromatin marks before exposure to external stimuli such as LPS. Constitutive enhancers are marked by both H3K4me1 and K3K27ac, while poised lack the acetylation mark, and gain it following activation. Latent enhancers are inaccessible and do not have active chromatin marks, but gain these characteristics upon stimulation. Following exposure to LPS, latent enhancers are bound by lineage specific transcription factors such as PU.1, and stimulus specific transcription factors such as STAT1/6. They are then given active chromatin marks and made accessible. Due to enhancers being enriched for the motifs and

binding of inflammatory associated transcription factors, and the wide-ranging changes in enhancer activation following LPS stimulation of macrophages, they are thought to be important for the regulation of inflammatory gene expression. Indeed the loss of STAT activation and subsequent loss of enhancer activation, leads to deregulated inflammatory gene expression (Ostuni et al., 2013).

The inducible inflammatory gene expression program in macrophages provides an ideal model to further our understanding of the role of 3D genome organisation in transcriptional regulation. The macrophage transcriptional response to LPS is highly stereotyped and understood, therefore changes can easily be identified. The same is true for changes to chromatin marks associated with enhancer activation and repression. The importance of enhancers in macrophage inducible gene expression, allows for the testing of the supposed role of 3D genome organisation in gene regulation. If cohesin is lost and therefore enhancer-promoter contacts disorganised, can enhancers still exert their regulatory function?

Work from the lab, has shown that cohesin is required for the inflammatory gene expression program in macrophages. In the absence of cohesin, inducible gene expression in response to LPS is deregulated. While constitutive enhancers are largely unaffected, enhancers that are either induced or repressed in response to LPS fail to be properly regulated. Those that fail are enriched for interferon stimulated response elements (ISRE) and PU.1 binding sites, while those for NF- κ B remain inducible. Inflammatory gene expression could be partly rescued by providing cohesin deficient macrophages with IFN during LPS stimulation (Cuartero et al., 2018). Work in this thesis discusses whether the effects on inducible gene expression observed in cohesin deficient macrophages is due to a direct or indirect role of cohesin. Does

cohesin directly regulate inducible gene expression, or is their deregulation a secondary result due to cohesin's requirement elsewhere in the cell?

1.2.3 Neuronal gene expression and function

Neurons are the primary cell of the brain. They drive behaviour through learning and memory, and are responsible for giving us the ability to think and exert higher cognitive function. There are hundreds of different neuronal subtypes in the human brain, localised to different regions, with specialised anatomies and functions. At their core though, neurons share a common theme. They are communicators, cells made up of a cell body and thousands of dendrites that branch out and interact with other surrounding cells through junctions known as synapses. Through synapses information is transmitted throughout the brain, as well as to muscles and glands. Synapses are highly malleable, and can be strengthened or weakened, and even removed entirely throughout life (Schafer et al., 2012, Citri and Malenka, 2008). Memory and learning is rooted in these synaptic changes, in a process known as synaptic plasticity. Synapses communicate through the release of neurotransmitters. Neurotransmitters are released from one neuron, at the pre-synaptic terminal, and diffuse across the synaptic cleft, where they bind to receptors of other neurons at the post-synaptic terminal.

Activation of synapses and the passing of information from one neuron to another triggers an event known as an action potential. Action potentials occur as a result of changes to the membrane potential of a neuron, and are electrical impulses that travel along the cell. Action

potentials are facilitated by the presence of ion channels on neurons, that allow for the movement of K^+ , Na^+ , Cl^- and Ca^{2+} in and out of cells.

Like all cells, in order to carry out their specific functions, neurons express a wide variety of cell-type specific genes. Genes responsible for synaptic plasticity, neurotransmitter production, neurotransmitter receptors, ion channels, and all the transcription factors that drive these genes are just some examples.

Germline mutations to cohesin proteins and associated factors lead to the disease known as Cornelia de Lange Syndrome (CdLS, discussed in further detail below). The mouse model of CdLS, a *Nipbl*^{+/-} mouse, display abnormal behaviour, such as increased repetitive behaviour, abnormal anxiety behaviours, and decreased learning and memory (Chien et al., 2011, Fujita et al., 2017). Cohesin deficient neurons also display abnormal dendritic and synaptic development (Fujita et al., 2017). NIPBL has also been shown to be important for neuronal cortical migration (van den Berg et al., 2017). The role of cohesin in neuronal gene expression however remains unclear. While differentially expressed genes were identified in *Smc3*^{+/-} mouse pups, this was confined to one time point, 7 days after birth (Fujita et al., 2017). Furthermore, RAD21 depletion by CRISPR in cerebellar neurons had a limited effect on gene expression, with few genes being deregulated and only in response to behavioural tests (Yamada et al., 2019). Constitutively cohesin depleted animal models, such as *Nipbl*^{+/-} and *Smc3*^{+/-} undergo abnormal development (Fujita et al., 2017, Kawauchi et al., 2009). The changes seen in neuronal gene expression in these mice do not reflect the ongoing requirement for cohesin in regulation neuronal gene expression. Defects in development and lineage commitment could well account for the changes observed. Gene expression changes

in CRISPR RAD21 depletion were mild, potentially due to the low level of cohesin depletion, and as such the role of cohesin in neuronal gene regulation remains unknown. Therefore, as part of my thesis, I sought to understand the role of cohesin neuronal gene regulation.

Speaking more broadly than just understanding the role for cohesin in neuronal gene expression, neurons provide an ideal model to understand what types of genes are regulated by cohesin and 3D genome organisation. They are primary cells, which provide a more realistic insight into 3D genome organisation and gene regulation than cell lines. They are an active cell that even when grown in culture are communicating with each other and undergoing synaptic plasticity. This can also be further manipulated through stimulation with molecules such as BDNF and KCl, as discussed previously. Neurons are also post-mitotic, therefore any changes in gene expression caused by cohesin deficiency cannot be due to defective cell cycle. Finally, we can generate primary neurons from the RAD21-TEV mouse. This allows us to achieve temporal control of cohesin depletion and understand the role of cohesin in neuronal gene expression.

1.2.4 Cornelia de Lange Syndrome

Diseases caused by germline mutations to proteins in and associated with the cohesin complex, are known as cohesinopathies. These include the Cornelia de Lange Syndrome (CdLS) and Roberts Syndrome.

Roberts Syndrome is caused by mutations to *ESCO2* which is required for cohesion establishment in the cell cycle. Patients with Roberts Syndrome display chromatid cohesion defects and aneuploidy most likely as a result from defective chromosome segregation during the cell cycle. Robert Syndrome cells also have abnormal cell cycle with extended metaphase and anaphase (Vega et al., 2005). The loss of correct cohesin function in the cell cycle due to *ESCO2* mutations is highly likely to lead to abnormal development and be the driver of the symptoms of Roberts Syndrome, including slow growth, cognitive impairment and limb defects.

CdLS is caused by mutations to genes including *NIPBL*, *SMC1A*, *SMC3*, *RAD21* and *HDAC8* (Krantz et al., 2004, Tonkin et al., 2004, Kline et al., 2018). The vast majority, around 70% of cases, are caused by mutations to *NIPBL*. The prevalence of CdLS is somewhere between 1:10,000 and 1:30,000 (Kline et al., 2007). It is clinically diagnosed based on a distinctive craniofacial appearance and the presence of limb malformations. As patients age they also display cognitive impairment. CdLS exists as a spectrum, and not all patients display the same phenotypes, or the same severity (Figure 1.7) (Kline et al., 2018). There are also patients that display CdLS-like phenotypes, but do not pass all the test to be defined as having CdLS. These are caused by mutations to genes not involved in cohesin function, such as *BRD4* and *ANKRD11* (Figure 1.7) (Olley et al., 2018, Parenti et al., 2016).

Patients with CdLS can display a wide range of symptoms. Patients can suffer from defects in the gastrointestinal tract, visual impairment, hearing loss, seizures as well as others. Symptoms are wide ranging with few being consistently found in all patients. However, one feature that is found across all CdLS patients is cognitive impairment. While there is some

evidence for structural brain abnormalities, the number of CdLS brains investigated remains too low to draw any significant conclusions (Schrier et al., 2011, Vuilleumier et al., 2002).

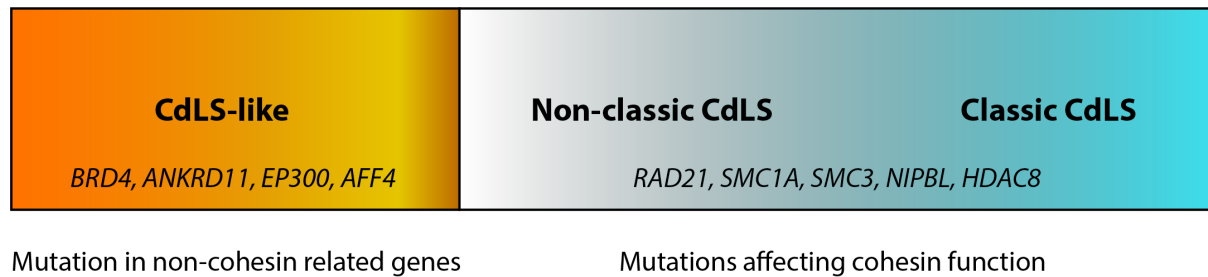


Figure 1.7 CdLS is a spectrum disease and can be caused by mutations to multiple genes. The spectrum of CdLS contains patients with classic CdLS where they have a loss of function mutation to a cohesin related gene, as well as patients who display several similar phenotypes but do not directly qualify as having CdLS (CdLS-like). Patients with cohesin mutations and CdLS can display all canonical symptoms or the minimal number of symptoms and to varying degrees of severity. Patients with CdLS-like symptoms have mutations in genes not associated with cohesin function, and also overlapping symptoms but not those required for CdLS diagnostic certainty.

Experimental perturbation of cohesin expression during development results in abnormal neuronal morphology, migration, defective gene expression and behavioural abnormalities (Fujita et al., 2017, van den Berg et al., 2017, Kawauchi et al., 2009, Remeseiro et al., 2012). However, it is not known whether defects in gene expression arise primarily from an essential role for cohesin in early development, during neuronal specification and lineage commitment, or whether cohesin is continually required by post-mitotic neurons. This is of interest because neurodevelopmental syndromes, such as Rett Syndrome, can benefit from postnatal interventions that correct aberrant gene expression (Cobb et al., 2010, Guy et al., 2007).

Identifying changes in gene expression following neuronal exit from the cell-cycle and development would allow us to ascertain if cohesin is continuously required in neuronal gene expression. Secondly, it remains unknown whether changes in gene expression caused by cohesin deficiency can be rescued. If there is a chance for postnatal intervention, this would need to be established first. Finally, there has been no characterisation of gene expression changes in patients with CdLS. If we want to understand the disease better, as well as the role of cohesin in neuronal gene expression, CdLS patient brain samples must undergo RNA-seq to determine their transcriptional profile. In this thesis I aim to address all three of these issues, to further our understanding of the role of cohesin in gene expression and the effect on CdLS patients.

1.3 Aims of this study

As our understanding of how chromatin is organised in 3D space grows, the physiological relevance of 3D genome organisation remains unknown. New techniques based on proteolytic cleavage systems allow for the interrogation of cohesin in gene regulation and cellular function. Furthermore, little is known about changes in gene regulation in real-life cohesin deficient scenarios, such as CdLS. Therefore, my thesis focuses on the establishment of a proteolytic cleavage system to remove cohesin function, and the use of this system to understand how and when cohesin is important for gene regulation.

Knowing that cohesin deficient macrophages are unable to regulate inducible gene expression, to what extent does this directly rely on cohesin function? Is cohesin required by every inducible gene to be expressed, or is it a key few regulators that determine the regulation of the whole system? I will investigate this using rapid cohesin degradation following by LPS stimulation and RNA-seq.

To understand the role of cohesin in neuronal gene expression, outside of its requirement during development, I will generate RAD21-TEV neurons before removing cohesin function. Any changes detected by RNA-seq will only be due to an ongoing requirement for cohesin, and will establish if cohesin is continually required by neurons. This data will also contribute to our understanding on 3D genome organisation in gene regulation of primary cells, in the absence of abnormal cell cycle.

Thirdly, understanding how permanent changes to gene expression are following the loss of cohesin function could be promising for the potential treatment of CdLS. If changes in gene expression caused by the loss of cohesin function, postnatal intervention, such as in Rett Syndrome could be developed (Cobb et al., 2010, Guy et al., 2007). Using the RAD21-TEV system, and the temporal control of cohesin depletion it offers, I will manipulate cohesin levels to deplete it followed by restoration, analysing gene expression changes at each stage.

Finally, our understanding of gene regulation in cohesin deficient humans remains woefully slim. No data has been collected on the transcriptional activity of humans with CdLS. Such insight is vital to further our understanding of CdLS and the role cohesin plays in human neuronal function. Using post-mortem tissue from patients with CdLS, I will analyse the transcriptome of CdLS neurons, and identify changes in gene expression.

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Antibodies

Rad21	Rabbit polyclonal to Rad21 (ab992; Abcam). Diluted 1:1000 for western blot.
c-Myc	Mouse monoclonal to c-Myc 9E10 (sc-40; Santa Cruz Biotechnology). Diluted 1:100 for western blot, 1:500 for immunofluorescence.
β - Actin	Mouse monoclonal to β-Actin AC-15 (sc-69879; Santa Cruz Biotechnology). Diluted 1:5000 for western blot.
v5	Mouse monoclonal to v5 tag clone V5-10 (V8012; Sigma-Aldrich Company). Diluted 1:250 for immunofluorescence.
TUJ1	Mouse monoclonal to Tubulin-Beta 3 (801213; BioLegend). Diluted 1:500 for immunofluorescence.
LAMINB	Goat polyclonal to Lamin B1 C20 (sc-6216; Santa Cruz Biotechnology). Diluted 1:2500 for western blot.
GAD67	Mouse monoclonal to glutamic acid decarboxylase (MAB5406; Millipore). Diluted 1:500 for immunofluorescence.
NEUN	Mouse monoclonal to NeuN (ab104224; Abcam). Diluted 1:1000 for immunofluorescence.
NEUN-488	Rabbit monoclonal to NeuN (ab190195, Abcam). Diluted 1:2000 for FACS.

GAPDH	Mouse monoclonal to glyceraldehyde-3-phosphate dehydrogenase (ab8245; Abcam). Diluted 1:15000 for western blot.
CD11b- FITC	Rat monoclonal conjugated with FITC to CD11b (561688; BD Biosciences). Diluted 1:100 for FACS.
F4/80-PE	Rat monoclonal conjugated with PE to F4/80 (12-4801-82; eBioscience). Diluted 1:200 for FACS.

2.1.2 Cell lines

HEK293T	Commercially available human embryonic kidney cell line used for lentivirus production.
HEK293T-GP	Commercially available human embryonic kidney cell line that stably expresses retroviral Gag-Pol, used for retrovirus production.
L929	Commercially available mouse fibroblast line used to generate conditioned media for macrophage differentiation.

2.1.3 Primers

Table 1. Primers for genotyping

Primers	Orientation	Sequence 5' – 3'
Scc1	Forward	KT315F: GAA GGT CAT GGT TGG CAG AT
	Reverse	KT630R: CAA TCC AAG CTT CAT TTG TAT TT
NEX	Multiplex	4: GAG TCC TGG AAT CAG TCT TTT TC
		5: AGA ATG TGG AGT AGG GTG
		6: CCG CAT AAC CAT TGA AAC AG

TEV	Forward	KT442FG: CAG CGA GAT GTC ATC GGT TA
	Reverse	KT443RG: TCA TTC TCC GAT CCA CTT CC

Table 2. Primers for cloning

Primers	Orientation	Sequence 5' – 3'
EcoR1 v5 Fwd	Forward	AAT GAA TTC GGT AAG CCT ATC CCT AAC CCT
EcoR1 Tev Rev	Reverse	ATT GAA TTC TCA CCC TTG CGA GTA CAC C
Lenti Backbone Fwd	Forward	GGT TCT ATC TCC TTC GAA GC
Lenti Backbone Rev	Reverse	TGA TGT ACA AGT AAA GCG GC
NLS Tev Fwd	Forward	ACA AAA AAG CAG GCT TCG AAG GAG ATA GAA CCG CCA CCA TGG CTA GCC CTA A
NLS Tev Rev	Reverse	GAA AGC TGG GTC CTA GCG GCC GCT TAA TCA ACT TTT C
UbcP IRES Fwd	Forward	TGA TTA AGC GGC CGC TAG GAC CCA GCT TTC TTG TAC
UbcP IRES Rev	Reverse	CTC GGA GGA GGC CAT GGT GGA CCG GTA AGC TTA TC
RFP Fwd	Forward	GCT TAC CGG TCC ACC ATG GCC TCC TCC GAG GAC GTC ATC AAG GAG TTC
RFP Rev	Reverse	TAG AGT CGC GGC CGC TTT ACT TGT ACA TCA GGC GCC GGT GGA GTG GCG GC

Table 3. Primers for Protocadherin B 4C

Locus	Primers	Orientation	Sequence 5' – 3'
Protocadherin Beta	C234 4C mHiNI HS1820R	Reverse	CAA GCA GAA GAC GGC ATA CGA GAT CTG TGT GAC TGG AGT TCA GAC GTG TGC TCT TCC GAT CAA TCC CAA CCT AAG ACA GC
	C233 4C mHiNI HS1820F1	Forward	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TAT TCA ATC AGG CCT TTT AAG CT
	C235 4C mHiNI HS1820F2	Forward	CAA GCA GAA GAC GGC ATA CGA GAT TGG CGT GAC TGG AGT TCA GAC GTG TGC TCT TCC GAT CAA TCC CAA CCT AAG ACA GC
	C23 4C mHiNI HS1820F3	Forward	CAA GCA GAA GAC GGC ATA CGA GAT TCT GGT GAC TGG AGT TCA GAC GTG TGC TCT TCC GAT CAA TCC CAA CCT AAG ACA GC

Table 4. Primers for *Ifnar1* 4C

Locus	Primers	Orientation	Sequence 5' – 3'
Ifnar1	C180 4C Ifnar1For	Forward	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TAG TCC ACT GGC TCG AAT CCT G
	C181 4C Ifnar1 ReIndex2	Reverse	CAA GCA GAA GAC GGC ATA CGA GAT ACA TGT GAC TGG AGT TCA GAC GTG TGC TCT TCC GAT CTT GGG TGC TAC CCT CAG CT
	C201 4C Ifnar1 ReIndex9	Reverse	CAA GCA GAA GAC GGC ATA CGA GAT CTG AGT GAC TGG AGT TCA GAC GTG TGC TCT TCC GAT CTT GGG TGC TAC CCT CAG CT
	C202 4C Ifnar1 ReIndex10	Reverse	CAA GCA GA AGA CGG CAT ACG AGA TAA GCG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT TGG GTG CTA CCC TCA GCT
	C203 4C Ifnar1 ReIndex11	Reverse	CAA GCA GAA GAC GGC ATA CGA GAT GTA GGT GAC TGG AGT TCA GAC GTG TGC TCT TCC GAT CTT GGG TGC TAC CCT CAG CT

Table 5. Primers for *Egr2* 4C

Locus	Primers	Orientation	Sequence 5' – 3'
Egr2	C184 4C Egr2For	Forward	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TGT TCG TTT CCT GGC TCC AAG
	C184 4C Egr2ReIn dex4	Reverse	CAA GCA GAA GAC GGC ATA CGA GAT CAC TGT GAC TGG AGT TCA GAC GTG TGC TCT TCC GAT CTG ACA GAA GCC ACC AAG ACA CTG
	C184 4C Egr2ReIn dex15	Reverse	CAA GCA GAA GAC GGC ATA CGA GAT CGT AGT GAC TGG AGT TCA GAC GTG TGC TCT TCC GAT CTG ACA GAA GCC ACC AAG ACA CTG
	C184 4C Egr2ReIn dex19	Reverse	CAA GCA GAA GAC GGC ATA CGA GAT CCA CGT GAC TGG AGT TCA GAC GTG TGC TCT TCC GAT CTG ACA GAA GCC ACC AAG ACA CTG
	C184 4C Egr2ReIn dex20	Reverse	CAA GCA GAA GAC GGC ATA CGA GAT ATC AGT GAC TGG AGT TCA GAC GTG TGC TCT TCC GAT CTG ACA GAA GCC ACC AAG ACA CTG

Table 6. Primers for *Cebpb* 4C

Locus	Primers	Orientation	Sequence 5' – 3'
Cebpb	C186 4C Cebpb For	Forward	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TGT TCG TTT CCT GGC TCC AAG
	C187 4C Cebpb ReIndex5	Reverse	CAA GCA GAA GAC GGC ATA CGA GAT CAC TGT GAC TGG AGT TCA GAC GTG TGC TCT TCC GAT CTG ACA GAA GCC ACC AAG ACA CTG
	C210 4C Cebpb ReIndex22	Reverse	CAA GCA GAA GAC GGC ATA CGA GAT CGT AGT GAC TGG AGT TCA GAC GTG TGC TCT TCC GAT CTG ACA GAA GCC ACC AAG ACA CTG
	C211 4C Cebpb ReIndex23	Reverse	CAA GCA GAA GAC GGC ATA CGA GAT CCA CGT GAC TGG AGT TCA GAC GTG TGC TCT TCC GAT CTG ACA GAA GCC ACC AAG ACA CTG
	C212 4C Cebpb ReIndex25	Reverse	CAA GCA GAA GAC GGC ATA CGA GAT ATC AGT GAC TGG AGT TCA GAC GTG TGC TCT TCC GAT CTG ACA GAA GCC ACC AAG ACA CTG

2.2 Methods

2.2.1 Genomic DNA extraction and genomic PCR

Mouse tails for genotyping were lysed in 100µl of lysis buffer (10mM TRIS (pH 7.5), 100mM NaCl, 10mM EDTA, 0.5%N-Lauroylsarcosine sodium salt), incubated at 55°C overnight.

All reagents used from here on were from New England Biolabs (New England Biolabs), unless otherwise specified. Genomic DNA was amplified using 0.625U of Standard Taq Polymerase in a final reaction volume of 25µl, with 1X Standard Taq Reaction buffer, 0.2mM dNTPs (Life

Technologies, Invitrogen), 0.2 μ M primers, and 1 μ L DNA as prepared above. Brought to a final volume of 25 μ L with nuclease free water. PCR amplification was performed using the following cycle: 3 minutes at 95°C, followed by 39 cycles of denaturation at 95°C for 30 seconds, primer annealing at 55°C for 30 seconds, and elongation at 72°C for 30 seconds. PCR products were separated by electrophoresis on 1% agarose gels and visualised by SYBR Safe staining (ThermoFisher). The one exception was for the PCR reaction to detect NEX, where the primer annealing was 58 °C.

2.2.2 Primary neuron cell culture

Plates were prepared day before dissection, coated overnight at room temperature with poly-D-lysine (A-003-E; Merck) (0.1mg/ml in ddH₂O,). Before the dissection, poly-D-lysine coated plates were washed three times with sterile water, and plating media (MEM Eagle's with Earle's BSS, 10% filtered foetal calf serum (FCS, Biosera), 0.45% glucose, 1mM sodium pyruvate, 2mM glutamine, antibiotics (100U/ml Penicillin and 100 μ g/ml Streptomycin) (all from Life Technologies, Invitrogen)) was added, and they were subsequently placed in an incubator at 37°C and 5% CO₂. For immunofluorescence, poly-D-lysine/laminin precoated glass coverslips (Corning, Cat#354087) were placed in 24-well plates and 500 μ L of plating media was added and placed in incubator prior to dissection. For generation of RAD21-TEV neurons, embryonic day 13.5 mice were removed from pregnant mothers after cervical dislocation, for all other genotypes mice were dissected at embryonic day 17.5. The embryos were removed and placed in a petri dish containing ice cold dissection medium (HBSS Ca²⁺

and Mg^{2+} free, 1mM sodium pyruvate, 0.1% glucose, 10mM HEPES (all from Life Technologies, Invitrogen)) on ice. Samples were taken from each embryo for genotyping. All dissection was done in ice cold dissection medium. Using scissors, the head was removed from the embryo. Forceps were then used to remove the skin, and developing skull. The brain was then removed from the head and the two hemispheres separated, and the cortex isolated from each hemisphere and placed together in a 15ml falcon tube containing 1ml of dissection medium. The cortex was then washed twice with 1ml of dissection medium. Cortex was then resuspended in 450 μ l of dissection medium and 50 μ l of 2.5% trypsin (Life Technologies, Invitrogen) added and mixed. The cortex in trypsin was incubated at 37°C for 20 minutes. Following trypsin treatment, 20 μ l of DNase (DN-25; Sigma Aldrich. 1% wt/vol) was added and tissue incubated for a further 3 minutes at room temperature. The trypsin/DNase solution was then removed and the tissue washed twice with 1ml room temperature plating medium. The cortices were then resuspended in 2ml of plating medium and dissociated using a plastic Pasteur pipette by going up and down between 6 and 8 times, or until a homogenous cell suspension was achieved. Cells were then strained using a 70 μ m cell strainer and counted. Isolated cortical neurons were plated at 125,000/cm². After 2 hours, plating medium was removed and replaced with pre-warmed maintenance medium (Neurobasal medium, 0.5X B-27 Supplement, 2mM glutamine and antibiotics (100U/ml Penicillin and 100 μ g/ml Streptomycin) (all from Life Technologies, Invitrogen)). Following three days in vitro, cells were transduced overnight with lentivirus at a multiplicity of infection of 1. Maintenance media was replaced the following morning with 50% fresh/50% conditioned maintenance media, and so on every third day. Two days after transduction 5 μ M AraC was added. For cells undergoing imaging, initial maintenance media was also supplemented with 2.5% FBS, and AraC was omitted.

2.2.3 Primary macrophage cell culture

For the generation of RAD21-TEV primary macrophages, embryonic day 13.5 mice were removed from pregnant mothers culled via cervical dislocation. Samples were taken from each embryo for genotyping. The foetal liver was removed from embryos in Ca^{2+} and Mg^{2+} free phosphate buffered saline (PBS, Invitrogen) supplemented with 10% foetal calf serum (FCS, Biosera). The removed foetal liver was then transferred to 2ml of room temperature macrophage media (Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 1mM sodium pyruvate, 2mM glutamine, antibiotics (100U/ml Penicillin and 100 $\mu\text{g}/\text{ml}$ Streptomycin), 0.05mM β -mercaptoethanol (all from Life Technologies, Invitrogen), and 20% L929 conditioned medium), and resuspended until a homogenous single cell solution had been achieved. Cells were then seeded at a density of 450,000/cm² on non-attaching plates. Two days after plating macrophage media was removed and cells were resuspended in 2ml of retroviral containing supernatant and 4 $\mu\text{g}/\text{ml}$ Polybrene (Merck), and centrifuged at 1,250g for 90 minutes, at 37°C. Following 6 to 8 days, 500nM 4-hydroxytamoxifen (4-OHT), or carrier (ethanol) was added for 8 hours. Where indicated, cells were treated with LPS (2h, 10ng/ml). In order to remove macrophages from plates, they were treated with accutase (Merck) for 5 minutes at room temperature. The top 20 – 30% GFP-expressing cells were sorted for RNA and protein analysis.

2.2.4 L929 conditioned media production

L929 cells were cultured in Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 1mM sodium pyruvate, 2mM glutamine, antibiotics (100U/ml Penicillin and 100µg/ml Streptomycin), and 0.05mM β -mercaptoethanol (all from Life Technologies, Invitrogen). To prepare L929 conditioned media for macrophage differentiation, L929 cells were grown until 80% confluent, and their media replaced. Conditioned media was collected 10 days later and filtered through a sterile 0.22µm filter before being stored at -20°C.

2.2.5 HEK 293 cell culture

HEK293T and HEK293T-GP were cultured in Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 1mM sodium pyruvate, 2mM glutamine, antibiotics (100U/ml Penicillin and 100µg/ml Streptomycin), and 0.05mM β -mercaptoethanol (all from Life Technologies, Invitrogen). HEK293 and HEK293T cells were passaged every two days to maintain the population at less than 80% confluence.

2.2.6 Lentivirus production

In order to generate lentivirus to make neurons express the ERT2-TEV inducible system, the FUG vector containing the ERT2-TEV insert, was co-transfected with CMV Δ 8.9 and CMV-VSV-

G into HEK293T cells using polyethylenimine (PEI, 24765-2, Polysciences Inc.) (Auer. et al., 2011). Briefly, 6×10^6 HEK293T cells were plated on a 0.0005% poly-L-lysine (Sigma Aldrich) coated 10cm^2 plate in HEK293T media (Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 1mM sodium pyruvate, 2mM glutamine, antibiotics (100U/ml Penicillin and 100 $\mu\text{g}/\text{ml}$ Streptomycin), and 0.05mM β -mercaptoethanol (all from Life Technologies, Invitrogen)). The following day, 12 μg vector DNA, 6.5 μg CMV Δ 8.9 and 3.5 μg CMV-VSV-G (all 1 $\mu\text{g}/\mu\text{l}$) were mixed with 51 μg of PEI (1 $\mu\text{g}/\mu\text{l}$) in 3000 μl of Optimem (Invitrogen) per 10cm^2 plate. HEK293T media was replaced with 5ml of Optimem, and the 3.1ml of DNA/PEI mix added, per 10cm^2 plate. After 7 hours HEK293T the transfection media was replaced with of 37°C fresh Optimem media. 48 hours later Optimem, now containing viral particles, was pooled and centrifuged at 1500g for 3 minutes, before being filtered through 0.45 μm PVDF syringe filters (SLHV033RS, Merck). The resulting supernatant was ultracentrifuged at 17000RPM for 3 hours (Beckman Coulter L-80 XP Ultracentrifuge, Beckman Coulter SW-28 Rotor) in Ultra-Clear tubes (344058, Beckman Coulter). Following ultracentrifugation the viral pellet was resuspended in 100 μl ice cold PBS on a shaker for 1 hour, before being aliquoted.

2.2.7 Retrovirus production

In order to generate retrovirus to make macrophages express the inducible ERT2-TEV system, the pMIG retrovirus vector containing the ERT2-TEV insert, was co-transfected with pCL-Eco packaging vector, into 293T-GP using a calcium phosphate protocol (Cobb. et al., 2000). Briefly, DNA-containing precipitates were formed by slowly adding 500 μl HEBS buffer (280mM NaCl, 10mM KCl, 1.5mM $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$, 12mM glucose, 50mM HEPES free acid, pH 7.05 in

distilled water) to 500µl of a 0.4M CaCl₂ solution containing 4µg of MSCV plasmid DNA and 4µg of vector DNA. 1ml of DNA precipitate was then slowly added to a 60-80% confluent 10cm² plate of HEK293T-GP cells cultured in 9ml of HEK293T-GP media. Cells were fed with fresh media 24 hours post-transfection, and supernatant containing retrovirus particles were harvested at 48, 60, 72 and 84 hours post-transfection and pooled.

2.2.8 Retrovirus transduction of foetal liver cells

After 2 days in culture, foetal liver cells were transduced by resuspending 4x10⁶ cells in 2ml of viral supernatant supplemented with 4µg/mL polybrene (Sigma) in a single well of a 6-well plate. Plates were then centrifuged at 900g for 90 minutes at 37°C, followed by incubation in viral supernatant for a further 3 hours at 37°C before viral supernatant was replaced with fresh macrophage media (HEK293T media supplemented with 20% L929 conditioned media).

2.2.9 Lentiviral transduction of primary neurons

After 3 days in culture, neurons were infected overnight with concentrated lentivirus at a MOI of 1.5. The following morning media was changed to 50% fresh neuronal maintenance media, and 50% conditioned maintenance media from cells of the same biological replicate.

2.2.10 Isolation of nuclei from post-mortem human tissue

In order to isolate nuclei from human brain for RNA-seq, human tissue had to be homogenised and nuclei purified. 50-250mg of human cortical tissue was homogenised in a dounce homogeniser containing 1ml of homogenisation buffer (250mM sucrose, 25mM KCl, 5mM MgCl₂, 10mM Tris buffer pH 8.0, 1μM DTT, 1X Proteinase Inhibitor w/o EDTA (Roche), 0.4U μl⁻¹ RNaseIn (AM2684, ThermoFisher) 0.2U μl⁻¹ Supersin (AM2696, ThermoFisher), 1μM DAPI). Tissue was homogenised using 5 strokes of the loose pestle, followed by 10 – 15 strokes of the tight pestle. An additional 2ml of homogenisation buffer was added to homogenised brain, and all 3ml of passed through a 0.45μm filter. Homogenate was transferred to Eppendorf tubes and centrifuged at 1000G for 8 minutes at 4°C.

Meanwhile 29% and 50% iodixanol solutions were prepared from iodixanol medium (250mM sucrose, 150mM KCl, 30mM MgCl₂, 60mM Tris buffer pH 8.0) and iodixanol 60% (D1556, Sigma). Supernatant was removed from pelleted tissue, and tissue was resuspend in homogenisation buffer to 250μl. A further 250μl of 50% iodixanol was added to the resuspended tissue and gently mixed. 500μl of 29% iodixanol was added to a fresh Eppendorf, and the iodixanol/tissue mix gently layered on top. Samples were then centrifuged at 13000G for 40 minutes at 4°C.

Supernatant was removed from pelleted nuclei using aspirator to remove all myelin. Nuclei were then resuspended in 1ml wash buffer (PBS, 1% BSA, 0.2U/μl RNaseIn (AM2684, ThermoFisher)) and passed through a 0.22μm filtered FACS tube. Washed nuclei were then

transferred to a fresh Eppendorf and pelleted by centrifugation at 500G, for 5 minutes at 4°C. This wash step and centrifugation was repeated once more.

Following the final wash, pelleted nuclei were resuspended in antibody staining solution (PBS, 1% BSA, 0.2U/μl RNaseIn (AM2684, ThermoFisher), 1:200 NeuN antibody). For negative control, antibody was excluded. Nuclei were then stained for one hour in cold room on rotation before being sorted on BD Fusion.

2.2.11 Fluorescent activated cell sorting – FACS

Cells were centrifuged at 1200rpm for 5 minutes, and 5×10^5 resuspended in 50μl of blocking buffer with primary antibody at the appropriate dilution. These were then incubated at room temperature for 20 minutes. Following staining, cells were centrifuged for 2 minutes at 500g and resuspended in 500μl FACS buffer (PBS Mg^{2+} and Ca^{2+} free, 2% FCS, 2mM EDTA). Cells were then analysed using the BD LSR II. Antibodies used for FACS are listed in section 2.1.1.

To sort RAD21-TEV macrophages that were GFP positive and therefore contain the Ert2-TEV-IRES-GFP construct, cells were centrifuged at 500g for 5 minutes, and resuspended in FACS buffer and submitted for FACS purification using the FACSaria or BD Fusion cell sorters (BD Biosciences).

2.2.12 Western blot

Cells were resuspended in 1ml of ice cold PBS with cOmplete proteinase inhibitor (Roche, Cat#18970600), and centrifuged at 3000g for 5 minutes at 4°C. Whole cell extracts were prepared by lysis of cells (5 minutes at 95°C and subsequent vortexing) in protein sample buffer (50mM Tris-HCl pH6.8, 1% SDS, 10% glycerol) followed by quantification using Qubit. Following quantification 0.001% Bromophenol blue and 5% beta-mercaptoethanol were added.

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out with the Bio-Rad minigel system. 20µg of protein sample and the benchmark pre-stained protein ladder (Invitrogen) were loaded and separated on a 7.5% acrylamide pre-cast resolving gel (Biorad) using Tris-glycine electrophoresis buffer (1.5% Tris, 7.2% glycine, 0.5% SDS). Resolved gels were blotted to a polyninylidene fluoride transfer membrane (Millipore) in transfer buffer (48mM Trizma base, 39mM glycine, 0.037% SDS and 20% methanol) using the trans-blot semi-dry electrophoretic transfer apparatus (BioRad). Membranes were incubated for 1 hour with fluorescent blocker (Millipore) followed by incubation with primary antibody diluted in blocker at an appropriate dilution overnight at 4°C, with agitation. After washing three times in wash buffer (1.2g/L Tris pH7.4, 8.75g/L NaCl) for 10 minutes, blots were incubated with secondary fluorescent antibodies at 1:10000 (anti-rabbit and anti-mouse antibodies purchased from Life Technologies) in blocker for 1 hour at room temperature, followed by three more 10-minute washes in wash buffer. Detection was measured using the LI-COR Odyssey imaging system and analysed on Image Studio Lite.

2.2.13 RNA extraction and sequencing

RNA extraction was performed using the PicoPure RNA Isolation Kit (ThermoFisher) and residual DNA eliminated using RNase-Free DNase Set (Qiagen). One brief amendment to the protocol was made for neuronal RNA extraction; cells were washed once with RNase-Free PBS before Extraction Buffer was added directly into the wells.

RNA sequencing was performed from three biological replicates per condition. For RAD21-TEV macrophages, 100ng of RNA were used to prepare libraries using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina using ribozero ribosomal RNA (rRNA) depletion. Libraries for RNA-sequencing of RAD21-TEV neurons were prepared using 200ng of RNA and the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina with polyA enrichment. Library quality and quantity were assessed on a bioanalyzer and Qubit, respectively. Libraries were sequenced on an Illumina HiSeq2500 (v4 chemistry), generating 40 million paired-end 100bp reads per sample, and analysed as stated below.

Paired-end 100bp sequencing reads were aligned against Ensembl mouse genome (NCBIM37) or human genome (hg38) with Tophat2 (2.0.11;(Kim et al., 2013). Reference sequence assembly and transcript annotation that obtained from Illumina iGenomes (https://support.illumina.com/sequencing/sequencing_software/igenome.html). Gene-based read counts were obtained using featureCounts function from Rsubread Bioconductor package (1.24.2; (Liao et al., 2013, Liao et al., 2014). Normalisation was performed in DESeq2

Bioconductor package (1.24.0; (Love et al., 2014a) and data was rlog transformed to allow for visualisation by PCA and heatmaps. Differentially expressed gene (DEG) analysis was also performed with DESeq2 and DEGs were defined with Benjamini-Hochberg adjusted P-value<0.05. Gene ontology analysis was performed with goseq Bioconductor package (1.24;(Young et al., 2010)). After converting mouse gene symbol to human gene symbol using the report of Human and Mouse Homology retrieved from Mouse Genome Informatics (MGI, <http://www.informatics.jax.org>), Gene Set Enrichment Analysis (GSEA 2.2.0;(Mootha et al., 2003, Subramanian et al., 2005)) was then performed with GseaPreranked tool using Hallmarks gene set (h.all.v5.2.symbols.gmt).

