Mechanisms of cohesin mediated regulation of gene expression

by

Preksha Gupta

A thesis submitted to Imperial College London

for the degree of Doctor of Philosophy

MRC Clinical Sciences Centre

Imperial College School of Medicine

October, 2014

I, Preksha Gupta, declare that the work presented in this thesis is my own, and that any work carried out by others has been acknowledged and appropriately referenced in the text.

Dedication

I dedicate this thesis to my teachers, with special thanks to Mummy, Papa, Indrani Bordoloi, Rajeev Kumar, Pankaj Dwivedi, Anita Desai, Rashmi Shrivastava, Dr. K.C. Sarath, Dr. A.K. Kavathekar, Dr. P. Dhanraj, Dr. Nandita Narayanswami, Dr. Manu Anantpadma, Dr. Gopal K. Chowdhary, Dr. D.P. Sarkar, Dr. Rohini Muthuswami and Dr. R.N.K. Bamezai.

Abstract

The cohesin complex is essential for proper sister chromatid segregation during cell division and post-replicative DNA repair. Cohesin is also important for the regulation of gene expression. However, the mechanisms by which cohesin impacts gene expression remain incompletely understood. Owing to its vital role in cell division and DNA repair, loss of cohesin can indirectly impact gene expression programme in cycling cells. Thus, in order to investigate cohesin's role in gene regulation, a conditional knockout system was used which allowed rapid depletion of the cohesin subunit RAD21 and avoided secondary stress-induced effects on gene expression. Acute depletion of cohesin in mouse embryonic stem (ES) cells did not lead to a global collapse in pluripotency. Instead, the impact of cohesin depletion was limited to about 600 genes and was locus-specific in terms of direction of deregulation. A subset of deregulated genes was selected based on the positioning of *cis*-regulatory elements and relevance to the pluripotent state and the role of cohesin in mediating long-range interactions was analysed using chromosome conformation capture (3C). Interestingly, cohesin binding, DNA looping and transcriptional changes were not always correlated. At some of the loci tested, these interactions were maintained after removal of cohesin, questioning models where cohesin regulates gene expression solely by mediating long-range interactions.

One of the pluripotency factors affected by cohesin depletion in ES cells was *Myc*. Experiments analysing the expression of *Myc* showed that it was post-transcriptionally upregulated, specifically in cohesin deficient ES cells growing in defined media supplemented with ERK and GSK3β inhibitors (2i media). Further investigation revealed that contrary to the previously reported downregulation of *Myc* upon cohesin depletion, cohesin was not essential for *Myc* expression in various cell types.

In separate experiments, I investigated if cohesin was required for the transcriptional activation of a silent gene in response to extracellular stimuli. Results from IFNγ induction of cohesin deficient non-cycling mouse embryonic fibroblasts (MEFs) showed that cohesin was important for the activation of MHC class II genes and their master regulator *Ciita*. The expression of MHC class I genes and the associated regulatory factors remained unaffected by cohesin depletion. Further evidence is provided for the involvement of cohesin in regulating transcription by modulating RNA polymerase processivity and through the action of PTIP subunit of the MLL

complexes. Altogether my work gives an insight into the role of cohesin in mediating long-range DNA interactions important for regulation of gene expression and explores novel mechanisms of gene activation by cohesin.

Acknowledgements

First and foremost, I would like to thank Matthias and Mandy for giving me this opportunity to be a part of exciting scientific research. I am particularly grateful to Matthias for his constant guidance and constructive criticism, stimulating discussions, and especially for the freedom and encouragement to explore new ideas. I want to thank Mandy for being such an amazing example of maintaining a successful work-life balance. I am also extremely thankful to the Commonwealth Scholarship Commission for the financial support and for making my stay in UK an enjoyable one.

I am grateful to every past and present member of the Lymphocyte Development group for their help and advice. I am thankful to Hegias for getting me interested into cohesin biology and for taking out time to troubleshoot my endless failed experiments; to Luke for all the easy discussions; and to Antoine and Isa for answering all those random emergency questions at random times. I am also grateful to Thais, Lesly, Stephan and Vlad for critical reading of this thesis. Hakan, thanks for constantly asking me “How are you?” on all those bad days. Thanks to Andy, Lee, Allifia and Hakan for all the ‘extracurricular’ knowledge I have gained in the PhD office and for all those precious moments of laughter. Anne, Lesly and Feng, thank you for being such amazing people. I feel blessed to have you as friends. Lesly, you have been an inspiration for me. Anna, Tathiana, Tom, Bryony, Rory, Francesco, Cynthia, Kotryna, Liz, Grainne, Ziwei, Matthew, Irene, Sergi, Ludovica, Jorge, Amelie, I thank you all for making this journey an interesting one.

Matteo, Joana, Silvia, Joao, Marie Therese, I am glad to have had your company as friends. Thanks for all the amazing house parties! Thanks to the Indian gang – Shichina, Prashant, Sanjay, Gopu, Kedar, Surender, Malika, for the delicious get-togethers, exciting discussions and friendly advice. Gopu, thanks also for all your help with the bioinformatics. Adi, thank you for hearing me out and being the persistent silent support. Thank you Goutam for inspiring me to constant perseverance, hard work and for never letting me feel lonely. Abhishek, I am thankful for our endless discussions and the happy times spent together.

Shalini mamiji, Sharad mamaji and Mahika, I am grateful to you for your constant support and being my family away from home. Mummy, Papa, Prakhar, thank you for your endless encouragement, love, care and affection. I am here only because you made it possible! Nanaji, thank you for your unwavering faith and confidence in me. I will always be grateful for your blessings.

Table of Contents

Figures and Tables

List of Figures

List of Tables

Abbreviations

µ	Micro
2i	Small molecule inhibitors of GSK3β and MEK signalling
3C	Chromosome conformation capture
4'-OHT	4'-Hydroxytamoxifen
AML	Acute myeloid leukaemia
APCs	Antigen presenting cells
ATP	Adenosine Triphosphate
BAC	Bacterial artificial chromosome
bp	Base pair
CdLS	Cornelia de Lange Syndrome
cDNA	Complementary DNA
ChIA-PET	Chromatin Interaction Analysis by Paired-End Tag Sequencing
ChIP	Chromatin immunoprecipitation
ChIP-seq	ChIP followed by high-throughput sequencing
CpG	Cytosine-guanine dinucleotide
CTCF	CCCTC-binding factor
CTD	Carboxy terminal domain
d	Day
DAPI	4',6-diamidino-2-phenylindole
DMEM	Dulbecco's Modified Eagle's Medium
DN	Double negative
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DP	Double positive
DTT	Dithiothreitol
dUTP	2´-Deoxyuridine, 5´-Triphosphate
EDTA	Ethylene diamine tetraacetic acid
EdU	5-ethynyl-2´-deoxyuridine
EGTA	Ethylene glycol tetraacetic acid
ERT2	Oestrogen receptor
ES	Embryonic stem
EtOH	Ethanol
FACS	Fluorescence activated cell sorting
FCS	Fetal calf serum
FGF	Fibroblast growth factor
g	Gram
GFP	Green fluorescent protein
GO Term	Gene Ontology Term
GTP	Guanosine-5'-triphosphate
H	Histone

h	Hour
hB	Epstein-Barr virus-transformed human B lymphocytes
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HLA	Human Leukocyte Antigen
ICM	Inner cell mass
IF	Immunofluorescence
IFNγ	Interferon gamma
IL	Interleukin
IMDM	Iscove's Modified Dulbecco's Medium
iPS	Induced pluripotent stem
K	Lysine
Kb	Kilobase pair
kDa	Kilo Dalton
KO	Knockout
L	Litre
LIF	Leukaemia inhibitory factor
lncRNA	Long noncoding RNA
m	Milli
M	Molar
MAPK	Mitogen activated kinase
me3	Trimethyl group
MEFs	Mouse embryonic fibroblasts
MHC I	Major Histocompatibility Complex class I genes
MHC II	Major Histocompatibility Complex class II genes
min	Minute
mRNA	Messenger RNA
NP-40	Nonylphenyl Polyethylene Glycol
NTD	Amino terminal domain
PBS	Phosphate buffered saline
PcG	Polycomb group proteins
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PI	Propidium iodide
PI	Propidium iodide
PIPES	Piperazine-N,N'-bis(2-ethanesulfonic acid)
Pol II	RNA polymerase II
Pre-RC	pre-replication complex
qPCR	Quantitative polymerase chain reaction
RBS	Roberts syndrome
RIPA	Radioimmunoprecipitation assay buffer
RNA	Ribonucleic acid
RNAi	RNA interference
RNA-seq	RNA extraction followed by high-throughput sequencing

rpm	Rotations per minute
S. cerevisiae	*Saccharomyces cerevisiae* (budding yeast)
S. pombe	*Saccharomyces pombe* (fission yeast)
SDS	Sodium dodecyl sulphate
Ser	Serine
SMC	Structural maintenance of chromosomes
SP	Single positive
STI	Imatinib (STI 571)
TF	Transcription factor
Tris	tris(hydroxymethyl)aminomethane
TSA	Trichostatin A
TSS	Transcription start site
U	Units
WT	Wildtype

Chapter 1 : Introduction

The genetic material present inside a cell forms the very basis of life ranging from unicellular to multicellular organisms. One of the key processes required for the continuation of life is the faithful transmission of this genetic material from one generation to the next through cell division. In eukaryotic cells, DNA has to be accurately replicated followed by a precise segregation of the resulting sister chromatids between the daughter cells. The cohesin protein complex, plays a critical role in this process by holding the two replicated sister chromatids together from the time of their synthesis. This facilitates efficient DNA repair by homologous recombination and proper chromosome alignment at the spindle apparatus until they are separated later in mitosis and meiosis (Hirano, 2006; Jeppsson et al., 2014). Another important attribute of the genome is its role in determining the cellular identity, a process which requires the establishment of specific gene expression programmes. Cohesin's role in this process of modulating gene expression is increasingly being appreciated now as it has emerged as an important contributor to the spatial organisation of chromatin within the nucleus (Merkenschlager, 2010). However, much of the mechanistic details of how cohesin is involved in the intricate regulation of gene expression remain to be elucidated and require further investigation. Given the extensive implications of this remarkable architectural complex in controlling genomic integrity and function, a thorough analysis of its mechanisms of action is of immense significance.

1.1 The cohesin protein complex and its role in the cell cycle

Structural maintenance of chromosomes (SMC) complexes, comprising of cohesin, condensin and the Smc5/6 complexes in eukaryotic cells , are central regulators of chromosome dynamics and are important for the control of sister chromatid cohesion, chromosome condensation, DNA replication, DNA repair (Jeppsson et al., 2014). The cohesin protein complex, formed of the Smc1-Smc3 heterodimer, establishes sister chromatid cohesion between duplicating sister chromatids in S phase. The bulk of these cohesin complexes are released at the onset of mitosis in prophase and the chromosomes undergo extensive condensation. Chromosome condensation is facilitated by the progressive loading of the condensin complexes composed of the Smc2-Smc4 subunits. This results in the formation of metaphase chromosomes with well-resolved sister chromatids due to compaction of each arm of the chromosome. The

condensed sister chromatids are finally segregated into daughter cells at the onset of anaphase triggered by their separation due to proteolytic cleavage of the centromeric cohesin (Figure 1.1). Thus, the cohesin and condensin SMC complexes each perform distinct functions to facilitate proper chromosome organisation and orientation essential for their faithful transmission into daughter cells (Shintomi and Hirano, 2010).

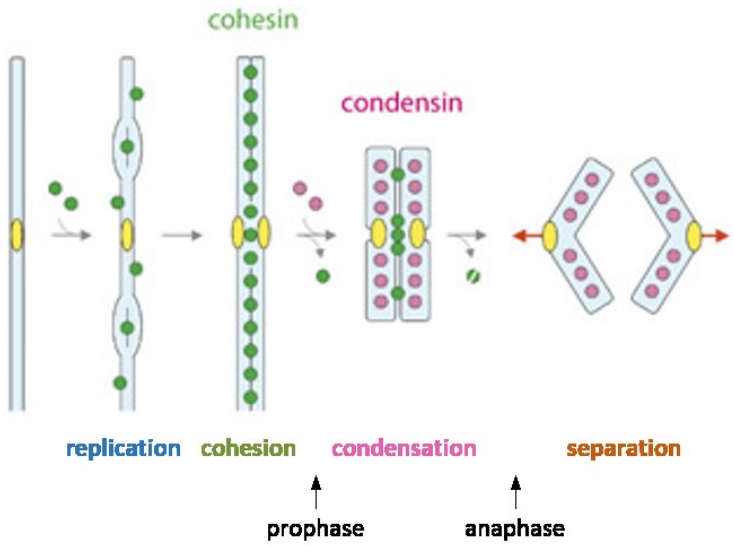


Figure 1.1 Schematic representation of the role of SMC complexes in sister chromatid segregation during cell cycle. Replicated sister chromatids are held together by the action of cohesin during S phase. At the onset of mitosis, bulk of the cohesin dissociates from chromosome arms whereas condensin associates with them to induce condensation. These processes lead to the formation of metaphase chromosomes in which sister chromatids are microscopically distinguishable from each other. In late mitosis, residual cohesin is cleaved, thereby facilitating separation of sister chromatids towards opposite spindle poles (Adapted from Shintomi and Hirano, 2010).

Proteins of the cohesin complex were identified in several independent genetic screens for mutants that were defective in chromosome segregation or DNA damage repair (Michaelis et al., 1997; Losada et al., 1998). The subunits of the core complex and its associated regulatory proteins which indirectly contribute to cohesion have been highly conserved during evolution and have several orthologs in animal cells (see Table 1 for nomenclature of subunits in eukaryotes) (Sumara et al., 2000; Remeseiro and Losada, 2013). Together they help the cohesin ring complex to mediate sister chromatid cohesion in a cell-cycle regulated manner to ensure proper segregation of chromosomes during cell division.

Table 1.1 Regulatory factors and subunits of the cohesin complex

	S. cerevisiae	*S. pombe*	*D. melanogaster*	*X. laevis*	*H. sapiens / M. musculus*
SMC	Smc1 Smc3	Psm1 Psm3	Smc1 Smc3	Smc1 Smc3	Smc1α Smc1β Smc3
α•Kleisin	Scc1/Mcd1 Rec8	Rad21 Rec8	Rad21 C(2)M	Rad21	Scc1/Rad21 Rad21L, Rec8
α•Kleisin interacting subunits	Scc3	Psc3 Rec11	SA	SA1, SA2	SA1/Stag1, SA2/Stag2 SA3/Stag3
Regulatory factors	Pds5 Rad61/Wapl -	Pds5 Wapl -	Pds5 Wapl Dalmatian	Pds5A, Pds5B Wapl Sororin	Pds5A, Pds5B/APRIN Wapl/Wapal Sororin
Cohesin loading complex	Scc4	Mis4 Ssl3	Nipped-B Scc4	Scc2 Scc4	Nipbl/Scc2 Mau2/Scc4
Cohesin Acetyl Transferases (CoATs)	Eco1/Ctf7	Eso1	Deco, San	Esco1, Esco2	Esco1, Esco2
Cohesin Deacetylases (CoDACs)	Hos1	-	-	-	HDAC8

Proteins in red correspond to meiotic isoforms (Remeseiro and Losada, 2013)

1.1.1 Molecular architecture of the core cohesin complex

The core cohesin complex consists of a heterodimer of two SMC (structural maintenance of chromosomes) proteins, SMC1A and SMC3, and two non-SMC proteins, RAD21 (Mcd1/Scc1 in *S. cerevisiae*) and STAG1 or STAG2 (Scc3 in *S. cerevisiae* and SA in *D. melanogaster*) (Figure 1.2). The SMC proteins are large ATPases possessing globular N- and C- terminal domains separated by a long amphipathic α-helix interrupted by a central globular domain. The polypeptide chains of SMC proteins fold back on themselves around the central “hinge” domain to form an intramolecular coiled-coil structure with the N- and C- terminal sequences forming a terminal ATPase “head”. The hinge domains can tightly bind each other and allow heterodimerization of the Smc1 and Smc3 subunits (Haering et al., 2002; Hirano and Hirano, 2002). Scc1, belonging to the kleisin family of proteins, connects the ATPase domains of Smc1 and Smc3 creating a tripartite ring. The N-terminus of Scc1 binds the ATPase domain of Smc3 and the C-terminus binds to Smc1 (Schleiffer et al., 2003). Scc1 is further associated with the fourth subunit Scc3 (Losada et al., 2000). The structure of Scc3 has not been determined yet and its functions are not well understood. It is thought to interact with several proteins including CTCF (Rubio et al., 2008).

High-resolution microscopy and biochemical studies show that the cohesin ring structure thus formed has a diameter of 40nm, considerably larger than an extended 10nm nucleosomal chromatin fibre. Additionally, the findings that opening of the cohesin ring by site-specific proteolytic cleavage of Scc1 or Smc3 is sufficient to release cohesin from chromosomes, suggest

that the ring structure can topologically embrace two sister chromatids (Haering et al., 2002; Gruber et al., 2003). The ring structure thus provides the vital feature of the cohesin complex which allows it to entrap DNA strands and mediate cohesion.

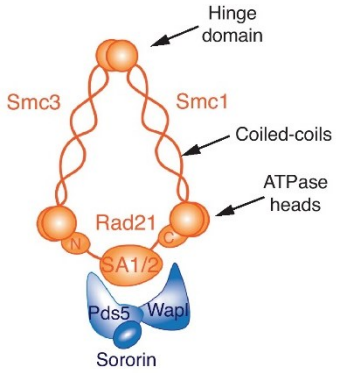


Figure 1.2 The molecular framework of the cohesin complex and its associated proteins. The core ring shaped structure is formed of four subunits (colored in orange) – Smc1, Smc3, Rad21 and SA1/2. The Smc1 and Smc3 polypeptide chains fold back on themselves to form anti-parallel coiled-coil with a 'hinge' domain at one end and a globular 'ATPase' head at the other. The hinge domains of Smc1 and Smc3 associate with each other through strong intermolecular interactions while the ATPase heads are bridged by the Rad21 subunit. Rad21 is further associated with the SA1/SA2 (Scc3) subunit. The binding of Pds5, Wapl and Sororin – the regulatory factors (in blue), helps modulate cohesin association with DNA (Losada, 2014).