2.2.14 DNA delivery into primary neurons

Neurons were transfected between 7 and 12 days in vitro with pcDNA3-mRFP, a gift from Doug Golenbock (Addgene #13032), or FUG-TEVP, using Lipofectamine 2000 (Invitrogen). For each well of 1.25×10^5 cells, $1 \mu\text{g}$ of chosen plasmid (at $1 \mu\text{g}/\mu\text{l}$), and $2 \mu\text{g}$ of Lipofectamine 2000 (Invitrogen) were mixed in $100 \mu\text{l}$ of Neurobasal media supplemented with 2mM glutamine and 0.5X B-27 (both Life Technologies, Invitrogen), and left to incubate at room temperature for 30 minutes. Following incubation, cell media was replaced with fresh Neurobasal media supplemented with 2mM glutamine and 0.5X B-27 (both Life Technologies, Invitrogen), and the $100 \mu\text{l}$ of transfection mix added to the wells. Cells were left in the incubator at 37°C for 2 hours before transfection reaction media was replaced with conditioned neuronal maintenance media.

2.2.15 Immunofluorescence (IF)

Antibodies for IF can be found in section 2.1.1. Samples were fixed with 500µl of warmed 4% sucrose and 4% paraformaldehyde (PFA) in PBS for 10 minutes at room temperature. Fixed samples were washed twice with PBS and then permeabilised with 500µl of permeabilization solution (0.3% Triton X-100 in PBS) for 10 minutes at room temperature. Samples were then washed with PBS and blocked in 500µl of blocking solution (0.1% Triton X-100, 10% Normal Goat Serum (Vector) in PBS) for 1 hour at room temperature. Primary antibody was diluted in staining solution (0.1% Triton X-100, 2% Normal Goat Serum in PBS) at specific dilutions and added to samples in a humidified chamber and incubated at 4°C overnight. Samples were washed three times for ten minutes with PBS. Following washes, samples were incubated with secondary antibodies coupled to appropriate fluorophores (Molecular Probes) diluted 1:500 in staining solution for 1 hour at room temperature in a dark humidified chamber. Samples were then washed three times for ten minutes with PBS and mounted in Vectashield with DAPI (Vector Laboratories,).

Samples were visualised using a TCS SP5 Leica laser scanning confocal microscope. Images were processed using Leica Confocal Software and FIJI.

2.2.16 Cell cycle analysis – Propidium Iodide (PI) staining

Cells were collected, counted and 5×10^5 centrifuged at 1200rpm for 5 minutes. Cells were then incubated in 1ml of PI-Saponin solution (2.5% RNase A (Invitrogen), 2.5% propidium iodide solution (Sigma), 0.03% Saponin (Sigma) in PBS), at room temperature in the dark for 30 minutes. Cells were then analysed using the BD LSR II (BD Biosciences). Data were analysed using the FloJo software.

2.2.17 Chromosome Conformation Capture – 4C

0.5×10^6 neurons or macrophages were used for each 4C procedure. Macrophages were sorted following treatment with 4-OHT to enrich for cohesin depleted cells, while neurons were scraped, both cell types were then pelleted by centrifugation at 300g for 5 minutes at room temperature. Cells were resuspended in 10ml of PBS and 670 μ l of 16% formaldehyde solution (Thermofisher, Cat# 28908), to bring to a final concentration of 1%, was added. Samples were mixed by inverting and incubated at room temperature for 10 minutes. Following formaldehyde treatment, 712.5 μ l of 2M glycine solution was added, to a final concentration of 0.125M, to quench the reaction. Samples were mixed by inversion and incubated at room temperature for 5 minutes. Samples were then centrifuged for 10 minutes, at 1000G and at 4°C, and supernatant was discarded. Cell pellets were now flash frozen in liquid nitrogen and stored at -80°C until all had been collected for downstream procedures.

Crosslinked cells were thawed on ice and resuspended by pipetting up and down 10 times in 950µl of ice-cold NP-40 lysis buffer (10mM TRIS-HCL, 150mM NaCl, 0.5% NP-40, 1X Roche PI). Samples were then incubated for 15 minutes on slow rotation in cold room, and then centrifuged at 2500G for 5 minutes at 4°C, and supernatant was discarded. Resuspension in NP-40 and centrifugation steps were repeated. Pelleted nuclei were then washed once with 500µl of 1X NEB Buffer 2 (NEB, Cat# 7002) and centrifuged at 2500G for 5 minutes at 4°C, and supernatant was removed. Pellet was gently resuspended in 50µl of sodium dodecyl sulfate (SDS) solution (1X NEB Buffer 2, 0.5% SDS) and incubated at 62°C for 8 minutes. Following incubation, 170µl of Triton X-100 solution (1X NEB Buffer 2, 1.47% Triton X-100) was added to quench SDS. Samples were mixed well and incubated at 37°C, at 900RPM for 15 minutes. Samples then undergo their first digestion. 30µl of 1X NEB Buffer 2 and 100U of restriction enzyme (HindIII (NEB, R0104) for neuronal PCDHB 4C, and MboI (NEB, R0147) for macrophage 4C) is added to samples and chromatin is digested overnight at 37°C at 900RPM. A further 100U of restriction enzyme is added the following morning and digestion continued for a further 2 hours. Samples were then incubated at 65°C for 20 minutes to inactivate restriction enzymes. Add 950µl of ligation reaction mix (120µl 10X NEB T4 DNA ligase buffer (NEB, B0202), 100µl 10% Triton X-100, 6µl of 20mg/ml Bovine Serum Albumin (BSA), 5µl of 400U/µl T4 DNA ligase (NEB, M0202) and 719µl of H₂O) to samples. Mix by inverting and incubate at room temperature for 6 hours on slow rotation. Degrade protein by adding 25µl of 30mg/ml proteinase K (NEB, P8107) and 120µl of 10% SDS, and incubating at 55°C for 30 minutes. Add 130µl 5M NaCl and incubate overnight at 65°C.

For DNA precipitation of digested DNA 3µl of 20mg/ml RNase A was added to samples and incubated at 37°C for 45 minutes. An equal volume of phenol-chloroform was added and

samples vortex for 1 minute. Samples were then centrifuged for 5 minutes at 3500G at room temperature and the supernatant was retained. 1/10 volume of 3M sodium acetate (pH 5.2) and 2.5X volume of ice cold 100% ethanol was then added to samples. Samples were then mixed vigorously by hand for 1 minute and placed at -80°C for 45 minutes. Samples were then centrifuged for 20 minutes at 14000G at 4°C and the supernatant was removed. 1ml of room temperature 75% ethanol was added to the pellet, and centrifuged again for 5 minutes at 14000G at 4°C. Supernatant was removed and the pellet air dried. DNA was then dissolved in 130µl of 10mM Tris Buffer (pH 7.5).

Following precipitation, DNA underwent a second round of digestion. 10µg of DNA in 36µl of H₂O was mixed with 46µl of digestion buffer (10µl 10X NEB CutSmart Buffer (NEB, B7204)) and second restriction enzyme added (30U of NlaIII (NEB, R0125) for neuron PCDHB and macrophage 4C). Samples were mixed gently and incubated overnight at 37°C.

Following second digestions samples undergo a second ligation followed by DNA purification. Restriction digest was inactivated by incubating samples at 65°C for 25 minutes. Samples were then incubated in 2.7ml of ligation mix (1X NEB T4 DNA ligase buffer, 2µl of 400U/µl T4 DNA ligase) at 16°C overnight. Following ligation reaction, 1/10 volume of 3M sodium acetate (pH 5.2) and 2.5X volume of ice cold 100% ethanol was added to the ligation reaction. Samples were then mixed vigorously and incubated at -80°C for 45 minutes. Samples were then centrifuged for 30 minutes at 14000G at 4°C. Supernatant was removed and 2ml of 75% ethanol was added. Samples were then centrifuged for 5 minutes at 14000G at 4°C, and the pellets dried at room temperature. DNA was then resuspended in 50µl of 10mM Tris buffer

(pH 7.5), and 4C templates purified using a Roche High-Pure PCR Product Purification kit (Roche, 11796828001).

DNA was sequenced using an Illumina HiSeq 2500 platform. Sequence data was analysed using the 4Cseqpipe software suite (van de Werken et al., 2012) as detailed below.

Data analysis for 4C-seq:

a) Sequence extraction for each 4C viewpoint was done using an in-house script (python v2.7.5) shown below:

```
#!/usr/bin/env python
import re
def extraction(INPUT_FILE,OUTPUT_FILE):
    fjob=open('%s' % OUTPUT_FILE,'w')
    for eachLine in open('%s' % INPUT_FILE):
        m1=re.search('^@HWI-',eachLine)
        m2=re.search('^\\+',eachLine)
        if m1 is not None or m2 is not None:
            fjob.write(eachLine);each=f.next();fjob.write('%s\\n' % each[16:46])
    fjob.close()
extraction(INPUT_FASTQ,OUTPUT_FASTQ)
```

b) Then, extracted sequences were analysed using the 4Cseqpipe software suite (Version 0.7), which is a user-friendly and complete software package for analyzing 4C-seq experiments containing sequence extraction, mapping, normalization, and generation of contact profiles around the viewpoint (van de Werken et al., 2012)

. The detailed setting of the 4Cseqpipe was:

1) Converting fastq files to "raw" format:

```
perl 4cseqpipe.pl -fastq2raw \
-ids ID1 \
-rawdir directory \
-fastq_fn input_file_name \
-convert_qual 1
```

2) Mapping valid 4C-Seq products to restriction fragments in the genome:

```
perl 4cseqpipe.pl -map \
-ids ID1 \
-trackset_name trackset_name \
-binsize bin_size \
-trackdb_root database_file_name \
-rawdir root_file_name
```

3) Normalizing and generating near-cis domainograms:

```
perl 4cseqpipe.pl -nearcis \  
-ids ID1  
-trackset_name trackset_name \  
-expname experiment_name \  
-stat_type mean \  
-trend_resolution 40000 \  
-plot_from 5'_start_coordinate \  
-plot_to 3'_start_coordinate \  
-figure_fn figure_file_name
```

"-stat_type" indicates the type of statistic to calculate and plot. In this work, 4Cseqpipe was used to calculate the mean of values in sliding windows of size 2-50kb. "-trend_resolution" shows the setting of resolution used (size of window in basepairs) at which the main trend will be plotted, with resolutions from 2-50kb plotted below the main trend using a color-code. Mus musculus genome assembly MGSCv37 (mm9) was used for the mapping.

Primers used for 4C can be found in tables 3, 4, 5 and 6.

2.2.18 Cloning strategy

For generation of the lentiviral plasmid FUG-TEVP, cloning was performed by PCR amplification from appropriate cDNA using Phusion Taq (Invitrogen) and primers listed in Table 2, restriction digestion, gel purification of fragments, and ligation using T4 DNA ligase (New England Biolabs) into the target vector. The v5-ERT2-TEVP cDNA was amplified from the ERT2-TEVP-ERT2-LL construct previously generated in the lab, with the addition of 5' and 3' EcoR1 site overhangs. Cas9 was removed from FUGt2aCas9, a gift from Dr. Bryan Luikart (Addgene plasmid #75346), using EcoR1 restriction enzymes, and the v5-ERT2-TEVP cDNA generated previously ligated in.

TETOnTEVP lentiviral plasmid was generated by Gibson Assembly. cDNA was generated using primers listed in Table 2, and Q5 Polymerase (New England Biolabs), before being assembled in Gibson Assembly Mastermix (New England Biolabs). The UbcTetON backbone components were amplified from Lenti-iCas9-Neo, a gift from Dr. Qin Yan (Addgene plasmid #85400). NLS-TEVP cDNA was generated from NLS-TEV-pRNA construct, a kind gift from Dr. Kikue Tachibana (IMBA, Austria).

All vectors were sequenced by GeneWiz (Beckman Coulter) to ensure correct insert orientation and that no mutations had occurred during the PCR reactions.

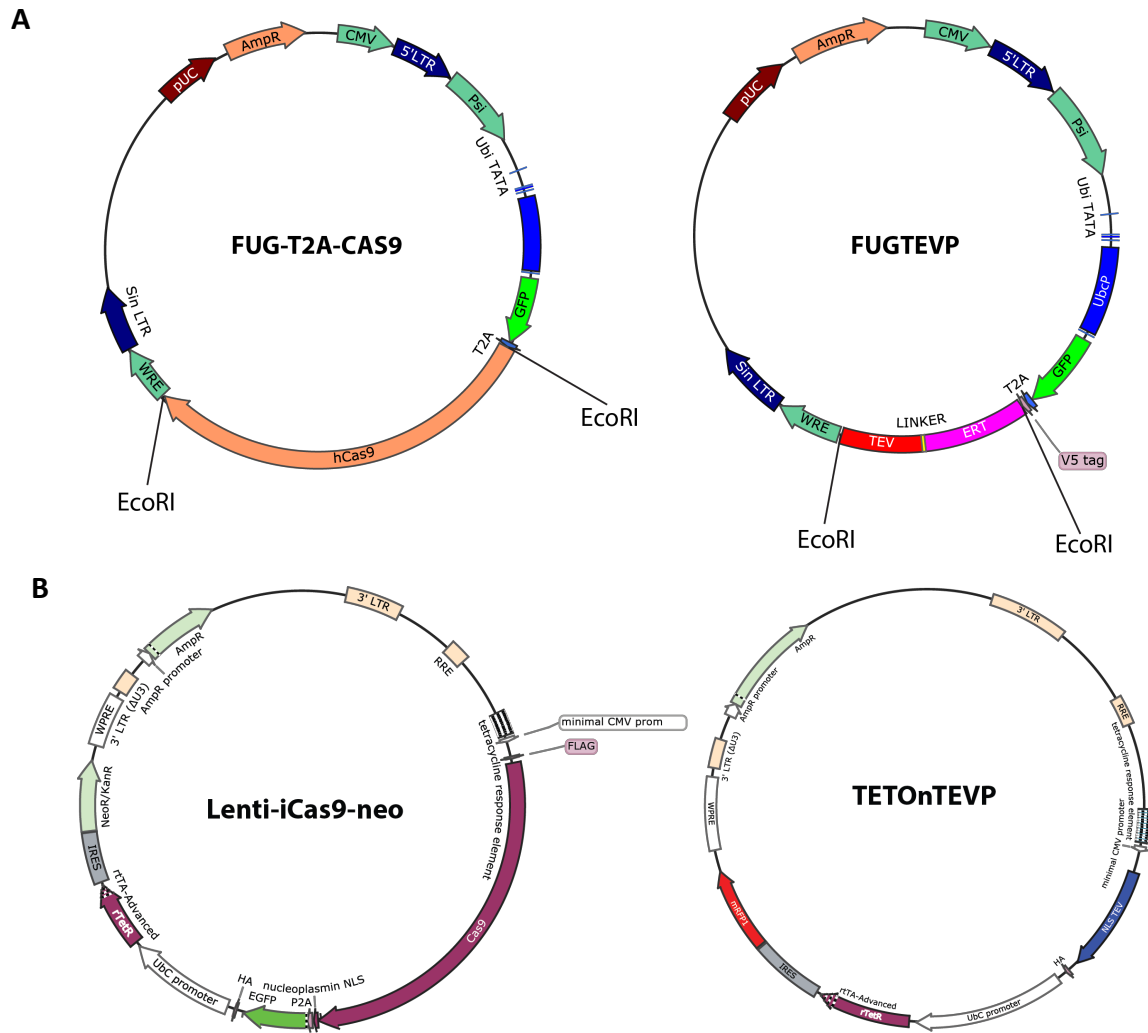


Figure 2.1 Vectors used and generated for inducible control of TEV protease (A) Amplified ERT2-TEV protease was inserted in place of hCas9, between two EcoRI sites, downstream of the T2A. **(B)** NLS-TEV protease was inserted downstream of the tetracycline response element, in place of Cas9, while mRFP1 was inserted in place of NeoR/KanR element downstream of the IRES. The EGFP and P2A sites were removed from Lenti-iCas9-Neo.

2.2.19 Plasmid construction and bacteria transformation

All enzymes and buffers used in this section were purchased from New England Biolabs (New England Biolabs Ltd, Herts, UK). For generation of the FUG-TEVP lentiviral vector, 2µg of v5-ERT2-TEV, and 2µg of the FUGt2acas9 vector were digested with EcoR1-HF enzyme for 1 hour at 37°C. Digested vector was treated with calf intestinal alkaline phosphatase (CIP) for 20 minutes at 37°C in the same digestion mix. Digested products were run on a 1% agarose gel and the correct size bands were gel purified using QIAquick Gel Extraction Kit (Qiagen) following manufacturer's instructions. Then 50ng of the vector and a three-molar excess of the insert were ligated with 1µl (400U) of T4 DNA Ligase in T4 DNA Ligase buffer for 20 minutes at room temperature.

In order to introduce the ligated FUG-TEVP vector into bacteria and use it to produce larger amounts of the construct, 10µl of the ligation product was transformed into 50µl of One Shot Stbl3 *E. coli* (ThermoFisher). The 10µl of the ligation product was mixed with 50µl of bacteria and placed on ice for 30 minutes. The transformation mixture was then incubated at 42°C for 45 seconds, before going back on ice for a further 5 minutes. Following these incubation steps, 250µl of room temperature S.O.C Medium (ThermoFisher) was added and the bacteria placed at 37°C on a horizontal shaker at 300rpm, for 1 hour.

To produce the TETOnTEVP lentiviral vector, the four fragments (lenti backbone, RFP, NLS-TEVP and Ubc-IRES) were generated by PCR reaction using Q5 polymerase according to manufacturer's instructions (New England Biolabs). Fragments were then run on a 1% agarose gel and gel excised using Qiaquick Gel Extraction Kit (Qiagen). Fragments were then incubated

in Gibson Assembly Mastermix (New England Biolabs) for 1 hour at 50°C with a three-molar excess of RFP, NLS-TEVP and Ubs-IRES fragments to 400ng of lenti backbone fragment.

In order to introduce the TETOnTEVP vector into bacteria and use it to produce larger amounts of the construct, the ligation product was added to 20µl of 5X KCM buffer (0.5M KCl, 0.15M CaCl₂, 0.25M MgCl₂) and the total volume was adjusted to 100µl with H₂O. 100µl of competent DHS5α bacteria (Invitrogen) was added to the mixture and incubated on ice for 20 minutes. 250µl of pre-warmed to 37°C Lysogeny Broth (LB) was added to the transformation mixture and bacteria were incubated at 37°C with agitation for 1 hour.

For both FUG-TEVP and TETOnTEVP vectors, transformed bacteria were plated into LB agar supplemented with 50µg/µl of ampicillin and incubated overnight at 37°C. Individual colonies were picked and transferred to 3ml of LB broth with 50µg/µl ampicillin and left overnight at 37°C with agitation. DNA was extracted from exponentially growing cultures with a miniprep (Qiagen) or maxiprep kit (Zymo) according to manufacturer's instructions.

Chapter 3: The role of cohesin in macrophage inflammatory gene expression

3.1 Introduction

Organisation of the genome in 3D space is thought to facilitate gene regulation by providing a framework through which regulatory elements such as enhancers can come into contact with and exert their regulatory effect on target genes (Dowen et al., 2014, Hnisz et al., 2016, Merckenschlager and Nora, 2016, Rao et al., 2014). The removal of architectural proteins, as well as CTCF binding sites and domain boundaries, results in the loss or alteration of 3D genome organisation (Rao et al., 2017, Schwarzer et al., 2017, Nora et al., 2017, Lupianez et al., 2015, Seitan et al., 2011). However, while some specific genes have been shown to require correct 3D genome organisation for expression, (Seitan et al., 2011, Narendra et al., 2015, Jiang et al., 2017), and there are other transcriptional changes caused by DNA damage due to cohesin's requirement for maintaining chromosome integrity during the cell cycle (Lavagnoli et al., 2015, Kagey et al., 2010), the loss of cohesin or CTCF has shown a limited impact on gene regulation. Even the total loss of cohesin or CTCF does not result in clear gene regulatory phenotypes (Rao et al., 2017, Nora et al., 2017).

This raises the possibility that the importance of 3D genome organisation has been overstated, or that current models of studying transcriptional regulation do not highlight or test the requirement for correct 3D genome organisation.

An inviting system that has yet to be studied is the role of 3D genome organisation on inducible gene expression. Models so far have looked at steady state changes in gene regulation in response to loss of 3D genome organisation (Nora et al., 2017, Rao et al., 2017), yet inducible responses, such as the myeloid inflammatory gene expression program have yet to be investigated. Here hundreds of genes, and thousands of regulatory elements, are dynamically controlled in response to external stimulus (Bhatt et al., 2012, Ostuni et al., 2013, Glass and Natoli, 2016). This requires an exceptionally high level of regulatory control.

In macrophages, activation of Toll-like receptor 4 (TLR4) by the bacterial cell wall component lipopolysaccharide (LPS) triggers an inducible gene expression program (Bhatt et al., 2012). Transcription factors and cytokines are activated early, and trigger the auto- and paracrine induction of secondary response genes (Bhatt et al., 2012). Previous work in the lab (Cuartero et al., 2018) showed that genetic depletion of cohesin led to the deregulation of the LPS responsive inflammatory gene expression program in macrophages (Figure 3.1A).

While it is clear cohesin is required for inducible gene expression in macrophages, it remains unknown whether cohesin is directly regulating inducible genes or their deregulation is a secondary effect of the loss of cohesin. In this chapter I use the RAD21-TEV proteolytic cleavage system in order to identify direct cohesin targets, and understand whether cohesin is directly regulating inducible gene expression.

3.2 Inducible genes are sensitive to cohesin dosage

In order to study the role of cohesin in gene regulation of myeloid cells, macrophages were generated from the bone marrow of *Rosa26^{ERT2-cre/ERT2-cre} Rad21^{lox/lox}* mice, where exons 5 and 6 of the gene coding for the essential cohesin subunit RAD21 are flanked by *loxP* sites. Addition of 4-hydroxy-tamoxifen (4-OHT) led to the removal of these exons and a reduction in RAD21 protein level over the course of 2 to 3 days (Figure 3.1A). As macrophages are quiescent, the loss of cohesin and any subsequent changes in gene regulation, are not caused by cohesin's role in the cell cycle.

In order to understand the role of cohesin in gene regulation, *Rosa26^{ERT2-cre/ERT2-cre} Rad21^{lox/lox}* macrophages were depleted of cohesin and analysed by RNA-seq. The overall effect on gene regulation was limited, as expected (Nora et al., 2017, Rao et al., 2017, Schwarzer et al., 2017, Seitan et al., 2011) with only 22% of constitutive genes effected by the loss of cohesin (Figure 3.1B). However, inducible genes were much more effected with over 50% deregulated even at baseline, preceding LPS treatment (Figure 3.1B). In LPS treated *Rad21*-deleted macrophages, the proportion of deregulated inducible genes increased to 58% following 2h LPS treatment, and 75% after 8h LPS treatment (Figure 3.1B, C).

Inducible genes in macrophages are largely proinflammatory (Bhatt et al., 2012, Chen et al., 2012b, Glass and Natoli, 2016), genes related to inflammatory expression were largely downregulated in *Rad21*-deleted macrophages (FDR = 0.0, Figure 3.1D).

The data show that cohesin is required for inducible gene expression in macrophages. However, it is unknown whether this requirement stems from inducible genes directly relying on cohesin for expression. In *Rad21*-deleted macrophages, cohesin takes 24h to be removed, and is absent for 2 to 3 days before gene expression is analysed (Cuartero et al., 2018). Therefore, inducible genes may be deregulated in the absence of cohesin due to secondary effects as opposed to being directly reliant on cohesin for expression. In order to understand what genes are directly reliant on cohesin for expression, and if this includes inducible genes, one needs to remove cohesin rapidly and in a temporally controlled manner.

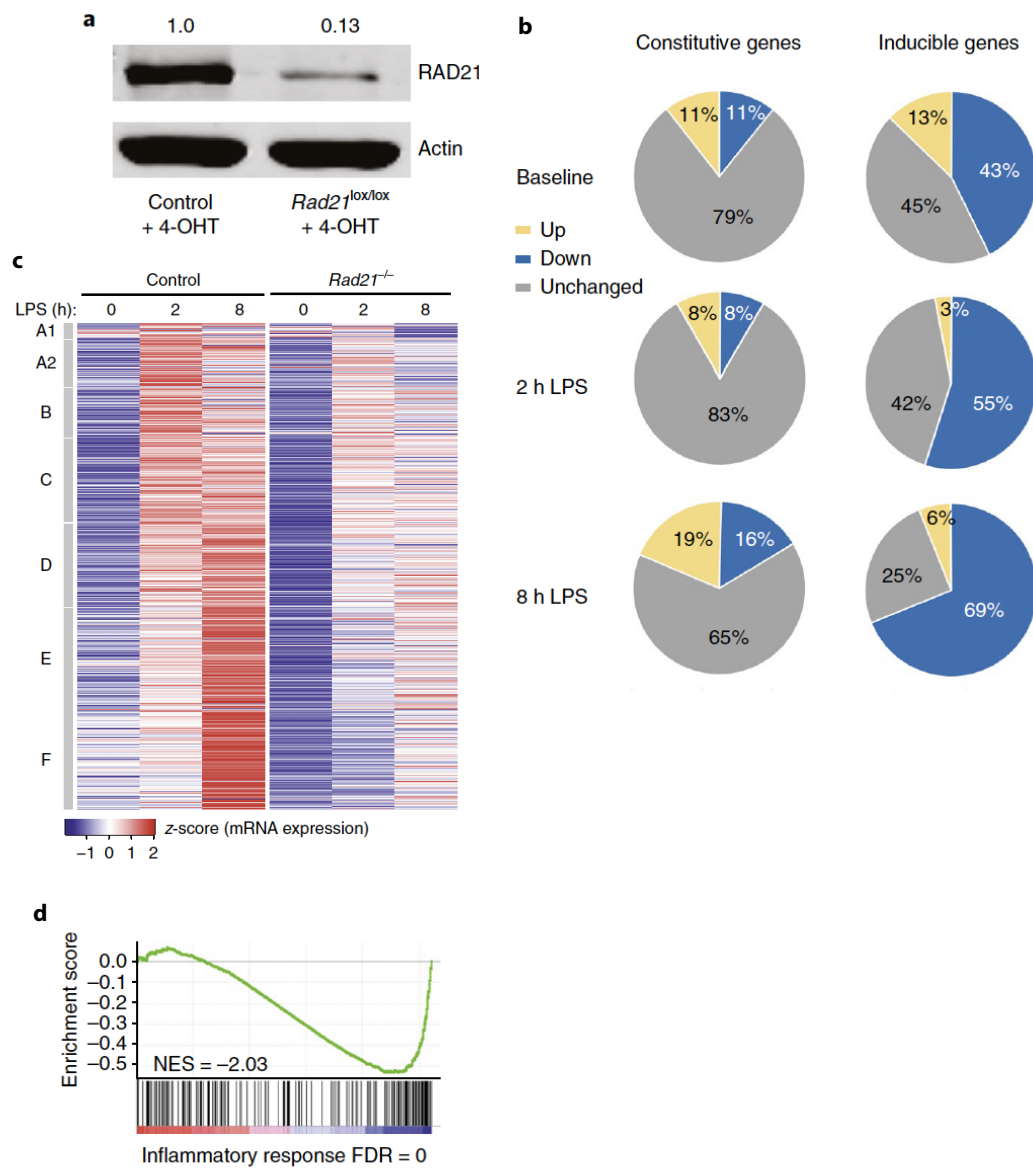


Figure 3.1 Cohesin is required for inflammatory gene expression (A) Immunoblot analysis of RAD21 protein expression following *Rad21* deletion in mature macrophages (B) Fraction of significantly deregulated ($p < 0.05$, DESeq2) LPS inducible genes (Bhatt et al., 2012) ($n = 560$) and constitutively expressed genes ($n = 10,780$) at baseline, and after 2h and 8h LPS induction (C) Heat map of inducible gene expression in control and *Rad21^{-/-}* macrophages at baseline, 0h and 8h LPS induction. Inducible gene classes indicated on the left (Bhatt et al., 2012) (D) GSEA of inflammatory response genes in *Rad21^{-/-}* macrophages (normalized enrichment score (NES) = -2.03, FDR = 0)

3.3 Generation of RAD21-TEV macrophages from foetal liver

In order to identify immediate targets of cohesin in macrophages, we used the RAD21-TEV mouse model developed and kindly provided by Kim Nasmyth (Tachibana-Konwalski et al., 2010). The RAD21-TEV mouse model was developed in order to allow for rapid and inducible cohesin depletion through proteolytic cleavage of the key cohesin subunit, RAD21. Tobacco etch virus (TEV) cleavage sites are short sequences that when integrated into proteins can be targeted by the TEV protease and lead to cleavage of the targeted peptide. Having been previously shown in yeast and fly that insertion of TEV cleavage sites into *Sccl/Rad21* leads to loss of cohesin function (Uhlmann et al., 2000, Oliveira et al., 2010) Tachibana et al. adapted the system to mouse. Three TEV cleavage sites were inserted into exon 8 of *Rad21*, as well as addition of a c-Myc tag at the C-terminal to allow for easy targeting in molecular biology techniques.

Mouse embryos where RAD21-TEV was the sole source of RAD21 were generated through animal breeding. This allowed for the isolation and culture of a wide range of cells in which to study the role of cohesin in gene regulation. Following on from previous work in the lab, we were interested in identifying genes directly regulated by cohesin in macrophages. In order to generate RAD21-TEV macrophages, fetal liver was removed from embryonic day 14.5 RAD21-TEV embryos and cultured in the presence of macrophage colony stimulating factor (MCSF) conditioned media. Macrophages can be identified through the combined expression of *Adrge1* (F4/80) and *Itgam* (CD11B), as well as no longer being mitotic.

In order to test that fetal liver cells explanted from RAD21-TEV embryos were capable of producing macrophages, cells were stained after 7 days in culture in the presence of MCSF with CD11b antibody conjugated to FITC, and F4/80 antibody conjugated to PE. They were then analysed using fluorescent automatic cell sorting (FACS), where 96% were found to be CD11b+ and F4/80+, compared to fewer than 1% in the isotype control (Figure 3.3A). In order to understand whether the explanted RAD21-TEV fetal liver cells were still proceeding through the cell cycle, or had become post-mitotic following 7 days in MCSF conditioned media, they were stained with propidium iodide (PI). Propidium iodide intercalates into DNA and fluoresces, allowing for the visualisation of cellular DNA content using FACS. Cells were found to be largely non-cycling as only 5% of cells were in S phase, and 7.2% in G2 (Figure 3.2B).

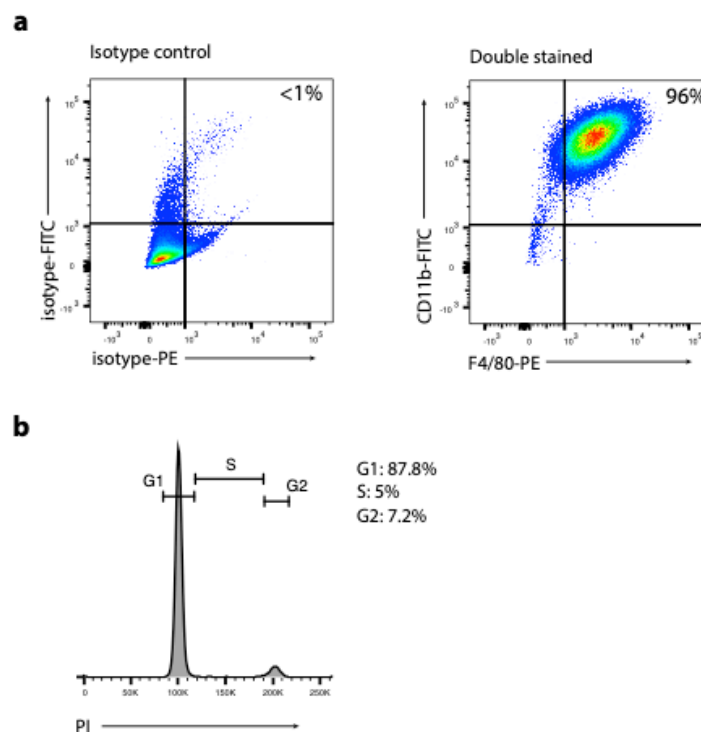


Figure 3.2 RAD21-TEV foetal liver produces macrophages (A) Flow cytometric analysis of macrophage markers CD11b and F4/80 on RAD21-TEV macrophages **(B)** DNA content of RAD21-TEV macrophages, and % of cells in G1, S and G2

These results show that foetal liver cells explanted from RAD21-TEV embryos and grown in the presence of MCSF conditioned media are capable of producing macrophages. They can therefore be used to study the role of cohesin in macrophage gene regulation.

3.4 Characterisation of RAD21-TEV proteolytic cleavage in macrophages

In order for RAD21-TEV to be cleaved and cohesin function impaired, the TEV protease needs to be introduced into the system. In order to temporally control the cleavage of cohesin a fusion protein had been developed in the lab previously. TEV protease is fused with a modified estrogen receptor (ERt2) (Figure 3.3A), where ERt2 is no longer responsive to its endogenous ligand, estradiol, and instead binds 4-hydroxytamoxifen (4-OHT) (Feil et al., 1997). In the absence of 4-OHT this fusion protein resides in the cytoplasm, however upon 4-OHT application the ERt2 fusion translocates to the nucleus. Due to cohesin only being present in the nucleus, this allows for the temporal control of RAD21-TEV cleavage. The ERt2-TEV protease fusion also contains a v5 epitope tag allowing for the tracking of the TEV protease. The fusion construct is in a retroviral vector, pMSCV-IRES-GFP (pMIG), where the ERt2-TEV protease expression is driven by the MSCV promoter, and infected cells also express GFP.

In order to only analyse changes to cells where TEV protease was expressed, cells were isolated using FACS, based on GFP expression, before downstream analysis (Figure 3.3B). First, we identified GFP⁺ cells from a pMIG transduced population, from these GFP⁺ cells, the top 30% based on GFP intensity were selected for analysis in order to select those with

maximal TEV protease expression (Figure 3.3B). Following treatment with 4-OHT for 8 hours, the top 30% GFP+ macrophages progressively lost RAD21-TEV protein expression until they had an average residual level of 10.25% (Figure 3.3C).

This demonstrates that RAD21-TEV macrophages can be inducibly and rapidly depleted of RAD21-TEV and cohesin function through activation of the Ert2 fused TEV protease, allowing for analysis of direct cohesin targets in gene expression.

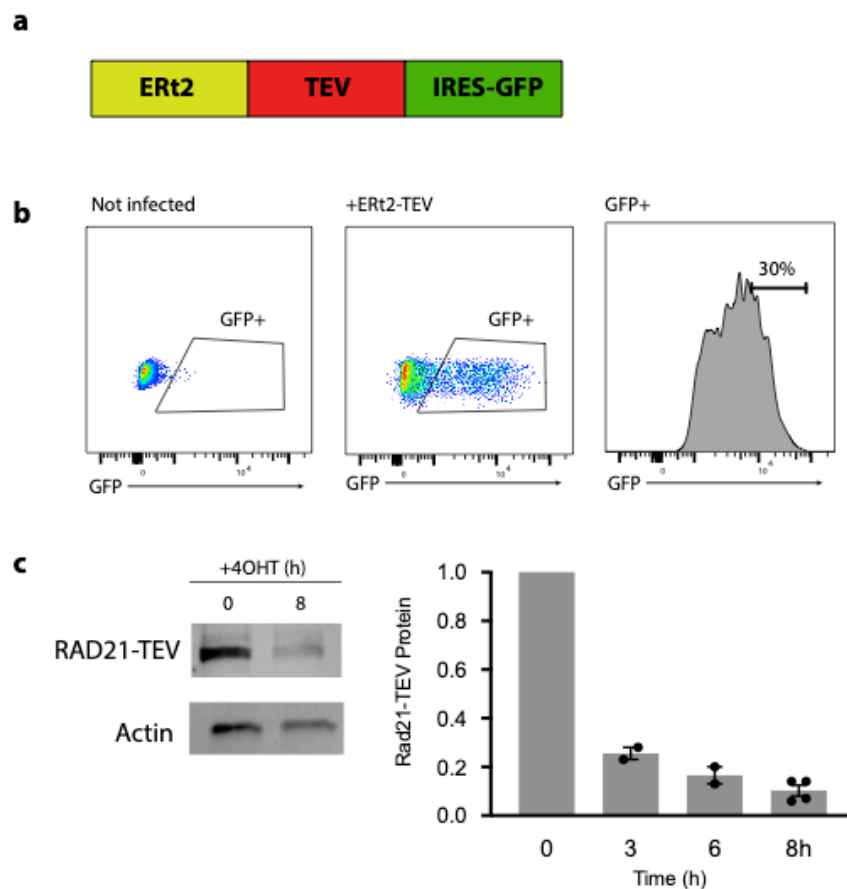


Figure 3.3 Depletion of RAD21-TEV in mature macrophages (A) Design of the pMIG retroviral insert containing TEV protease fused to Ert2 used to infect cells (B) Flow cytometric analysis of macrophages without infection, following infection, and the top 30% GFP+ used for downstream analysis (C) Immunoblot analysis of RAD21-TEV protein expression in top 30% GFP+ macrophages after 0h and 8h 4-OHT treatment, bar graph showing quantification of RAD21-TEV protein expression based on fluorescent immunoblotting after 0h, 2h, 4h and 8h 4-OHT treatment (mean $\pm$ s.e.m of 2 -4 biological replicates per time point)

3.5 Rapid RAD21-TEV cleavage leads to transcriptional deregulation

In order to identify direct cohesin targets, RAD21-TEV was depleted from macrophages for 8 hours, and RAD21-TEV depleted macrophages were then subjected to RNA-seq. Following cohesin depletion, 1034 genes were significantly deregulated (adj $P < 0.05$ based on DEseq2 analysis of 3 biological replicates), with 331 up- and 703 downregulated (Figure 3.4A). The majority of these genes (557 of 1034) were also deregulated by *Rad21* deletion ($P < 2.22 \times 10^{-16}$, odds ratio = 2.65). Acute cohesin depletion resulted in the deregulation of genes associated with processes including signaling (adj $P = 3.33 \times 10^{-8}$), inflammatory response (adj $P = 7.93 \times 10^{-7}$), immune system (adj $P = 9.23 \times 10^{-6}$) and inflammation mediated by chemokine and cytokine signaling (adj $P = 2.44 \times 10^{-8}$). Acute cohesin depletion also resulted in the preferential deregulation of inducible (Bhatt et al., 2012) over constitutive genes ($P < 2.22 \times 10^{-16}$, odds ratio = 4.44, Figure 3.4B). A 2-hour pulse of LPS further deregulated genes in macrophages acutely depleted of cohesin ($P < 2.22 \times 10^{-16}$, odds ratio = 3.54, Figure 3.4B). Inflammatory response genes were preferentially downregulated (Figure 3.4C). Transcripts downregulated following acute cohesin cleavage included key regulators of inducible gene expression such as transcription factors (*Fos*, *Jun*, *Irf2*, *Myc*, *Ets2*, *Prdm1/Blimp1*, *Egr2*, *Cebpa*, *Cebpb*), cytokines (*Il1b*), chemokines (*Ccl3*, *Ccl7*, *Ccl9*), chemokine receptors (*Ccr1*, *Ccr3*, *Ccr5*) and receptors for inflammatory mediators (*Ifnar1*, *Ifnar2*, *Ifngr1*). These data show the cohesin directly regulates inducible gene expression in macrophages, and is required for the expression of genes associated with immune system and inflammatory response.