1.1.2 Loading, establishment and removal of the cohesin complex during cell cycle

The association of the cohesin complex subunits with chromatin is highly regulated and involves the interplay between the activities of several associated regulatory proteins. In budding yeast, cohesin is loaded onto chromosomes at the end of G1 phase (Guacci et al., 1997; Michaelis et al., 1997). In vertebrates, however, the loading is initiated already in telophase following reformation of the nuclear envelope (Losada et al., 1998; Sumara et al., 2000). In *S. cerevisiae* this process requires the activity of the heterodimeric cohesin-loading factor Scc2-Scc4 (Ciosk et al., 2000) and the ability of Smc1-Smc3 to bind and hydrolyse ATP (Arumugam et al., 2003; Weitzer et al., 2003). In addition, it has been proposed that the Smc dimer has to be opened at the hinge region in order to permit DNA entry (Gruber et al., 2006). It is therefore possible that the Scc2/Scc4 complex promotes loading of cohesin onto DNA by stimulating its ATPase activity, which might in turn allow transient opening of the hinge domain. Human cohesin also requires NIPBL (yeast Scc2 homolog) and its partner MAU2 (yeast Scc4 homolog) for chromatin loading (Seitan et al., 2006; Watrin et al., 2006). In *Xenopus* egg extracts, Scc2/Scc4 is recruited to the assembly of pre-replicative complexes (pre-RCs) on DNA (Takahashi et al., 2004). In budding yeast, however, the pre-RC protein Cdc6 is dispensable for cohesin loading (Uhlmann and Nasmyth, 1998), suggesting that Scc2/Scc4 recruitment to DNA may occur by different mechanisms in different species. In fact, recent biochemical studies reconstituting the loading reaction onto

naked DNA indicate that cohesin has an intrinsic ability to load topologically on DNA but the process is inefficient unless NIPBL is present (Murayama and Uhlmann, 2014).

Cohesin loading, however, does not ensure its stable association with DNA. Live cell imaging studies suggest that cohesin is constantly being exchanged from chromatin throughout interphase with a residence time of less than 25 minutes. During S phase, the equilibrium shifts towards a more stable association with chromatin and the half-life of cohesin binding increases considerably (Gerlich et al., 2006). The rapid turnover of bound cohesin complexes has been attributed to the 'anti-establishment' activity of WAPL (wings apart-like protein homolog) and PDS5. Wapl was identified as a regulator of mitotic chromosome morphology in *Drosophila* (Vernì et al., 2000). Two recent studies also showed that Wapl inactivation stabilized cohesin on chromosomes in interphase and cells displayed chromosome segregation errors (Haarhuis et al., 2013; Tedeschi et al., 2013). It is thought that Wapl releases cohesin from chromatin by transiently opening the ring gate between Smc3 and Rad21 by binding to the ATPase head of the Smc3 subunit (Chan et al., 2012; Buheitel and Stemmann, 2013; Chatterjee et al., 2013; Eichinger et al., 2013). As a result, the fraction of cohesin that is bound to chromatin is an outcome of the opposing actions of NIPBL-MAU2 and PDS5-WAPL.

During S phase, the cohesin complex entraps the replicated DNA strands and establishes stable cohesion. In order to do so, cells need to antagonise the cohesin destabilising activity of WAPL. This process requires the acetylation of two lysine residues in the SMC3 head domain by the cohesin acetyl-transferases (CoATs) ESCO1 and ESCO2 (Eco1 in yeast), as well as the binding of the protein sororin to PDS5 (Hou and Zou, 2005; Rolef Ben-Shahar et al., 2008; Unal et al., 2008; Zhang et al., 2008; Nishiyama et al., 2010). The binding of Sororin to PDS5 has been proposed to displace WAPL, thereby preventing its unloading action. Smc3 acetylation also facilitates DNA replication fork progression. The restriction of cohesion establishment to S-phase can thus be attributed to the cell cycle regulation of CoAT and its interaction with the components of the DNA replication machinery (Lyons and Morgan, 2011; Higashi et al., 2012).

Most cohesin is released from chromatin at the onset of mitosis by the prophase dissociation pathway. This requires the activity of three protein kinases – cyclin-dependent kinase 1 (CDK1), aurora kinase B (AURKB) and polo-like kinase 1 (PLK1). CDK1 and AURKB phosphorylate sororin to drive its dissociation from PDS5, thus allowing PDS5-WAPL to unload cohesin. PLK1 phosphorylates the SA subunit, further facilitating cohesin release (Shintomi and Hirano, 2010).

However, a small proportion of cohesin, mostly enriched at centromeres is protected from the prophase pathway by the action of shugoshin 1 (SGO1) bound to protein phosphatase 2A (PP2A). SGO1-PP2A recognises cohesin bound sororin and prevents its phosphorylation. This centromeric cohesin allows the alignment of chromosomes at the metaphase plate. Activation of anaphase promoting complex (APC) at the onset of anaphase leads to the degradation of securin and activation of separase. Separase cleaves the Rad21 subunit, thereby destroying the integrity of the ring and allows separation of the sister chromatids (Gutiérrez-Caballero et al., 2012). The cohesin complexes released during mitosis by the prophase pathway can be reused in the following G1 phase. This, however, requires cohesin to be deacetylated and the task is performed by cohesin deacetylases (CoDACs) Hos1 in yeast and HDAC8 in humans (Beckouët et al., 2010; Borges et al., 2010; Deardorff et al., 2012a). This whole process of cohesin loading, establishment, unloading and reuse can be depicted in the form of a regulatory cycle as in Figure 1.3.

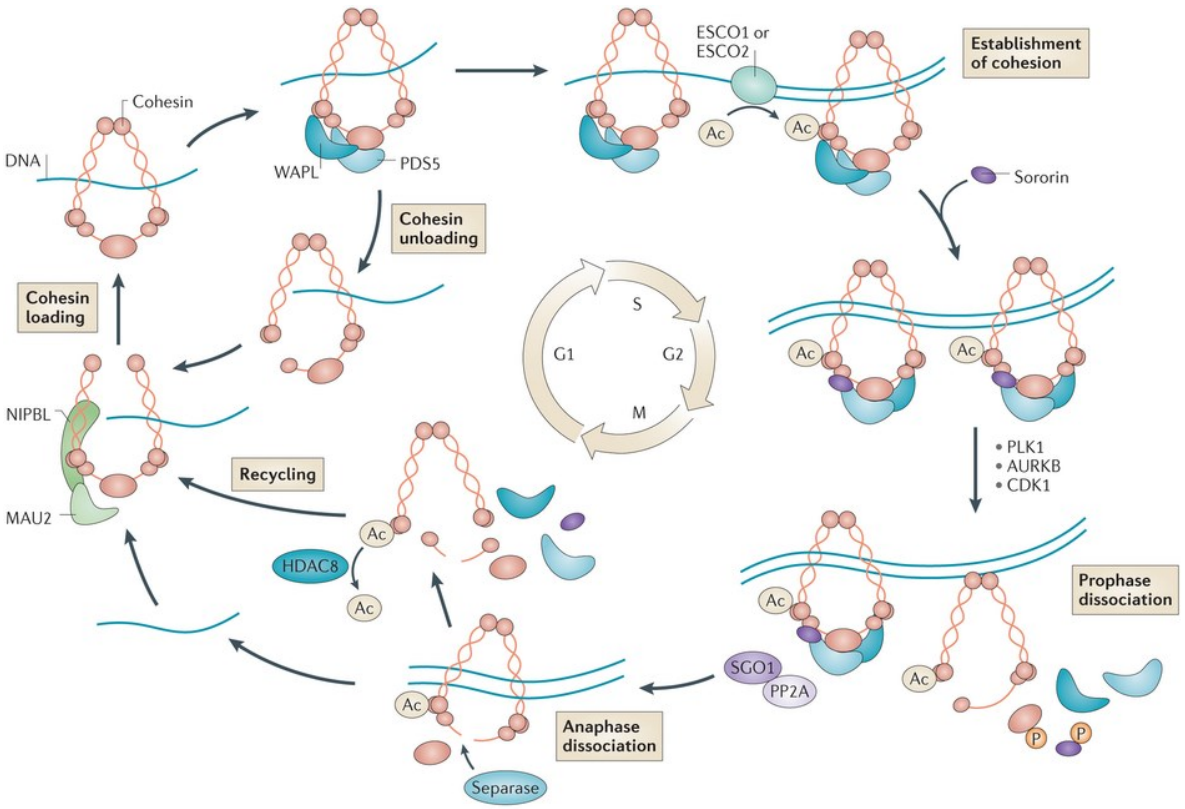


Figure 1.3 The cohesin cycle. Cohesin is loaded onto chromatin during late mitosis and early G1 phase by the NIPBL-MAU2 heterodimer and is maintained in a dynamic equilibrium by the unloading action of PDS5 and WAPL. Upon initiation of DNA replication, ESCO1/ESCO2 acetylates SMC3 leading to the recruitment of sororin to PDS5. Sororin displaces WAPL and establishes stable cohesion between replicated sister chromatids. In prophase, most of the sororin, except that protected by SGO1-PP2A at centromeres, is released and cohesin is unloaded. At the onset of anaphase, separase cleaves the centromeric RAD21 and allows sister chromatid segregation. The released cohesin can then be reused after deacetylation by HDAC8 (Losada, 2014).

1.1.3 Canonical cohesin functions

The primary functions of the cohesin complex are attributed to its ability to entrap DNA strands within the ring structure and provide cohesion. The precise regulation of its association with chromatin during the cell cycle allows cohesin to facilitate faithful chromosome segregation. During cell division, cohesin holds the newly replicated sister chromatids together. This helps to achieve a proper back-to-back orientation of the sister-kinetochores at the metaphase plate and their attachment to microtubules from opposite spindle poles. Cohesion prevents premature separation of sister chromatids under the pulling forces of spindle fibres until all chromosomes achieve bipolar attachment (biorientation). At the onset of anaphase, cohesin rings are opened, dissolving the cohesive forces which allows the segregation of one copy of the replicated DNA to each daughter cell (Losada, 2014).

Cohesin also has an important role in maintaining genome stability through post-replicative DNA double-strand break (DSB) repair in mitotic and meiotic cells (Klein et al., 1999; Sjögren and Nasmyth, 2001). In mitotic cells, cohesion can be established in response to DNA damage in G2 phase in the absence of DNA replication. Cohesion then facilitates the use of sister chromatid as the template for the repair of the double strand break through homologous recombination. In meiotic cells, cohesin holds the bivalent chromosomes together during chiasmata formation (reciprocal recombination event) where programmed DNA double-strand breaks are repaired preferentially using non-sister chromatids (Peters et al., 2008; Nasmyth and Haering, 2009). In addition to homologous recombination mediated DNA repair, cohesin is also involved in DNA damage checkpoint activation during S and G2/M phase transitions (Wu and Yu, 2012; Ball et al., 2014). Increasing evidence suggests that cohesin also plays a role in stabilizing stalled DNA replication forks at regions which are difficult to replicate, such as telomeres and centromeres, and promotes their restart (Remeseiro et al., 2012; Carretero et al., 2013).

1.2 Non-canonical cohesin functions and their role in disease and development

As cohesin is essential for chromosome segregation during cell division, a homozygous null mutation in any of the core complex subunits or the regulatory factors would be deleterious for a cell and result in embryonic lethality. However, hypomorphic and/or heterozygous mutations in cohesin subunits and its regulators have been observed in several human malignancies and genetic disorders collectively known as cohesinopathies. It is also thought that infertility and increased frequency of children with birth defects in ageing women could be due to aneuploidy resulting from reduced cohesion in oocytes that have been arrested in G2 phase for decades (Jessberger, 2012). It is suspected that increased chromosomal instability may favour cancerous growth of cells, but interestingly, patient cells with cohesinopathies show limited cohesion defects and instead display altered transcriptional profiles. Studies in several model organisms have demonstrated that cohesin plays an important role in regulating gene expression. Nonetheless, much remains to be understood about how mutations in cohesin affect the transcriptional activity of cells leading to developmental defects. In order to be able to deduce the mechanistic links, features of some cohesinopathies and cohesin associated cancers along with studies in model organisms will be discussed below in detail.

1.2.1 Cohesinopathies

Human genetic disorders related to dysfunction of cohesin and its associated regulators are collectively known as cohesinopathies. They are characterised by both physical and mental developmental anomalies. The most prevalent cohesinopathy is Cornelia de Lange Syndrome (CdLS). CdLS is a congenital multi-system disorder and has an incidence of between 1:10,000 and 1:30,000 live births. Classical CdLS patients show pre- and postnatal growth retardation, microcephaly, developmental delay, cognitive impairment, facial dysmorphia, hirsutism and upper limb defects ranging from small hands to severe forms of oligodactyly and truncation of the forearms. Typical features include fine arched eyebrows, low-set posteriorly rotated ears, long philtrum, thin upper lip, depressed nasal bridge and anteverted nares (Figure 1.4) (Mannini et al., 2013). Patients are also reported to have recurrent infections at high frequency accompanied by a decrease in T regulatory cells, T follicular cells and antibody deficiency (Jyonouchi et al., 2013).

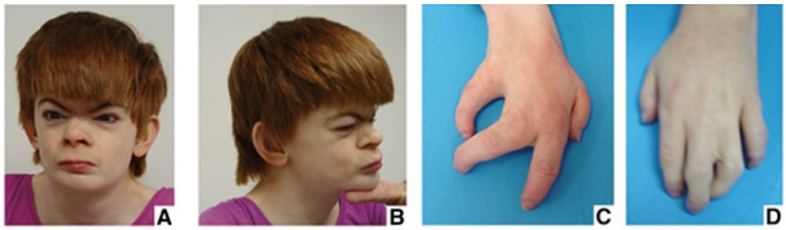


Figure 1.4 Phenotypic characteristics of CdLS patients. (A-D) 28 year old girl with truncating mutations in *NIPBL* showing typical developmental limb defects and craniofacial abnormalities.

CdLS is a genetically heterogenous diagnosis with heterozygous mutations in *NIPBL* found in at least 60% of CdLS probands. Patients with frameshift or nonsense mutations of *NIPBL* that result in NIPBL haploinsufficiency often exhibit more severe phenotypes compared to missense mutations (Gillis et al., 2004). Mutations in X-linked *SMC1* and *SMC3* on chromosome 10 are also found in a minor subset of clinically milder CdLS cases (~5% and <1% respectively). SMC mutations are mostly missense mutations and patients show mental retardation as the primary symptom, with other abnormalities being fewer and/or milder (Musio et al., 2006; Deardorff et al., 2007). Non-sense or missense mutations that cause a loss of HDAC8 activity with a concomitant SMC3 hyperacetylation and increased chromatin retention of cohesin during mitosis, have also been found in a subset of CdLS patients with phenotype similar to those of NIPBL mutations (Deardorff et al., 2012a). Unlike SMC mutations, CdLS patients with *RAD21* mutations exhibit classical CdLS phenotypic characteristics but have milder cognitive impairment (Deardorff et al., 2012b). Overall, CdLS patients with *NIPBL, SMC1/3, RAD21, HDAC8* display specific yet overlapping clinical features encompassing neuro-developmental defects (Ball et al., 2014). Cells from CdLS patients do not display cohesion defects, although they do show increased sensitivity to DNA damage (Dorsett and Ström, 2012) and show changes in gene expression (Liu et al., 2009).

Another cohesinopathy, Roberts Syndrome (RBS) or SC Phocomelia, is a rare disorder caused by homozygous mutations in *ESCO2* (Vega et al., 2005). RBS patients have a wide range of clinical phenotypes that include upper and lower limb defects, growth retardation, craniofacial anomalies and mental retardation with limited similarity to CdLS patients. RBS chromosomes exhibit premature centromere separation and heterochromatin puffing, indicative of a sister chromatid cohesion defect (Tomkins et al., 1979). Additionally, RBS cells show increased apoptosis and also display slower replication fork progression, which may contribute to reduced proliferation of critical progeniters (Terret et al., 2009). More recently, it was observed that *Eco1*, and to a lesser extent, *Smc1* mutations (but not *Scc2* mutations), lead to decreased rRNA production and ribosomal biogenesis resulting in translational defects. Similar defects were seen

in RBS patients with *ESCO2* mutations (Bose et al., 2012). Increasing evidence now also points towards involvement of transcriptional dysregulation in RBS as ESCO2 has been implicated in recruitment of chromatin modifiers apart from its acetyltransferase activity (Kim et al., 2008a; Choi et al., 2010).

Genome-wide transcriptional analysis of 16 mutant cell lines from severely affected CdLS probands was used to identify a unique profile of dysregulated gene expression. This was further validated in 101 patient samples and serves as a diagnostic and classification tool (Liu et al., 2009). This gene set was also shown to be deregulated in two tested RBS probands indicating an overlap in transcriptional abnormalities, consistent with the similarity of developmental defects observed in the two diseases. Furthermore, it is speculated that other genetic mutation disorders displaying similar phenotypes as CdLS and increased genotoxic sensitivity, might fall under the same umbrella of cohesin associated birth defects and may help explain the occurrence of CdLS in the remaining 35% of the probands (Skibbens et al., 2013).

1.2.2 Cohesin mutations in human malignancies

Mutations in genes encoding cohesin subunits and its regulators have been identified in several types of tumours. Initially, *NIPBL* mutations were identified in colorectal cancer (Barber et al., 2008) and later *STAG2* mutations were found in glioblastoma, Ewing's sarcoma and melanoma. *STAG2* mutations are most common in urothelial bladder cancer (Solomon et al., 2013). In AML, Down syndrome related acute megakaryocytic leukaemia and other myeloid neoplasms, however, mutations across most cohesin subunits have been described (Kon et al., 2013). In a recent exome sequencing study of 4,742 human cancer samples across 21 cancer types, *STAG2* was identified as one of the 12 genes that are mutated at substantial frequencies in at least four tumour types (Lawrence et al., 2014). Although most identified mutations are heterozygous, the presence of *SMC1A* and *STAG2* genes on the X-chromosome can make their mutation functionally homozygous, at least in males and in somatic cells in females with randomly inactivated X-chromosome. Since a single hit is sufficient for the loss of STAG2 function and STAG1 might partially compensate for its loss, *STAG2* mutations might be observed at a higher rate (Losada, 2014).

Cohesin dysfunction could affect tumorigenesis by increasing genomic instability due to faulty DNA replication and/or repair and chromosome segregation. Even though aneuploidy and genomic instability are detrimental to cell survival, they can favour tumour formation. However,

an association between aneuploidy and cohesin mutations in cancer has only been reported in some studies but not in others (Solomon et al., 2011; Balbás-Martínez et al., 2013). The contribution of cohesin mutations in deregulating the expression of crucial tumour suppressors or oncogenes could be important but remains to be investigated.