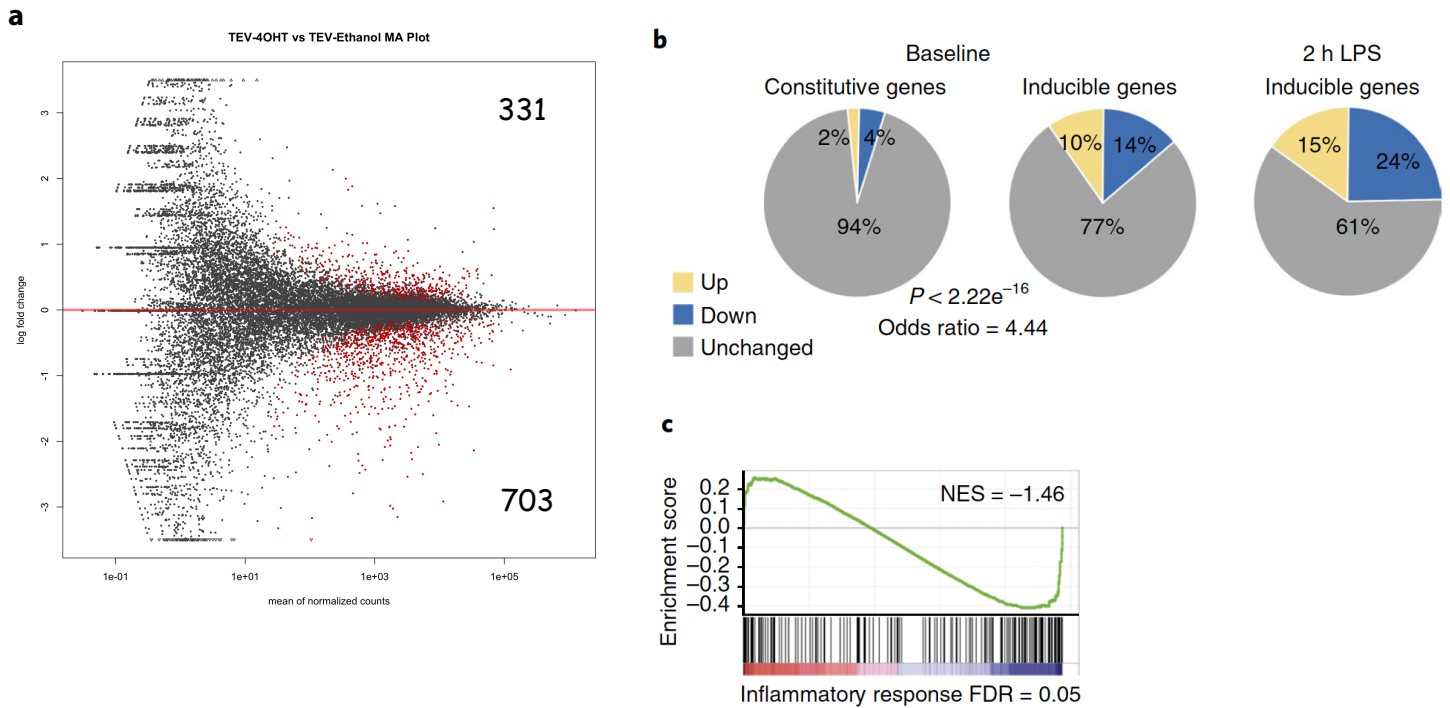


Figure 3.4 Rapid cohesin depletion leads to transcriptional deregulation, affecting inducible genes and inflammatory response (A) Volcano plot of significantly deregulated genes ($p < 0.05$, DESeq2, 3 biological replicates) in RAD21-TEV macrophages following 8h of 4-OHT treatment and RAD21-TEV depletion **(B)** Changes in constitutive and inducible gene expression in response to RAD21-TEV cleavage, and changes in inducible gene expression following 8h RAD21-TEV cleavage and an additional 2h LPS treatment, identified by DESeq2 analysis of RNA-seq ($p < 0.05$, 3 biological replicates per treatment and time point) **(C)** GSEA of inflammatory response genes at baseline following 8h RAD21-TEV depletion (NES = -1.46, FDR = 0.05)

3.6 Mechanisms of cohesin regulation of the inflammatory response

Cohesin dependent genes have previously been identified through an enrichment for cohesin binding (Seitan et al., 2011) as well as proximity to enhancers and superenhancers (Rao et al., 2017, Ing-Simmons et al., 2015, Schwarzer et al., 2017). In order to identify whether genes deregulated following the acute depletion of cohesin were indeed direct cohesin targets, we assessed whether they were enriched for cohesin binding and enhancer proximity. Cohesin

binding was defined where RAD21 peaks were identified within 10kb of the gene body, based on previously published macrophage RAD21 ChIP-seq (Cuartero et al., 2018). Enhancer and superenhancer proximity were defined by a gene body being within 40kb of the enhancer or superenhancer, which were themselves defined by previously published H3K27ac in macrophages (Cuartero et al., 2018). These features of cohesin dependent genes can be seen by the IFN receptor genes *Ifnar1* and *Ifnar2* (Figure 3.5A). Genes deregulated after acute cohesin depletion in RAD21-TEV macrophages were enriched for cohesin binding ($P < 0.01$, Figure 3.5B). Deregulated inducible genes were more likely than deregulated constitutive genes to overlap with cohesin binding sites (Figure 3.5B). Genes deregulated after acute cohesin depletion in RAD21-TEV macrophages also displayed an enrichment for enhancer proximity ($P < 10^{-9}$, Figure 3.5C), with deregulated inducible genes showing greater enrichment for enhancer proximity than deregulated constitutive genes (Figure 3.5C).

Local contacts at inducible genes in macrophages, such as *Egr2*, are known to be reconfigured following stimulation with LPS (Cuartero et al., 2018). This is thought to play a role in the dynamic regulation of these inducible genes. In order to assess whether inducible genes that were identified as being direct cohesin targets based on RNA-seq data, as well as cohesin binding and enhancer proximity, had aberrant interactions following acute cohesin depletion we performed circular chromosome conformation capture (4C) analysis. 4C analysis of the inducible genes *Egr2*, *Ifnar1*, and *Cebpb* showed a striking reduction of local interactions between the genes and surrounding regions which includes enhancers and superenhancers (Figure 3.5D) following acute cohesin depletion. This loss of interactions at baseline, and the inability to generate new interactions in the absence of cohesin, contributes to the

transcriptional deregulation observed at these genes at baseline, and following LPS stimulation, after acute cohesin depletion.

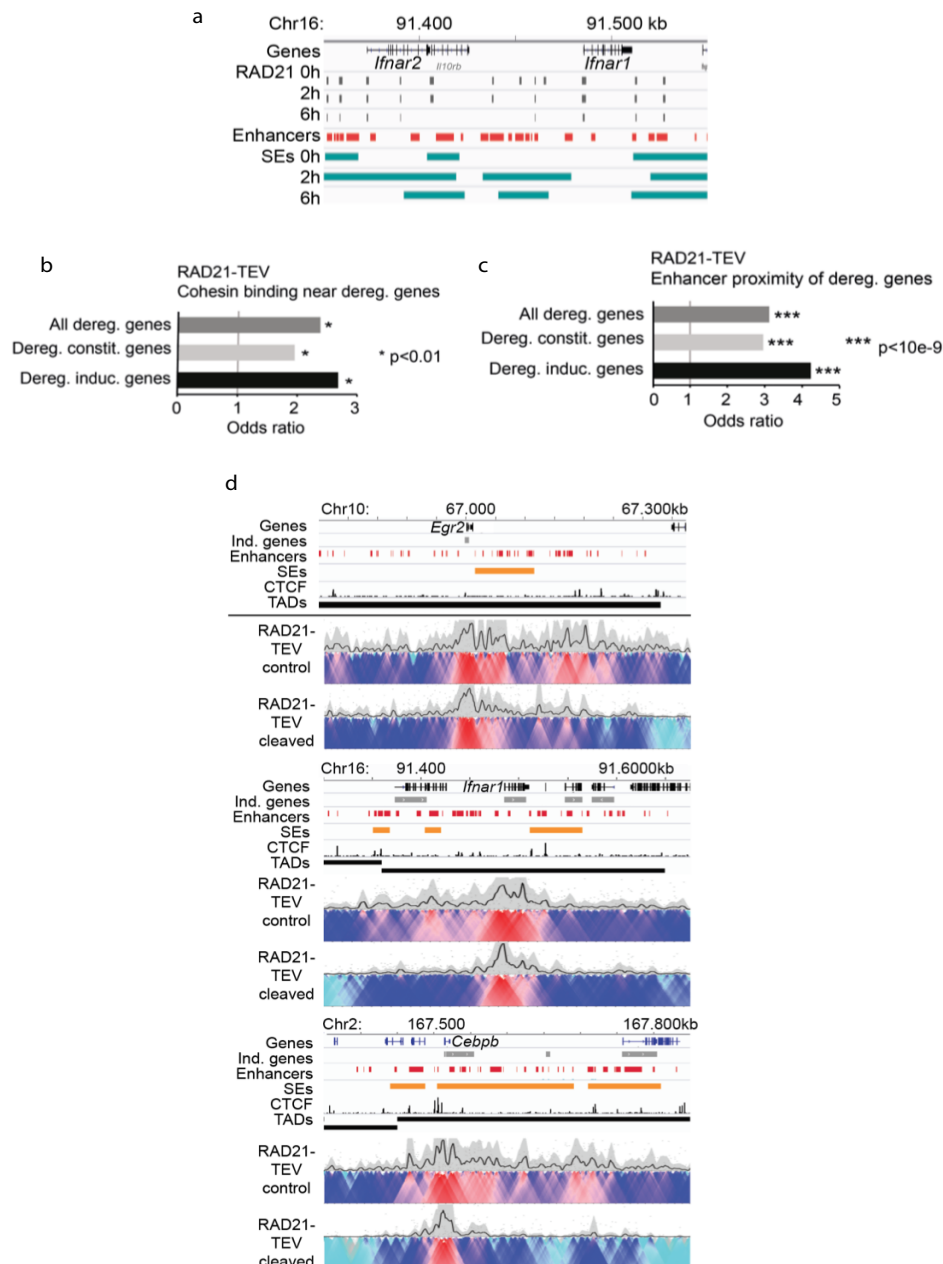


Figure 3.5 Features of cohesin dependent genes and 4C analysis of chromatin contacts (A) Characteristic features of deregulated cohesin dependent genes, that are involved in the regulation of inducible gene expression, examples are in bold. Figure also shows RAD21 ChIP-seq peaks, enhancers, and super-enhancers (SEs) mapped at indicated time points following LPS treatment **(B)** Deregulated constitutive and inducible genes are enriched for cohesin binding (RAD21 peaks within 10kb at baseline) **(C)** Deregulated constitutive and inducible genes are enriched for enhancer proximity (within 40kb of macrophage SE) **(D)** Chromosomal coordinates and deregulated genes (in bold), as well as inducible gene marker, enhancers, SEs, CTCF binding sites and TADs. Interactions between *Egr2*, *Ifnar1* and *Cebpb* promoters before and after 8h RAD21-TEV cleavage (n = 3)

3.7 How do direct targets of cohesin regulation contribute to the loss of LPS responsive gene regulatory network

Inducible gene expression in response to LPS is organised in a hierarchical network (Figure 3.6A). Early events such as the induction of transcription factors and cytokines, are required for the correct regulation of downstream genes, which in turn can affect even more downstream targets (Bhatt et al., 2012, Chen et al., 2012b, Glass and Natoli, 2016). As a result of this expanding downward propagation of signals, hierarchical transcriptional systems are susceptible to failure if key upstream regulators are disrupted. In macrophage LPS response these key upstream regulators include TLRs and IFN receptors, and transcription factors.

Longer periods of cohesin deficiency in macrophages leads to an increase in the percentage of deregulated inducible genes (Figure. 3.6B). In response to acute cohesin depletion in RAD21-TEV macrophages, 24% of inducible genes were deregulated at baseline. This fraction increased to 56% following two days of cohesin depletion in *Rad21*-deleted macrophages. These differences were mirrored following 2h treatment with LPS, with 39% of inducible genes being deregulated in acutely cohesin depleted macrophages, compared to 58% in *Rad21*-deleted macrophages (Figure 3.6B). Key upstream inducible genes, including *Ifnar1*, *Il1b*, were deregulated following acute cohesin depletion.

Deregulated inducible genes following acute cohesin depletion in RAD21-TEV macrophages were enriched for the markers of cohesin dependency; cohesin binding and enhancer proximity (Figure 3.6C). However, as deregulation of inducible genes spread to include most

inducible genes following *Rad21* deletion in macrophages, it was no longer focused on cohesin bound and enhancer proximal genes (Figure 3.6C).

These finding are consistent with the logic of inducible gene expression program discussed above. Key regulators of the inducible gene program are directly reliant on cohesin for expression, and as they fail, the knock-on effect leads to eventual collapse of the majority of the LPS transcriptional response.

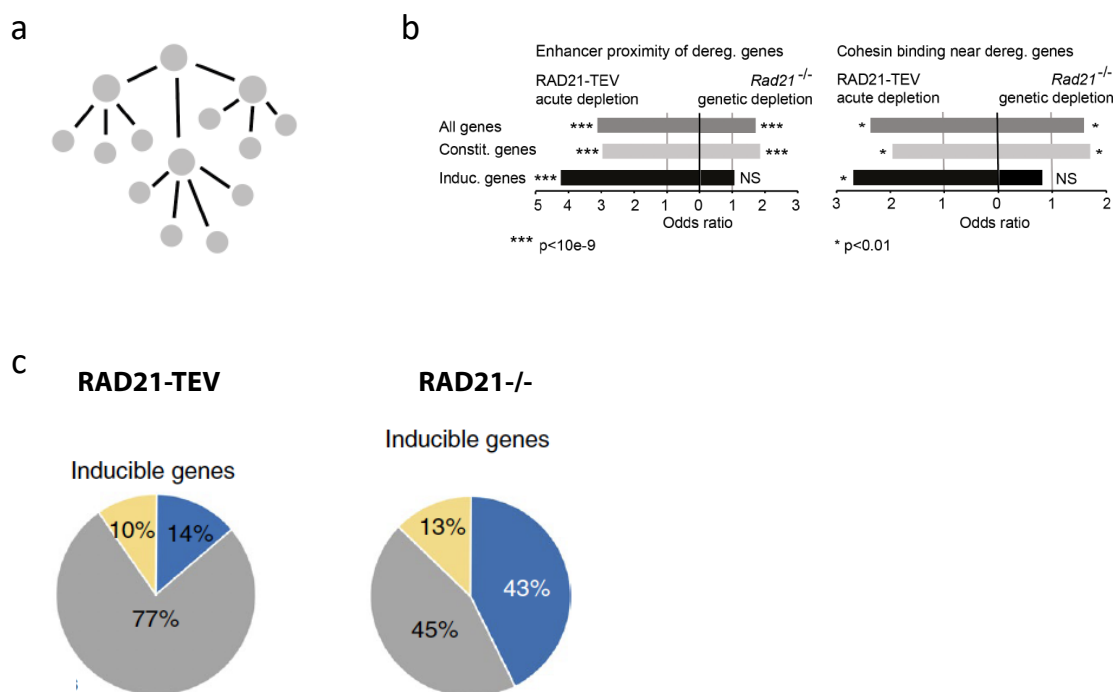


Figure 3.6 Hierarchical nature of inducible gene expression and the effect on longer term cohesin depletion (A) Model illustrating the hierarchical nature of inducible gene expression. Nodes represent genes and their products, edges indicate regulatory relationships **(B)** Proportion of significantly deregulated inducible genes following 8h of RAD21-TEV cleavage and genetic deletion of *Rad21* **(C)** Features of immediate cohesin targets are diluted as deregulation is prolonged in the extended absence of cohesin

3.8 Discussion

The loss or disruption of 3D genome organisation has previously shown a limited impact on gene expression (Rao et al., 2017, Schwarzer et al., 2017, Ing-Simmons et al., 2015). This brought into question the importance of genome organisation in gene regulation. However, work from the lab identified that cohesin, and correct 3D genome organisation, is essential for inflammatory gene expression in macrophages (Cuartero et al., 2018). This suggests that 3D genome organisation is important for changes in cell state and dramatic shifts in gene expression that require a complex change in regulatory process. The role of 3D genome organisation has therefore been underestimated by other studies focusing on cell lines in steady-state conditions.

Despite showing that cohesin and 3D organisation are important for inducible gene expression in macrophages in response to LPS, whether the process was directly controlled by cohesin remained unknown. In order to question this, RNA-seq needed to be carried out after rapid cohesin depletion. Here I show that the RAD21-TEV system can be used to generate macrophages from foetal liver progenitors, and that cohesin depletion can be achieved rapidly with temporal control.

Following the rapid loss of cohesin, RNA-seq confirmed that a selection of key upstream inducible genes are indeed directly regulated by cohesin. The hierarchical nature of the inducible inflammatory gene expression program means that deregulation of these key upstream regulators leads to widespread deregulation following longer cohesin depletion (Cuartero et al., 2018).

Inducible genes deregulated following the loss of cohesin display altered enhancer interactions, supporting the idea that 3D genome organisation affects gene expression through generating a framework by which enhancers can interact with their target genes and exert regulatory effect.

Although cohesin is required for cell proliferation (Nasmyth and Haering, 2009), many cancers accumulate cohesin mutations, including human acute myeloid leukaemia (AML) (Kon et al., 2013, Cancer Genome Atlas Research et al., 2013, Thota et al., 2014). Cells are still able to cycle as cohesin is in excess in cells of what is required for cell cycle progression (Heidinger-Pauli et al., 2010), however modest reduction in cohesin function affect the development and function of multiple organ systems in humans (Deardorff et al., 1993). This suggests correct gene expression is highly sensitive to cohesin dosage.

In haematopoietic stem and progenitor cells (HSPCs), reduced cohesin function shifts the balance of the system between self-renewal differentiation, and leads to the increased proliferation of immature progenitors (Mazumdar et al., 2015, Mullenders et al., 2015, Viny et al., 2015, Galeev et al., 2016). Inflammatory signals induce differentiation in HSPCs (Baldridge et al., 2010, Essers et al., 2009, Mossadegh-Keller et al., 2013, Nagai et al., 2006, Zhao et al., 2014). Here, along with previous work in the lab (Cuartero et al., 2018), I show that the requirement for cohesin for expression of inflammatory gene expression may explain the inability for cohesin deficient HSPCs to differentiate, as cohesin deficient myeloid cells are unable to respond to and induce inflammatory gene expression program. This may explain the significant presence of cohesin mutations in human AML.

Chapter 4: The role of cohesin in neuronal gene regulation and function

4.1 Introduction

Disruption to 3D DNA organisation through genomic deletions, or due to mutations to key genome organiser proteins such as cohesin and CTCF can cause human disease (Lupianez et al., 2015, Spielmann et al., 2018, Deardorff et al., 1993). Heterozygous or hypomorphic germline mutations to cohesin and associated co-factors causes the disease CdLS (Deardorff et al., 1993). Patients with CdLS suffer from a variety of symptoms including abnormal growth, facial abnormalities and organ defects (Deardorff et al., 1993). In addition, all CdLS patients show a degree of intellectual disability, and 60 – 65% have autism spectrum disorders (Moss et al., 2012).

Transcriptional deregulation, caused by potential abnormalities in 3D genome organisation in CdLS patients, is likely to contribute significantly to the neuronal dysfunction observed. Neurons formed during development and young age do not divide again, and a very small number are replaced in specific regions (Aimone et al., 2014). It is not known whether defects in gene expression in CdLS arise primarily from an essential role for cohesin early in development, or whether cohesin is required continuously by post-mitotic neurons. This is of interest because neurodevelopmental syndromes, such as Rett Syndrome, can benefit from postnatal interventions that correct deregulated gene expression (Cobb et al., 2010, Guy et al., 2007).

In order to identify what genes and functions are deregulated in the neurons of a mouse model of CdLS (Kawauchi et al., 2009), previous work in the lab looked at changes in gene regulation of *Nipbl*^{+/-} neurons. In addition, to understand the potential contribution of the continuous requirement of cohesin in neuronal function and gene regulation, previous work utilised the *Nex*^{Cre} mediated conditional deletion of *Rad21* in immature post-mitotic neurons (Goebbels et al., 2006, Hirayama et al., 2012) (Figure. 4.1, 4.2, 4.3).

As explained below, these results show that cohesin is required for the expression of cell-type specific genes in neurons, both after lifetime and post-mitotic depletion. In addition, cohesin is required for correct neuronal morphological maturation after exiting the cell-cycle.

The similarities in gene expression between constitutively cohesin deficient *Nipbl*^{+/-} neurons and conditionally cohesin depleted *Nex*^{Cre}*Rad21*^{lox/lox} suggested the possibility of a cohesin dependent gene expression program in neurons.

In order to address this, we used the RAD21-TEV proteolytic cleavage system (Tachibana-Konwalski et al., 2010). Controlling the time of cohesin depletion allowed for the removal of cohesin and study of transcriptional change in the absence of confounding factors in morphologically mature neurons. Using the RAD21-TEV system, we are able to address whether cohesin is required for neuronal gene expression following maturation, and if there is a cohesin dependent gene expression program in neurons.

In addition, the ability to remove cohesin in morphologically identical neurons allows us to question the role of cohesin in immediate early gene (IEG) regulation. IEGs are a group of genes transcriptionally activated in response to external stimuli to neurons. In the absence of IEG induction following stimulation, memory formation is inhibited (Frey et al., 1996, Nguyen et al., 1994). Activation of IEGs and correct regulation has previously been associated with enhancer activation and recruitment (Kim et al., 2010, Joo et al., 2016), therefore we wanted to understand the potential role for cohesin in IEG transcription.

This chapter describes unpublished work from the lab done previously, which shows shared changes in gene regulation at different stages of cohesin depletion, suggesting a cohesin dependent gene expression program in neurons. It also shows the establishment of the RAD21-TEV system in neurons, its effectiveness, and use in identifying cohesin dependent gene expression programs in neurons as well as the studying the role of cohesin in IEG regulation.

4.2 NIPBL^{+/-} neurons display abnormal gene regulation

In order to better understand the role of NIPBL, in neuronal gene expression we established explant neuronal cultures from *Nipbl*^{+/-} mice, an established mouse model of CdLS (Kawauchi et al., 2009) . *Nipbl*^{+/-} neurons display a mild reduction in *Nipbl* transcript, and protein (Fig 4.1A, B). In order to understand the role of *Nipbl* in neuronal gene regulation, primary cortical neurons from *Nipbl*^{+/-} mice were cultured and subjected to RNA-seq. RNA-seq showed significant deregulation (adjp < 0.05) of 330 genes, most of which were downregulated (251

down- and 69 upregulated) (Fig 4.1C). Downregulated genes in *Nipbl*^{+/-} neurons were related to neuronal processes including synaptic transmission, connectivity, neuronal development, and signalling as shown by gene ontology analysis. (Fig. 4.1D). Gene ontologies from upregulated genes showed a much lower significance value than downregulated and no real functional enrichment or relationship with neurons (Fig. 4.1D). The data show that NIPBL, and therefore cohesin, is required in neurons for the expression of genes related specifically to neuronal function. Genes related to neuronal function are downregulated after a partial reduction in NIPBL.

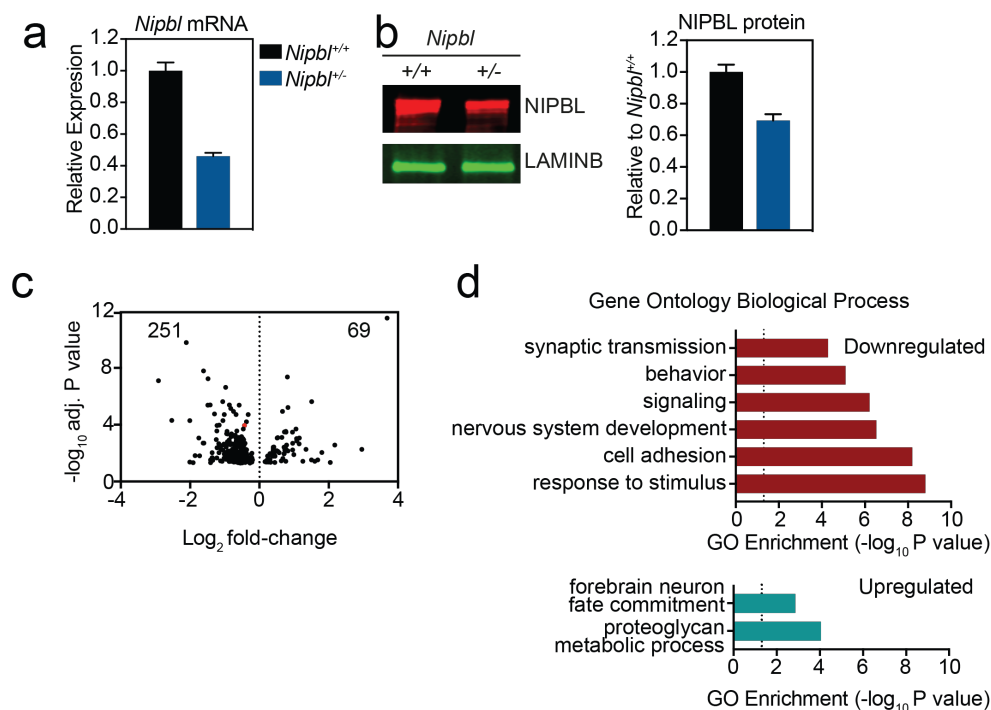


Figure 4.1. *Nipbl*^{+/-} neurons display abnormal gene regulation. (A) *Nipbl* mRNA levels in wild-type and *Nipbl*^{+/-} neurons determined by qRT-PCR and represented as relative to wild-type levels. *Hprt* and *Ubc* were used for normalisation. (B) Fluorescent western blots for NIPBL and LAMINB in *Nipbl*^{+/-} and wild-type neurons (left). Quantification of NIPBL protein levels based on western blot analysis (right). LAMINB was used for normalisation. (C) Volcano plot of significantly deregulated genes (adjP < 0.05) in *Nipbl*^{+/-} neurons analysed by RNA-seq after 10 days in culture (n = 3 biological replicated). (D) Gene ontology analysis of biological functions for up- and downregulated genes in *Nipbl*^{+/-} neurons

4.3 Conditional cohesin deletion in post-mitotic neurons

Cohesin is required for key processes in the cell cycle, such as sister chromatid cohesion, chromosome segregation, and DNA repair (Nasmyth and Haering, 2009). As a result, changes in gene expression caused by constitutive cohesin depletion, such as those observed in *Nipbl*^{+/-} neurons, may arise due to abnormalities in cell cycle or lineage commitment. To resolve this, and to understand the requirement for cohesin in neuronal gene regulation, cohesin function must be selectively abolished in cells following their exit from the cell cycle. Therefore, we combined *Rad21*^{lox}, a conditional allele of the gene encoding key cohesin subunit RAD21 (Seitan et al., 2011), with *Nex*^{Cre}, where *cre* expression is limited to post-mitotic neurons (Goebbels et al., 2006, Hirayama et al., 2012).

Explant cortical cultures of E18.5 *Nex*^{Cre/+} *Rad21*^{lox/lox} mice contained >95% neurons, as identified by pan-neuronal marker MAP2, and <1% GFAP+ astrocytes, and IBA1+ microglia (Figure 4.2A). *Rad21* mRNA expression was reduced by ≈75% (75.3 ±0.1%, n=18) in *Nex*^{Cre/+} *Rad21*^{lox/lox} explant cortical cultures (Figure 4.2B), and immunoblotting showed a corresponding reduction in Rad21 protein (76.9 ±0.1% n=6, Figure 4.3B).

In order to isolate RNA specifically from cells where CRE was expressed, we combined *Nex*^{Cre} mediated deletion of *Rad21* with *Nex*^{Cre} activated expression of an epitope-tagged ribosomal subunit (Rpl22-HA RiboTag), followed by high throughput sequencing of Rpl22-HA Ribotag associated mRNA (RiboTag RNA-seq). This approach allowed us to identify 1028 down- and 572 upregulated genes in *Nex*^{Cre/+} *Rad21*^{lox/lox} cortical neurons (Figure 4.2C). To ask what functions are associated with differentially expressed genes in cohesin deficient neurons, we

carried out gene ontology analysis (Figure 4.2D). Genes important for signalling, nervous system development, synaptic transmission and neuronal function were preferentially downregulated (Figure 4.2D). No comparable function enrichment was apparent for upregulated genes (Figure 4.2D).

Genes deregulated in *Nipbl*^{+/-} cortical neurons were also significantly likely to be deregulated in *Nex*^{Cre/+} *Rad21*^{lox/lox} neurons (Figure 4.2E), especially those which were downregulated (Figure 4.2E). Overlapping genes and pathways affected by life-long (*Nipbl*^{+/-}) and post-mitotic (*Nex*^{Cre/+} *Rad21*^{lox/lox}) cohesin depletion included neurotransmitter receptors (GABAergic, glutamatergic and others), components of the Wnt, ephrin and netrin signalling pathways, potassium channels, and key mediators of cell adhesion such as cadherins and contactins (Figure 4.2F). Genes important for dendrite and synapse development and maintenance were also affected, as well as a selection of transcription factors (Figure 2.3G). Some of these genes, shown in red, have previously been implicated in intellectual disability and autism based on the SFARI database (Figure 4.2F).

These data show that cohesin is required for neuronal gene expression, and that changes in gene expression caused by post-mitotic loss of cohesin are highly similar to those seen in life-long cohesin deficiency. Genes deregulated by the loss of cohesin are associated with cell-type specific processes and most are downregulated, with upregulated genes showing no functional enrichment. Cohesin therefore appears required for the expression of cell-type specific genes in neurons.

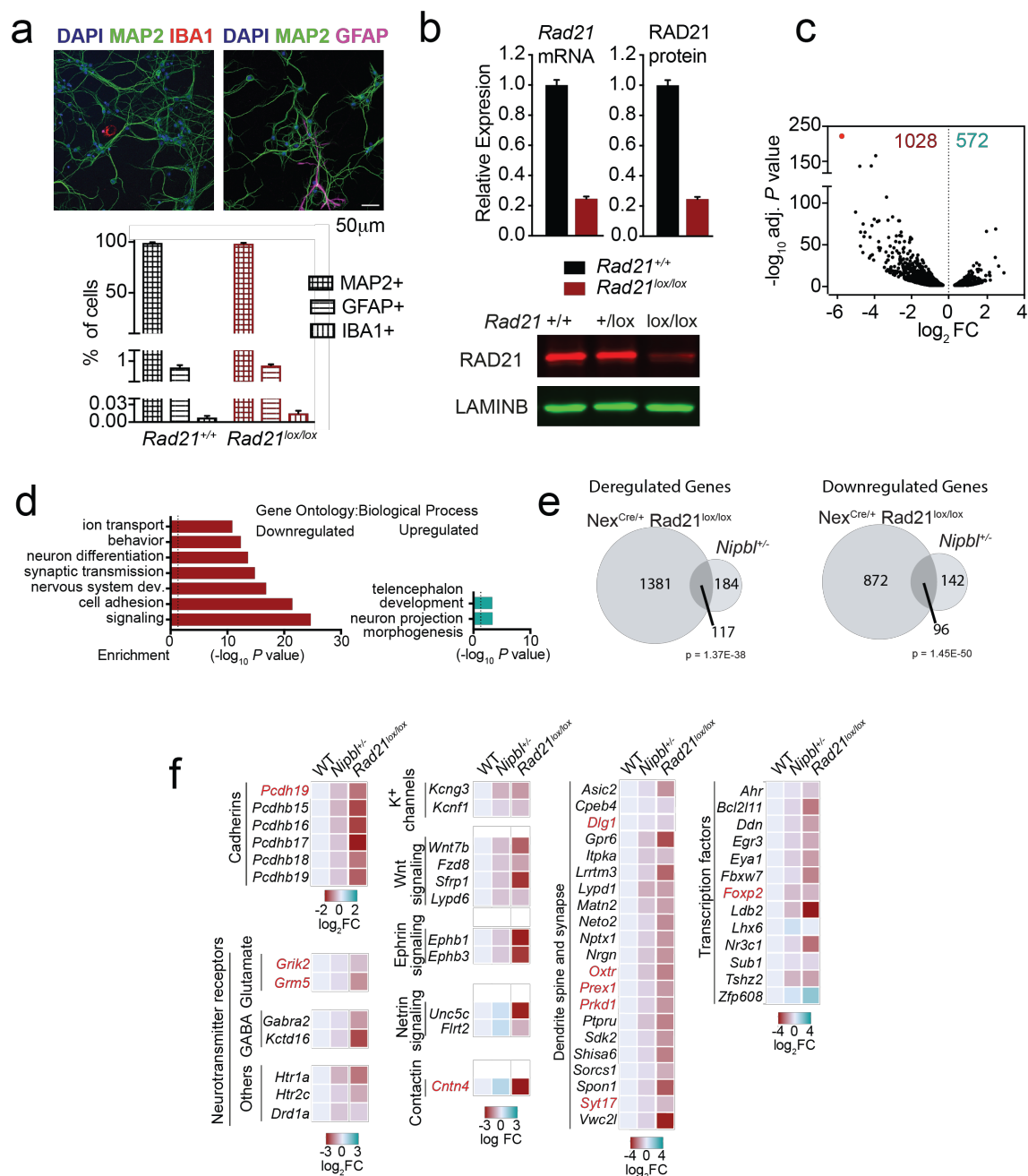


Figure 4.2 Cohesin deletion in post-mitotic neurons and changes in gene expression. (A) Neuronal cultures from E17.5 – E18.5 mouse embryos were stained with neuronal (MAP2), astrocyte (GFAP) and microglia (IBA1) markers. Quantification of markers showed >95% neuronal purity. (B) qRT-PCR analyses of *Rad21* mRNA expression in *Nex^{Cre} Rad21^{lox/lox}* and *Rad21^{+/+}* cultures. *Hprt* and *Ubc* were used for normalisation (to, left). RAD21 protein quantification in *Nex^{Cre} Rad21^{lox/lox}* and *Rad21^{+/+}* cultures, LAMINB was used for normalisation (top, right). Immunoblot of RAD21 protein levels in *Nex^{Cre} Rad21^{lox/lox}*, *Rad21^{lox/+}* and *Rad21^{+/+}* cultures. (C) Volcano plot showing log2 fold change (FC) versus significance ($-\log_{10}$ of adjP values) of significantly deregulated genes based on Ribotag RNA-seq from *Nex^{Cre} Rad21^{lox/lox}* versus *Nex^{Cre} Rad21^{+/+}* neurons. *Rad21* is indicated in red. (D) Gene ontology analysis of biological functions of deregulated genes in *Nex^{Cre} Rad21^{lox/lox}* neurons (E) Overlap of significantly deregulated genes (left) and only significantly downregulated genes (right) between *Nex^{Cre} Rad21^{lox/lox}* and *Nipbl^{+/-}* neurons (F) Examples of shared significantly deregulated genes between *Nex^{Cre} Rad21^{lox/lox}* and *Nipbl^{+/-}* neurons, and their associated functions. Genes associated with autism spectrum disorder are highlighted in red.

4.4 Role of cohesin in neuronal maturation

Many downregulated genes in *Nex*^{Cre/+} *Rad21*^{lox/lox} cortical neurons have essential roles in axon growth and guidance, dendritic and spine development, and synapse assembly, function and plasticity. To assess the impact of deregulated gene expression on the morphology of cohesin-deficient neurons we cultured *Nex*^{Cre/+} *Rad21*^{lox/lox} and *Nex*^{Cre/+} *Rad21*^{+/+} cortical neurons in the presence of wild-type glia. We sparsely labelled cultures with GFP to visualise processes of individual GAD67- neurons (Fig. 2.4A) and used Sholl (Sholl, 1953) analysis to quantitate the number of axonal crossing, the length of dendrites, number of terminal and branch points, as well as number of spines (Fig. 2.4B, C). Based on these measures, cohesin deficient *Nex*^{Cre/+} *Rad21*^{lox/lox} neurons showed significantly impaired morphological complexity.