1.2.3 Non-mitotic cohesin functions in gene regulation and development

Cohesin functions beyond its primary role in sister chromatid cohesion were first suggested based on the observation that in vertebrate cells cohesin is loaded onto DNA during telophase (i.e. long before cohesion is established) and the bulk of it dissociates again in prophase (i.e. before cohesion is dissolved) (Losada et al., 1998; Sumara et al., 2000; Gerlich et al., 2006). Cohesin was also found associated with DNA in post-mitotic neurons which do not replicate DNA again and hence don't require cohesion (Wendt et al., 2008). The idea that these non-canonical functions might be related to chromatin structure and gene expression were first sparked by genetic experiments in yeast and *Drosophila.* In *S. cerevisiae*, mutations in *Smc1* and *Smc3* inactivated the boundary elements that prevent the spread of heterochromatin from the silent HMR locus into neighbouring regions (Donze et al., 1999). In *Drosophila* wing margin cells, Nipped-B, the fly ortholog of Scc2, was discovered as a protein required for the activation of *cut* and *Ultrabithorax* homeobox genes and was speculated to facilitate enhancer-promoter communication (Rollins et al., 1999). Later, *Drosophila Wapl* mutants were identified which showed defects in position effect variegation (Vernì et al., 2000). Further evidence for cohesin's role in gene regulation came from studies which reported developmental defects in model organisms due to mutations in cohesin and cohesin regulators. MAU-2 mutants were found to be defective in axon guidance (Bénard et al., 2004; Seitan et al., 2006) while *Rad21* mutants in *Drosophila* displayed severe axon pruning defects during nervous system development (Pauli et al., 2008). In zebrafish, cohesin is required for the transcription of runx transcription factors and hematopoiesis (Horsfield et al., 2007) while reduced dosage of Nipbl lead to multiple heart and gut defects with no chromosome segregation defects (Muto et al., 2011).

To increase the understanding of cohesin's role in development, several mouse models have been used. Mice partially deficient for *Nipbl* (~30% reduction) recapitulate several features of CdLS and display modest but significant transcriptional deregulation of many genes (Kawauchi et al., 2009). Neural crest cell-specific inactivation of *Nipbl* or *Mau2* during mouse development results in craniofacial defects (Smith et al., 2014). Mouse embryos lacking the SA1 subunit show

a clear developmental delay and die before birth. They display defects in telomere cohesion along with altered transcriptional profiles related to CdLS (Remeseiro et al., 2012). Mice lacking PDS5 also die perinatally and show developmental defects resembling CdLS pathology (Zhang et al., 2007). Additionally, ESCO2 deficiency in mice results in very early embryonic lethality (Whelan et al., 2012). Similarly, *Wapl*$^{-/-}$ mice die before birth (Tedeschi et al., 2013).

As most cohesin deficient mouse models show early lethality, scope for detailed analysis is limited. Nonetheless, strategies have been developed to study locus-specific effects of cohesin depletion by knockdown or conditional knockout methods. These studies have reinforced the role of cohesin in transcription which will be discussed in detail later. Experiments also show that different cohesin functions require different amounts of cohesin. In budding yeast, even 13% of normal cohesin levels are enough to support sister chromatid cohesion but cells show defects in DNA repair and chromosome condensation (Heidinger-Pauli et al., 2010). Likewise, reduction of cohesin levels by 80% in *Drosophila* cells has dramatic effects on gene expression, but has no significant effect on cohesion or chromosome segregation (Schaaf et al., 2009). These studies along with observations in CdLS patients, suggest that gene transcription is more sensitive to perturbations in cohesin levels while the more conserved cohesin functions in cohesion are more resistant to cohesin dosage.

An important consideration while studying effects of cohesin depletion on gene expression is the dissociation of its functions in gene regulation from its essential functions in cell cycle and sister chromatid cohesion. As depicted in Figure 1.5, gene expression changes observed can be a direct or an indirect consequence of cohesin depletion.

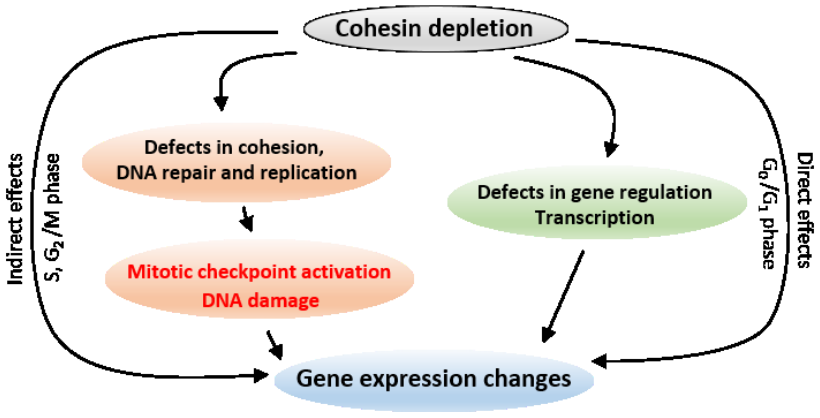


Figure 1.5 Direct and indirect effects of cohesin depletion. Cohesin depletion in cycling cells may lead to indirect effects on gene expression due to activation of stress response pathways upon perturbation of cohesin's mitotic functions. A more direct effect of cohesin on transcription can instead be studied during interphase.

In cycling cells, cohesin deficiency can lead to incomplete DNA replication, accumulation of DNA damage and prolonged activation of cell cycle checkpoints, causing the cells to initiate the expression of stress response genes. Thus, the gene expression changes observed can be an indirect effect of stress signals instead of the absence of cohesin. As a result, it is important to study the role of cohesin in transcription either in non-dividing cells or at early time points before secondary effects due to stress response pathways set in. To date, studies in post-mitotic *Drosophila* neurons have provided the clearest distinction between cohesin's cell division-related and cell division-independent functions where cohesin deficient neurons showed defective axon pruning due to deregulated expression of ecdysone receptor (Pauli et al., 2008, 2010; Schuldiner et al., 2008). Studies in non-cycling mouse thymocytes further demonstrated that cohesin is required for the rearrangement of the T cell receptor alpha locus (Seitan et al., 2011) and the regulation of approximately 1000 other genes (Seitan et al., 2013).

1.3 Cohesin and transcription

Cohesin binds to DNA in association with CTCF and other tissue-specific transcription factors. It then facilitates long-range chromatin interactions, a function attributed to the ability of cohesin ring to entrap DNA strands. These prevailing concepts are discussed further below. However, much remains to be understood of this process in order to explain the impact of cohesin deficiency on gene expression.

1.3.1 Cohesin localisation in the genome

In budding yeast *S. cerevisiae* and fission yeast *S. pombe*, cohesin is primarily located downstream of active genes at sites of convergent transcription (Lengronne et al., 2004; Gullerova and Proudfoot, 2008). In budding yeast, it is believed that cohesin is loaded at active gene promoters by Scc2 and then slides along the DNA, possibly pushed by RNA polymerase. In fission yeast, however, bidirectional transcription at convergent genes causes RNAi-dependent formation of heterochromatin proteins and the recruitment of cohesin. This mechanism is thought to be important for the correct termination of transcription (Gullerova and Proudfoot, 2008). In *Drosophila*, cohesin binding almost completely co-localises with the Scc2 homolog Nipped-B. Here Nipped-B and cohesin binding is enriched at a subset of active genes as well as DNA replication origins, but largely excluded from silenced genes (Misulovin et al., 2008).

In mammalian cells, two distinct types of cohesin binding sites have been described. At active promoters and enhancers, cohesin co-localises with NIPBL, Mediator and cell-type specific transcription factors (Kagey et al., 2010; Schmidt et al., 2010; Nitzsche et al., 2011; Faure et al., 2012; Prickett et al., 2013). For example, cohesin co-localises with estrogen receptor binding in MCF7 breast cancer cells and with liver-specific transcription factors in HepG2 hepatocellular carcinoma cells (Schmidt et al., 2010). The strongest cohesin binding sites, however, show an enrichment for the consensus sequence motif of the DNA binding protein CTCF. Cohesin co-localises extensively with CTCF, and siRNA-mediated knockdown of CTCF abolishes cohesin recruitment (Parelho et al., 2008a; Rubio et al., 2008; Stedman et al., 2008; Wendt et al., 2008). It was further shown that the cohesin subunit Scc3 interacts directly with CTCF (Rubio et al., 2008). Together, these studies provide the first mechanism for the sequence-specific localisation of cohesin along the mammalian chromosome arms. CTCF functions as a transcriptional regulator and as an architectural protein at insulators, boundary elements. It also acts as a genome organiser by formation of chromatin loops (Ong and Corces, 2014). Knockdown studies indicated

the requirement of cohesin for CTCF functions and demonstrated a functional link between CTCF and cohesin (Parelho et al., 2008a; Wendt et al., 2008; Nativio et al., 2009) thus providing the first rationale for non-canonical cohesin functions. CTCF binding itself is regulated not just by DNA sequence but also by the epigenetic state of chromatin, for example DNase I hypersensitivity and DNA methylation. As a result CTCF binding and cohesin localisation on DNA is cell-type specific (Parelho et al., 2008b; Nativio et al., 2009). Together these mechanisms of cohesin localisation on DNA allow for the integration of genetic and epigenetic information to achieve highly cell type specific binding patterns.