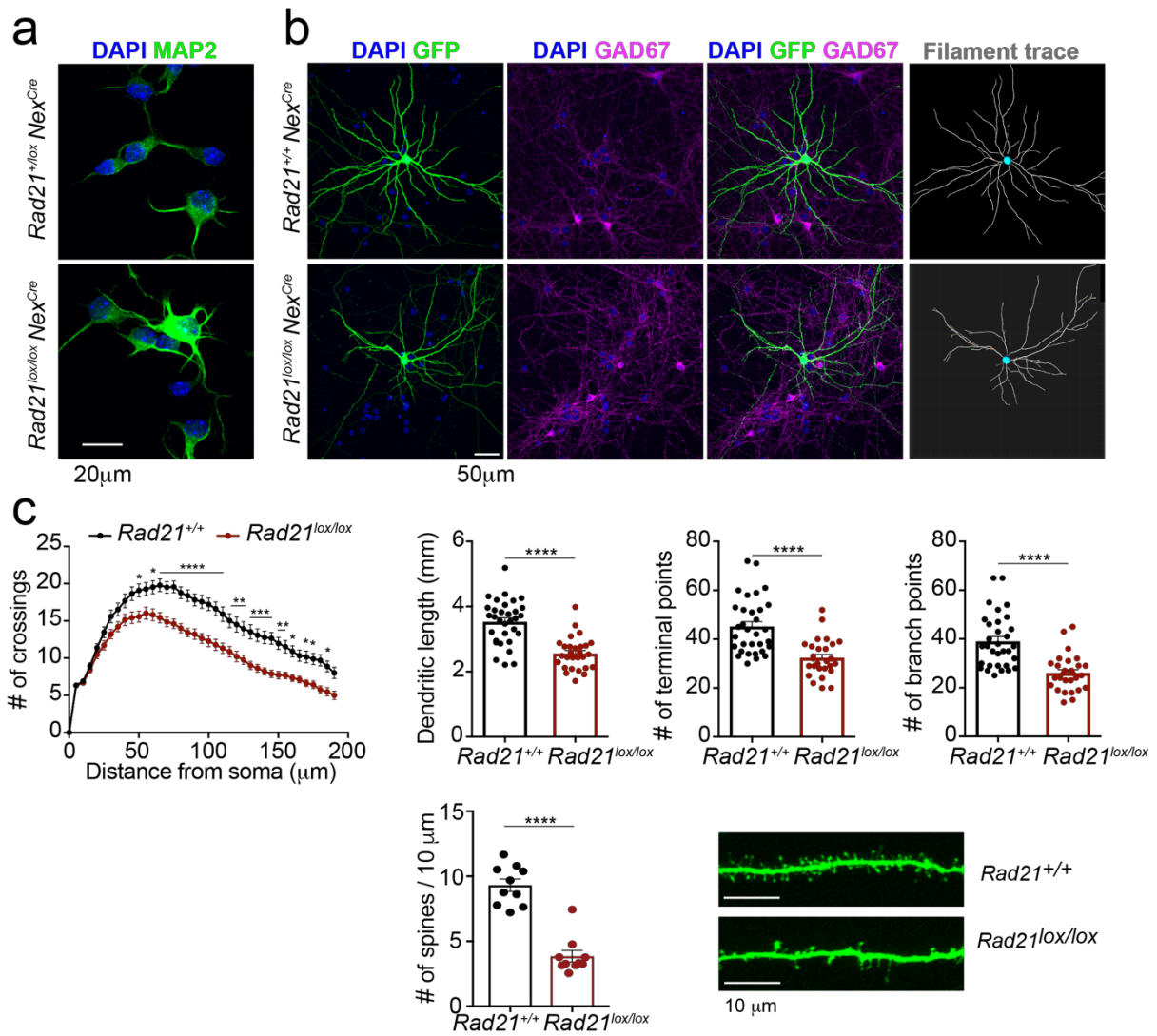


Figure 4.3 Cohesin is required for correct neuronal maturation. (A) Immunofluorescence staining of freshly explanted E18.5 neurons 1D in culture. (B) Morphology of *Nex^{Cre/+} Rad21^{lox/lox}* and *Nex^{Cre/+} Rad21^{+/+}* neurons in explant culture. Cultures were sparsely labelled with GFP to visualise processes of individual cells, and stained with GAD67 to exclude RAD21 containing GABAergic cells. (C) Sholl analysis of *Nex^{Cre/+} Rad21^{lox/lox}* and *Nex^{Cre/+} Rad21^{+/+}* neurons in culture. Sholl analysis was used to determine the number of crossings, dendritic length, number of terminal points, number of branch points and number of spines. * adj. $P < 0.05$, ** adj. $P < 0.01$, *** adj. $P < 0.001$, **** adj. $P < 0.0001$.

4.5 Generation of RAD21-TEV neurons from foetal brain

The similarities in gene expression between constitutively cohesin deficient Nipbl^{+/-} neurons and conditionally cohesin depleted Nex^{Cre/+} Rad21^{lox/lox} neurons suggested the possibility of a cohesin dependent gene expression programme in neurons. Following the discovery that cohesin is essential for correct neuronal morphological maturation, cohesin needed to be dynamically removed in more mature neurons in order to identify possible cohesin dependent gene expression programs in neurons. To achieve temporal control of cohesin depletion, we returned to the RAD21-TEV system described in the previous chapter (Tachibana-Konwalski et al., 2010). This system allows us to culture neurons with cohesin function, before removing it at a later time point following maturation.

Neurons from the cortex of embryos at E14.5, where RAD21-TEV was the sole source of RAD21, were grown in explant culture (Figure 4.4A). Neurons developed and were maintained in culture over the course of 10 days and longer. Explant cultures were treated with araC, leading to the depletion of dividing cells and enriching the cultures for neurons (Figure 4.4B). Overall, cultures were 90% neuronal, based on NeuN staining.

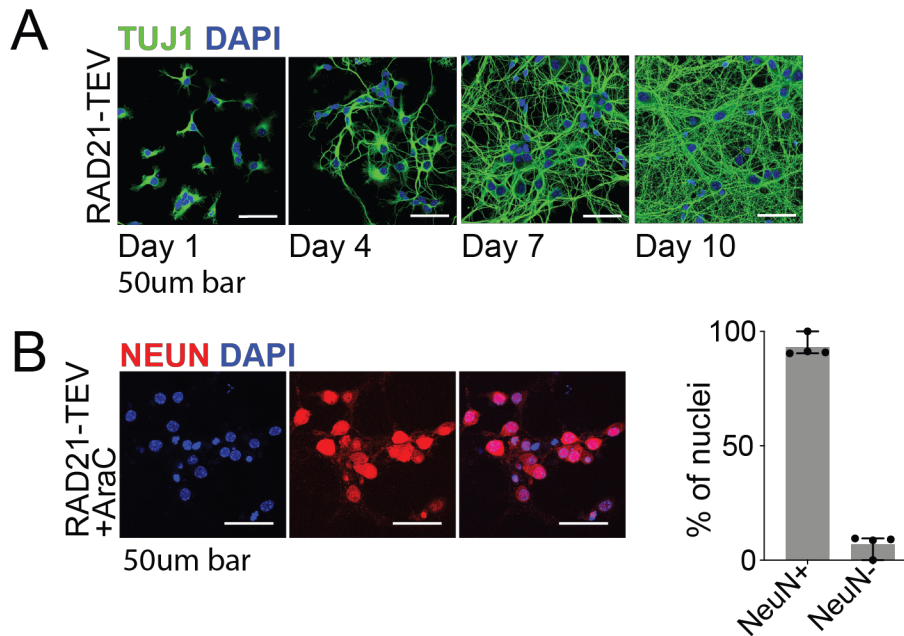


Figure 4.4 Neuronal cultures generated from E14.5 embryos. (A) Immunofluorescence of neurons explanted from E14.5 RAD21-TEV embryos and grown in culture over 10 days. (B) Immunofluorescence of neurons stained with neuronal nuclei marker (NeuN) and DNA stain (DAPI) to identify neuronal cells in culture. Quantification of immunofluorescent images to determine percentage of neuronal cells in culture, analysis shows on average 90% of cells were NeuN+ (n=4, error bars = range).

4.6 Characterisation of RAD21-TEV depletion in neurons

In order for RAD21-TEV to be cleaved and cohesin function impaired in neurons, the TEV protease needs to be expressed in those cells. Due to neurons being post-mitotic we were unable to use the retrovirus described earlier. Instead a lentiviral based system was developed where FUG-T2A-CAS9 lentivirus (Williams et al., 2016) was modified so the CAS9 was replaced with an ERT2 fused TEV protease (FUGTEVP). ERT2 fused TEV protease was therefore expressed under control of the ubiquitin promoter (Figure 4.5A). Once again, the ERT2-TEV protease fusion also contained a v5 epitope tag, allowing for tracking of the TEV protease. Downstream of the TEV protease was a t2a fused GFP. This allowed for the

ratiometric expression of TEV protease and GFP, allowing to identify successfully infected and therefore TEV protease expressing cells.

In order to cleave RAD21-TEV, cells were infected with FUGTEVP and then treated with 4-OHT. This caused the translocation of ERt2 fused TEV protease to the nucleus from the cytoplasm (Figure 4.5B, C), causing rapid cleavage of RAD21-TEV (Figure 4.5D). After only 8h of 4-OHT treatment, RAD21-TEV levels were significantly reduced, and this was sustained for 24h (Figure 4.5D). Following the addition of 4-OHT, RAD21-TEV levels were on average 15% compared to cells not treated with 4-OHT (Figure 4.5E).

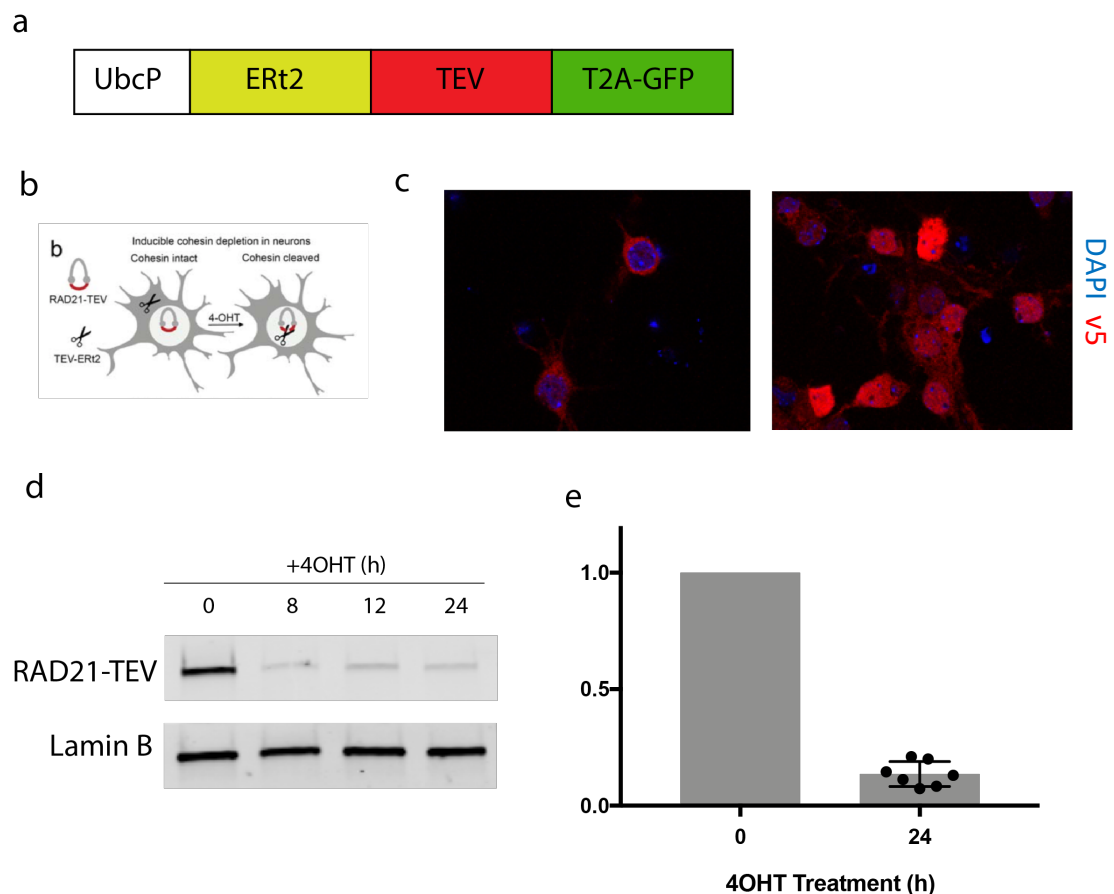


Figure 4.5 Characterisation of RAD21-TEV cleavage in neurons. (A) Representation of the FUGTEVP lentivirus design in order to achieve temporal control of RAD21-TEV cleavage. (B) Schematic image of how RAD21-TEV cleavage can be achieved in neurons infected FUGTEVP following 4-OHT treatment. (C) Immunofluorescent staining for the v5 TEV protease tag, and DNA marker DAPI following 0 or 15m of 4-OHT treatment, showing translocation of TEV protease from the cytoplasm to the nucleus. (D) Immunoblot of RAD21-TEV and LAMINB following 4-OHT treatment, showing the rapid loss of RAD21-TEV protein expression. (E) Quantification of RAD21-TEV protein levels, based on immunoblots, at 0h and 24h of 4-OHT treatment. RAD21-TEV amount represented against 0h levels. LAMIN B was used for normalisation (n = 7 biological replicates, bars = S.E.M)

4.7 Cohesin is required for maintaining 3D interactions in mature neurons

The protocadherin gene cluster is known to be regulated by 3D interactions in neurons between *Pcdh* genes and their enhancers, HS5-HS1 (Hirayama et al., 2012, Guo et al., 2015). Previous work has shown that *Pcdh* genes are sensitive to the loss of cohesin function (Kawauchi et al., 2009). In order to identify whether RAD21-TEV depletion led to a loss of interactions, we carried out a 4C analysis of the *Pcdhb* locus. Interactions at the *Pcdhb* locus, and those between the HS5 enhancer and *Pcdh* genes were significantly reduced following the cleavage of RAD21-TEV (Figure 4.6A). Cohesin is required to maintain 3D genome interactions in post-mitotic neurons, and RAD21-TEV neurons significantly lose 3D genome organisation following the loss of cohesin function.

These results demonstrate that RAD21-TEV can be inducibly and rapidly depleted in RAD21-TEV neurons through a lentiviral based system, leading to a reduction in 3D genome organisation, and allowing for the study of cohesin based gene regulation in neurons.

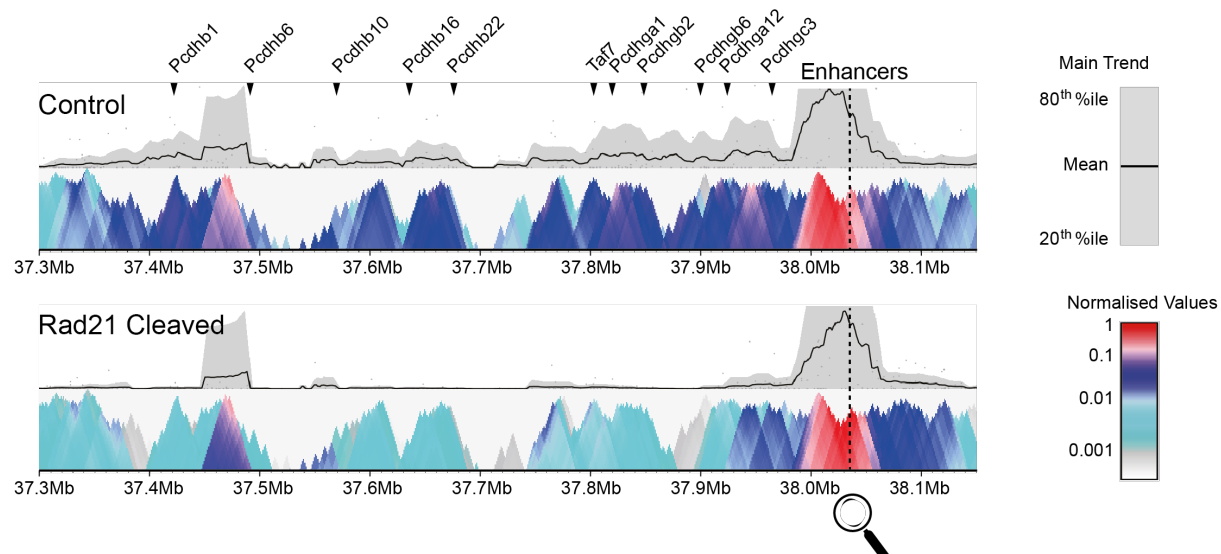


Figure 4.6 Interactions at the *Pcdhb* locus following RAD21-TEV cleavage. (A) Contact profiles of the enhancers HS18-20 in control cells (top panel) and in Rad21 cleaved cells (bottom panel). Black arrowheads indicate promoter sites. Dashed line and looking glass indicate the enhancer site (viewpoint). Grey band displays the 20%-80% percentiles and black line within the grey band shows mean values that were calculated for 40 kb windows. Colour panel below the grey band shows means of contact intensities for multiple window sizes (from 2 kb to 50 kb) (n = 4 biological replicates).

4.8 Cohesin is required in mature neurons to maintain neuronal gene expression

In order to identify genes that were reliant on cohesin for gene expression in neurons, RAD21-TEV was depleted from neurons for 24H and up to 72H (Figure 4.7A), and RNA-seq was performed. RNA-seq showed that 759 genes were deregulated (287 up- and 472 downregulated) within 24H of TEV induction (Figure 4.7B), with deregulated expression maintained following 72H of TEV induction with 1321 genes significantly deregulated (460 up- and 861 downregulated, Figure 4.7B). Gene ontology analysis showed selective downregulation of genes related to neuronal development and function, including synapse formation and transmission, cell adhesion, signalling, and gated channel activity 24H after

TEV induction (Figure 4.7C). Gene ontology analysis of upregulated genes showed reduced significance and functional enrichment (Figure 4.7C). As mentioned earlier, cohesin dependent genes have previously been identified through an enrichment for cohesin binding (Seitan et al., 2011) as well as proximity to enhancers and superenhancers (Rao et al., 2017, Ing-Simmons et al., 2015, Schwarzer et al., 2017). Genes deregulated following 24H TEV induction in neurons showed a significant enrichment for both enhancer proximity and RAD21 binding compared to all genes (Figure 4.7D).

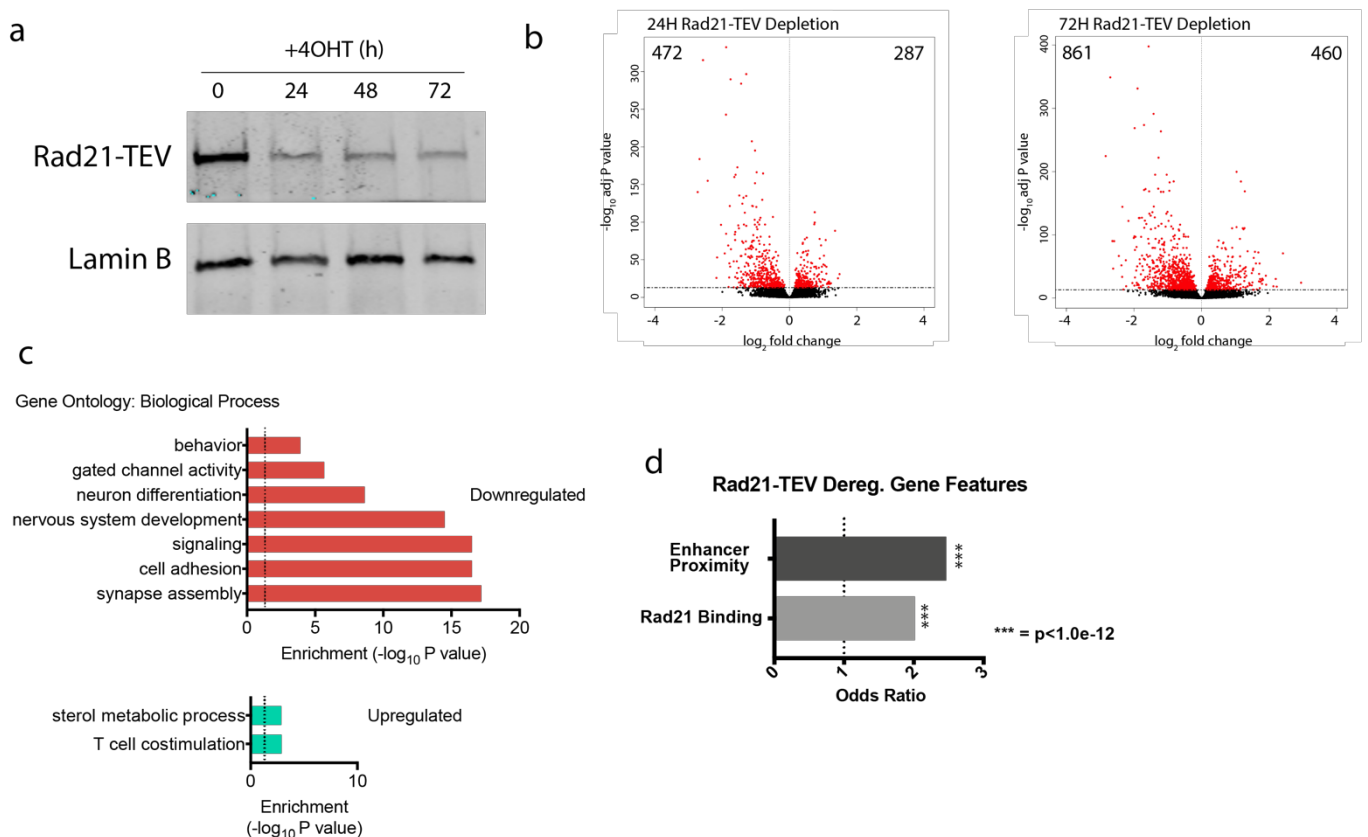


Figure 4.7 Cohesin sustains neuronal gene expression in mature neurons. (A) Immunoblot of RAD21-TEV and LAMINB in RAD21-TEV neurons following 0h, 24h, 48h and 72h of 4-OHT treatment. **(B)** Volcano plot showing log₂ fold change (FC) and -log₁₀ adj P value of gene expression in neurons following 24H and 72H of 4-OHT treatment. Significantly deregulated genes (adj P < 0.05) are marked in red (n = 3). **(C)** Gene ontology analysis of biological processes from significantly deregulated genes following 24H of RAD21-TEV cleavage. **(D)** Analysis of enhancer proximity, (genes within 40kB of previously identified neuronal enhancers (Malik et al., 2014)), and RAD21 binding (genes within 10kB of previously identified brain RAD21 binding sites (Fujita et al., 2017)) at significantly deregulated genes following 24H of RAD21-TEV cleavage.

4.9 Cohesin dependent gene regulation in neurons

In order to identify whether there is a cohesin dependent gene expression program in neurons, we compared changes in neuronal gene expression between three different cohesin depletion models; life-long cohesin depletion in neurons cultured from *Nipbl*^{+/-} mice, cohesin depletion following exit from the cell cycle using *Nex*^{Cre/+} *Rad21*^{lox/lox} cortical neurons, and finally cohesin depletion in more mature neurons through depletion of RAD21-TEV.

To visualise similarities in processes effected by the loss of cohesin across the different time points, we grouped affected gene ontologies into broad categories (Figure 4.8A). Gene ontologies affected across all time points of cohesin depletion are broadly shared, with enrichment for the deregulation of neuron specific processes (Figure 4.8A). The same processes are not always affected when known regulators of transcription are disrupted in neurons. For example, genes deregulated in *Mecp2*-deficient cortical neurons (Kriaucionis et al., 2006) are related to processes such as cellular secretion, homeostasis and cell cycle (Figure 4.8A). These data demonstrate that reduced cohesin function resulted in the downregulation of specific gene sets related to neuronal function, regardless of when cohesin function is lost.

To further examine the expression changes of individual genes in the absence of cohesin, we visualised expression values of individual genes within GO terms related to adhesion, synapse and neuronal and nervous system development (Figure 4.8B). Gene expression was highly similar between conditions of acute (RAD21-TEV) and chronic (*Nex*^{Cre/+} *Rad21*^{lox/lox}) cohesin depletion, and the majority of genes were regulated in the same direction (Figure 4.8B). While

not all genes that were significantly deregulated in one cohesin depletion setting necessarily passed the significance in the other, they overwhelmingly showed the same trend (Figure 4.8B) (Synapse: $p < 2.22e-16$ odds ratio = 11.26, $R_s = 0.62$; Adhesion: $p < 2.22e-16$ odds ratio = 15.21, $R_s = 0.6$; Neuronal and nervous system development: $p < 2.22e-16$ odds ratio = 9.92, $R_s = 0.58$).

We used GSEA to further compare the expression of genes deregulated across cohesin depletion models (Figure 4.8C, D). The set of 472 genes downregulated 24H after TEV induction in RAD21-TEV neurons (DEseq2 adj. $p < 0.05$) was also downregulated in the other two cohesin depletion settings (*Nex*^{Cre/+} *Rad21*^{lox/lox}: NES -3.61, FDR = 0; *Nipbl*^{+/-}: NES -2.91, FDR = 0) (Figure 4.8C). We then asked the reciprocal question by examining the impact of acute cohesin depletion in RAD21-TEV neurons on gene sets that were found downregulated in *Nex*^{Cre/+} *Rad21*^{lox/lox} and *Nipbl*^{+/-} neurons in explant culture. Genes downregulated in *Nex*^{Cre/+} *Rad21*^{lox/lox} and *Nipbl*^{+/-} neurons were also downregulated in RAD21-TEV neurons (*Nex*^{Cre/+} *Rad21*^{lox/lox}: NES -2.95, FDR=0; *Nipbl*^{+/-}: NES -2.8; FDR = 0) (Figure 4.8D).

These findings indicate that acute (RAD21-TEV), chronic (*Nex*^{Cre/+} *Rad21*^{lox/lox}) and life-long (*Nipbl*^{+/-}) cohesin deficiency in cortical neurons results in the coordinate deregulation of highly similar gene sets and genes related to synaptic transmission, connectivity, neuronal development and signalling (Figure 4.8E). Cohesin therefore does regulate specific gene expression programs in neurons, and those are related to cell-type specific function.

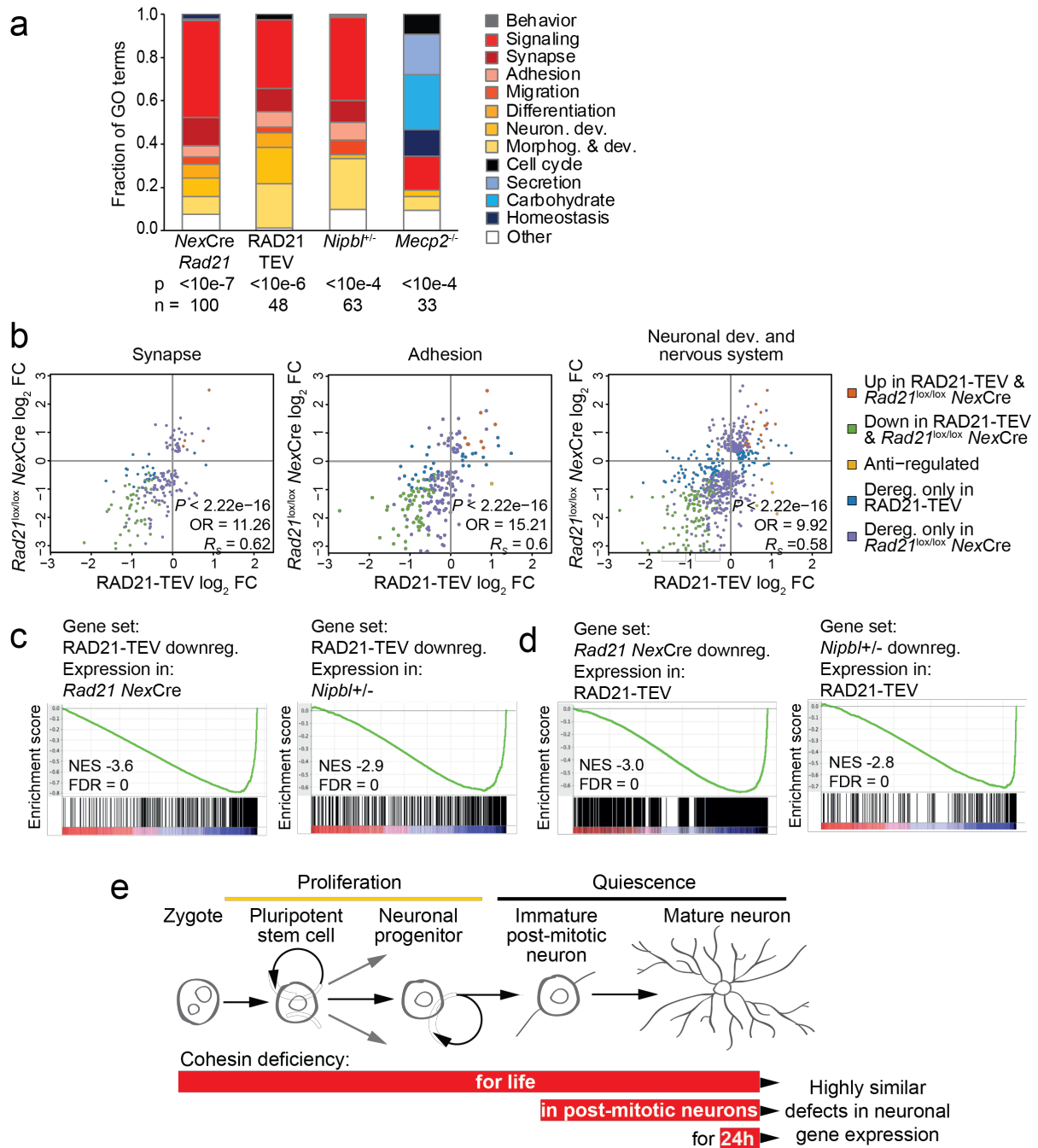


Figure 4.8 Cohesin dependent neuronal gene regulation. (A) Overlap of represented gene ontologies based on deregulated genes in *Nex^{Cre} Rad21^{lox/lox}*, *RAD21-TEV*, *Nipbl^{+/-}*, and *Mecp2*-deficient neurons (Kriaucionis et al., 2006). (B) Scatter plots of gene expression within aggregate GO terms, comparing *RAD21-TEV* with *Nex^{Cre} Rad21^{lox/lox}* neurons. Genes that were found deregulated in at least one of the genotypes are shown. P-values and odds ratios refer to the probability of finding the observed patterns of co-regulation by chance. R_s : Spearman's rank coefficient. (C) GSEA showing the expression of the gene set downregulated in *RAD21-TEV* neurons in *Nex^{Cre} Rad21^{lox/lox}* and *Nipbl^{+/-}* neurons. (D) GSEA showing the expression of genes downregulated in *Nex^{Cre} Rad21^{lox/lox}* and *Nipbl^{+/-}* neurons in *RAD21-TEV* neurons. (E) Similar gene sets are deregulated in cohesin deficient neurons regardless of when cohesin function is impaired.

4.10 The role of cohesin in activity dependent gene expression

Neuronal synaptic plasticity is the cellular basis for learning and memory. In order for synaptic plasticity to occur, neurons need to be able to regulate transcription in response to stimuli. A key component of this transcriptional regulation is the activation of IEGs.

By using the RAD21-TEV system we could eliminate potential confounders on IEG regulation caused by long-term cohesin deficiency, such as abnormal morphology, and instead look at the more direct role for cohesin in neuronal gene expression.

In order to define our set of IEGs, RAD21-TEV neurons were cultured for 10 days, in the absence of TEV induction so cohesin function was maintained, and then treated for 30 and 120 minutes with brain derived neurotrophic factor (BDNF) to elicit IEG transcription. Treatment with BDNF for 30 minutes led to the significant upregulation of 24 genes ($p < 0.05$, $\log FC > 0$), including canonical IEGs such as *Arc*, *Fos*, *Jun* and *Nr4a1* (Figure 4.9A)

Following 24H of RAD21-TEV cleavage 11 out of 24 IEGs were significantly deregulated at baseline (Figure 4.9B) and all of these were significantly downregulated. Following 30 minutes of BDNF stimulation however, all deregulated genes were rescued (Figure 4.9B) and expressed at normal levels. Treatment with BDNF for a further 90 minutes led to 3 genes becoming significantly deregulated (Figure 4.9B), interestingly these three were all upregulated compared to neurons with cohesin function.

Deregulated IEGs at baseline in cohesin deficient cells were able to be expressed at normal levels following 30 minutes of BDNF treatment as they were over induced compared to controls (Figure 4.9C). This shows that cohesin is not required for the correct function of feedback mechanisms that determine IEG expression nor gene induction following BDNF treatment.

This shows that cohesin is not required for the induction of IEGs, but instead the fine tuning of expression. Cohesin deficient cells are able to correctly express IEGs despite starting incorrectly at baseline. They have the ability to over induce and compensate for the lower baseline expression, and achieve close to correct levels of expression, that are not significantly different, in the absence of cohesin.

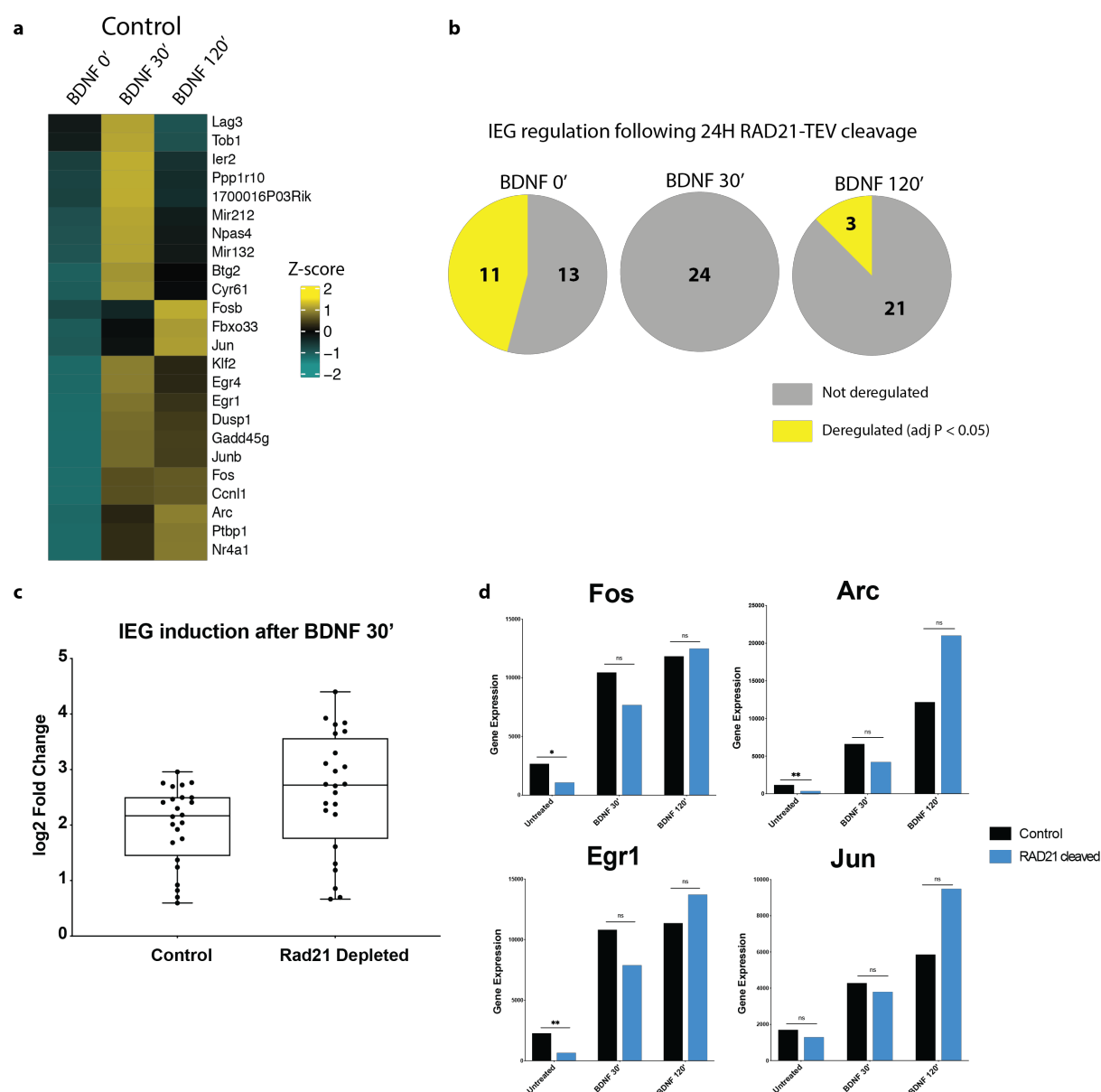


Figure 4.9 IEG regulation in cohesin deficient neurons. (A) IEG expression values at baseline, 30 minutes BDNF treatment, and 120 minutes BDNF treatment in control neurons. (B) Proportion of IEGs deregulated in RAD21-TEV cleaved neurons at baseline, 30 minutes BDNF treatment, and 120 minutes BDNF treatment. (C) Induction of IEGs in control and RAD21-TEV cleaved neurons following 30 minutes of BDNF treatment. (D) Examples of IEGs expressed in control and RAD21-TEV cleaved neurons. Gene expression values are normalised counts based on RNA-seq from 3 biological replicates. * adj. P < 0.05, ** adj. P < 0.01

4.11 Discussion

CdLS is caused by mutations to the cohesin complex and associated co-factors. These mutations compromise cohesin function throughout development. While previous work has focused on the role of cohesin in early development (Fujita et al., 2017, van den Berg et al., 2017, Kawauchi et al., 2009, Remeseiro et al., 2012, Bottai et al., 2019) , the continuous requirement for cohesin in neuronal gene expression and function remains unknown. In this chapter, I, along with previous work in the lab, have shown that cohesin is required for the expression of cell type specific genes, that cohesin is continuously required for neuronal gene expression in neurons, and that cohesin is not required for the induction, but for the fine tuning of IEG expression.

Our data show that neuronal life-long cohesin deficiency, in the *Nipbl*^{+/-} mouse model of CdLS, leads to deregulation of cell-type specific genes. These changes in gene expression are also reflected in post-mitotic genetic deletion of cohesin using the *Nes*^{Cre} *Rad21*^{lox} system, as well as following cohesin cleavage using RAD21-TEV in mature neurons. Regardless of when cohesin is lost, the same genes and the same processes are affected. Genes deregulated by the loss of cohesin have important roles in maintaining neuronal function through synapse and dendrite development, as well as maintaining circuit function through signalling. These processes are required throughout neuronal life. Therefore cohesin, and correct 3D organisation, is essential for life-long neuronal function.

The data show that 3D genome organisation is essential for cell-type specific gene expression in neurons. This has been further backed up by data from CTCF deficient neurons where

neuron specific genes were also preferentially deregulated (Hirayama et al., 2012). This contributes to identifying a gene expression phenotype of cells with improper or lack of 3D genome organisation. It also highlights why 3D genome organisation is important, not for the regulation of common genes expressed by all cells such as housekeeping genes, but for regulation of genes that give specific cells their own functions.

In addition, the similarities in transcriptional deregulation between lifelong cohesin deficiency and removal in mature neurons, suggests that defective gene expression during development caused by cohesin deficiency may not have long lasting effects on neuronal gene expression. Therefore, defective gene regulation in neurons of CdLS patients may be a result of an ongoing requirement for cohesin as opposed to critical requirement during development. In addition, it suggests that the cognitive impairment in those patients may also partly arise from the ongoing requirement for cohesin. This is of particular interest as it suggests that being able to rescue changes in gene expression caused by cohesin deficiency, even in adult neurons, maybe lead to some benefit. Much like what has previously been reported in models of Rett Syndrome (Cobb et al., 2010, Guy et al., 2007).

Induction of IEGs in response to external stimuli can be accompanied by recruitment of RAD21 and Mediator, as well as increased 3C interactions (Schaukowitch et al., 2014). It has also been suggested that cohesin and CTCF may provide the specific enhancer-promoter interactions that enable the inducible expression of activity-dependent genes (Schaukowitch et al., 2014, Rajarajan et al., 2016, Sams et al., 2016). My data shows that cohesin is in fact not required for IEG induction, but instead for the fine tuning of expression. IEGs are not deregulated following stimulation in RAD21-TEV cohesin depleted cells. Work shown here is

the first study to investigate the role of cohesin in IEG regulation in rapidly cohesin depleted cells. This allows for the identification of the direct requirement for cohesin in IEG regulation, in the absence of changes in neuronal morphology and function caused by prolonged deficiencies of architectural proteins and the resulting changes in gene expression (Fujita et al., 2017, Hirayama et al., 2012, Sams et al., 2016). Cohesin is required for correct baseline expression of IEGs, but inducibility is not impaired. The role of 3D genome organisation on IEG regulation is more nuanced. It appears to be required for the fine tuning of expression. Further work using chromosome conformation techniques in wild-type and cohesin deficient neurons is required in order to identify whether there are differences in enhancer-promoter interactions at IEGs following activation.

Overall this chapter has shown that cohesin is required for the regulation of cell-type specific genes in neurons. That cohesin is continuously required to maintain neuronal gene expression in neurons, and regardless of when cohesin is lost changes in gene expression are highly comparable. Cohesin is not required for the induction of IEGs in response to stimuli but is required for the appropriate expression. In the absence of cohesin, and by extension 3D genome organisation, neurons are unable to regulate the genes that provide the genetic foundation for learning and memory.

Chapter 5: Can changes in gene expression caused by the loss of cohesin function be rescued?

5.1 Introduction

Disruption to 3D genome organisation caused by cohesin depletion can be rescued (Rao et al., 2017), however whether this also leads to a concurrent rescue in gene expression remains unknown. Work has previously shown that the removal followed by reintroduction of CTCF can lead to rescue of gene expression (Nora et al., 2017), suggesting that defective gene expression caused by aberrant 3D genome organisation is not permanent. The changes in 3D genome organisation however, compared between the loss of CTCF and the loss of cohesin, are highly different. Where the loss of CTCF leads to a weakening of TAD boundaries and the loss of loop formation (Nora et al., 2017), the loss of cohesin has a more widespread effect with the loss of TADs, contact domains and chromatin loops (Rao et al., 2017). Therefore, it is not a given that rescuing cohesin will have the same effect and lead to a rescue of transcriptional deregulation.

The overlap between changes in gene regulation caused by lifelong and transient cohesin depletions, shows that there is a cohesin dependent gene expression in neurons. In addition, the overlap suggests that defective gene expression during development may have a limited effect on gene expression in developed neurons. Therefore, the continuous requirement for

cohesin on neuronal gene expression may play an important role in the cognitive impairment and autism spectrum disorder observed in CdLS patients.

As mentioned in the previous chapter, Rett Syndrome models can benefit from postnatal interventions to correct deregulated gene expression (Cobb et al., 2010, Guy et al., 2007) . Therefore, if changes in gene expression caused by defective 3D genome organisation can be rescued, it is an inviting prospect to apply similar approaches as those in the Rett Syndrome field to models of CdLS.

In order to identify whether changes in gene regulation caused by defective gene expression could be rescued, the RAD21-TEV system was once again utilised. The proteolytic targeting nature of the system allows for reversibility, unlike genetic modifications such as Cre recombinase targeting to *loxP* flanked *Rad21* exons.

In this chapter, I show how the RAD21-TEV system was adapted to being reversible, and how loss of RAD21 followed by its reintroduction affected neuronal gene expression.

5.2 Establishing a proteolytic based cohesin depletion and rescue system

In order to test whether changes in gene regulation caused by defective 3D genome organisation can be rescued, I developed a reversible doxycycline inducible TEV protease based system (Figure 5.1A). A lentiviral construct was developed from a previously designed inducible Cas9 system (Cao et al., 2016) where rTetR is constitutively expressed under control

of the Ubc promoter, followed by IRES-RFP. Upon treatment with doxycycline, rTetR binds to the Tet response element (TRE), which drives expression of a nuclear localisation sequence (NLS) fused TEV protease. This lentivirus was known as (FUGdoxTEV). NLS-TEV protease translocates to the nucleus where it cleaves RAD21-TEV leading to reduction of RAD21-TEV protein levels, and the loss of cohesin function (Figure 5.1B). RAD21-TEV protein levels are rapidly depleted following treatment with doxycycline, with maximum depletion occurring after 24 hours. RAD21-TEV levels remain low for several days before returning to normal 7 days after treatment (Figure 5.1B).

This system allows for the analysis of changes in gene expression following cohesin depletion and rescue.

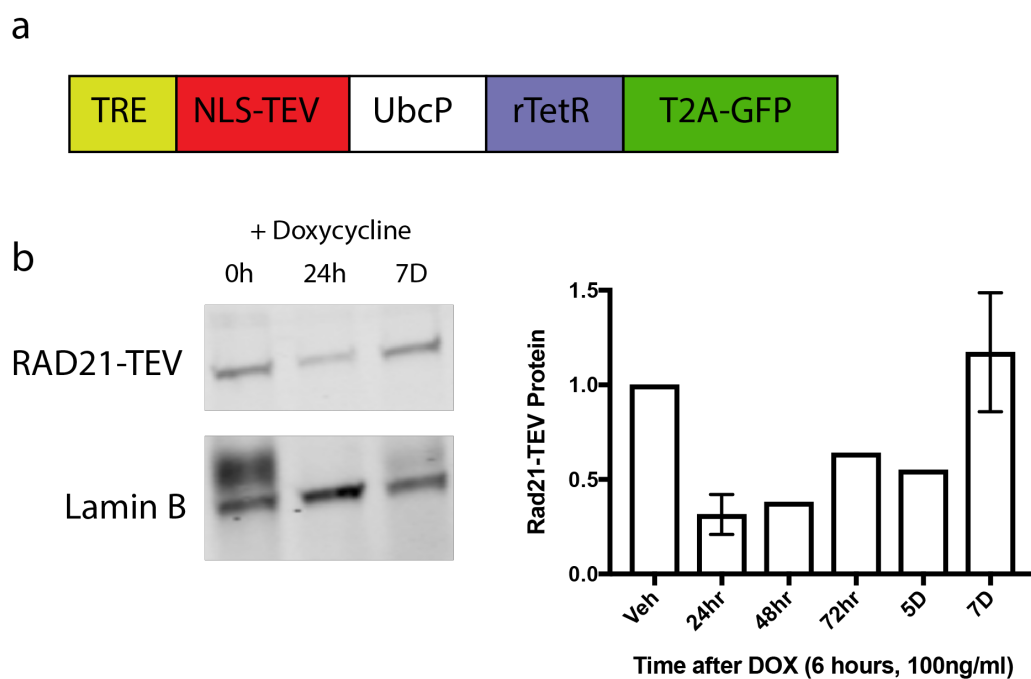


Figure 5.1 Doxycycline based control of cohesin expression. (A) Schematic representation of the lentivirus design of FUGdoxTEV, where doxycycline controls the expression of NLS fused TEV protease. (B) Immunoblot of RAD21-TEV and LAMINB following doxycycline treatment and wash off, at 0h, 24h and 7D post-treatment (left). Quantification of RAD21-TEV protein expression at sequential time points following doxycycline treatment and wash off. Values are normalised to LAMINB expression (veh, 24h and 7D are based on 4 biological replicates, 48h, 72h and 5D based on 1 biological replicate, error bars = range).

5.3 24 hours of doxycycline controlled RAD21-TEV cleavage leads to changes in gene expression

Following 24H of doxycycline induced RAD21-TEV cleavage, RNA-seq showed the deregulation of 720 genes (170 up- and 570 downregulated, adj $p < 0.05$ based on DEseq2 analysis of 3 biological replicates, Figure 5.2A). As seen before in cohesin depleted neurons, downregulated genes were associated with cell-type specific processes when analysed by gene ontology (Figure 5.2B), whereas upregulated genes showed little functional enrichment and much lower significance (Figure 5.2B).

In order to assess whether the new inducible doxycycline inducible RAD21-TEV cleavage system was comparable with the previously established 4-OHT system we compared changes in gene expression between the two systems (Figure 5.2C). A significant number of deregulated genes overlapped between the two systems (OR = 36.21, adj $P = 8.0 \times 10^{-268}$). The overlap in deregulated genes was even more pronounced when focused on downregulated genes (OR = 61.88, adj $P < 2.0 \times 10^{-307}$, Figure 5.2C).

To better understand changes in gene expression between 4-OHT and doxycycline induced RAD21-TEV cleavage in the absence of a significance threshold, we plotted genes which were significantly deregulated in one setting but not necessarily the other (Figure 5.2D, E). These data show a high correlation in changes in gene expression between 4-OHT and doxycycline induced 24H of RAD21-TEV cleavage ($r^2 = 0.76$, $p = 6.41 \times 10^{-225}$).

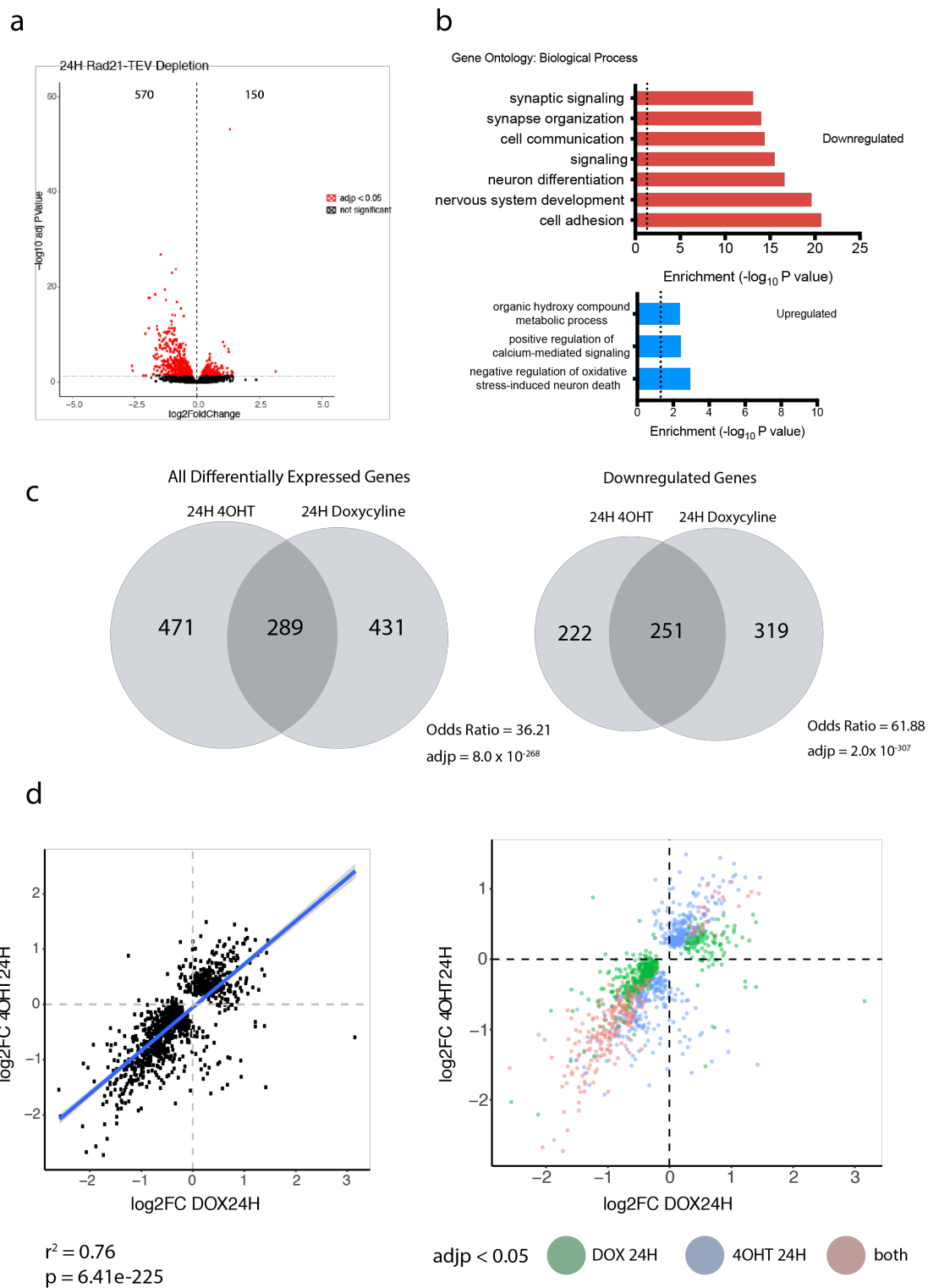


Figure 5.2 Doxycycline mediated loss of cohesin leads to changes in gene expression. (A) Volcano plot showing $\log_2$ fold change and $-\log_{10}$ adj P value of genes expressed in neurons following 24H of doxycycline mediated RAD21-TEV cleavage ($n = 3$ biological replicates). Significantly deregulated genes compared to RAD21-TEV uncleaved neurons are shown in red. **(B)** Gene ontology analysis of biological processes represented by significantly deregulated genes following doxycycline mediated RAD21-TEV cleavage. **(C)** Venn diagrams showing the overlap of significantly deregulated genes between 4-OHT and doxycycline mediated RAD21-TEV cleavage.

Figure 5.2 continued. (D) Scatter plots of genes significantly deregulated in at least one of the settings shown. There is a high degree of correlation in changes in gene expression between the two systems of RAD21-TEV cleavage (left). Genes significantly deregulated in doxycycline, 4-OHT or both systems are shown by the indicated colour (right).

5.4 Changes in gene expression following the removal and rescue of cohesin expression

As mentioned previously, following 24H of doxycycline inducible RAD21-TEV cleavage, 720 genes were significantly deregulated (170 up- and 570 downregulated, adj $p < 0.05$ based on DEseq2 analysis of 3 biological replicates, Figure 5.3A). After allowing RAD21-TEV levels to return to normal over the following 6 days, RNA-seq was performed again to see how many genes were significantly deregulated. Following cohesin rescue, the number of significantly deregulated genes dropped from 720 to 87 (44 up- and 43 downregulated, adj $p < 0.05$ based on DEseq2 analysis of 3 biological replicates, Figure 5.3A). Genes that were significantly deregulated, either up or down, after 24H RAD21-TEV cleavage were able to be rescued (Figure 5.3B), including individual genes that were highly deregulated, such as those with $\log_2FC > 3$, and $\log_2FC < -2$ (Figure 5.3B).

Of the 87 genes that were significantly deregulated after rescue, only 25 of them were initially deregulated after 24H of RAD21-TEV cleavage (Figure 5.3C). Therefore only 3.5% of genes deregulated after the loss of cohesin were unable to be rescued. Genes not rescued show no functional enrichment based on gene ontology, nor do they belong to any one pathway (Figure 5.3C). The only noticeable feature of those 25 genes is that all of them except *Rgs13*,

were downregulated following RAD21-TEV cleavage, yet were deregulated in both up and down directions following the rescue.

The data show that changes in gene expression caused by the loss of cohesin can be comprehensively rescued following the restoration of cohesin function. Genes that are both up- and downregulated can be restored, and a very small percentage persist in their deregulation. Those that do may either need more time for the rescue to take effect, or are a result of a disrupted feedback mechanism that has led to their continued deregulation.

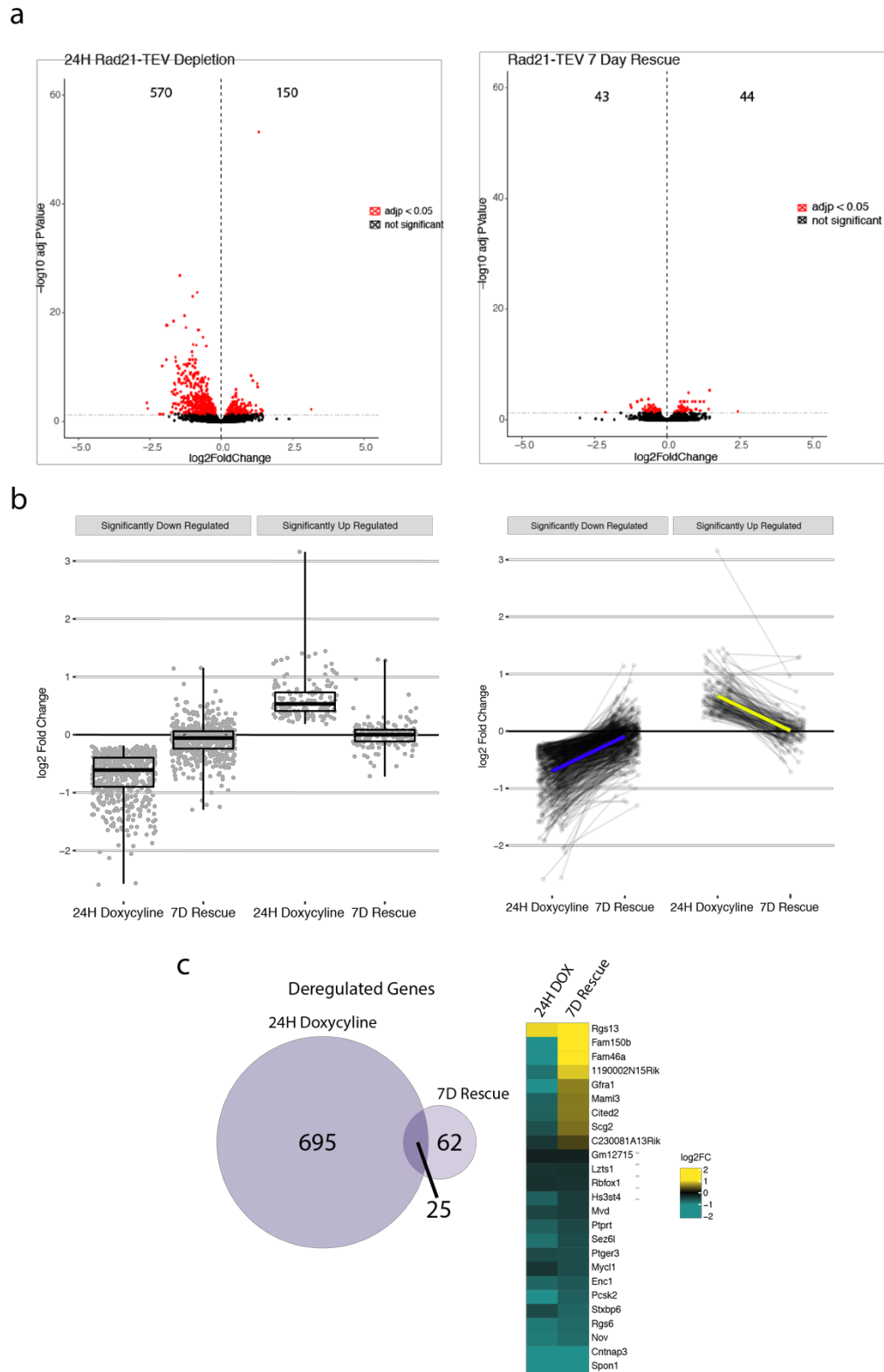


Figure 5.3 Rescue of changes in gene expression following cohesin depletion. (A) Volcano plot showing log₂ fold change and -log₁₀ adj P value of genes expressed in neurons following 24H of doxycycline mediated RAD21-TEV cleavage (left) and after rescue (right). Significantly deregulated genes are shown in red (n = 3 biological replicates).

Figure 5.3 continued. (B) Box plots of individual genes down- and upregulated following 24H doxycycline mediated RAD21-TEV cleavage. Plots show their log2 fold change and condition, bars indicate range. The solid bar in each box gives the mean log2 Fold Change of gene expression in that condition (left). Line plots track individual genes following RAD21-TEV cleavage and rescue. Blue and yellow bars indicate mean log2 Fold Change, for significantly down and upregulated genes following 24H RAD21-TEV cleavage respectively. **(C)** Venn diagram showing the overlap of significantly deregulated genes between 24H of doxycycline mediated RAD21-TEV cleavage, and rescue (left). Only 25 genes are significantly deregulated in both. Heatmap showing those 25 genes and their log2 Fold Change compared to control cells in each condition (right).

5.5 Discussion

Work in the previous chapter showed that changes in gene regulation caused by cohesin removal are highly comparable regardless of the developmental stage of depletion. Therefore, it was inviting to understand whether changes caused by 3D genome disruption through cohesin depletion could be rescued.

Whether changes in gene expression caused by cohesin deficiency could be rescued remained unknown. Here we show that transcriptional deregulation following the removal and reintroduction of cohesin can indeed be rescued. While some genes remain deregulated, and others are newly deregulated, the vast majority return to normal levels.

The RAD21-TEV system once again shows its robustness and ability to be manipulated in a temporal manner, not just in speed of cohesin depletion but also in rescue. Further adaptations of the system, such as time and concentration of doxycycline treatment, allows for the further manipulation of degree and time of cohesin depletion.

The next stage is to test whether rescue can occur in cells which suffer from a life-long deficiency of cohesin, such as those in the *Nipbl*^{+/-} mouse model of CdLS. Especially inviting is the ability to test this in vivo and extend the form of rescue not just from gene expression, but also to behavioural abnormalities of the CdLS mouse model identified by our lab and others (Fujita et al., 2017)

Chapter 6: Transcriptional changes in the neurons of Cornelia de Lange Syndrome (CdLS) patients

6.1 Introduction

Germline mutations to cohesin and associated proteins result in the developmental disease known as Cornelia de Lange Syndrome (CdLS). CdLS is a rare disease affecting between 1:10,000 and 1:30,000 live births (Kline et al., 2007). It is a multisystem disorder with physical and cognitive symptoms. Diagnosis is made based on a scoring system where cardinal features are rated higher than suggestive features. This system allows for the identification of classical CdLS and non-classical CdLS, a requirement due to the spectrum nature of CdLS. Cardinal features include synophrys, short nose, concave nasal ridge, thin upper vermillion and hand oligodactyly and/or adactyly, as well as other symptoms. Highly common suggestive features include small hands, hirsutism, growth retardation and intellectual disability (Kline et al., 2018). Patients with CdLS also suffer from other conditions as a result of their disease, such as heart malformations, epilepsy, and impaired vital organ functions (Deardorff et al., 1993). The prevalence of these however is not uniform across patients, and they are not considered when giving a diagnosis.

CdLS is most commonly caused by heterozygous or hypomorphic mutations to the *NIPBL* gene. However, mutations to core cohesin subunits such as *SMC1A*, *RAD21* and *SMC3*, as well as

the deacetylase *HDAC8*, can also give rise to CdLS (Krantz et al., 2004, Tonkin et al., 2004, Kline et al., 2018). Mutations to genes such as *ANKRD11* and *BRD4* can cause diseases that bare significant resemblance to CdLS (Olley et al., 2018, Parenti et al., 2016).

CdLS has been studied largely in animal models and cell lines from human patients. Previous work, both in this thesis and from other published papers, has identified role for cohesin in neuronal gene regulation and function in these systems (Fujita et al., 2017, Kawauchi et al., 2009, van den Berg et al., 2017, Remeseiro et al., 2012). In order to better understand the emergence and maintenance of CdLS symptoms in humans however, molecular biology techniques utilised in cell lines and animal models need to transfer to patient samples.

Utilising our ability to sort neuronal nuclei from post-mortem brain tissue, transcriptomic analysis of neurons from CdLS patients might allow us to better understand why patients with CdLS suffer from intellectual disability, and often autism. In addition, it will contribute to our understanding of the role of cohesin in gene regulation and cell function in humans *in vivo*. Finally, it will provide a comparative framework to understand how relatable animal models of CdLS are to the actual disease.

6.2 Identification of CdLS patients and samples

In order to generate meaningful transcriptomic data from human CdLS neurons, it was essential to identify the maximum number of CdLS samples available. However, due to the relative rarity of the disease, publicly available tissue samples were limited. Following

extensive searching, five CdLS patients were identified from which frozen post-mortem brain tissue was available (Table 7). Grey cortical matter from frontal lobe was requested from all patients. Patients included both males and females, and an age range from 4 to 48 years of age. Three age matched controls were obtained for each age group. Samples were obtained from the National Institutes of Health (NIH) Neurobiobank in Maryland, USA and from Prof. Ian Krantz's lab at the Children's Hospital of Philadelphia (CHOP), Pennsylvania, USA.

All patients had been diagnosed as having CdLS, and the samples from CHOP were defined as having definite, or classical, CdLS (Table 7). Clinical severity was largely unknown, apart from two samples, where one patient presented with moderate symptoms, and the other severe. In addition, genetic diagnoses had only been made on one of these patients. This 19-year-old male developed CdLS due to an insertion of an adenine between base pairs 2959 and 2960 (Table 7) of the *NIPBL* gene. This leads to a frame shift and the early termination of the NIPBL protein in translation.

Age	Sex	Mutation	Clinical Severity	Tissue bank
4	Female	unknown	CdLS - unknown	NIH
48	Female	unknown	CdLS - unknown	NIH
19	Male	2959_2960insA	Definite - severe	CHOP
24	Female	unknown	Definite - moderate	CHOP
31	Female	unknown	Definite - unknown	CHOP

Table 7. CdLS patients for neuronal analysis. Known parameters of CdLS patients from which brain tissue samples were obtained to perform neuronal nuclei RNA-seq.

6.3 Development of flow cytometry based neuronal nuclei purification

The human brain contains many different cell types including neurons, macro and micro glia, endothelial cells, and blood and lymph circulating cells. Therefore, RNA samples purified from brain without any form of separation do not allow for the attribution of changes in gene expression to specific cell types. Neurons are large myelinated cells that can span significant distances in the brain. As such, isolation by FACS, as done on other cell types from frozen human samples, is not possible. Instead, the nuclei of neurons can be isolated by flow cytometry, from which RNA can be purified, giving a representation of the transcriptomic profile of the cells (Figure 6.1A) (Krishnaswami et al., 2016).

Following tissue homogenisation and nuclei enrichment using a sucrose gradient, samples still contained a mixture of nuclei and debris. In order to separate nuclei from debris, samples were stained with DAPI, allowing for identification of nuclei (Figure 6.1B). For identification of neuronal nuclei, samples were additionally stained with a conjugated anti-NEUN antibody. In order to separate out neuronal and non-neuronal nuclei therefore, samples stained with DAPI and NEUN antibody were first gated for DAPI+, and then identified by their peaks for being NEUN- or NEUN+ (Figure 6.1B).

In order to validate that the samples determined by FACS as being NEUN- and NEUN+ were indeed non-neuronal and neuronal respectively, expression of cell type specific marker genes was quantified. Based on RNA-seq analysis, markers of astrocytes (*GFAP*), microglia (*TMEM119*) and oligodendrocytes (*NKX2.2*) were all significantly upregulated in the NEUN- samples, indicating an enrichment for glia, and therefore non-neuronal cells. Neuron specific

markers *SLC17A7* and *RBFOX3* were both significantly upregulated in the NEUN+ samples, indicating an enrichment for neurons (Figure 6.1C).

Highly pure neuronal nuclei populations can therefore be isolated from human cortical grey matter and subjected to RNA-seq using flow cytometry.

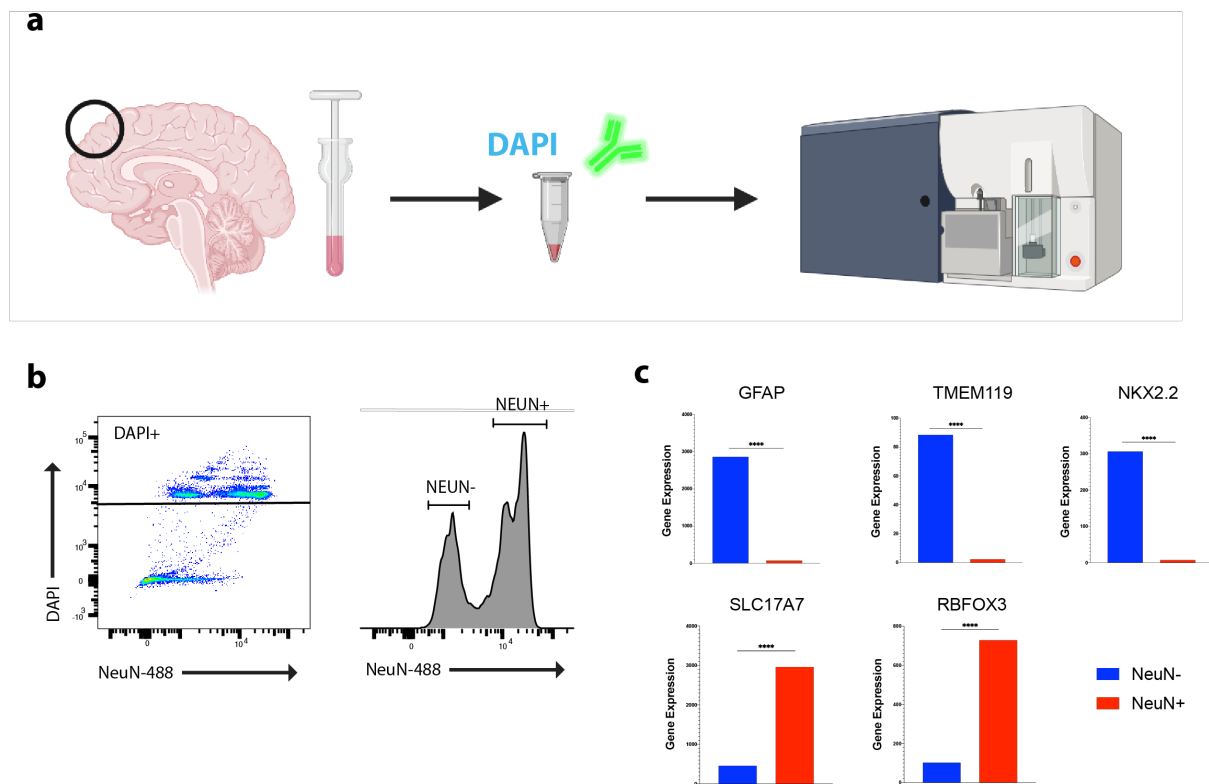


Figure 6.1. Purification and validation of neuronal nuclei isolation by flow cytometry. (A) Schematic of isolation of neuronal nuclei from human cortical grey matter. (B) Flow cytometry analysis of DAPI and NEUN-488 stained nuclei from homogenised and purified human cortical grey matter. DAPI+ nuclei are clearly separated from other debris (left). NEUN- and NEUN+ nuclei are clearly separated (right). (C) Analysis of glial and neuronal markers in NEUN- and NEUN+ nuclei. *GFAP* (astrocytes), *TMEM119* (microglia), *NKX2.2* (oligodendrocyte), and *SLC17A7* and *RBFOX3* (neurons). Based on RNA-seq from 3 biological replicates. * adj. $P < 0.05$, ** adj. $P < 0.01$, *** adj. $P < 0.001$, **** adj. $P < 0.0001$.

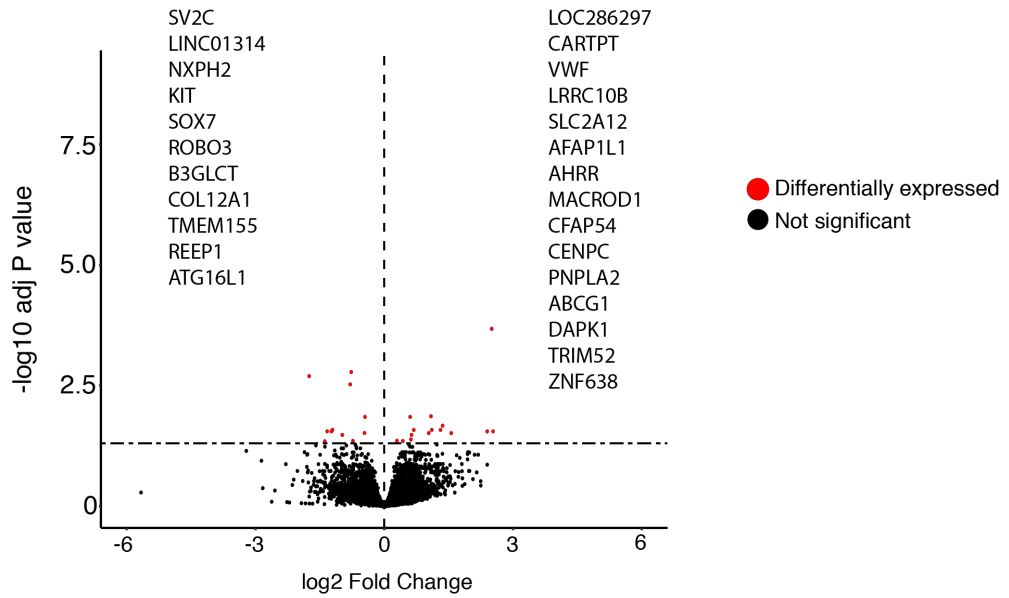
6.4 CdLS patients display abnormal neuronal gene regulation

Following purification and RNA-seq of neuronal nuclei from healthy controls and CdLS patients, changes in gene expression were measured. Overall, 26 genes were significantly differentially expressed, 11 down- and 15 upregulated (adj $P < 0.05$) between controls and CdLS patients (Figure 6.2A). Deregulated genes contribute to a variety of cellular functions including synapse formation and function (*NPXH2*, *ROBO3*, *SV2C*), extracellular matrix (*COL12A1*, *AFAP1L1*), organelle function (*REEP1*), autophagy (*ATG16L1*), and metabolism (*PNPLA2*) among others. None have been previously linked substantially to autism or cognitive impairment, both features of CdLS.

In order to identify cellular processes that are affected in CdLS patients, GSEA analysis was performed using biological processes sets from gene ontology. Biological processes associated with chromatin, including “chromatin assembly or disassembly”, “chromatin remodeling” and “DNA packaging” had highly significant positive normalized enrichment scores (NES) in CdLS neurons (Figure 6.2B). A significant contribution to the positive NESs in chromatin related functions in CdLS patients was the upregulation of histone genes, and others important for histone exchange and chromatin folding. In addition, other genes known to contribute to DNA damage and repair such as *HMGB2* and *HELLS* increased in expression in CdLS patients.

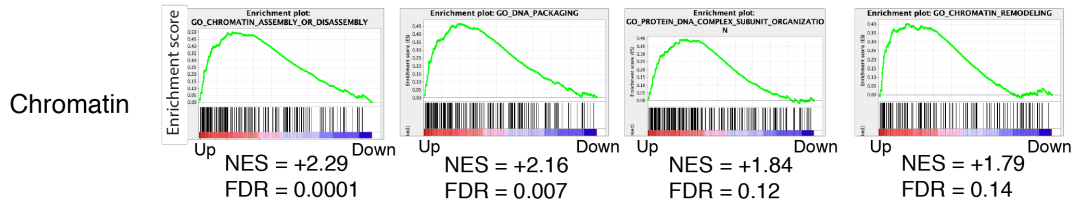
a

Cornelia de Lange Syndrome



b

Upregulated Processes



c

Downregulated Processes

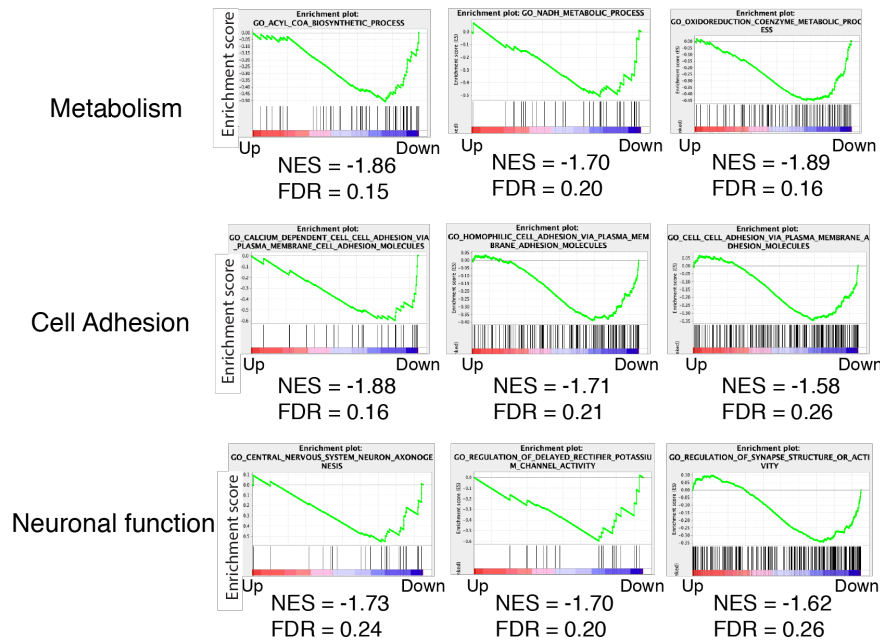


Figure 6.2 (previous page). Transcriptional deregulation in CdLS patient neurons. (A) Volcano plot showing log₂ fold change and -log₁₀ adj P value of gene expression in neurons in CdLS patients vs. healthy controls. Significantly deregulated genes (adj P < 0.05) are marked in red. Significantly deregulated genes both down (left) and up (right) are listed in descending order of strength of log₂ fold change. GSEA of gene expression in CdLS patients using gene sets derived from gene ontology. (B) GSEA of processes associated with chromatin have positive significant NESs. (C) GSEA of processes associated with metabolism, cell adhesion and neuronal function have negative NESs

Biological processes that showed the most significant negative NES covered a variety of processes including metabolism, cell adhesion and neuronal functions (Figure 6.2C). Downregulated genes contributing to the negative NESs of cell adhesion includes many protocadherin genes. These genes have been previously identified both in this thesis (Figure 4.8) and in published work as being reliant on 3D genome organization and cohesin function for correct expression (Guo et al., 2015). Genes related to neuronal function, including potassium channels, axon guidance and synapse function genes also showed downward trends (Figure 6.2C). This was of particular interest as data covered in chapter 4 of this thesis identifies cohesin as being important for cell type specific gene regulation in neurons (Figure 4.8).

6.5 Cohesin dependent gene regulation in CdLS neurons

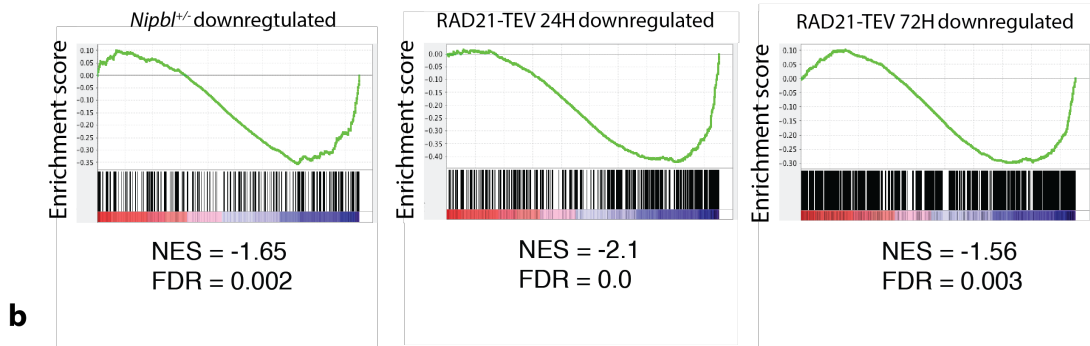
CdLS patients display deregulation of many processes based on GSEA (Figure 6.2). Several of these processes have previously been identified as being cohesin dependent in cultured *Nipbl*^{+/-} and RAD21-TEV cohesin depleted neurons.