1.3.2 Cohesin as a mediator of long-range chromatin interactions

Transcriptional regulation requires the cooperation of sequence-specific factors, chromatin modifiers and long-range interactions between gene regulatory elements. In the past, most genome organisation studies relied exclusively on the use of microscopy based techniques which lack the resolution necessary to observe individual physical interactions between DNA regulatory elements. These microscopy techniques are now complemented by the molecular technique of chromosome conformation capture (3C) (Dekker et al., 2002). The basic methodology involves the use of chemical crosslinking to secure 3D contacts between genomic loci occurring in live cells. The crosslinked chromatin is then digested with a restriction enzyme and religated in a dilute solution so that only loci that were in contact in vivo (and thus fixed together by crosslinking) will be ligated together. Ideally, each ligation product should correspond to a pair of loci that were in contact in vivo at the time of crosslinking. These ligation products can then be assayed to quantify the frequency of contacts between specific loci. Several variations of the original 3C technique (4C, 5C, Hi-C) have now been developed which allow the measurement of genomic contacts with varying scope and throughput (Figure 1.6). A related technique is ChIA-PET, which couples ChIP with 3C to focus on interactions between genomic loci mediated by a protein of interest (Gorkin et al., 2014). Collectively, 3C based technologies have allowed an unprecedented view of genomic interactions and their role in regulating transcription. These studies have also established cohesin as an important contributor to long-range DNA interactions and genome organisation (Merkenschlager and Odom, 2013).

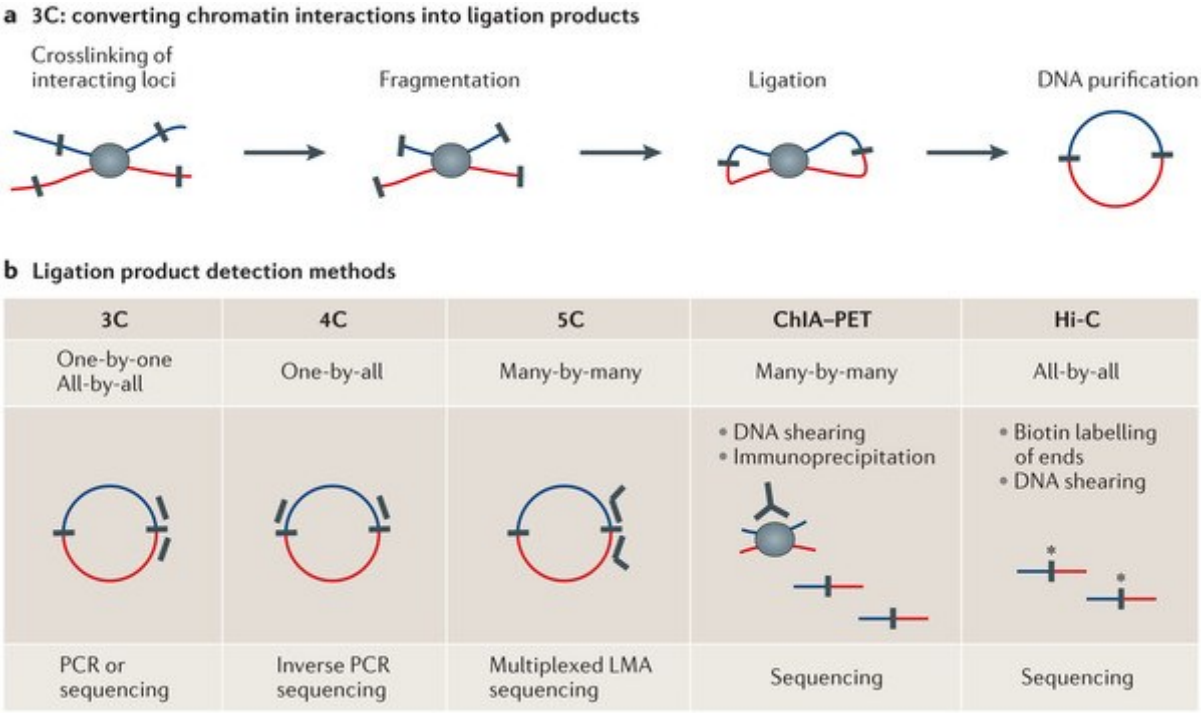


Figure 1.6 Detection of chromatin interactions by 3C based methods. a) Formaldehyde crosslinked chromatin is digested with restriction enzyme or sonication. Fragments are then religated in a diluted solution which prevents random ligation and favours ligation of crosslinked products in close proximity. The purified DNA can then be analysed as detailed in **b)**. In classical 3C experiments, single ligation products are detected by PCR using locus-specific primers. 4C uses inverse PCR to generate genome-wide interaction profiles for a single loci. 5C identifies several million interactions in parallel between two large sets of loci. Hi-C is an unbiased genome-wide adaptation of 3C where the staggered DNA ends are filled by biotinylated nucleotides and the ligation products are directly sequenced. The resolution of the map depends on the depth of sequencing. Adapted from Dekker et al., 2013.

Based on the discovery of cohesin's association with CTCF, it was hypothesised that the cohesin ring structure can connect distant CTCF associated genomic regions and form chromatin loops by entrapping the DNA strands. This dependence of CTCF-based long-range interaction on cohesin was first demonstrated for the mouse *Ifng* locus. A conserved CTCF binding site 60-70 kb upstream of the *Ifng* coding region contacts two other CTCF sites, one in the first intron of *Ifng* and the other about 100 kb downstream of the locus selectively in T helper 1 cells which inducibly express *Ifng.* Both CTCF and cohesin are required for these interactions and reduced interactions correlated with reduced expression. But while cohesin knockdown abolished these interactions, CTCF binding at these sites remained relatively unaffected. It was concluded that CTCF recruits cohesin but the local chromosome conformation of *Ifng* is defined by cohesin, not CTCF (Hadjur et al., 2009).

Several other genomic loci have since been shown to have cohesin-dependent chromatin interactions. At the imprinted *H19/IGF2* locus, cohesin is required for the CTCF-mediated chromatin loops that separate the genes into active and inactive domains. The *IGF2/H19* imprinting control region (ICR) comprises a cluster of CTCF binding sites. ICR is selectively methylated in sperm, but not in ova. Consequently, CTCF selectively binds at the unmethylated ICR of the maternal allele and acts as an insulator where it blocks the interaction of a distal enhancer with the *IGF2* promoter. This restricts the *IGF2* promoter in an inactive domain and prevents its expression. Methylation of the paternal ICR precludes CTCF binding and abrogates insulator activity. The distal enhancer can now interact with *IGF2* promoter so that paternal *IGF2* is expressed (Murrell et al., 2004). Deletion of cohesin leads to disruption of long-range interactions and changes expression levels of the IGF2 gene (Nativio et al., 2009).

Another example is the β-globin locus where cohesin is involved in both CTCF-dependent insulator interactions and CTCF-independent enhancer-promoter interactions. CTCF binds to the 5' and 3' boundaries of the locus forming a loop while the distal enhancer present in the locus control region (LCR) interacts with the globin genes present inside the loop in a developmental-stage specific manner (Splinter et al., 2006). This process requires the expression of lineage-specific transcription factors and is associated with increased binding of Nipbl and cohesin at the interaction sites of β-globin but not the foetal globin genes upon erythroid differentiation. Depletion of Nipbl or cohesin decreased both the insulator interaction and LCR enhancer-promoter interaction, but CTCF depletion only affected the insulator interaction. In accordance with this, cohesin depletion and not CTCF depletion, lead to decreased β-globin expression (Chien et al., 2011).

The examples discussed above suggest that cohesin and CTCF-mediated interactions are important for the regulation of complex loci with multicluster genes. This is also the case for the proto-cadherin loci (Kawauchi et al., 2009; Hirayama et al., 2012; Monahan et al., 2012; Remeseiro et al., 2012), the MHC class II gene cluster (Majumder and Boss, 2011), the apolipoprotein gene cluster (Mishiro et al., 2009), the *HoxA* locus (Kim et al., 2011), X chromosome inactivation region (Spencer et al., 2011) and the T cell receptor alpha (*Tcra*), immunoglobulin κ light chain, immunoglobulin heavy chain (*Igh*) lymphocyte receptor loci (Degner et al., 2011; Guo et al., 2011; Ribeiro de Almeida et al., 2011; Seitan et al., 2011). At these loci, cohesin and CTCF orchestrate and fine tune the expression of genes in a cell-type specific manner by mediating long-range DNA interactions and partitioning gene clusters into chromatin

domains. Seitan et al. (2011) first demonstrated a cell division-independent role of cohesin in transcriptional regulation in mammalian cells (Figure 1.7). They analysed the role of cohesin in the rearrangement of the *Tcrα* locus. Lymphocyte receptor loci like the *Tcrα*, contain hundreds of coding elements arranged over large genomic regions. To make functional receptors, these regions have to be rearranged. This process of somatic recombination is mediated by Rag1 and Rag2 recombinases. The recruitment of Rag2 is coupled to transcription of *Tcrα*. This provides a mechanism to rearrange the receptor in the appropriate cells at the appropriate time (Merkenschlager and Odom, 2013). Loss of the Rad21 subunit in non-dividing $CD4^+CD8^+$ double positive thymocytes (DPTs) reduced interactions between the *Tcra* gene promoter (TEA) and enhancer (Eα), thereby reducing *Tcra* transcription and rearrangement which further impaired thymocyte differentiation (Seitan et al., 2011).