In order to determine how previously identified cohesin dependent genes behaved in CdLS patients, we performed GSEA analysis (Figure 6.3A). Gene sets were generated from genes downregulated in *Nipbl*^{+/-} and RAD21-TEV cohesin depleted neurons, and then analysed in CdLS patients by GSEA. Gene sets of downregulated genes in *Nipbl*^{+/-}, 24-hour RAD21-TEV cohesin depleted and 72-hour RAD21-TEV cohesin depleted neurons all showed a significant downregulation in neurons from CdLS patients (Figure 6.3A). Gene sets of upregulated genes in *Nipbl*^{+/-} neurons showed a significant downregulation CdLS patient neurons (NES: 1.37, FDR: 0.012), while those upregulated in 24-hour RAD21-TEV neurons showed no significant trend in either direction. A gene set of upregulated genes in 72-hour RAD21-TEV cohesin depletion was significantly likely to be upregulated (NES: 1.2, FDR: 0.12) in CdLS patient neurons. This is partly due to the significant abundance of histone genes that are upregulated in both systems.

Previously identified cohesin dependent genes were significantly downregulated in CdLS patient neurons (Figure 6.3A). From these genes we were interested in identifying which processes were the most affected. Genes related to processes including neuronal development, synaptic function and cell communication were all downregulated in models of cohesin depletion, and these gene sets also showed a significant downregulation in CdLS patient neurons (Figure 6.3B). Other processes, such as cell adhesion, ion transport, and the MAPK cascade, also showed common directionality changes in gene expression between model systems and CdLS patients (Figure 6.3B). Processes that were downregulated in model systems but not significantly up or down regulated in CdLS samples included terms related to homeostasis, wound healing and cell death.

a

Expression in CDLS neurons:



b

Significantly overlapping GOs with CDLS neurons:

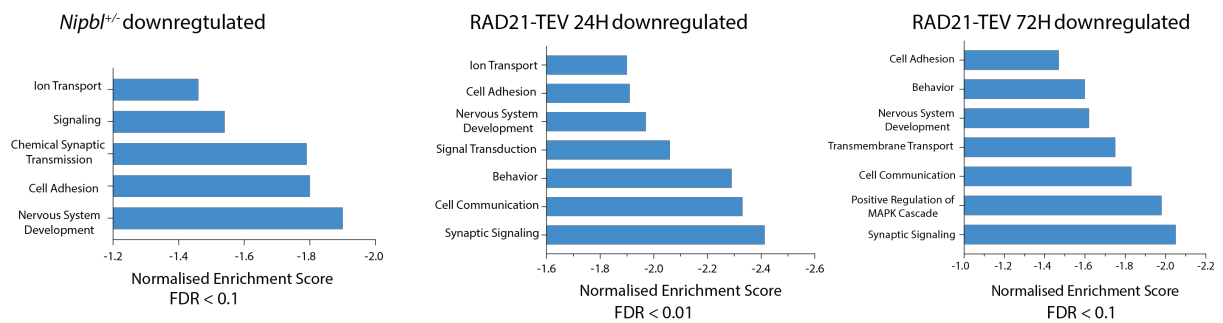


Figure 6.3 Changes in gene expression are consistent between cohesin depletion model systems and human CdLS patients (A) GSEA analysis of gene sets generated from genes downregulated in *Nipbl*^{+/−} mouse explant neurons, and RAD21-TEV 24 and 72-hour cohesin depleted explant neurons. **(B)** GSEA analysis in CdLS patient humans of gene sets generated from significantly differentially expressed genes (adj P < 0.05) in previously mentioned model cohesin depletion systems, based on biological process categories of gene ontology.

6.6 Discussion

Our understanding of the role of cohesin in neuronal gene regulation has so far only been based on data from cell lines and animal models. In addition, how cohesin deficiency leads to CdLS and its associated symptoms has not been addressed in humans. Isolation and RNA sequencing of cohesin deficient neurons from CdLS patients allows us to build on both of these topics.

RNA sequencing of human CdLS neuronal nuclei revealed a limited number of significantly deregulated genes between patients and controls. These genes covered a variety of processes including neuronal specific ones such as synaptic activity as well as fundamental cellular processes including autophagy and environment sensing. Having such a limited number of differentially expressed genes between CdLS patients and controls could partly be explained by the limited number of available CdLS samples. Human RNA-seq is often highly variable and so human neuronal nuclei RNA-seq experiments generally involve large cohorts of patients and controls, with often more than 40 samples for each group. Brains were frozen at various times post-mortem. Therefore, often the significant changes observed are not those contributing most to the disease.

Analysing expression data by GSEA allows for the identification of processes which are deregulated in disease samples compared to controls in the absence of highly stringent individual gene significance cut offs. The identification of the upregulation of processes associated with DNA packaging and chromatin had not been previously reported in cohesin deficient neurons. In addition, some upregulated genes have been associated with DNA repair. Neurons are thought to undergo DNA damage linked to transcription, and it may be that cohesin is required for the correct DNA damage response. This is partly substantiated by the recent finding that cohesin is required for correct DNA repair in B cells following V(D)J recombination.

The downregulation of cellular adhesion processes, largely due to the deregulation of protocadherin genes, supports previously identified roles of 3D genome organization and cohesin in their transcriptional control. The downregulation of processes associated with

neuronal function genes in CdLS patients reinforces the previously identified role of cohesin being important for the regulation of cell type genes in neurons. This is further backed up by data showing genes which are downregulated in culture-based models of neuronal cohesin depletion are significantly downregulated in CdLS patients, and that the most affected processes are neuron specific. Such a strong relationship in gene expression changes is remarkable given the stark differences in systems (cultured mouse neurons compared to human samples).

The data uncovered here support the previously identified role of cohesin in neuronal gene regulation, and also contribute to deregulated processes that could potentially contribute to the cognitive impairment observed in CdLS patients.

Chapter 7. General discussion and future perspectives

The field of 3D genome organisation and gene regulation has progressed rapidly over the past decade. Different structures of basic 3D chromatin architecture are determined and facilitated by a group of architectural proteins including cohesin and CTCF. The loss of each of these leads to distinct changes in at the compartment, domain and loop level (Rao et al., 2017, Schwarzer et al., 2017, Nora et al., 2017). While changes in 3D genome organisation have been observed in development and disease, its role in gene regulation remains relatively unknown. In this thesis I have attempted to elucidate the role of cohesin, and by extension the 3D genome organisation it facilitates, in transcriptional control across different cell types and disease.

Cohesin has been shown to be essential for the inflammatory response in macrophages. In the absence of a functioning cohesin complex, macrophages are unable to properly induce the inflammatory transcriptional program in response to LPS stimulation. Work so far however has been unable to elucidate whether the inflammatory response is directly regulated by cohesin, or whether changes following cohesin deficiency are a secondary effect. To identify primary and secondary targets of cohesin in the inflammatory response, I rapidly depleted macrophages of RAD21 and assayed changes in gene expression. Key upstream regulators of the inflammatory response were deregulated upon rapid cohesin depletion, suggesting they directly rely on cohesin for gene expression. Furthermore, deregulated inducible inflammatory genes following proteolytic RAD21 depletion were much more enriched for hallmarks for cohesin dependency, such as enhancer proximity and RAD21

binding, than long term cohesin depleted macrophages. Consistent with the current model of cohesin function, that it works to maintain local chromatin contacts, there was a loss of 3D genome organisation at key inflammatory regulators following proteolytic RAD21 depletion. This led to deregulated enhancer-promoter contacts and could explain the mechanism by which these genes directly rely on cohesin for expression.

Due to the high prevalence of cohesin mutations in human AML, developing our understanding of the role of cohesin in myeloid cell gene regulation is essential. This is particularly true for cohesin control of inflammatory genes. Myeloid progenitors are maintained in a balance of proliferation vs. differentiation that allows for the maintenance of the progenitor pool as well as differentiated myeloid cells such as monocytes and macrophages required for the fighting of infections and repairing of tissue damage. This balance between differentiation and proliferation is partly controlled by inflammatory genes and the ability of progenitors to respond to inflammatory signals, which triggers differentiation. In the absence of cohesin function, cells are unable to respond to inflammatory signals. While in macrophages this could impede their ability to carry out their tissue defence roles, in cohesin deficient myeloid progenitors this could lead to a change in the balance of proliferation vs. differentiation, in favour of proliferation. This would then lead to an increased number of myeloid progenitors and the development of AML, explaining the significant presence of cohesin mutations in human AML.

Future work will focus on using cohesin deficient AML cell lines and identifying small molecules that can normalise the inflammatory gene response of cohesin deficient progenitor cells. This could one day have an impact in the clinic as patients who are identified as having

cohesin mutated AML could be treated with any compounds identified as being able to normalise the proliferation and differentiation balance of cohesin deficient progenitors.

Multiple studies have shown that cohesin deficiency causes transcriptional deregulation in neurons (Kawauchi et al., 2009, Fujita et al., 2017) . However, these studies have used genetic models which feature long term cohesin deficiency. Therefore, these have failed to address whether gene deregulation in cohesin deficient neurons arises as a consequence of aberrant development or a continuous requirement for cohesin. In order to address this, I established a proteolytic cleavage system in primary mouse neurons that allows for the inducible, rapid and reversible depletion of RAD21 and therefore cohesin function.

Firstly, we identified that neurons do rely on continuous cohesin function for maintaining neuronal gene expression. Following RAD21 depletion in mature neurons, genes related to neuronal functions such as synaptic transmission and cell communication become deregulated. Along with those associated with cell adhesion and metabolism. Interestingly, we found that the same genes and functions were disrupted regardless of whether neurons are transiently depleted from cohesin for 24 hours, or are life-long cohesin deficient as in *Nipbl*^{+/-} mice that are used to model the cohesin deficient syndrome CdLS. This implies that neurons are able to form relatively normally in the absence of full cohesin function, and that the transcriptional changes observed in models of CdLS at least partly result from a continuous requirement for cohesin control of gene regulation. We then showed that deregulated genes in transiently cohesin depleted neurons were enriched for previously established markers of cohesin dependency, namely enhancer proximity and RAD21 binding (Ing-Simmons et al., 2015, Seitan et al., 2011). Furthermore, consistent with the current

model that cohesin effects gene regulation through maintaining local chromatin contacts, allowing regulatory elements such as enhancers and promoters to interact, we found a decrease in 3D genome organisation in transiently cohesin depleted neurons at the *Pcdh* locus.

We then wanted to examine whether changes in gene regulation caused by cohesin deficiency could be rescued. While it has previously been shown cohesin depletion and rescue can restore changes in 3D genome organisation, the same had not been done for gene regulation. To this end we depleted neurons of cohesin for 24 hours before allowing cohesin levels to return to normal. We found that the vast majority of genes deregulated following cohesin depletion in neurons are able to be rescued.

Proving that deregulated genes caused by cohesin deficiency can be rescued through the reintroduction of cohesin to cells has major implications. Firstly, it proves the cohesin dependency of said genes. Identification of cohesin dependency has relied on hallmarks such as enhancer proximity and RAD21 binding, but now I show that genes associated with these hallmarks are indeed cohesin dependent as proven by the rescue. Secondly, mouse models and gene regulation in neurons of other cognitive diseases such as Rett Syndrome have benefited from adult intervention and the restoring of protein function after development leading to alleviation of symptoms (Cobb et al., 2010, Guy et al., 2007). Having shown that cohesin dependent gene regulation deficiencies can be rescued in a mouse neuronal model system the question is whether the same can be done in human models of CdLS.

Future work would focus on whether CdLS patients and animal models can benefit from post-development intervention, and whether animal behavioural deficient and human neuronal gene deregulation can benefit from rescue of cohesin function. In order to address these issues, one could use models such as *NIPBL*-flex mice, which have a reversible *Nipbl* loss of function allele (Santos et al., 2016). When exposed to *cre*, the allele undergoes recombination to generate a wild-type *Nipbl* gene from a mutant. This would allow for the rescue of *Nipbl* expression, and therefore cohesin function, in neurons following development (using the *Nex-cre* line mentioned earlier) and observing whether changes in gene regulation caused by life-long cohesin deficiency can be rescued in post-mitotic neurons *in vivo*. Secondly, an inviting prospect is to see whether changes in gene regulation in human neurons can be rescued. iPSC lines from CdLS patients have been generated and can be successfully differentiated into an ectodermal lineage (Mills et al., 2018). Further differentiation into neurons followed by infection with a lentivirus containing wild-type *NIPBL*, rescuing deficient *NIPBL* expression, would allow one to assess whether changes in neuronal gene regulation in CdLS can be rescued following development.

While animal models are a valuable resource in better understanding biological mechanisms that underlie disease, where possible the use of human tissue provides much greater insight. To this end I identified a number of post-mortem brain samples from CdLS patients and controls to assess changes in CdLS patient neuronal gene regulation.

The identification of a number of significant deregulated genes between CdLS patients and controls is an important first step in understanding the potential defects in CdLS neurons. Immediate work must be to identify the mutations in the three currently unaccounted for

patients. Following that, further optimisation of RNAseq analysis is required to see whether more differentially expressed genes can be identified. Currently the samples represent a wide age range (4 to 48), and removal of the 4-year-old samples may yield different results. Further techniques such as batch correction using RUVseq may also lead to identification of more differences.

Next steps will involve identifying cohesin dependent gene regulation in CdLS patients. This could be done by overlapping data obtained from RAD21-TEV depleted neurons, and looking for previously established cohesin dependent mechanisms associated with deregulated genes, such as enhancer-promoter contacts and changes to 3D genome organisation in the patients, which have previously been identified in cell line and animal models.

Further work could also look at changes in gene regulation of CdLS patients in the brain in cells other than neurons. Neurons only make up 50% of cells in the brain, the others are a mixture of glial cells, mainly astrocytes followed by oligodendrocytes and microglia, endothelial cells and circulating blood cells. How their gene expression is altered in the CdLS could also yield some interesting results. Increased inflammatory gene expression in the glial cells of ASD patients has previously been identified, and it would be interesting to see if the same holds true here.

Overall, the data uncovered here support our analysis of cohesin as being essential for cell-type specific gene expression in neurons, but also put forward novel processes that may be implicated in the cognitive symptoms of CDLS.

Bibliography

- ANDERSON, E., DEVENNEY, P. S., HILL, R. E. & LETTICE, L. A. 2014. Mapping the Shh long-range regulatory domain. *Development*, 141, 3934-43.
- BALDRIDGE, M. T., KING, K. Y., BOLES, N. C., WEKSBERG, D. C. & GOODELL, M. A. 2010. Quiescent haematopoietic stem cells are activated by IFN-gamma in response to chronic infection. *Nature*, 465, 793-7.
- BANANI, S. F., LEE, H. O., HYMAN, A. A. & ROSEN, M. K. 2017. Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol*, 18, 285-298.
- BARTMAN, C. R., HSU, S. C., HSIUNG, C. C., RAJ, A. & BLOBEL, G. A. 2016. Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Mol Cell*, 62, 237-247.
- BEAGAN, J. A., DUONG, M. T., TITUS, K. R., ZHOU, L., CAO, Z., MA, J., LACHANSKI, C. V., GILLIS, D. R. & PHILLIPS-CREMINS, J. E. 2017. YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Res*, 27, 1139-1152.
- BHATT, D. M., PANDYA-JONES, A., TONG, A. J., BAROZZI, I., LISSNER, M. M., NATOLI, G., BLACK, D. L. & SMALE, S. T. 2012. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell*, 150, 279-90.
- BINTU, B., MATEO, L. J., SU, J. H., SINNOTT-ARMSTRONG, N. A., PARKER, M., KINROT, S., YAMAYA, K., BOETTIGER, A. N. & ZHUANG, X. 2018. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, 362.
- BOIJA, A., KLEIN, I. A., SABARI, B. R., DALL'AGNESE, A., COFFEY, E. L., ZAMUDIO, A. V., LI, C. H., SHRINIVAS, K., MANTEIGA, J. C., HANNETT, N. M., ABRAHAM, B. J., AFEYAN, L. K., GUO, Y. E., RIMEL, J. K., FANT, C. B., SCHUIJERS, J., LEE, T. I., TAATJES, D. J. & YOUNG, R. A. 2018. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell*, 175, 1842-1855 e16.
- BUHEITEL, J. & STEMMANN, O. 2013. Prophase pathway-dependent removal of cohesin from human chromosomes requires opening of the Smc3-Scc1 gate. *EMBO J*, 32, 666-76.
- CANCER GENOME ATLAS RESEARCH, N., LEY, T. J., MILLER, C., DING, L., RAPHAEL, B. J., MUNGALL, A. J., ROBERTSON, A., HOADLEY, K., TRICHE, T. J., JR., LAIRD, P. W., BATY, J. D., FULTON, L. L., FULTON, R., HEATH, S. E., KALICKI-VEIZER, J., KANDOTH, C., KLCO, J. M., KOBOLDT, D. C., KANCHI, K. L., KULKARNI, S., LAMPRECHT, T. L., LARSON, D. E., LIN, L., LU, C., MCLELLAN, M. D., MCMICHAEL, J. F., PAYTON, J., SCHMIDT, H., SPENCER, D. H., TOMASSON, M. H., WALLIS, J. W., WARTMAN, L. D., WATSON, M. A., WELCH, J., WENDL, M. C., ALLY, A., BALASUNDARAM, M., BIROL, I., BUTTERFIELD, Y., CHIU, R., CHU, A., CHUAH, E., CHUN, H. J., CORBETT, R., DHALLA, N., GUIN, R., HE, A., HIRST, C., HIRST, M., HOLT, R. A., JONES, S., KARSAN, A., LEE, D., LI, H. I., MARRA, M. A., MAYO, M., MOORE, R. A., MUNGALL, K., PARKER, J., PLEASANCE, E., PLETTNER, P., SCHEIN, J., STOLL, D., SWANSON, L., TAM, A., THIESSEN, N., VARHOL, R., WYE, N., ZHAO, Y., GABRIEL, S., GETZ, G., SOUGNEZ, C., ZOU, L., LEISERSON, M. D., VANDIN, F., WU, H. T., APPLEBAUM, F., BAYLIN, S. B., AKBANI, R., BROOM, B. M., CHEN, K., MOTTER, T. C., NGUYEN, K., WEINSTEIN, J. N., ZHANG, N., FERGUSON, M. L., ADAMS, C., BLACK, A., BOWEN, J., GASTIER-FOSTER, J., GROSSMAN, T., LICHTENBERG, T., WISE, L., DAVIDSEN, T., DEMCHOK, J. A., SHAW, K. R., SHETH, M., SOFIA, H. J., YANG, L., DOWNING, J. R., et al. 2013. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*, 368, 2059-74.

- CATARINO, R. R. & STARK, A. 2018. Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. *Genes Dev*, 32, 202-223.
- CHANG, J., ZHANG, B., HEATH, H., GALJART, N., WANG, X. & MILBRANDT, J. 2010. Nicotinamide adenine dinucleotide (NAD)-regulated DNA methylation alters CCCTC-binding factor (CTCF)/cohesin binding and transcription at the BDNF locus. *Proc Natl Acad Sci U S A*, 107, 21836-41.
- CHEN, X., BAROZZI, I., TERMANINI, A., PROSPERINI, E., RECCHIUTI, A., DALLI, J., MIETTON, F., MATTEOLI, G., HIEBERT, S. & NATOLI, G. 2012. Requirement for the histone deacetylase Hdac3 for the inflammatory gene expression program in macrophages. *Proc Natl Acad Sci U S A*, 109, E2865-74.
- CHEN, H., TIAN, Y., SHU, W., BO, X. & WANG, S. 2012. Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. *PLoS One*, 7, e41374.
- CHIEN, R., ZENG, W., KAWAUCHI, S., BENDER, M. A., SANTOS, R., GREGSON, H. C., SCHMIESING, J. A., NEWKIRK, D. A., KONG, X., BALL, A. R., JR., CALOF, A. L., LANDER, A. D., GROUDINE, M. T. & YOKOMORI, K. 2011. Cohesin mediates chromatin interactions that regulate mammalian beta-globin expression. *J Biol Chem*, 286, 17870-8.
- CHO, W. K., SPILLE, J. H., HECHT, M., LEE, C., LI, C., GRUBE, V. & CISSE, II 2018. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science*, 361, 412-415.
- CHOW, J. C., YOUNG, D. W., GOLENBOCK, D. T., CHRIST, W. J. & GUSOVSKY, F. 1999. Toll-like receptor-4 mediates lipopolysaccharide-induced signal transduction. *J Biol Chem*, 274, 10689-92.
- CITRI, A. & MALENKA, R. C. 2008. Synaptic plasticity: multiple forms, functions, and mechanisms. *Neuropsychopharmacology*, 33, 18-41.
- COBB, S., GUY, J. & BIRD, A. 2010. Reversibility of functional deficits in experimental models of Rett syndrome. *Biochem Soc Trans*, 38, 498-506.
- CONSORTIUM, E. P. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489, 57-74.
- CUARTERO, S., WEISS, F. D., DHARMALINGAM, G., GUO, Y., ING-SIMMONS, E., MASELLA, S., ROBLES-REBOLLO, I., XIAO, X., WANG, Y. F., BAROZZI, I., DJEGHLOUL, D., AMANO, M. T., NISKANEN, H., PETRETTO, E., DOWELL, R. D., TACHIBANA, K., KAIKKONEN, M. U., NASMYTH, K. A., LENHARD, B., NATOLI, G., FISHER, A. G. & MERKENSCHLAGER, M. 2018. Control of inducible gene expression links cohesin to hematopoietic progenitor self-renewal and differentiation. *Nat Immunol*, 19, 932-941.
- DEARDORFF, M. A., NOON, S. E. & KRANTZ, I. D. 1993. Cornelia de Lange Syndrome. In: ADAM, M. P., ARDINGER, H. H., PAGON, R. A., WALLACE, S. E., BEAN, L. J. H., STEPHENS, K. & AMEMIYA, A. (eds.) *GeneReviews((R))*. Seattle (WA).
- DE WIT, E., BOUWMAN, B. A., ZHU, Y., KLOUS, P., SPLINTER, E., VERSTEGEN, M. J., KRIJGER, P. H., FESTUCCIA, N., NORA, E. P., WELLING, M., HEARD, E., GEIJSEN, N., POOT, R. A., CHAMBERS, I. & DE LAAT, W. 2013. The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature*, 501, 227-31.

- DE WIT, E., VOS, E. S., HOLWERDA, S. J., VALDES-QUEZADA, C., VERSTEGEN, M. J., TEUNISSEN, H., SPLINTER, E., WIJCHERS, P. J., KRIJGER, P. H. & DE LAAT, W. 2015. CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell*, 60, 676-84.
- DEKKER, J. & HEARD, E. 2015. Structural and functional diversity of Topologically Associating Domains. *FEBS Lett*, 589, 2877-84.
- DEKKER, J. & MIRNY, L. 2016. The 3D Genome as Moderator of Chromosomal Communication. *Cell*, 164, 1110-1121.
- DEKKER, J., RIPPE, K., DEKKER, M. & KLECKNER, N. 2002. Capturing chromosome conformation. *Science*, 295, 1306-11.
- DENG, W., RUPON, J. W., KRIVEGA, I., BREDA, L., MOTTA, I., JAHN, K. S., REIK, A., GREGORY, P. D., RIVELLA, S., DEAN, A. & BLOBEL, G. A. 2014. Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell*, 158, 849-860.
- DENHOLTZ, M., BONORA, G., CHRONIS, C., SPLINTER, E., DE LAAT, W., ERNST, J., PELLEGRINI, M. & PLATH, K. 2013. Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. *Cell Stem Cell*, 13, 602-16.
- DIXON, J. R., JUNG, I., SELVARAJ, S., SHEN, Y., ANTOSIEWICZ-BOURGET, J. E., LEE, A. Y., YE, Z., KIM, A., RAJAGOPAL, N., XIE, W., DIAO, Y., LIANG, J., ZHAO, H., LOBANENKOV, V. V., ECKER, J. R., THOMSON, J. A. & REN, B. 2015. Chromatin architecture reorganization during stem cell differentiation. *Nature*, 518, 331-6.
- DIXON, J. R., SELVARAJ, S., YUE, F., KIM, A., LI, Y., SHEN, Y., HU, M., LIU, J. S. & REN, B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485, 376-80.
- DOSTIE, J., RICHMOND, T. A., ARNAOUT, R. A., SELZER, R. R., LEE, W. L., HONAN, T. A., RUBIO, E. D., KRUMM, A., LAMB, J., NUSBAUM, C., GREEN, R. D. & DEKKER, J. 2006. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res*, 16, 1299-309.
- DOWEN, J. M., FAN, Z. P., HNISZ, D., REN, G., ABRAHAM, B. J., ZHANG, L. N., WEINTRAUB, A. S., SCHUIJERS, J., LEE, T. I., ZHAO, K. & YOUNG, R. A. 2014. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell*, 159, 374-387.
- DRYDEN, N. H., BROOME, L. R., DUDBRIDGE, F., JOHNSON, N., ORR, N., SCHOENFELDER, S., NAGANO, T., ANDREWS, S., WINGETT, S., KOZAREWA, I., ASSIOTIS, I., FENWICK, K., MAGUIRE, S. L., CAMPBELL, J., NATRAJAN, R., LAMBROS, M., PERRAKIS, E., ASHWORTH, A., FRASER, P. & FLETCHER, O. 2014. Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C. *Genome Res*, 24, 1854-68.
- ENGEL, N., WEST, A. G., FELSENFELD, G. & BARTOLOMEI, M. S. 2004. Antagonism between DNA hypermethylation and enhancer-blocking activity at the H19 DMD is uncovered by CpG mutations. *Nat Genet*, 36, 883-8.
- ESSERS, M. A., OFFNER, S., BLANCO-BOSE, W. E., WAIBLER, Z., KALINKE, U., DUCHOSAL, M. A. & TRUMPP, A. 2009. IFN α activates dormant haematopoietic stem cells in vivo. *Nature*, 458, 904-8.
- FEIL, R., WAGNER, J., METZGER, D. & CHAMBON, P. 1997. Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochem Biophys Res Commun*, 237, 752-7.
- FREY, U., FREY, S., SCHOLLMEIER, F. & KRUG, M. 1996. Influence of actinomycin D, a RNA synthesis inhibitor, on long-term potentiation in rat hippocampal neurons in vivo and in vitro. *J Physiol*, 490 (Pt 3), 703-11.

- FUDENBERG, G., IMAKAEV, M., LU, C., GOLOBORODKO, A., ABDENNUR, N. & MIRNY, L. A. 2016. Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep*, 15, 2038-49.
- FUJITA, Y., MASUDA, K., BANDO, M., NAKATO, R., KATOU, Y., TANAKA, T., NAKAYAMA, M., TAKAO, K., MIYAKAWA, T., TANAKA, T., AGO, Y., HASHIMOTO, H., SHIRAHIGE, K. & YAMASHITA, T. 2017. Decreased cohesin in the brain leads to defective synapse development and anxiety-related behavior. *J Exp Med*, 214, 1431-1452.
- FURLONG, E. E. M. & LEVINE, M. 2018. Developmental enhancers and chromosome topology. *Science*, 361, 1341-1345.
- GALEEV, R., BAUDET, A., KUMAR, P., RUNDBERG NILSSON, A., NILSSON, B., SONEJI, S., TORNGREN, T., BORG, A., KVIST, A. & LARSSON, J. 2016. Genome-wide RNAi Screen Identifies Cohesin Genes as Modifiers of Renewal and Differentiation in Human HSCs. *Cell Rep*, 14, 2988-3000.
- GANDHI, R., GILLESPIE, P. J. & HIRANO, T. 2006. Human Wapl is a cohesin-binding protein that promotes sister-chromatid resolution in mitotic prophase. *Curr Biol*, 16, 2406-17.
- GASZNER, M. & FELSENFELD, G. 2006. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet*, 7, 703-13.
- GHISLETTI, S., BAROZZI, I., MIETTON, F., POLLETTI, S., DE SANTA, F., VENTURINI, E., GREGORY, L., LONIE, L., CHEW, A., WEI, C. L., RAGOUSSIS, J. & NATOLI, G. 2010. Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. *Immunity*, 32, 317-28.
- GLASS, C. K. & NATOLI, G. 2016. Molecular control of activation and priming in macrophages. *Nat Immunol*, 17, 26-33.
- GRUBER, S., ARUMUGAM, P., KATOU, Y., KUGLITSCH, D., HELMHART, W., SHIRAHIGE, K. & NASMYTH, K. 2006. Evidence that loading of cohesin onto chromosomes involves opening of its SMC hinge. *Cell*, 127, 523-37.
- GRUBER, S., HAERING, C. H. & NASMYTH, K. 2003. Chromosomal cohesin forms a ring. *Cell*, 112, 765-77.
- GUO, Y., XU, Q., CANZIO, D., SHOU, J., LI, J., GORKIN, D. U., JUNG, I., WU, H., ZHAI, Y., TANG, Y., LU, Y., WU, Y., JIA, Z., LI, W., ZHANG, M. Q., REN, B., KRAINER, A. R., MANIATIS, T. & WU, Q. 2015. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*, 162, 900-10.
- GUY, J., GAN, J., SELFRIDGE, J., COBB, S. & BIRD, A. 2007. Reversal of neurological defects in a mouse model of Rett syndrome. *Science*, 315, 1143-7.
- HABERLE, V. & STARK, A. 2018. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol*, 19, 621-637.
- HADJUR, S., WILLIAMS, L. M., RYAN, N. K., COBB, B. S., SEXTON, T., FRASER, P., FISHER, A. G. & MERKENSCHLAGER, M. 2009. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature*, 460, 410-3.
- HEIDINGER-PAULI, J. M., MERT, O., DAVENPORT, C., GUACCI, V. & KOSHLAND, D. 2010. Systematic reduction of cohesin differentially affects chromosome segregation, condensation, and DNA repair. *Curr Biol*, 20, 957-63.
- HEINZ, S., BENNER, C., SPANN, N., BERTOLINO, E., LIN, Y. C., LASLO, P., CHENG, J. X., MURRE, C., SINGH, H. & GLASS, C. K. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*, 38, 576-89.
- HEINZ, S., ROMANOSKI, C. E., BENNER, C. & GLASS, C. K. 2015. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol*, 16, 144-54.

- HNISZ, D., DAY, D. S. & YOUNG, R. A. 2016. Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. *Cell*, 167, 1188-1200.
- ING-SIMMONS, E., SEITAN, V. C., FAURE, A. J., FLICEK, P., CARROLL, T., DEKKER, J., FISHER, A. G., LENHARD, B. & MERKENSCHLAGER, M. 2015. Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin. *Genome Res*, 25, 504-13.
- JHUNJHUNWALA, S., VAN ZELM, M. C., PEAK, M. M. & MURRE, C. 2009. Chromatin architecture and the generation of antigen receptor diversity. *Cell*, 138, 435-48.
- JIANG, Y., LOH, Y. E., RAJARAJAN, P., HIRAYAMA, T., LIAO, W., KASSIM, B. S., JAVIDFAR, B., HARTLEY, B. J., KLEOFAS, L., PARK, R. B., LABONTE, B., HO, S. M., CHANDRASEKARAN, S., DO, C., RAMIREZ, B. R., PETER, C. J., C, W. J., SAFAIE, B. M., MORISHITA, H., ROUSSOS, P., NESTLER, E. J., SCHAEFER, A., TYCKO, B., BRENNAND, K. J., YAGI, T., SHEN, L. & AKBARIAN, S. 2017. The methyltransferase SETDB1 regulates a large neuron-specific topological chromatin domain. *Nat Genet*, 49, 1239-1250.
- JOO, J. Y., SCHAUKOWITCH, K., FARBIAK, L., KILARU, G. & KIM, T. K. 2016. Stimulus-specific combinatorial functionality of neuronal c-fos enhancers. *Nat Neurosci*, 19, 75-83.
- KAAIJ, L. J. T., MOHN, F., VAN DER WEIDE, R. H., DE WIT, E. & BUHLER, M. 2019. The ChAHP Complex Counteracts Chromatin Looping at CTCF Sites that Emerged from SINE Expansions in Mouse. *Cell*, 178, 1437-1451 e14.
- KAGEY, M. H., NEWMAN, J. J., BILODEAU, S., ZHAN, Y., ORLANDO, D. A., VAN BERKUM, N. L., EBMEIER, C. C., GOOSSENS, J., RAHL, P. B., LEVINE, S. S., TAATJES, D. J., DEKKER, J. & YOUNG, R. A. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, 467, 430-5.
- KAIKKONEN, M. U., SPANN, N. J., HEINZ, S., ROMANOSKI, C. E., ALLISON, K. A., STENDER, J. D., CHUN, H. B., TOUGH, D. F., PRINJHA, R. K., BENNER, C. & GLASS, C. K. 2013. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol Cell*, 51, 310-25.
- KAWAUCHI, S., CALOF, A. L., SANTOS, R., LOPEZ-BURKS, M. E., YOUNG, C. M., HOANG, M. P., CHUA, A., LAO, T., LECHNER, M. S., DANIEL, J. A., NUSSENZWEIG, A., KITZES, L., YOKOMORI, K., HALLGRIMSSON, B. & LANDER, A. D. 2009. Multiple organ system defects and transcriptional dysregulation in the Nipbl(+/-) mouse, a model of Cornelia de Lange Syndrome. *PLoS Genet*, 5, e1000650.
- KIM, J. H., REGE, M., VALERI, J., DUNAGIN, M. C., METZGER, A., TITUS, K. R., GILGENAST, T. G., GONG, W., BEAGAN, J. A., RAJ, A. & PHILLIPS-CREMINS, J. E. 2019. LADL: light-activated dynamic looping for endogenous gene expression control. *Nat Methods*, 16, 633-639.
- KIM, S. & SHENDURE, J. 2019. Mechanisms of Interplay between Transcription Factors and the 3D Genome. *Mol Cell*.
- KIM, T. K., HEMBERG, M., GRAY, J. M., COSTA, A. M., BEAR, D. M., WU, J., HARMIN, D. A., LAPTEWICZ, M., BARBARA-HALEY, K., KUERSTEN, S., MARKENSCOFF-PAPADIMITRIOU, E., KUHL, D., BITO, H., WORLEY, P. F., KREIMAN, G. & GREENBERG, M. E. 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature*, 465, 182-7.
- KLINE, A. D., GRADOS, M., SPONSELLER, P., LEVY, H. P., BLAGOWIDOW, N., SCHOEDEL, C., RAMPOLLA, J., CLEMENS, D. K., KRANTZ, I., KIMBALL, A., PICHARD, C. & TUCHMAN, D. 2007. Natural history of aging in Cornelia de Lange syndrome. *Am J Med Genet C Semin Med Genet*, 145C, 248-60.
- KLINE, A. D., MOSS, J. F., SELICORNI, A., BISGAARD, A. M., DEARDORFF, M. A., GILLET, P. M., ISHMAN, S. L., KERR, L. M., LEVIN, A. V., MULDER, P. A., RAMOS, F. J., WIERZBA, J.,

- AJMONE, P. F., AXTELL, D., BLAGOWIDOW, N., CEREDA, A., COSTANTINO, A., CORMIER-DAIRE, V., FITZPATRICK, D., GRADOS, M., GROVES, L., GUTHRIE, W., HUISMAN, S., KAISER, F. J., KOEKKOEK, G., LEVIS, M., MARIANI, M., MCCLEERY, J. P., MENKE, L. A., METRENA, A., O'CONNOR, J., OLIVER, C., PIE, J., PIENING, S., POTTER, C. J., QUAGLIO, A. L., REDEKER, E., RICHMAN, D., RIGAMONTI, C., SHI, A., TUMER, Z., VAN BALKOM, I. D. C. & HENNEKAM, R. C. 2018. Diagnosis and management of Cornelia de Lange syndrome: first international consensus statement. *Nat Rev Genet*, 19, 649-666.
- KOJIC, A., CUADRADO, A., DE KONINCK, M., GIMENEZ-LLORENTE, D., RODRIGUEZ-CORSINO, M., GOMEZ-LOPEZ, G., LE DILY, F., MARTI-RENOM, M. A. & LOSADA, A. 2018. Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization. *Nat Struct Mol Biol*, 25, 496-504.
- KON, A., SHIH, L. Y., MINAMINO, M., SANADA, M., SHIRAISHI, Y., NAGATA, Y., YOSHIDA, K., OKUNO, Y., BANDO, M., NAKATO, R., ISHIKAWA, S., SATO-OTSUBO, A., NAGAE, G., NISHIMOTO, A., HAFERLACH, C., NOWAK, D., SATO, Y., ALPERMANN, T., NAGASAKI, M., SHIMAMURA, T., TANAKA, H., CHIBA, K., YAMAMOTO, R., YAMAGUCHI, T., OTSU, M., OBARA, N., SAKATA-YANAGIMOTO, M., NAKAMAKI, T., ISHIYAMA, K., NOLTE, F., HOFMANN, W. K., MIYAWAKI, S., CHIBA, S., MORI, H., NAKAUCHI, H., KOEFFLER, H. P., ABURATANI, H., HAFERLACH, T., SHIRAHIGE, K., MIYANO, S. & OGAWA, S. 2013. Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. *Nat Genet*, 45, 1232-7.
- KRANTZ, I. D., MCCALLUM, J., DESCIPIO, C., KAUR, M., GILLIS, L. A., YAEGER, D., JUKOFSKY, L., WASSERMAN, N., BOTTANI, A., MORRIS, C. A., NOWACZYK, M. J., TORIELLO, H., BAMSHAD, M. J., CAREY, J. C., RAPPAPORT, E., KAWAUCHI, S., LANDER, A. D., CALOF, A. L., LI, H. H., DEVOTO, M. & JACKSON, L. G. 2004. Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of *Drosophila melanogaster* Nipped-B. *Nat Genet*, 36, 631-5.
- KRISHNASWAMI, S. R., GRINDBERG, R. V., NOVOTNY, M., VENEPALLY, P., LACAR, B., BHUTANI, K., LINKER, S. B., PHAM, S., ERWIN, J. A., MILLER, J. A., HODGE, R., MCCARTHY, J. K., KELDER, M., MCCORRISON, J., AEVERMANN, B. D., FUERTES, F. D., SCHEUERMANN, R. H., LEE, J., LEIN, E. S., SCHORK, N., MCCONNELL, M. J., GAGE, F. H. & LASKEN, R. S. 2016. Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat Protoc*, 11, 499-524.
- KUENG, S., HEGEMANN, B., PETERS, B. H., LIPP, J. J., SCHLEIFFER, A., MECHTLER, K. & PETERS, J. M. 2006. Wapl controls the dynamic association of cohesin with chromatin. *Cell*, 127, 955-67.
- LAI, A. Y., FATEMI, M., DHASARATHY, A., MALONE, C., SOBOL, S. E., GEIGERMAN, C., JAYE, D. L., MAV, D., SHAH, R., LI, L. & WADE, P. A. 2010. DNA methylation prevents CTCF-mediated silencing of the oncogene BCL6 in B cell lymphomas. *J Exp Med*, 207, 1939-50.
- LEVINE, M. 2010. Transcriptional enhancers in animal development and evolution. *Curr Biol*, 20, R754-63.
- LIEBERMAN-AIDEN, E., VAN BERKUM, N. L., WILLIAMS, L., IMAKAEV, M., RAGOCZY, T., TELLING, A., AMIT, I., LAJOIE, B. R., SABO, P. J., DORSCHNER, M. O., SANDSTROM, R., BERNSTEIN, B., BENDER, M. A., GROUDINE, M., GNIRKE, A., STAMATOYANNOPOULOS, J., MIRNY, L. A., LANDER, E. S. & DEKKER, J. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326, 289-93.

- LAVAGNOLLI, T., GUPTA, P., HORMANSEDER, E., MIRA-BONTENBAL, H., DHARMALINGAM, G., CARROLL, T., GURDON, J. B., FISHER, A. G. & MERKENSCHLAGER, M. 2015. Initiation and maintenance of pluripotency gene expression in the absence of cohesin. *Genes Dev*, 29, 23-38.
- LONG, H. K., PRESCOTT, S. L. & WYSOCKA, J. 2016. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell*, 167, 1170-1187.
- LOVE, P. E., WARZECHA, C. & LI, L. 2014. Ldb1 complexes: the new master regulators of erythroid gene transcription. *Trends Genet*, 30, 1-9.
- LUPIANEZ, D. G., KRAFT, K., HEINRICH, V., KRAWITZ, P., BRANCATI, F., KLOPOCKI, E., HORN, D., KAYSERILI, H., OPITZ, J. M., LAXOVA, R., SANTOS-SIMARRO, F., GILBERT-DUSSARDIER, B., WITTLER, L., BORSCHIWER, M., HAAS, S. A., OSTERWALDER, M., FRANKE, M., TIMMERMAN, B., HECHT, J., SPIELMANN, M., VISEL, A. & MUNDLOS, S. 2015. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, 161, 1012-1025.
- MALIK, A. N., VIERBUCHEN, T., HEMBERG, M., RUBIN, A. A., LING, E., COUCH, C. H., STROUD, H., SPIEGEL, I., FARH, K. K., HARMIN, D. A. & GREENBERG, M. E. 2014. Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. *Nat Neurosci*, 17, 1330-9.
- MAZUMDAR, C., SHEN, Y., XAVY, S., ZHAO, F., REINISCH, A., LI, R., CORCES, M. R., FLYNN, R. A., BUENROSTRO, J. D., CHAN, S. M., THOMAS, D., KOENIG, J. L., HONG, W. J., CHANG, H. Y. & MAJETI, R. 2015. Leukemia-Associated Cohesin Mutants Dominantly Enforce Stem Cell Programs and Impair Human Hematopoietic Progenitor Differentiation. *Cell Stem Cell*, 17, 675-688.
- MEDZHITOV, R. & HORNG, T. 2009. Transcriptional control of the inflammatory response. *Nat Rev Immunol*, 9, 692-703.
- MERKENSCHLAGER, M. & NORA, E. P. 2016. CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annu Rev Genomics Hum Genet*, 17, 17-43.
- MOSSADEGH-KELLER, N., SARRAZIN, S., KANDALLA, P. K., ESPINOSA, L., STANLEY, E. R., NUTT, S. L., MOORE, J. & SIEWEKE, M. H. 2013. M-CSF instructs myeloid lineage fate in single haematopoietic stem cells. *Nature*, 497, 239-43.
- MULLENDERS, J., ARANDA-ORGILLES, B., LHOUMAUD, P., KELLER, M., PAE, J., WANG, K., KAYEMBE, C., ROCHA, P. P., RAVIRAM, R., GONG, Y., PREMSRIRUT, P. K., TSIRIGOS, A., BONNEAU, R., SKOK, J. A., CIMMINO, L., HOEHN, D. & AIFANTIS, I. 2015. Cohesin loss alters adult hematopoietic stem cell homeostasis, leading to myeloproliferative neoplasms. *J Exp Med*, 212, 1833-50.
- NAGAI, Y., GARRETT, K. P., OHTA, S., BAHRUN, U., KOURO, T., AKIRA, S., TAKATSU, K. & KINCADE, P. W. 2006. Toll-like receptors on hematopoietic progenitor cells stimulate innate immune system replenishment. *Immunity*, 24, 801-12.
- NAIR, S. J., YANG, L., MELUZZI, D., OH, S., YANG, F., FRIEDMAN, M. J., WANG, S., SUTER, T., ALSHAREEDAH, I., GAMLIEL, A., MA, Q., ZHANG, J., HU, Y., TAN, Y., OHGI, K. A., JAYANI, R. S., BANERJEE, P. R., AGGARWAL, A. K. & ROSENFELD, M. G. 2019. Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly. *Nat Struct Mol Biol*, 26, 193-203.
- NARENDRA, V., ROCHA, P. P., AN, D., RAVIRAM, R., SKOK, J. A., MAZZONI, E. O. & REINBERG, D. 2015. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science*, 347, 1017-21.

- NASMYTH, K. & HAERING, C. H. 2009. Cohesin: its roles and mechanisms. *Annu Rev Genet*, 43, 525-58.
- NATSUME, T., KIYOMITSU, T., SAGA, Y. & KANEMAKI, M. T. 2016. Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. *Cell Rep*, 15, 210-218.
- NGUYEN, P. V., ABEL, T. & KANDEL, E. R. 1994. Requirement of a critical period of transcription for induction of a late phase of LTP. *Science*, 265, 1104-7.
- NITZSCHE, A., PASZKOWSKI-ROGACZ, M., MATARESE, F., JANSSEN-MEGENS, E. M., HUBNER, N. C., SCHULZ, H., DE VRIES, I., DING, L., HUEBNER, N., MANN, M., STUNNENBERG, H. G. & BUCHHOLZ, F. 2011. RAD21 cooperates with pluripotency transcription factors in the maintenance of embryonic stem cell identity. *PLoS One*, 6, e19470.
- NORA, E. P., GOLOBORODKO, A., VALTON, A. L., GIBBUS, J. H., UEBERSOHN, A., ABDENNUR, N., DEKKER, J., MIRNY, L. A. & BRUNEAU, B. G. 2017. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*, 169, 930-944 e22.
- NORA, E. P., LAJOIE, B. R., SCHULZ, E. G., GIORGETTI, L., OKAMOTO, I., SERVANT, N., PIOLOT, T., VAN BERKUM, N. L., MEISIG, J., SEDAT, J., GRIBNAU, J., BARILLOT, E., BLUTHGEN, N., DEKKER, J. & HEARD, E. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 485, 381-5.
- NORD, A. S., BLOW, M. J., ATTANASIO, C., AKIYAMA, J. A., HOLT, A., HOSSEINI, R., PHOUANENAVONG, S., PLAJSER-FRICK, I., SHOUKRY, M., AFZAL, V., RUBENSTEIN, J. L., RUBIN, E. M., PENNACCHIO, L. A. & VISEL, A. 2013. Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell*, 155, 1521-31.
- OHLSSON, R., RENKAWITZ, R. & LOBANENKOV, V. 2001. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet*, 17, 520-7.
- OLIVEIRA, R. A., HAMILTON, R. S., PAULI, A., DAVIS, I. & NASMYTH, K. 2010. Cohesin cleavage and Cdk inhibition trigger formation of daughter nuclei. *Nat Cell Biol*, 12, 185-92.
- OLLEY, G., ANSARI, M., BENGANI, H., GRIMES, G. R., RHODES, J., VON KRIEGSHEIM, A., BLATNIK, A., STEWART, F. J., WAKELING, E., CARROLL, N., ROSS, A., PARK, S. M., DECIPHERING DEVELOPMENTAL DISORDERS, S., BICKMORE, W. A., PRADEEPA, M. M. & FITZPATRICK, D. R. 2018. BRD4 interacts with NIPBL and BRD4 is mutated in a Cornelia de Lange-like syndrome. *Nat Genet*, 50, 329-332.
- ONG, C. T. & CORCES, V. G. 2014. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet*, 15, 234-46.
- OSTERWALDER, M., BAROZZI, I., TISSIERES, V., FUKUDA-YUZAWA, Y., MANNION, B. J., AFZAL, S. Y., LEE, E. A., ZHU, Y., PLAJSER-FRICK, I., PICKLE, C. S., KATO, M., GARVIN, T. H., PHAM, Q. T., HARRINGTON, A. N., AKIYAMA, J. A., AFZAL, V., LOPEZ-RIOS, J., DICKEL, D. E., VISEL, A. & PENNACCHIO, L. A. 2018. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature*, 554, 239-243.
- OSTUNI, R., PICCOLO, V., BAROZZI, I., POLLETTI, S., TERMANINI, A., BONIFACIO, S., CURINA, A., PROSPERINI, E., GHISLETTI, S. & NATOLI, G. 2013. Latent enhancers activated by stimulation in differentiated cells. *Cell*, 152, 157-71.
- PALSTRA, R. J., TOLHUIS, B., SPLINTER, E., NIJMEIJER, R., GROSVELD, F. & DE LAAT, W. 2003. The beta-globin nuclear compartment in development and erythroid differentiation. *Nat Genet*, 35, 190-4.
- PARELHO, V., HADJUR, S., SPIVAKOV, M., LELEU, M., SAUER, S., GREGSON, H. C., JARMUZ, A., CANZONETTA, C., WEBSTER, Z., NESTEROVA, T., COBB, B. S., YOKOMORI, K., DILLON,

- N., ARAGON, L., FISHER, A. G. & MERKENSCHLAGER, M. 2008. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell*, 132, 422-33.
- PARENTI, I., GERVASINI, C., POZOJEVIC, J., GRAUL-NEUMANN, L., AZZOLLINI, J., BRAUNHOLZ, D., WATRIN, E., WENDT, K. S., CEREDA, A., CITTARO, D., GILLESSEN-KAESBACH, G., LAZAREVIC, D., MARIANI, M., RUSSO, S., WERNER, R., KRAWITZ, P., LARIZZA, L., SELICORNI, A. & KAISER, F. J. 2016. Broadening of cohesinopathies: exome sequencing identifies mutations in ANKRD11 in two patients with Cornelia de Lange-overlapping phenotype. *Clin Genet*, 89, 74-81.
- PAULI, A., ALTHOFF, F., OLIVEIRA, R. A., HEIDMANN, S., SCHULDINER, O., LEHNER, C. F., DICKSON, B. J. & NASMYTH, K. 2008. Cell-type-specific TEV protease cleavage reveals cohesin functions in *Drosophila* neurons. *Dev Cell*, 14, 239-51.
- PERRY, M. W., BOETTIGER, A. N. & LEVINE, M. 2011. Multiple enhancers ensure precision of gap gene-expression patterns in the *Drosophila* embryo. *Proc Natl Acad Sci U S A*, 108, 13570-5.
- RAO, S. S., HUNTLEY, M. H., DURAND, N. C., STAMENOVA, E. K., BOCHKOV, I. D., ROBINSON, J. T., SANBORN, A. L., MACHOL, I., OMER, A. D., LANDER, E. S. & AIDEN, E. L. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159, 1665-80.
- RAO, S. S. P., HUANG, S. C., GLENN ST HILAIRE, B., ENGREITZ, J. M., PEREZ, E. M., KIEFFER-KWON, K. R., SANBORN, A. L., JOHNSTONE, S. E., BASCOM, G. D., BOCHKOV, I. D., HUANG, X., SHAMIM, M. S., SHIN, J., TURNER, D., YE, Z., OMER, A. D., ROBINSON, J. T., SCHLICK, T., BERNSTEIN, B. E., CASELLAS, R., LANDER, E. S. & AIDEN, E. L. 2017. Cohesin Loss Eliminates All Loop Domains. *Cell*, 171, 305-320 e24.
- REMESEIRO, S., CUADRADO, A., GOMEZ-LOPEZ, G., PISANO, D. G. & LOSADA, A. 2012. A unique role of cohesin-SA1 in gene regulation and development. *EMBO J*, 31, 2090-102.
- ROWLEY, M. J. & CORCES, V. G. 2018. Organizational principles of 3D genome architecture. *Nat Rev Genet*, 19, 789-800.
- RUBIO, E. D., REISS, D. J., WELCSH, P. L., DISTECHE, C. M., FILIPPOVA, G. N., BALIGA, N. S., AEBERSOLD, R., RANISH, J. A. & KRUMM, A. 2008. CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci U S A*, 105, 8309-14.
- SCHAFER, D. P., LEHRMAN, E. K., KAUTZMAN, A. G., KOYAMA, R., MARDINLY, A. R., YAMASAKI, R., RANSOHOFF, R. M., GREENBERG, M. E., BARRES, B. A. & STEVENS, B. 2012. Microglia sculpt postnatal neural circuits in an activity and complement-dependent manner. *Neuron*, 74, 691-705.
- SCHMITT, A. D., HU, M. & REN, B. 2016. Genome-wide mapping and analysis of chromosome architecture. *Nat Rev Mol Cell Biol*, 17, 743-755.
- SCHOENFELDER, S. & FRASER, P. 2019. Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet*.
- SCHOENFELDER, S., SEXTON, T., CHAKALOVA, L., COPE, N. F., HORTON, A., ANDREWS, S., KURUKUTI, S., MITCHELL, J. A., UMLAUF, D., DIMITROVA, D. S., ESKIW, C. H., LUO, Y., WEI, C. L., RUAN, Y., BIEKER, J. J. & FRASER, P. 2010. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet*, 42, 53-61.
- SCHRIER, S. A., SHERER, I., DEARDORFF, M. A., CLARK, D., AUDETTE, L., GILLIS, L., KLINE, A. D., ERNST, L., LOOMES, K., KRANTZ, I. D. & JACKSON, L. G. 2011. Causes of death and

- autopsy findings in a large study cohort of individuals with Cornelia de Lange syndrome and review of the literature. *Am J Med Genet A*, 155A, 3007-24.
- SCHWARZER, W., ABDENNUR, N., GOLOBORODKO, A., PEKOWSKA, A., FUDENBERG, G., LOE-MIE, Y., FONSECA, N. A., HUBER, W., HAERING, C. H., MIRNY, L. & SPITZ, F. 2017. Two independent modes of chromatin organization revealed by cohesin removal. *Nature*, 551, 51-56.
- SEITAN, V. C., HAO, B., TACHIBANA-KONWALSKI, K., LAVAGNOLLI, T., MIRA-BONTENBAL, H., BROWN, K. E., TENG, G., CARROLL, T., TERRY, A., HORAN, K., MARKS, H., ADAMS, D. J., SCHATZ, D. G., ARAGON, L., FISHER, A. G., KRANGEL, M. S., NASMYTH, K. & MERKENSCHLAGER, M. 2011. A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. *Nature*, 476, 467-71.
- SEXTON, T., YAFFE, E., KENIGSBERG, E., BANTIGNIES, F., LEBLANC, B., HOICHMAN, M., PARRINELLO, H., TANAY, A. & CAVALLI, G. 2012. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, 148, 458-72.
- STADHOUDERS, R., FILION, G. J. & GRAF, T. 2019. Transcription factors and 3D genome conformation in cell-fate decisions. *Nature*, 569, 345-354.
- STANHOPE-BAKER, P., HUDSON, K. M., SHAFFER, A. L., CONSTANTINESCU, A. & SCHLISSEL, M. S. 1996. Cell type-specific chromatin structure determines the targeting of V(D)J recombinase activity in vitro. *Cell*, 85, 887-97.
- STROM, A. R., EMEYANOV, A. V., MIR, M., FYODOROV, D. V., DARZACQ, X. & KARPEN, G. H. 2017. Phase separation drives heterochromatin domain formation. *Nature*, 547, 241-245.
- TACHIBANA-KONWALSKI, K., GODWIN, J., VAN DER WEYDEN, L., CHAMPION, L., KUDO, N. R., ADAMS, D. J. & NASMYTH, K. 2010. Rec8-containing cohesin maintains bivalents without turnover during the growing phase of mouse oocytes. *Genes Dev*, 24, 2505-16.
- TERAKAWA, T., BISHT, S., EEFTENS, J. M., DEKKER, C., HAERING, C. H. & GREENE, E. C. 2017. The condensin complex is a mechanochemical motor that translocates along DNA. *Science*, 358, 672-676.
- THOTA, S., VINY, A. D., MAKISHIMA, H., SPITZER, B., RADIVOYEVITCH, T., PRZYCHODZEN, B., SEKERES, M. A., LEVINE, R. L. & MACIEJEWSKI, J. P. 2014. Genetic alterations of the cohesin complex genes in myeloid malignancies. *Blood*, 124, 1790-8.
- TOLHUIS, B., PALSTRA, R. J., SPLINTER, E., GROSVELD, F. & DE LAAT, W. 2002. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell*, 10, 1453-65.
- TONKIN, E. T., WANG, T. J., LISGO, S., BAMSHAD, M. J. & STRACHAN, T. 2004. NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. *Nat Genet*, 36, 636-41.
- UHLMANN, F., WERNIC, D., POUPART, M. A., KOONIN, E. V. & NASMYTH, K. 2000. Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast. *Cell*, 103, 375-86.
- VAN DEN BERG, D. L. C., AZZARELLI, R., OISHI, K., MARTYNOGA, B., URBAN, N., DEKKERS, D. H. W., DEMMERS, J. A. & GUILLEMOT, F. 2017. Nipbl Interacts with Zfp609 and the Integrator Complex to Regulate Cortical Neuron Migration. *Neuron*, 93, 348-361.
- AIMONE, J. B., LI, Y., LEE, S. W., CLEMENSON, G. D., DENG, W. & GAGE, F. H. 2014. Regulation and function of adult neurogenesis: from genes to cognition. *Physiol Rev*, 94, 991-1026.

- ANDERSON, E., DEVENNEY, P. S., HILL, R. E. & LETTICE, L. A. 2014. Mapping the Shh long-range regulatory domain. *Development*, 141, 3934-43.
- BALDRIDGE, M. T., KING, K. Y., BOLES, N. C., WEKSBERG, D. C. & GOODELL, M. A. 2010. Quiescent haematopoietic stem cells are activated by IFN-gamma in response to chronic infection. *Nature*, 465, 793-7.
- BANANI, S. F., LEE, H. O., HYMAN, A. A. & ROSEN, M. K. 2017. Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol*, 18, 285-298.
- BARTMAN, C. R., HSU, S. C., HSIUNG, C. C., RAJ, A. & BLOBEL, G. A. 2016. Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Mol Cell*, 62, 237-247.
- BEAGAN, J. A., DUONG, M. T., TITUS, K. R., ZHOU, L., CAO, Z., MA, J., LACHANSKI, C. V., GILLIS, D. R. & PHILLIPS-CREMINS, J. E. 2017. YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Res*, 27, 1139-1152.
- BHATT, D. M., PANDYA-JONES, A., TONG, A. J., BAROZZI, I., LISSNER, M. M., NATOLI, G., BLACK, D. L. & SMALE, S. T. 2012. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell*, 150, 279-90.
- BINTU, B., MATEO, L. J., SU, J. H., SINNOTT-ARMSTRONG, N. A., PARKER, M., KINROT, S., YAMAYA, K., BOETTIGER, A. N. & ZHUANG, X. 2018. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, 362.
- BOIJA, A., KLEIN, I. A., SABARI, B. R., DALL'AGNESE, A., COFFEY, E. L., ZAMUDIO, A. V., LI, C. H., SHRINIVAS, K., MANTEIGA, J. C., HANNETT, N. M., ABRAHAM, B. J., AFEYAN, L. K., GUO, Y. E., RIMEL, J. K., FANT, C. B., SCHUIJERS, J., LEE, T. I., TAATJES, D. J. & YOUNG, R. A. 2018. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell*, 175, 1842-1855 e16.
- BOTTAI, D., SPREAFICO, M., PISTOCCHI, A., FAZIO, G., ADAMI, R., GRAZIOLI, P., CANU, A., BRAGATO, C., RIGAMONTI, S., PARODI, C., CAZZANIGA, G., BIONDI, A., COTELLI, F., SELICORNI, A. & MASSA, V. 2019. Modeling Cornelia de Lange syndrome in vitro and in vivo reveals a role for cohesin complex in neuronal survival and differentiation. *Hum Mol Genet*, 28, 64-73.
- BUHEITEL, J. & STEMMANN, O. 2013. Prophase pathway-dependent removal of cohesin from human chromosomes requires opening of the Smc3-Scc1 gate. *EMBO J*, 32, 666-76.
- CANCER GENOME ATLAS RESEARCH, N., LEY, T. J., MILLER, C., DING, L., RAPHAEL, B. J., MUNGALL, A. J., ROBERTSON, A., HOADLEY, K., TRICHE, T. J., JR., LAIRD, P. W., BATY, J. D., FULTON, L. L., FULTON, R., HEATH, S. E., KALICKI-VEIZER, J., KANDOTH, C., KLCO, J. M., KOBOLDT, D. C., KANCHI, K. L., KULKARNI, S., LAMPRECHT, T. L., LARSON, D. E., LIN, L., LU, C., MCLELLAN, M. D., MCMICHAEL, J. F., PAYTON, J., SCHMIDT, H., SPENCER, D. H., TOMASSON, M. H., WALLIS, J. W., WARTMAN, L. D., WATSON, M. A., WELCH, J., WENDL, M. C., ALLY, A., BALASUNDARAM, M., BIROL, I., BUTTERFIELD, Y., CHIU, R., CHU, A., CHUAH, E., CHUN, H. J., CORBETT, R., DHALLA, N., GUIN, R., HE, A., HIRST, C., HIRST, M., HOLT, R. A., JONES, S., KARSAN, A., LEE, D., LI, H. I., MARRA, M. A., MAYO, M., MOORE, R. A., MUNGALL, K., PARKER, J., PLEASANCE, E., PLETTNER, P., SCHEIN, J., STOLL, D., SWANSON, L., TAM, A., THIESSEN, N., VARHOL, R., WYE, N., ZHAO, Y., GABRIEL, S., GETZ, G., SOUGNEZ, C., ZOU, L., LEISERSON, M. D., VANDIN, F., WU, H. T., APPLEBAUM, F., BAYLIN, S. B., AKBANI, R., BROOM, B. M., CHEN, K., MOTTER, T. C., NGUYEN, K., WEINSTEIN, J. N., ZHANG, N., FERGUSON, M. L., ADAMS, C., BLACK, A., BOWEN, J., GASTIER-FOSTER, J., GROSSMAN, T., LICHTENBERG, T., WISE, L., DAVIDSEN, T., DEMCHOK, J. A., SHAW, K. R., SHETH, M., SOFIA, H. J., YANG, L.,

- DOWNING, J. R., et al. 2013. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*, 368, 2059-74.
- CAO, J., WU, L., ZHANG, S. M., LU, M., CHEUNG, W. K., CAI, W., GALE, M., XU, Q. & YAN, Q. 2016. An easy and efficient inducible CRISPR/Cas9 platform with improved specificity for multiple gene targeting. *Nucleic Acids Res*, 44, e149.
- CATARINO, R. R. & STARK, A. 2018. Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. *Genes Dev*, 32, 202-223.
- CHANG, J., ZHANG, B., HEATH, H., GALJART, N., WANG, X. & MILBRANDT, J. 2010. Nicotinamide adenine dinucleotide (NAD)-regulated DNA methylation alters CCCTC-binding factor (CTCF)/cohesin binding and transcription at the BDNF locus. *Proc Natl Acad Sci U S A*, 107, 21836-41.
- CHEN, H., TIAN, Y., SHU, W., BO, X. & WANG, S. 2012a. Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. *PLoS One*, 7, e41374.
- CHEN, X., BAROZZI, I., TERMANINI, A., PROSPERINI, E., RECCHIUTI, A., DALLI, J., MIETTON, F., MATTEOLI, G., HIEBERT, S. & NATOLI, G. 2012b. Requirement for the histone deacetylase Hdac3 for the inflammatory gene expression program in macrophages. *Proc Natl Acad Sci U S A*, 109, E2865-74.
- CHIEN, R., ZENG, W., KAWAUCHI, S., BENDER, M. A., SANTOS, R., GREGSON, H. C., SCHMIESING, J. A., NEWKIRK, D. A., KONG, X., BALL, A. R., JR., CALOF, A. L., LANDER, A. D., GROUDINE, M. T. & YOKOMORI, K. 2011. Cohesin mediates chromatin interactions that regulate mammalian beta-globin expression. *J Biol Chem*, 286, 17870-8.
- CHO, W. K., SPILLE, J. H., HECHT, M., LEE, C., LI, C., GRUBE, V. & CISSE, II 2018. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science*, 361, 412-415.
- CHOW, J. C., YOUNG, D. W., GOLENBOCK, D. T., CHRIST, W. J. & GUSOVSKY, F. 1999. Toll-like receptor-4 mediates lipopolysaccharide-induced signal transduction. *J Biol Chem*, 274, 10689-92.
- CITRI, A. & MALENKA, R. C. 2008. Synaptic plasticity: multiple forms, functions, and mechanisms. *Neuropsychopharmacology*, 33, 18-41.
- COBB, S., GUY, J. & BIRD, A. 2010. Reversibility of functional deficits in experimental models of Rett syndrome. *Biochem Soc Trans*, 38, 498-506.
- CONSORTIUM, E. P. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489, 57-74.
- CUARTERO, S., WEISS, F. D., DHARMALINGAM, G., GUO, Y., ING-SIMMONS, E., MASELLA, S., ROBLES-REBOLLO, I., XIAO, X., WANG, Y. F., BAROZZI, I., DJEGHLOUL, D., AMANO, M. T., NISKANEN, H., PETRETTO, E., DOWELL, R. D., TACHIBANA, K., KAIKKONEN, M. U., NASMYTH, K. A., LENHARD, B., NATOLI, G., FISHER, A. G. & MERKENSCHLAGER, M. 2018. Control of inducible gene expression links cohesin to hematopoietic progenitor self-renewal and differentiation. *Nat Immunol*, 19, 932-941.
- DE WIT, E., BOUWMAN, B. A., ZHU, Y., KLOUS, P., SPLINTER, E., VERSTEGEN, M. J., KRIJGER, P. H., FESTUCCIA, N., NORA, E. P., WELLING, M., HEARD, E., GEIJSEN, N., POOT, R. A., CHAMBERS, I. & DE LAAT, W. 2013. The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature*, 501, 227-31.

- DE WIT, E., VOS, E. S., HOLWERDA, S. J., VALDES-QUEZADA, C., VERSTEGEN, M. J., TEUNISSEN, H., SPLINTER, E., WIJCHERS, P. J., KRIJGER, P. H. & DE LAAT, W. 2015. CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell*, 60, 676-84.
- DEARDORFF, M. A., NOON, S. E. & KRANTZ, I. D. 1993. Cornelia de Lange Syndrome. In: ADAM, M. P., ARDINGER, H. H., PAGON, R. A., WALLACE, S. E., BEAN, L. J. H., STEPHENS, K. & AMEMIYA, A. (eds.) *GeneReviews*((R)). Seattle (WA).
- DEKKER, J. & HEARD, E. 2015. Structural and functional diversity of Topologically Associating Domains. *FEBS Lett*, 589, 2877-84.
- DEKKER, J. & MIRNY, L. 2016. The 3D Genome as Moderator of Chromosomal Communication. *Cell*, 164, 1110-1121.
- DEKKER, J., RIPPE, K., DEKKER, M. & KLECKNER, N. 2002. Capturing chromosome conformation. *Science*, 295, 1306-11.
- DENG, W., RUPON, J. W., KRIVEGA, I., BREDA, L., MOTTA, I., JAHN, K. S., REIK, A., GREGORY, P. D., RIVELLA, S., DEAN, A. & BLOBEL, G. A. 2014. Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell*, 158, 849-860.
- DENHOLTZ, M., BONORA, G., CHRONIS, C., SPLINTER, E., DE LAAT, W., ERNST, J., PELLEGRINI, M. & PLATH, K. 2013. Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. *Cell Stem Cell*, 13, 602-16.
- DIXON, J. R., JUNG, I., SELVARAJ, S., SHEN, Y., ANTOSIEWICZ-BOURGET, J. E., LEE, A. Y., YE, Z., KIM, A., RAJAGOPAL, N., XIE, W., DIAO, Y., LIANG, J., ZHAO, H., LOBANENKOV, V. V., ECKER, J. R., THOMSON, J. A. & REN, B. 2015. Chromatin architecture reorganization during stem cell differentiation. *Nature*, 518, 331-6.
- DIXON, J. R., SELVARAJ, S., YUE, F., KIM, A., LI, Y., SHEN, Y., HU, M., LIU, J. S. & REN, B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485, 376-80.
- DOSTIE, J., RICHMOND, T. A., ARNAOUT, R. A., SELZER, R. R., LEE, W. L., HONAN, T. A., RUBIO, E. D., KRUMM, A., LAMB, J., NUSBAUM, C., GREEN, R. D. & DEKKER, J. 2006. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res*, 16, 1299-309.
- DOWEN, J. M., FAN, Z. P., HNISZ, D., REN, G., ABRAHAM, B. J., ZHANG, L. N., WEINTRAUB, A. S., SCHUIJERS, J., LEE, T. I., ZHAO, K. & YOUNG, R. A. 2014. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell*, 159, 374-387.
- DRYDEN, N. H., BROOME, L. R., DUDBRIDGE, F., JOHNSON, N., ORR, N., SCHOENFELDER, S., NAGANO, T., ANDREWS, S., WINGETT, S., KOZAREWA, I., ASSIOTIS, I., FENWICK, K., MAGUIRE, S. L., CAMPBELL, J., NATRAJAN, R., LAMBROS, M., PERRAKIS, E., ASHWORTH, A., FRASER, P. & FLETCHER, O. 2014. Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C. *Genome Res*, 24, 1854-68.
- ENGEL, N., WEST, A. G., FELSENFELD, G. & BARTOLOMEI, M. S. 2004. Antagonism between DNA hypermethylation and enhancer-blocking activity at the H19 DMD is uncovered by CpG mutations. *Nat Genet*, 36, 883-8.
- ESSERS, M. A., OFFNER, S., BLANCO-BOSE, W. E., WAIBLER, Z., KALINKE, U., DUCHOSAL, M. A. & TRUMPP, A. 2009. IFN α activates dormant haematopoietic stem cells in vivo. *Nature*, 458, 904-8.
- FEIL, R., WAGNER, J., METZGER, D. & CHAMBON, P. 1997. Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochem Biophys Res Commun*, 237, 752-7.

- FREY, U., FREY, S., SCHOLLMEIER, F. & KRUG, M. 1996. Influence of actinomycin D, a RNA synthesis inhibitor, on long-term potentiation in rat hippocampal neurons in vivo and in vitro. *J Physiol*, 490 (Pt 3), 703-11.
- FUDENBERG, G., IMAKAEV, M., LU, C., GOLOBORODKO, A., ABDENNUR, N. & MIRNY, L. A. 2016. Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep*, 15, 2038-49.
- FUJITA, Y., MASUDA, K., BANDO, M., NAKATO, R., KATOU, Y., TANAKA, T., NAKAYAMA, M., TAKAO, K., MIYAKAWA, T., TANAKA, T., AGO, Y., HASHIMOTO, H., SHIRAHIGE, K. & YAMASHITA, T. 2017. Decreased cohesin in the brain leads to defective synapse development and anxiety-related behavior. *J Exp Med*, 214, 1431-1452.
- FURLONG, E. E. M. & LEVINE, M. 2018. Developmental enhancers and chromosome topology. *Science*, 361, 1341-1345.
- GALEEV, R., BAUDET, A., KUMAR, P., RUNDBERG NILSSON, A., NILSSON, B., SONEJI, S., TORNGREN, T., BORG, A., KVIST, A. & LARSSON, J. 2016. Genome-wide RNAi Screen Identifies Cohesin Genes as Modifiers of Renewal and Differentiation in Human HSCs. *Cell Rep*, 14, 2988-3000.
- GANDHI, R., GILLESPIE, P. J. & HIRANO, T. 2006. Human Wapl is a cohesin-binding protein that promotes sister-chromatid resolution in mitotic prophase. *Curr Biol*, 16, 2406-17.
- GASZNER, M. & FELSENFELD, G. 2006. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet*, 7, 703-13.
- GHISLETTI, S., BAROZZI, I., MIETTON, F., POLLETTI, S., DE SANTA, F., VENTURINI, E., GREGORY, L., LONIE, L., CHEW, A., WEI, C. L., RAGOISSIS, J. & NATOLI, G. 2010. Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. *Immunity*, 32, 317-28.
- GLASS, C. K. & NATOLI, G. 2016. Molecular control of activation and priming in macrophages. *Nat Immunol*, 17, 26-33.
- GOEBBELS, S., BORMUTH, I., BODE, U., HERMANSON, O., SCHWAB, M. H. & NAVE, K. A. 2006. Genetic targeting of principal neurons in neocortex and hippocampus of NEX-Cre mice. *Genesis*, 44, 611-21.
- GRUBER, S., ARUMUGAM, P., KATOU, Y., KUGLITSCH, D., HELMHART, W., SHIRAHIGE, K. & NASMYTH, K. 2006. Evidence that loading of cohesin onto chromosomes involves opening of its SMC hinge. *Cell*, 127, 523-37.
- GRUBER, S., HAERING, C. H. & NASMYTH, K. 2003. Chromosomal cohesin forms a ring. *Cell*, 112, 765-77.
- GUO, Y., XU, Q., CANZIO, D., SHOU, J., LI, J., GORKIN, D. U., JUNG, I., WU, H., ZHAI, Y., TANG, Y., LU, Y., WU, Y., JIA, Z., LI, W., ZHANG, M. Q., REN, B., KRAINER, A. R., MANIATIS, T. & WU, Q. 2015. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*, 162, 900-10.
- GUY, J., GAN, J., SELFRIDGE, J., COBB, S. & BIRD, A. 2007. Reversal of neurological defects in a mouse model of Rett syndrome. *Science*, 315, 1143-7.
- HABERLE, V. & STARK, A. 2018. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol*, 19, 621-637.
- HADJUR, S., WILLIAMS, L. M., RYAN, N. K., COBB, B. S., SEXTON, T., FRASER, P., FISHER, A. G. & MERKENSCHLAGER, M. 2009. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature*, 460, 410-3.
- HEIDINGER-PAULI, J. M., MERT, O., DAVENPORT, C., GUACCI, V. & KOSHLAND, D. 2010. Systematic reduction of cohesin differentially affects chromosome segregation, condensation, and DNA repair. *Curr Biol*, 20, 957-63.

- HEINZ, S., BENNER, C., SPANN, N., BERTOLINO, E., LIN, Y. C., LASLO, P., CHENG, J. X., MURRE, C., SINGH, H. & GLASS, C. K. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*, 38, 576-89.
- HEINZ, S., ROMANOSKI, C. E., BENNER, C. & GLASS, C. K. 2015. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol*, 16, 144-54.
- HIRAYAMA, T., TARUSAWA, E., YOSHIMURA, Y., GALJART, N. & YAGI, T. 2012. CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons. *Cell Rep*, 2, 345-57.
- HNISZ, D., DAY, D. S. & YOUNG, R. A. 2016. Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. *Cell*, 167, 1188-1200.
- ING-SIMMONS, E., SEITAN, V. C., FAURE, A. J., FLICEK, P., CARROLL, T., DEKKER, J., FISHER, A. G., LENHARD, B. & MERKENSCHLAGER, M. 2015. Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin. *Genome Res*, 25, 504-13.
- JHUNJHUNWALA, S., VAN ZELM, M. C., PEAK, M. M. & MURRE, C. 2009. Chromatin architecture and the generation of antigen receptor diversity. *Cell*, 138, 435-48.
- JIANG, Y., LOH, Y. E., RAJARAJAN, P., HIRAYAMA, T., LIAO, W., KASSIM, B. S., JAVIDFAR, B., HARTLEY, B. J., KLEOFAS, L., PARK, R. B., LABONTE, B., HO, S. M., CHANDRASEKARAN, S., DO, C., RAMIREZ, B. R., PETER, C. J., C, W. J., SAFAIE, B. M., MORISHITA, H., ROUSSOS, P., NESTLER, E. J., SCHAEFER, A., TYCKO, B., BRENNAND, K. J., YAGI, T., SHEN, L. & AKBARIAN, S. 2017. The methyltransferase SETDB1 regulates a large neuron-specific topological chromatin domain. *Nat Genet*, 49, 1239-1250.
- JOO, J. Y., SCHAUKOWITCH, K., FARBIAK, L., KILARU, G. & KIM, T. K. 2016. Stimulus-specific combinatorial functionality of neuronal c-fos enhancers. *Nat Neurosci*, 19, 75-83.
- KAAIJ, L. J. T., MOHN, F., VAN DER WEIDE, R. H., DE WIT, E. & BUHLER, M. 2019. The ChAHP Complex Counteracts Chromatin Looping at CTCF Sites that Emerged from SINE Expansions in Mouse. *Cell*, 178, 1437-1451 e14.
- KAGEY, M. H., NEWMAN, J. J., BILODEAU, S., ZHAN, Y., ORLANDO, D. A., VAN BERKUM, N. L., EBMEIER, C. C., GOOSSENS, J., RAHL, P. B., LEVINE, S. S., TAATJES, D. J., DEKKER, J. & YOUNG, R. A. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, 467, 430-5.
- KAIKKONEN, M. U., SPANN, N. J., HEINZ, S., ROMANOSKI, C. E., ALLISON, K. A., STENDER, J. D., CHUN, H. B., TOUGH, D. F., PRINJHA, R. K., BENNER, C. & GLASS, C. K. 2013. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol Cell*, 51, 310-25.
- KAWAUCHI, S., CALOF, A. L., SANTOS, R., LOPEZ-BURKS, M. E., YOUNG, C. M., HOANG, M. P., CHUA, A., LAO, T., LECHNER, M. S., DANIEL, J. A., NUSSENZWEIG, A., KITZES, L., YOKOMORI, K., HALLGRIMSSON, B. & LANDER, A. D. 2009. Multiple organ system defects and transcriptional dysregulation in the Nipbl(+/-) mouse, a model of Cornelia de Lange Syndrome. *PLoS Genet*, 5, e1000650.
- KIM, D., PERTEA, G., TRAPNELL, C., PIMENTEL, H., KELLEY, R. & SALZBERG, S. L. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*, 14, R36.
- KIM, J. H., REGE, M., VALERI, J., DUNAGIN, M. C., METZGER, A., TITUS, K. R., GILGENAST, T. G., GONG, W., BEAGAN, J. A., RAJ, A. & PHILLIPS-CREMINS, J. E. 2019. LADL: light-activated dynamic looping for endogenous gene expression control. *Nat Methods*, 16, 633-639.