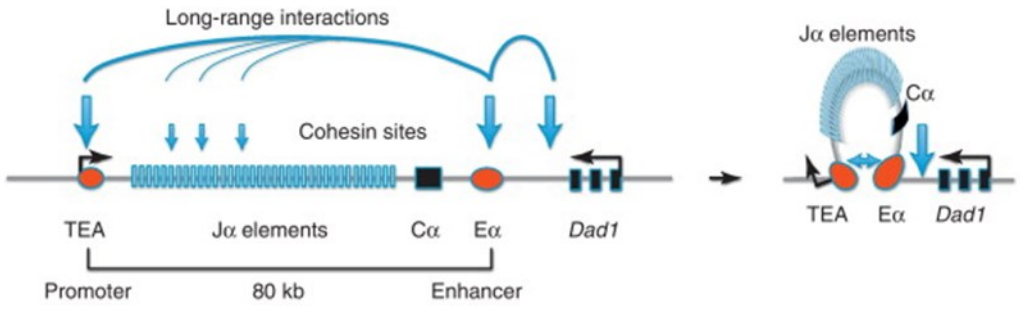


Figure 1.7 Cohesin and CTCF link regulatory elements at the Tcra locus. Cohesin binding sites flank the major regulatory elements of Tcra, the TEA promoter and the Eα enhancer. Cohesin facilitates enhancer-promoter interactions over a distance of 80kb and promotes Tcra transcription and rearrangement. Jα: Joining region with T-cell receptor elements, Cα: constant region, Dad1: neighbouring housekeeping gene, Arrows indicate long-range interactions between cohesion binding sites (Seitan et al., 2012).

In addition to its role in association with CTCF, cohesin co-localises with the loading factor Nipbl and Mediator components at enhancer elements in mouse embryonic stem (ES) cells. Cohesin bound enhancers of key pluripotency genes like *Nanog* were found to interact with their respective promoters (Kagey et al., 2010). Similarly, cohesin was found associated with a large fraction of cis-regulatory modules (CRMs) defined by the binding of multiple tissue-specific transcription factors in mouse liver cells. Additionally, Pol II was detected at nearly half of all predicted extragenic CRMs, even though most of these were not transcribed. This suggests that the Pol II at extragenic CRMs is transcriptionally engaged at a promoter, and interacts with the CRMs through cohesin-mediated looping. This interpretation is supported by the observation that

Pol II ChIP signals are reduced at a quarter of the predicted CRMs upon cohesin depletion, which is higher than the frequency of Pol II reduction at active promoters (Faure et al., 2012). Several other studies (discussed in section 1.4.2) have further elucidated cell-type specific binding of cohesin at enhancers and its role in enhancer-promoter interactions.

Collectively, a common theme emerges from these studies which suggests that cohesin contributes to gene regulation by mediating chromosomal interactions in *cis*. It can facilitate CTCF insulator function by forming a DNA loop and separating genes into an inactive domain. Conversely, it may help bring genes and their enhancers in the same loop to promote gene activation while excluding the insulators or silencers away from the gene. It is also directly associated with enhancer-promoter interactions. Loss of such cohesin-mediated interactions can thus affect gene expression causing downregulation or upregulation of the genes involved depending on the genomic context. However, much of the global gene expression changes observed upon cohesin depletion remain unexplained and are difficult to predict because of the complexity and multiplicity of these interactions at a given genetic loci. Additionally, it remains to be established whether the observed changes in DNA interaction upon cohesin depletion are the cause or a consequence of changes in gene transcription.

Genome-wide studies have further provided a more global view of cohesin localisation at *cis*-regulatory DNA elements involved in long-range interactions. Cohesin binds at active promoters marked by H3K4me3 modifications, and at putative enhancer elements, that can be identified by the presence of H3K4me1, p300, low H3K4me3 with or without H3K27ac mark (active or poised enhancers respectively) (Heintzman et al., 2009; Creyghton et al., 2010; Rada-Iglesias et al., 2011). These binding sites are usually CTCF-independent and are often associated with NIPBL and Mediator (Kagey et al., 2010). Activities of promoters and enhancers as well as their association with cohesin are highly cell-type specific and respond to external stimuli or developmental cues. Insulators, on the other hand display CTCF-cohesin binding which is largely invariant across diverse cell types (Heintzman et al., 2009; Faure et al., 2012; Phillips-Cremins et al., 2013) (Table 1.2). Such diverse cohesin based interactions are observed in developing mouse limbs as shown by SMC1A ChIA-PET (DeMare et al., 2013). Additionally, WAPL depletion in cells, which 'locks' cohesin onto chromatin, results in chromatin compaction and formation of axial (vermicelli) structures in interphase chromosomes that can be visualised by light microscopy (Tedeschi et al., 2013). This information along with observations from Hi-C based genome interactome maps of several cell-types have implicated cohesin mediated long-range DNA

interactions in not only local gene regulation but also as architectural components determining global genome organisation, albeit with distinct combinations of CTCF and other tissue-specific transcription factors (Gorkin et al., 2014).

Table 1.2 Cohesin associated genomic regulatory elements

Genomic feature	Characteristic properties	Associated chromatin features
Promoter	Located near gene TSS and directly regulates transcription	H3K4me3
Enhancer	Located distal to genes and may require juxtaposition onto respective gene to enhance transcription	H3K4me1, p300, low H3K4me3, with/without H3K27ac (active or inactive/poised respectively) (often with Nipbl. Mediator, tissue specific TFs)
Insulator	Separates chromatin or gene expression into active and inactive domains	CTCF
TAD boundaries	Cell-type invariant structural feature showing sharp changes in contact frequencies between adjacent TADs	CTCF

Genome-wide Hi-C interactome maps have revealed that the genome is organised into compartments of either active or repressed chromatin. These compartments are further separated into genomic units known as 'Topologically Associated Domains' (TADs) (Dixon et al., 2012; Nora et al., 2012). TADs are a structural unit of chromatin organisation comprising of regions which show high local contact frequency that are separated by sharp boundaries, across which contacts are relatively infrequent. The average size of a TAD is approximately 1Mb and each TAD contains several genes and enhancers. It is hypothesised that these self-interaction domains constrain looping interactions between enhancers and promoters and set the boundaries for coordinated gene regulation. The boundaries between TADs are largely invariant across cell types and are also highly conserved between mouse and human. This suggests that such physical partitioning of the genome is a fundamental principle of the hierarchical genome organisation(Gorkin et al., 2014). Moreover, TADs are not detectable during mitosis, indicating that their function is specific to interphase when transcription is most active (Naumova et al., 2013).

CTCF binding sites are highly enriched at TAD boundaries (Dixon et al., 2012) and deletion of a specific TAD boundary containing CTCF binding sites led to increased interactions between adjacent TADs (Nora et al., 2012). At a global scale, knockdown of CTCF expression leads to an increase in inter-domain interactions but does not completely abrogate TAD boundaries (Zuin et al., 2013). Likewise, loss of cohesin also leads to increased inter-domain interactions (Sofueva et

al., 2013), however, the impact is less than that of CTCF loss (Zuin et al., 2013). Interestingly, depletion of cohesin in non-cycling thymocytes did not affect architectural compartments (Seitan et al., 2013). Nonetheless, cohesin was required for specific long-range interactions within the compartments. Lineage-specific genes like those involved in haematopoiesis and lymphocyte activation were specifically deregulated upon cohesin depletion. Intriguingly, gene expression was perturbed across the whole range of expression spectrum where genes with low levels of expression were more often up-regulated whereas genes with high expression were more often down-regulated. This systematic skewing lead to compression of the dynamic range of gene expression away from the extremes towards a more average expression (Seitan et al., 2013). Rapid cleavage of the Rad21 subunit in HEK293 cells and cohesin depletion in postmitotic astrocytes also resulted in a global loss of intra-TAD interactions (Sofueva et al., 2013; Zuin et al., 2013). Together these studies signify the role of cohesin as a genome organiser where it helps demarcate the TAD boundaries along with coordinating gene expression within TADs while restricting interactions across TAD boundaries.

1.3.3 Cohesin mediated regulation of the transcriptional machinery and associated components

Apart from its role in mediating long-range DNA interactions, cohesin can directly affect the activity of RNA polymerase. Transcription is tightly regulated at the stages of initiation, elongation and termination. The basic steps constitute the formation of pre-initiation complex (PIC) containing RNA Pol II and several general transcription factors (GTFs). This is followed by the phosphorylation of Ser5 residues on YSPTSPS heptapeptide repeat consensus sequence present in the CTD of Pol II. Ser5 phosphorylation allows promoter escape for the Pol II and marks the initiation of transcription. Pol II then pauses in the 5’ region of the transcription unit and only progresses into productive elongation on stimulation by appropriate signals, a phenomenon known as promoter-proximal pausing (or transcriptional stalling). Pol II pausing acts as an important checkpoint before Pol II is committed to transcribe the gene. The pausing action is attributed to the activity of pause factors, DRB-sensitivity inducing factor (DSIF) and negative elongation factor (NELF), which remain bound to Pol II at the pause site downstream of the TSS. In the presence of appropriate signals, positive transcription elongation factor b (P-TEFb) is recruited to the genes. P-TEFb then phosphorylates the DSIF and NELF subunits along with the Ser2 residues in Pol II CTD licensing the Pol II into productive elongation. The CTD of Pol II can thus be used as an indicator of Polymerase activity. At active genes, Ser5 phosphorylation is high at

TSS and is present along the gene body. Ser2 phosphorylation on the other hand is enriched in the gene body and at transcription termination sites. Each of these mentioned steps can be rate limiting and provide several avenues for regulation of the transcriptional output (Saunders et al., 2006; Zhou et al., 2012).