- KIM, S. & SHENDURE, J. 2019. Mechanisms of Interplay between Transcription Factors and the 3D Genome. *Mol Cell*.
- KIM, T. K., HEMBERG, M., GRAY, J. M., COSTA, A. M., BEAR, D. M., WU, J., HARMIN, D. A., LAPTEWICZ, M., BARBARA-HALEY, K., KUERSTEN, S., MARKENSCOFF-PAPADIMITRIOU, E., KUHL, D., BITO, H., WORLEY, P. F., KREIMAN, G. & GREENBERG, M. E. 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature*, 465, 182-7.
- KLINE, A. D., GRADOS, M., SPONSELLER, P., LEVY, H. P., BLAGOWIDOW, N., SCHOEDEL, C., RAMPOLLA, J., CLEMENS, D. K., KRANTZ, I., KIMBALL, A., PICHARD, C. & TUCHMAN, D. 2007. Natural history of aging in Cornelia de Lange syndrome. *Am J Med Genet C Semin Med Genet*, 145C, 248-60.
- KLINE, A. D., MOSS, J. F., SELICORNI, A., BISGAARD, A. M., DEARDORFF, M. A., GILLET, P. M., ISHMAN, S. L., KERR, L. M., LEVIN, A. V., MULDER, P. A., RAMOS, F. J., WIERZBA, J., AJMONE, P. F., AXTELL, D., BLAGOWIDOW, N., CEREDA, A., COSTANTINO, A., CORMIER-DAIRE, V., FITZPATRICK, D., GRADOS, M., GROVES, L., GUTHRIE, W., HUISMAN, S., KAISER, F. J., KOEKKOEK, G., LEVIS, M., MARIANI, M., MCCLEERY, J. P., MENKE, L. A., METRENA, A., O'CONNOR, J., OLIVER, C., PIE, J., PIENING, S., POTTER, C. J., QUAGLIO, A. L., REDEKER, E., RICHMAN, D., RIGAMONTI, C., SHI, A., TUMER, Z., VAN BALKOM, I. D. C. & HENNEKAM, R. C. 2018. Diagnosis and management of Cornelia de Lange syndrome: first international consensus statement. *Nat Rev Genet*, 19, 649-666.
- KOJIC, A., CUADRADO, A., DE KONINCK, M., GIMENEZ-LLORENTE, D., RODRIGUEZ-CORSINO, M., GOMEZ-LOPEZ, G., LE DILY, F., MARTI-RENOM, M. A. & LOSADA, A. 2018. Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization. *Nat Struct Mol Biol*, 25, 496-504.
- KON, A., SHIH, L. Y., MINAMINO, M., SANADA, M., SHIRAISHI, Y., NAGATA, Y., YOSHIDA, K., OKUNO, Y., BANDO, M., NAKATO, R., ISHIKAWA, S., SATO-OTSUBO, A., NAGAE, G., NISHIMOTO, A., HAFERLACH, C., NOWAK, D., SATO, Y., ALPERMANN, T., NAGASAKI, M., SHIMAMURA, T., TANAKA, H., CHIBA, K., YAMAMOTO, R., YAMAGUCHI, T., OTSU, M., OBARA, N., SAKATA-YANAGIMOTO, M., NAKAMAKI, T., ISHIYAMA, K., NOLTE, F., HOFMANN, W. K., MIYAWAKI, S., CHIBA, S., MORI, H., NAKAUCHI, H., KOEFFLER, H. P., ABURATANI, H., HAFERLACH, T., SHIRAHIGE, K., MIYANO, S. & OGAWA, S. 2013. Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. *Nat Genet*, 45, 1232-7.
- KRANTZ, I. D., MCCALLUM, J., DESCIPIO, C., KAUR, M., GILLIS, L. A., YAEGER, D., JUKOFSKY, L., WASSERMAN, N., BOTTANI, A., MORRIS, C. A., NOWACZYK, M. J., TORIELLO, H., BAMSHAD, M. J., CAREY, J. C., RAPPAPORT, E., KAWAUCHI, S., LANDER, A. D., CALOF, A. L., LI, H. H., DEVOTO, M. & JACKSON, L. G. 2004. Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of *Drosophila melanogaster* Nipped-B. *Nat Genet*, 36, 631-5.
- KRIAUCIONIS, S., PATERSON, A., CURTIS, J., GUY, J., MACLEOD, N. & BIRD, A. 2006. Gene expression analysis exposes mitochondrial abnormalities in a mouse model of Rett syndrome. *Mol Cell Biol*, 26, 5033-42.
- KRISHNASWAMI, S. R., GRINDBERG, R. V., NOVOTNY, M., VENEPALLY, P., LACAR, B., BHUTANI, K., LINKER, S. B., PHAM, S., ERWIN, J. A., MILLER, J. A., HODGE, R., MCCARTHY, J. K., KELDER, M., MCCORRISON, J., AEVERMANN, B. D., FUERTES, F. D., SCHEUERMANN, R. H., LEE, J., LEIN, E. S., SCHORK, N., MCCONNELL, M. J., GAGE, F. H. & LASKEN, R. S.

2016. Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat Protoc*, 11, 499-524.
- KUENG, S., HEGEMANN, B., PETERS, B. H., LIPP, J. J., SCHLEIFFER, A., MECHTLER, K. & PETERS, J. M. 2006. Wapl controls the dynamic association of cohesin with chromatin. *Cell*, 127, 955-67.
- LAI, A. Y., FATEMI, M., DHASARATHY, A., MALONE, C., SOBOL, S. E., GEIGERMAN, C., JAYE, D. L., MAV, D., SHAH, R., LI, L. & WADE, P. A. 2010. DNA methylation prevents CTCF-mediated silencing of the oncogene BCL6 in B cell lymphomas. *J Exp Med*, 207, 1939-50.
- LAVAGNOLLI, T., GUPTA, P., HORMANSEDER, E., MIRA-BONTENBAL, H., DHARMALINGAM, G., CARROLL, T., GURDON, J. B., FISHER, A. G. & MERKENSCHLAGER, M. 2015. Initiation and maintenance of pluripotency gene expression in the absence of cohesin. *Genes Dev*, 29, 23-38.
- LEVINE, M. 2010. Transcriptional enhancers in animal development and evolution. *Curr Biol*, 20, R754-63.
- LIAO, Y., SMYTH, G. K. & SHI, W. 2013. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res*, 41, e108.
- LIAO, Y., SMYTH, G. K. & SHI, W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30, 923-30.
- LIEBERMAN-AIDEN, E., VAN BERKUM, N. L., WILLIAMS, L., IMAKAEV, M., RAGOCZY, T., TELLING, A., AMIT, I., LAJOIE, B. R., SABO, P. J., DORSCHNER, M. O., SANDSTROM, R., BERNSTEIN, B., BENDER, M. A., GROUDINE, M., GNIRKE, A., STAMATOYANNOPOULOS, J., MIRNY, L. A., LANDER, E. S. & DEKKER, J. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326, 289-93.
- LONG, H. K., PRESCOTT, S. L. & WYSOCKA, J. 2016. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell*, 167, 1170-1187.
- LOVE, M. I., HUBER, W. & ANDERS, S. 2014a. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, 15, 550.
- LOVE, P. E., WARZECHA, C. & LI, L. 2014b. Ldb1 complexes: the new master regulators of erythroid gene transcription. *Trends Genet*, 30, 1-9.
- LUPIANEZ, D. G., KRAFT, K., HEINRICH, V., KRAWITZ, P., BRANCATI, F., KLOPOCKI, E., HORN, D., KAYSERILI, H., OPITZ, J. M., LAXOVA, R., SANTOS-SIMARRO, F., GILBERT-DUSSARDIER, B., WITTLER, L., BORSCHIWER, M., HAAS, S. A., OSTERWALDER, M., FRANKE, M., TIMMERMAN, B., HECHT, J., SPIELMANN, M., VISEL, A. & MUNDLOS, S. 2015. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, 161, 1012-1025.
- MALIK, A. N., VIERBUCHEN, T., HEMBERG, M., RUBIN, A. A., LING, E., COUCH, C. H., STROUD, H., SPIEGEL, I., FARH, K. K., HARMIN, D. A. & GREENBERG, M. E. 2014. Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. *Nat Neurosci*, 17, 1330-9.
- MAZUMDAR, C., SHEN, Y., XAVY, S., ZHAO, F., REINISCH, A., LI, R., CORCES, M. R., FLYNN, R. A., BUENROSTRO, J. D., CHAN, S. M., THOMAS, D., KOENIG, J. L., HONG, W. J., CHANG, H. Y. & MAJETI, R. 2015. Leukemia-Associated Cohesin Mutants Dominantly Enforce Stem Cell Programs and Impair Human Hematopoietic Progenitor Differentiation. *Cell Stem Cell*, 17, 675-688.

- MEDZHITOV, R. & HORNG, T. 2009. Transcriptional control of the inflammatory response. *Nat Rev Immunol*, 9, 692-703.
- MERKENSCHLAGER, M. & NORA, E. P. 2016. CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annu Rev Genomics Hum Genet*, 17, 17-43.
- MILLS, J. A., HERRERA, P. S., KAUR, M., LEO, L., MCELDREW, D., TINTOS-HERNANDEZ, J. A., RAJAGOPALAN, R., GAGNE, A., ZHANG, Z., ORTIZ-GONZALEZ, X. R. & KRANTZ, I. D. 2018. NIPBL(+/-) haploinsufficiency reveals a constellation of transcriptome disruptions in the pluripotent and cardiac states. *Sci Rep*, 8, 1056.
- MOOHTHA, V. K., LINDGREN, C. M., ERIKSSON, K. F., SUBRAMANIAN, A., SIHAG, S., LEHAR, J., PUIGSERVER, P., CARLSSON, E., RIDDERSTRALE, M., LAURILA, E., HOUSTIS, N., DALY, M. J., PATTERSON, N., MESIROV, J. P., GOLUB, T. R., TAMAYO, P., SPIEGELMAN, B., LANDER, E. S., HIRSCHHORN, J. N., ALTSHULER, D. & GROOP, L. C. 2003. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*, 34, 267-73.
- MOSS, J., HOWLIN, P., MAGIATI, I. & OLIVER, C. 2012. Characteristics of autism spectrum disorder in Cornelia de Lange syndrome. *J Child Psychol Psychiatry*, 53, 883-91.
- MOSSADEGH-KELLER, N., SARRAZIN, S., KANDALLA, P. K., ESPINOSA, L., STANLEY, E. R., NUTT, S. L., MOORE, J. & SIEWEKE, M. H. 2013. M-CSF instructs myeloid lineage fate in single haematopoietic stem cells. *Nature*, 497, 239-43.
- MULLENDERS, J., ARANDA-ORGILLES, B., LHOUMAUD, P., KELLER, M., PAE, J., WANG, K., KAYEMBE, C., ROCHA, P. P., RAVIRAM, R., GONG, Y., PREMSRIRUT, P. K., TSIRIGOS, A., BONNEAU, R., SKOK, J. A., CIMMINO, L., HOEHN, D. & AIFANTIS, I. 2015. Cohesin loss alters adult hematopoietic stem cell homeostasis, leading to myeloproliferative neoplasms. *J Exp Med*, 212, 1833-50.
- NAGAI, Y., GARRETT, K. P., OHTA, S., BAHRUN, U., KOURO, T., AKIRA, S., TAKATSU, K. & KINCADE, P. W. 2006. Toll-like receptors on hematopoietic progenitor cells stimulate innate immune system replenishment. *Immunity*, 24, 801-12.
- NAIR, S. J., YANG, L., MELUZZI, D., OH, S., YANG, F., FRIEDMAN, M. J., WANG, S., SUTER, T., ALSHAREEDAH, I., GAMLIEL, A., MA, Q., ZHANG, J., HU, Y., TAN, Y., OHGI, K. A., JAYANI, R. S., BANERJEE, P. R., AGGARWAL, A. K. & ROSENFELD, M. G. 2019. Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly. *Nat Struct Mol Biol*, 26, 193-203.
- NARENDRA, V., ROCHA, P. P., AN, D., RAVIRAM, R., SKOK, J. A., MAZZONI, E. O. & REINBERG, D. 2015. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science*, 347, 1017-21.
- NASMYTH, K. & HAERING, C. H. 2009. Cohesin: its roles and mechanisms. *Annu Rev Genet*, 43, 525-58.
- NATSUME, T., KIYOMITSU, T., SAGA, Y. & KANEMAKI, M. T. 2016. Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. *Cell Rep*, 15, 210-218.
- NGUYEN, P. V., ABEL, T. & KANDEL, E. R. 1994. Requirement of a critical period of transcription for induction of a late phase of LTP. *Science*, 265, 1104-7.
- NITZSCHE, A., PASZKOWSKI-ROGACZ, M., MATARESE, F., JANSSEN-MEGENS, E. M., HUBNER, N. C., SCHULZ, H., DE VRIES, I., DING, L., HUEBNER, N., MANN, M., STUNNENBERG, H. G. & BUCHHOLZ, F. 2011. RAD21 cooperates with pluripotency transcription factors in the maintenance of embryonic stem cell identity. *PLoS One*, 6, e19470.

- NORA, E. P., GOLOBORODKO, A., VALTON, A. L., GIBCUS, J. H., UEBERSOHN, A., ABDENNUR, N., DEKKER, J., MIRNY, L. A. & BRUNEAU, B. G. 2017. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*, 169, 930-944 e22.
- NORA, E. P., LAJOIE, B. R., SCHULZ, E. G., GIORGETTI, L., OKAMOTO, I., SERVANT, N., PIOLOT, T., VAN BERKUM, N. L., MEISIG, J., SEDAT, J., GRIBNAU, J., BARILLOT, E., BLUTHGEN, N., DEKKER, J. & HEARD, E. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 485, 381-5.
- NORD, A. S., BLOW, M. J., ATTANASIO, C., AKIYAMA, J. A., HOLT, A., HOSSEINI, R., PHOUANENAVONG, S., PLAJSER-FRICK, I., SHOUKRY, M., AFZAL, V., RUBENSTEIN, J. L., RUBIN, E. M., PENNACCHIO, L. A. & VISEL, A. 2013. Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell*, 155, 1521-31.
- OHLSSON, R., RENKAWITZ, R. & LOBANENKOV, V. 2001. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet*, 17, 520-7.
- OLIVEIRA, R. A., HAMILTON, R. S., PAULI, A., DAVIS, I. & NASMYTH, K. 2010. Cohesin cleavage and Cdk inhibition trigger formation of daughter nuclei. *Nat Cell Biol*, 12, 185-92.
- OLLEY, G., ANSARI, M., BENGANI, H., GRIMES, G. R., RHODES, J., VON KRIEGSHEIM, A., BLATNIK, A., STEWART, F. J., WAKELING, E., CARROLL, N., ROSS, A., PARK, S. M., DECIPHERING DEVELOPMENTAL DISORDERS, S., BICKMORE, W. A., PRADEEPA, M. M. & FITZPATRICK, D. R. 2018. BRD4 interacts with NIPBL and BRD4 is mutated in a Cornelia de Lange-like syndrome. *Nat Genet*, 50, 329-332.
- ONG, C. T. & CORCES, V. G. 2014. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet*, 15, 234-46.
- OSTERWALDER, M., BAROZZI, I., TISSIERES, V., FUKUDA-YUZAWA, Y., MANNION, B. J., AFZAL, S. Y., LEE, E. A., ZHU, Y., PLAJSER-FRICK, I., PICKLE, C. S., KATO, M., GARVIN, T. H., PHAM, Q. T., HARRINGTON, A. N., AKIYAMA, J. A., AFZAL, V., LOPEZ-RIOS, J., DICKEL, D. E., VISEL, A. & PENNACCHIO, L. A. 2018. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature*, 554, 239-243.
- OSTUNI, R., PICCOLO, V., BAROZZI, I., POLLETTI, S., TERMANINI, A., BONIFACIO, S., CURINA, A., PROSPERINI, E., GHISLETTI, S. & NATOLI, G. 2013. Latent enhancers activated by stimulation in differentiated cells. *Cell*, 152, 157-71.
- PALSTRA, R. J., TOLHUIS, B., SPLINTER, E., NIJMEIJER, R., GROSVELD, F. & DE LAAT, W. 2003. The beta-globin nuclear compartment in development and erythroid differentiation. *Nat Genet*, 35, 190-4.
- PARELHO, V., HADJUR, S., SPIVAKOV, M., LELEU, M., SAUER, S., GREGSON, H. C., JARMUZ, A., CANZONETTA, C., WEBSTER, Z., NESTEROVA, T., COBB, B. S., YOKOMORI, K., DILLON, N., ARAGON, L., FISHER, A. G. & MERKENSCHLAGER, M. 2008. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell*, 132, 422-33.
- PARENTI, I., GERVASINI, C., POZOJEVIC, J., GRAUL-NEUMANN, L., AZZOLLINI, J., BRAUNHOLZ, D., WATRIN, E., WENDT, K. S., CEREDA, A., CITTARO, D., GILLESSEN-KAESBACH, G., LAZAREVIC, D., MARIANI, M., RUSSO, S., WERNER, R., KRAWITZ, P., LARIZZA, L., SELICORNI, A. & KAISER, F. J. 2016. Broadening of cohesinopathies: exome sequencing identifies mutations in ANKRD11 in two patients with Cornelia de Lange-overlapping phenotype. *Clin Genet*, 89, 74-81.
- PAULI, A., ALTHOFF, F., OLIVEIRA, R. A., HEIDMANN, S., SCHULDINER, O., LEHNER, C. F., DICKSON, B. J. & NASMYTH, K. 2008. Cell-type-specific TEV protease cleavage reveals cohesin functions in Drosophila neurons. *Dev Cell*, 14, 239-51.

- PERRY, M. W., BOETTIGER, A. N. & LEVINE, M. 2011. Multiple enhancers ensure precision of gap gene-expression patterns in the *Drosophila* embryo. *Proc Natl Acad Sci U S A*, 108, 13570-5.
- RAJARAJAN, P., GIL, S. E., BRENNAND, K. J. & AKBARIAN, S. 2016. Spatial genome organization and cognition. *Nat Rev Neurosci*, 17, 681-691.
- RAO, S. S., HUNTLEY, M. H., DURAND, N. C., STAMENOVA, E. K., BOCHKOV, I. D., ROBINSON, J. T., SANBORN, A. L., MACHOL, I., OMER, A. D., LANDER, E. S. & AIDEN, E. L. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159, 1665-80.
- RAO, S. S. P., HUANG, S. C., GLENN ST HILAIRE, B., ENGREITZ, J. M., PEREZ, E. M., KIEFFER-KWON, K. R., SANBORN, A. L., JOHNSTONE, S. E., BASCOM, G. D., BOCHKOV, I. D., HUANG, X., SHAMIM, M. S., SHIN, J., TURNER, D., YE, Z., OMER, A. D., ROBINSON, J. T., SCHLICK, T., BERNSTEIN, B. E., CASELLAS, R., LANDER, E. S. & AIDEN, E. L. 2017. Cohesin Loss Eliminates All Loop Domains. *Cell*, 171, 305-320 e24.
- REMESEIRO, S., CUADRADO, A., GOMEZ-LOPEZ, G., PISANO, D. G. & LOSADA, A. 2012. A unique role of cohesin-SA1 in gene regulation and development. *EMBO J*, 31, 2090-102.
- ROWLEY, M. J. & CORCES, V. G. 2018. Organizational principles of 3D genome architecture. *Nat Rev Genet*, 19, 789-800.
- RUBIO, E. D., REISS, D. J., WELCSH, P. L., DISTECHE, C. M., FILIPPOVA, G. N., BALIGA, N. S., AEBERSOLD, R., RANISH, J. A. & KRUMM, A. 2008. CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci U S A*, 105, 8309-14.
- SAMS, D. S., NARDONE, S., GETSELTHER, D., RAZ, D., TAL, M., RAYI, P. R., KAPHZAN, H., HAKIM, O. & ELLIOTT, E. 2016. Neuronal CTCF Is Necessary for Basal and Experience-Dependent Gene Regulation, Memory Formation, and Genomic Structure of BDNF and Arc. *Cell Rep*, 17, 2418-2430.
- SANTOS, R., KAWAUCHI, S., JACOBS, R. E., LOPEZ-BURKS, M. E., CHOI, H., WIKENHEISER, J., HALLGRIMSSON, B., JAMNICZKY, H. A., FRASER, S. E., LANDER, A. D. & CALOF, A. L. 2016. Conditional Creation and Rescue of Nipbl-Deficiency in Mice Reveals Multiple Determinants of Risk for Congenital Heart Defects. *PLoS Biol*, 14, e2000197.
- SCHAFER, D. P., LEHRMAN, E. K., KAUTZMAN, A. G., KOYAMA, R., MARDINLY, A. R., YAMASAKI, R., RANSOHOFF, R. M., GREENBERG, M. E., BARRES, B. A. & STEVENS, B. 2012. Microglia sculpt postnatal neural circuits in an activity and complement-dependent manner. *Neuron*, 74, 691-705.
- SCHAUKOWITCH, K., JOO, J. Y., LIU, X., WATTS, J. K., MARTINEZ, C. & KIM, T. K. 2014. Enhancer RNA facilitates NELF release from immediate early genes. *Mol Cell*, 56, 29-42.
- SCHMITT, A. D., HU, M. & REN, B. 2016. Genome-wide mapping and analysis of chromosome architecture. *Nat Rev Mol Cell Biol*, 17, 743-755.
- SCHOENFELDER, S. & FRASER, P. 2019. Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet*.
- SCHOENFELDER, S., SEXTON, T., CHAKALOVA, L., COPE, N. F., HORTON, A., ANDREWS, S., KURUKUTI, S., MITCHELL, J. A., UMLAUF, D., DIMITROVA, D. S., ESKIW, C. H., LUO, Y., WEI, C. L., RUAN, Y., BIEKER, J. J. & FRASER, P. 2010. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet*, 42, 53-61.
- SCHRIER, S. A., SHERER, I., DEARDORFF, M. A., CLARK, D., AUDETTE, L., GILLIS, L., KLINE, A. D., ERNST, L., LOOMES, K., KRANTZ, I. D. & JACKSON, L. G. 2011. Causes of death and

- autopsy findings in a large study cohort of individuals with Cornelia de Lange syndrome and review of the literature. *Am J Med Genet A*, 155A, 3007-24.
- SCHWARZER, W., ABDENNUR, N., GOLOBORODKO, A., PEKOWSKA, A., FUDENBERG, G., LOE-MIE, Y., FONSECA, N. A., HUBER, W., HAERING, C. H., MIRNY, L. & SPITZ, F. 2017. Two independent modes of chromatin organization revealed by cohesin removal. *Nature*, 551, 51-56.
- SEITAN, V. C., HAO, B., TACHIBANA-KONWALSKI, K., LAVAGNOLLI, T., MIRA-BONTENBAL, H., BROWN, K. E., TENG, G., CARROLL, T., TERRY, A., HORAN, K., MARKS, H., ADAMS, D. J., SCHATZ, D. G., ARAGON, L., FISHER, A. G., KRANGEL, M. S., NASMYTH, K. & MERKENSCHLAGER, M. 2011. A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. *Nature*, 476, 467-71.
- SEXTON, T., YAFFE, E., KENIGSBERG, E., BANTIGNIES, F., LEBLANC, B., HOICHMAN, M., PARRINELLO, H., TANAY, A. & CAVALLI, G. 2012. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, 148, 458-72.
- SHOLL, D. A. 1953. Dendritic organization in the neurons of the visual and motor cortices of the cat. *J Anat*, 87, 387-406.
- SPIELMANN, M., LUPIANEZ, D. G. & MUNDLOS, S. 2018. Structural variation in the 3D genome. *Nat Rev Genet*, 19, 453-467.
- STADHOUDERS, R., FILION, G. J. & GRAF, T. 2019. Transcription factors and 3D genome conformation in cell-fate decisions. *Nature*, 569, 345-354.
- STANHOPE-BAKER, P., HUDSON, K. M., SHAFFER, A. L., CONSTANTINESCU, A. & SCHLISSEL, M. S. 1996. Cell type-specific chromatin structure determines the targeting of V(D)J recombinase activity in vitro. *Cell*, 85, 887-97.
- STROM, A. R., EMELIANOV, A. V., MIR, M., FYODOROV, D. V., DARZACQ, X. & KARPEN, G. H. 2017. Phase separation drives heterochromatin domain formation. *Nature*, 547, 241-245.
- SUBRAMANIAN, A., TAMAYO, P., MOOTHA, V. K., MUKHERJEE, S., EBERT, B. L., GILLETTE, M. A., PAULOVICH, A., POMEROY, S. L., GOLUB, T. R., LANDER, E. S. & MESIROV, J. P. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, 102, 15545-50.
- TACHIBANA-KONWALSKI, K., GODWIN, J., VAN DER WEYDEN, L., CHAMPION, L., KUDO, N. R., ADAMS, D. J. & NASMYTH, K. 2010. Rec8-containing cohesin maintains bivalents without turnover during the growing phase of mouse oocytes. *Genes Dev*, 24, 2505-16.
- TERAKAWA, T., BISHT, S., EEFTEENS, J. M., DEKKER, C., HAERING, C. H. & GREENE, E. C. 2017. The condensin complex is a mechanochemical motor that translocates along DNA. *Science*, 358, 672-676.
- THOTA, S., VINY, A. D., MAKISHIMA, H., SPITZER, B., RADIVOYEVITCH, T., PRZYCHODZEN, B., SEKERES, M. A., LEVINE, R. L. & MACIEJEWSKI, J. P. 2014. Genetic alterations of the cohesin complex genes in myeloid malignancies. *Blood*, 124, 1790-8.
- TOLHUIS, B., PALSTRA, R. J., SPLINTER, E., GROSVELD, F. & DE LAAT, W. 2002. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell*, 10, 1453-65.
- TONKIN, E. T., WANG, T. J., LISGO, S., BAMSHAD, M. J. & STRACHAN, T. 2004. NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. *Nat Genet*, 36, 636-41.

- UHLMANN, F., WERNIC, D., POUPART, M. A., KOONIN, E. V. & NASMYTH, K. 2000. Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast. *Cell*, 103, 375-86.
- VAN DE WERKEN, H. J., LANDAN, G., HOLWERDA, S. J., HOICHMAN, M., KLOUS, P., CHACHIK, R., SPLINTER, E., VALDES-QUEZADA, C., OZ, Y., BOUWMAN, B. A., VERSTEGEN, M. J., DE WIT, E., TANAY, A. & DE LAAT, W. 2012. Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat Methods*, 9, 969-72.
- VAN DEN BERG, D. L. C., AZZARELLI, R., OISHI, K., MARTYNOGA, B., URBAN, N., DEKKERS, D. H. W., DEMMERS, J. A. & GUILLEMOT, F. 2017. Nipbl Interacts with Zfp609 and the Integrator Complex to Regulate Cortical Neuron Migration. *Neuron*, 93, 348-361.
- VEGA, H., WAISFISZ, Q., GORDILLO, M., SAKAI, N., YANAGIHARA, I., YAMADA, M., VAN GOSLIGA, D., KAYSERILI, H., XU, C., OZONO, K., JABS, E. W., INUI, K. & JOENJE, H. 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. *Nat Genet*, 37, 468-70.
- VIAN, L., PEKOWSKA, A., RAO, S. S. P., KIEFFER-KWON, K. R., JUNG, S., BARANELLO, L., HUANG, S. C., EL KHATTABI, L., DOSE, M., PRUETT, N., SANBORN, A. L., CANELA, A., MAMAN, Y., OKSANEN, A., RESCH, W., LI, X., LEE, B., KOVALCHUK, A. L., TANG, Z., NELSON, S., DI PIERRO, M., CHENG, R. R., MACHOL, I., ST HILAIRE, B. G., DURAND, N. C., SHAMIM, M. S., STAMENOVA, E. K., ONUCHIC, J. N., RUAN, Y., NUSSENZWEIG, A., LEVENS, D., AIDEN, E. L. & CASELLAS, R. 2018. The Energetics and Physiological Impact of Cohesin Extrusion. *Cell*, 173, 1165-1178 e20.
- VINY, A. D., OTT, C. J., SPITZER, B., RIVAS, M., MEYDAN, C., PAPALEXI, E., YELIN, D., SHANK, K., REYES, J., CHIU, A., ROMIN, Y., BOYKO, V., THOTA, S., MACIEJEWSKI, J. P., MELNICK, A., BRADNER, J. E. & LEVINE, R. L. 2015. Dose-dependent role of the cohesin complex in normal and malignant hematopoiesis. *J Exp Med*, 212, 1819-32.
- VUILLEUMIER, N., KOVARI, E., MICHON, A., HOF, P. R., MENTENOPOULOS, G., GIANNAKOPOULOS, P. & BOURAS, C. 2002. Neuropathological analysis of an adult case of the Cornelia de Lange syndrome. *Acta Neuropathol*, 104, 327-32.
- WEINTRAUB, A. S., LI, C. H., ZAMUDIO, A. V., SIGOVA, A. A., HANNETT, N. M., DAY, D. S., ABRAHAM, B. J., COHEN, M. A., NABET, B., BUCKLEY, D. L., GUO, Y. E., HNISZ, D., JAENISCH, R., BRADNER, J. E., GRAY, N. S. & YOUNG, R. A. 2017. YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell*, 171, 1573-1588 e28.
- WEITZER, S., LEHANE, C. & UHLMANN, F. 2003. A model for ATP hydrolysis-dependent binding of cohesin to DNA. *Curr Biol*, 13, 1930-40.
- WENDT, K. S., YOSHIDA, K., ITOH, T., BANDO, M., KOCH, B., SCHIRGHUBER, E., TSUTSUMI, S., NAGAE, G., ISHIHARA, K., MISHIRO, T., YAHATA, K., IMAMOTO, F., ABURATANI, H., NAKAO, M., IMAMOTO, N., MAESHIMA, K., SHIRAHIGE, K. & PETERS, J. M. 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*, 451, 796-801.
- WILLIAMS, M. R., FRICANO-KUGLER, C. J., GETZ, S. A., SKELTON, P. D., LEE, J., RIZZUTO, C. P., GELLER, J. S., LI, M. & LUIKART, B. W. 2016. A Retroviral CRISPR-Cas9 System for Cellular Autism-Associated Phenotype Discovery in Developing Neurons. *Sci Rep*, 6, 25611.
- WUTZ, G., VARNAL, C., NAGASAKA, K., CISNEROS, D. A., STOCSITS, R. R., TANG, W., SCHOENFELDER, S., JESSBERGER, G., MUHAR, M., HOSSAIN, M. J., WALTHER, N., KOCH, B., KUEBLBECK, M., ELLENBERG, J., ZUBER, J., FRASER, P. & PETERS, J. M. 2017.

- Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J*, 36, 3573-3599.
- XIAO, T., WALLACE, J. & FELSENFELD, G. 2011. Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. *Mol Cell Biol*, 31, 2174-83.
- YAMADA, T., YANG, Y., VALNEGRI, P., JURIC, I., ABNOUSI, A., MARKWALTER, K. H., GUTHRIE, A. N., GODEC, A., OLDENBORG, A., HU, M., HOLY, T. E. & BONNI, A. 2019. Sensory experience remodels genome architecture in neural circuit to drive motor learning. *Nature*, 569, 708-713.
- YOUNG, M. D., WAKEFIELD, M. J., SMYTH, G. K. & OSHLACK, A. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol*, 11, R14.
- ZHAO, J. L., MA, C., O'CONNELL, R. M., MEHTA, A., DILORETO, R., HEATH, J. R. & BALTIMORE, D. 2014. Conversion of danger signals into cytokine signals by hematopoietic stem and progenitor cells for regulation of stress-induced hematopoiesis. *Cell Stem Cell*, 14, 445-459.
- ZHAO, Z., TAVOOSIDANA, G., SJOLINDER, M., GONDOR, A., MARIANO, P., WANG, S., KANDURI, C., LEZCANO, M., SANDHU, K. S., SINGH, U., PANT, V., TIWARI, V., KURUKUTI, S. & OHLSSON, R. 2006. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet*, 38, 1341-7.
- ZHENG, H. & XIE, W. 2019. The role of 3D genome organization in development and cell differentiation. *Nat Rev Mol Cell Biol*.
- ZLATANOVA, J. & CAIAFA, P. 2009. CTCF and its protein partners: divide and rule? *J Cell Sci*, 122, 1275-84.
- VEGA, H., WAISFISZ, Q., GORDILLO, M., SAKAI, N., YANAGIHARA, I., YAMADA, M., VAN GOSLIGA, D., KAYSERILI, H., XU, C., OZONO, K., JABS, E. W., INUI, K. & JOENJE, H. 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. *Nat Genet*, 37, 468-70.
- VIAN, L., PEKOWSKA, A., RAO, S. S. P., KIEFFER-KWON, K. R., JUNG, S., BARANELLO, L., HUANG, S. C., EL KHATTABI, L., DOSE, M., PRUETT, N., SANBORN, A. L., CANELA, A., MAMAN, Y., OKSANEN, A., RESCH, W., LI, X., LEE, B., KOVALCHUK, A. L., TANG, Z., NELSON, S., DI PIERRO, M., CHENG, R. R., MACHOL, I., ST HILAIRE, B. G., DURAND, N. C., SHAMIM, M. S., STAMENOVA, E. K., ONUCHIC, J. N., RUAN, Y., NUSSENZWEIG, A., LEVENS, D., AIDEN, E. L. & CASELLAS, R. 2018. The Energetics and Physiological Impact of Cohesin Extrusion. *Cell*, 173, 1165-1178 e20.
- VINY, A. D., OTT, C. J., SPITZER, B., RIVAS, M., MEYDAN, C., PAPALEXI, E., YELIN, D., SHANK, K., REYES, J., CHIU, A., ROMIN, Y., BOYKO, V., THOTA, S., MACIEJEWSKI, J. P., MELNICK, A., BRADNER, J. E. & LEVINE, R. L. 2015. Dose-dependent role of the cohesin complex in normal and malignant hematopoiesis. *J Exp Med*, 212, 1819-32.
- VUILLEUMIER, N., KOVARI, E., MICHON, A., HOF, P. R., MENTENOPOULOS, G., GIANNAKOPOULOS, P. & BOURAS, C. 2002. Neuropathological analysis of an adult case of the Cornelia de Lange syndrome. *Acta Neuropathol*, 104, 327-32.
- WEINTRAUB, A. S., LI, C. H., ZAMUDIO, A. V., SIGOVA, A. A., HANNETT, N. M., DAY, D. S., ABRAHAM, B. J., COHEN, M. A., NABET, B., BUCKLEY, D. L., GUO, Y. E., HNISZ, D., JAENISCH, R., BRADNER, J. E., GRAY, N. S. & YOUNG, R. A. 2017. YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell*, 171, 1573-1588 e28.

- WEITZER, S., LEHANE, C. & UHLMANN, F. 2003. A model for ATP hydrolysis-dependent binding of cohesin to DNA. *Curr Biol*, 13, 1930-40.
- WENDT, K. S., YOSHIDA, K., ITOH, T., BANDO, M., KOCH, B., SCHIRGHUBER, E., TSUTSUMI, S., NAGAE, G., ISHIHARA, K., MISHIRO, T., YAHATA, K., IMAMOTO, F., ABURATANI, H., NAKAO, M., IMAMOTO, N., MAESHIMA, K., SHIRAHIGE, K. & PETERS, J. M. 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*, 451, 796-801.
- WUTZ, G., VARNAI, C., NAGASAKA, K., CISNEROS, D. A., STOCITS, R. R., TANG, W., SCHOENFELDER, S., JESSBERGER, G., MUHAR, M., HOSSAIN, M. J., WALTHER, N., KOCH, B., KUEBLBECK, M., ELLENBERG, J., ZUBER, J., FRASER, P. & PETERS, J. M. 2017. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J*, 36, 3573-3599.
- XIAO, T., WALLACE, J. & FELSENFELD, G. 2011. Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. *Mol Cell Biol*, 31, 2174-83.
- YAMADA, T., YANG, Y., VALNEGRI, P., JURIC, I., ABNOUSI, A., MARKWALTER, K. H., GUTHRIE, A. N., GODEC, A., OLDENBORG, A., HU, M., HOLY, T. E. & BONNI, A. 2019. Sensory experience remodels genome architecture in neural circuit to drive motor learning. *Nature*, 569, 708-713.
- ZHANG, Y., ZHANG, X., BA, Z., LIANG, Z., DRING, E. W., HU, H., LOU, J., KYRITSIS, N., ZURITA, J., SHAMIM, M. S., PRESSER AIDEN, A., LIEBERMAN AIDEN, E. & ALT, F. W. 2019. The fundamental role of chromatin loop extrusion in physiological V(D)J recombination. *Nature*.
- ZHAO, Z., TAVOOSIDANA, G., SJOLINDER, M., GONDOR, A., MARIANO, P., WANG, S., KANDURI, C., LEZCANO, M., SANDHU, K. S., SINGH, U., PANT, V., TIWARI, V., KURUKUTI, S. & OHLSSON, R. 2006. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet*, 38, 1341-7.
- ZHAO, J. L., MA, C., O'CONNELL, R. M., MEHTA, A., DILORETO, R., HEATH, J. R. & BALTIMORE, D. 2014. Conversion of danger signals into cytokine signals by hematopoietic stem and progenitor cells for regulation of stress-induced hematopoiesis. *Cell Stem Cell*, 14, 445-459.
- ZHENG, H. & XIE, W. 2019. The role of 3D genome organization in development and cell differentiation. *Nat Rev Mol Cell Biol*.
- ZLATANOVA, J. & CAIAFA, P. 2009. CTCF and its protein partners: divide and rule? *J Cell Sci*, 122, 1275-84.