Cohesin can aid transcription initiation by facilitating the binding of transcription factors at sub-optimal sequence motifs. This is suggested by a recent genome-wide study which analysed the co-localisation of cohesin with a large set of tissue-specific transcription factors, RNA Pol II and histone modifications in mouse liver cells. They observed that binding of tissue-specific transcription factors, master regulators and enhancer-associated chromatin marks at sites with low motif scores, coincided with strong cohesin-non-CTCF binding sites. The authors further showed that specific transcription factor modules with lower motif scores were preferentially destabilised in *Rad21* haploinsufficient cells (Faure et al., 2012). Based on these observations, the authors suggest that cohesin can stabilise large protein-DNA complexes and allow efficient transcription factor binding even at sites with low motif scores.

In *Drosophila*, cohesin preferentially binds genes with promoter-proximal paused RNA polymerase (Fay et al., 2011; Schaaf et al., 2013). However, cohesin depletion can have opposing impact on the transcription of these genes (Schaaf et al., 2009). As such these studies indicate a complex relationship between cohesin binding and transcriptional elongation in *Drosophila*. Similar studies on the global effects of cohesin on Pol II pausing and elongation have not yet been conducted in mammalian cells. However, in one study on mouse ES cells, knockdown of Smc3 expression resulted in reduced promoter-proximal Pol II occupancy at many Ell3 (a Pol II elongation factor) responsive genes (Lin et al., 2012a). It remains to be seen whether the observed changes in Pol II occupancy were caused by reduced gene expression or were a direct consequence of cohesin depletion.

Cohesin has also been shown to affect the processivity of RNA Polymerase in association with CTCF, possibly by physically stalling the movement of Pol II. The latency transcripts of Kaposi's sarcoma-associated herpesvirus (KSHV) contain cohesin-CTCF binding sites in the first intron. During latency a paused form of Pol II is loosely associated with the promoter region but is converted into an active elongating form upon reactivation induced by Rad21 depletion. Similar effects on pausing and transcription were seen at the endogenous *c-Myc* gene which also contains a cohesin-CTCF site in the first intron. These findings suggest that RNA Pol II pauses at intragenic

cohesin-CTCF binding sites which regulates the transcriptional status of the gene (Kang and Lieberman, 2011; Chen et al., 2012). This is also true for the p53 target PUMA gene where cohesin-CTCF binding acts as a transcriptional block and prevents read-through of the full length PUMA gene (Gomes and Espinosa, 2010). The rate of transcriptional elongation is also known to impact the process of alternative splicing. At the CD45 locus as well as genome-wide, intragenic CTCF binding promotes the inclusion of weak exons by mediating local RNA Pol II pausing at alternatively spliced sites (Shukla et al., 2011).

Moreover, cohesin localisation at sites of convergent transcription in budding and fission yeast reveal that cohesin is associated with transcriptional termination in these organisms (Lengronne et al., 2004; Gullerova and Proudfoot, 2008).

In addition to regulating Pol II processivity, cohesin and CTCF also control the transcription of non-coding RNAs as seen during *Igh* rearrangement, *ataxin-7* transcription and the transcription of *Xist*, *Tsix* involved in X-chromosome inactivation (Degner et al., 2011; Sopher et al., 2011; Spencer et al., 2011). At the chicken lysozyme gene promoter, CTCF binding at the -2.4kb insulator element blocks enhancer function. Upon induction by LPS, CTCF is evicted by the transcription of a non-coding RNA originating from the -2.4/-2.7kb region thus facilitating enhancer-promoter communication and transcription (Lefevre et al., 2008). CTCF and cohesin also regulate ribosomal DNA (rDNA) transcription and ribosomal biogenesis (van de Nobelen et al., 2010; Bose et al., 2012). rDNA locus contains hundreds of copies of the rDNA genes only some of which are actively transcribed. The activity of spacer promoters, which give rise to non-coding RNAs required for rDNA transcription, is regulated by CTCF binding. Cohesin is also reported to bind to non-coding RNAs transcribed from active enhancers (eRNAs) which is important for the upregulation of the target genes (Wang et al., 2011a).

Altogether, these examples indicate that apart from mediating long-range DNA interactions, cohesin can modulate gene expression by the regulation of RNA polymerase processivity and production of non-coding regulatory RNAs. However, much of this aspect remains to be explored in further detail.

1.3.4 Cohesin and the MLL complex

So far most studies have investigated how cohesin impacts the formation of long-range DNA interactions between enhancers and promoters. However, it is noteworthy that one of the prerequisites for promoter and enhancer activation is the methylation of lysine 4 residue on histone H3 - H3K4me3 and H3K4me1 modifications at promoters and enhancers respectively (Heintzman et al., 2009; Creyghton et al., 2010; Rada-Iglesias et al., 2011). Even though the importance of H3K4me1 modifications for enhancer function has been recognised, the molecular effectors of these cell-type specific histone marks are yet to be determined. Given the extensive binding of cohesin at cell-type specific active enhancers and promoters, it is possible that cohesin plays a role in the establishment of these chromatin modifications. One can speculate that the cohesin protein complex brings about this effect by interacting with and recruiting histone modifying enzymes like the histone methyltransferases to the site of an enhancer.

Mammalian cells contain six major methyltransferases belonging to the MLL/Set1 family of proteins (Figure 1.8). Although all the six family members are capable of H3K4 methylation in vitro, their activity requires the presence of additional accessory subunits, some of which are differentially associated with distinct family members.

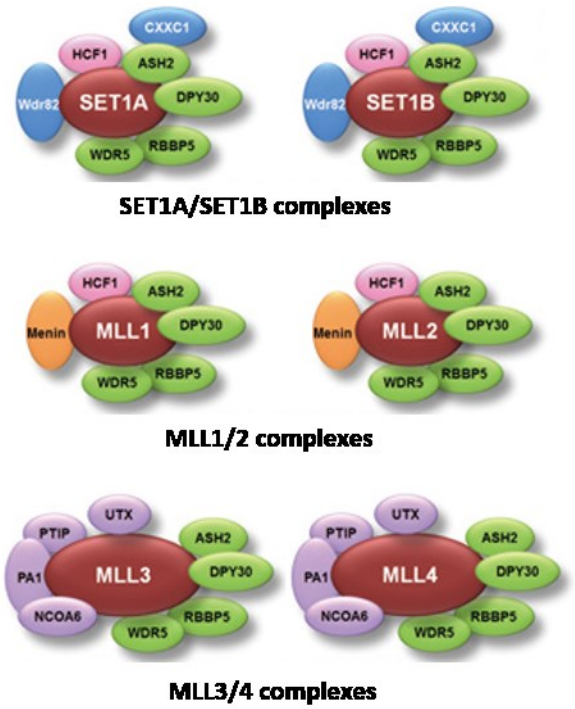


Figure 1.8 H3K4 methyltransferases in mammals. Mammalian cells have three major complexes with methyltransferase activity composed of the main catalytic subunits SET1A/SET1B, MLL1/MLL2 or MLL3/MLL4 along with accessory subunits. Paxip1 (or PTIP) specifically associates with the MLL3/4 complexes. Subunits common to all complexes are shown in green; complex-specific subunits are shown in blue, orange, pink and magenta (Hu et al., 2013).

Set1a and Set1b (homologous to yeast and *Drosophila* Set1) are responsible for bulk of the H3K4me3 in metazoan cells. Evidence suggests that Set1a/b complexes are recruited directly by the Pol II machinery and its occupancy is restricted to TSS by the activity of Cfp1 (CXXC1) (Shilatifard, 2012; Calo and Wysocka, 2013). Enhancer-associated H3K4me1 modifications are attributed to the activity of the MLL3/4 complex. In *Drosophila*, depletion of Trr, a *Drosophila* homolog of MLL3/4 affects global H3K4me1 levels in vivo, whereas knockdown of Set1 (or Trx) has a minor effect (Ardehali et al., 2011; Herz et al., 2012). Furthermore, MLL complex binding is not restricted to promoters but is also present at intergenic regions and gene bodies with many of the intergenic sites overlapping enhancers (Herz et al., 2012). More concrete evidence in mammalian cells, comes from a recent study which shows that MLL4 is essential for enhancer activation during cell differentiation (Lee et al., 2013). MLL4 exhibited cell-type- and differentiation-stage-specific genomic binding and was predominantly localized on enhancers along with lineage-determining transcription factors. Depletion of MLL4 markedly decreased H3K4me1/2, H3K27ac, Mediator and Polymerase II levels on enhancers and led to severe defects in cell-type-specific gene expression during adipogenesis and myogenesis (Lee et al., 2013). In fact, PTIP (Pax interaction with transcription-activation domain protein-1, also known as Paxip1), a subunit specifically associated with the MLL3/4 complex also plays a critical role in adipogenesis (Cho et al., 2009). The association of MLL1/2 with specific enhancers has only been documented in individual studies so far (Jeong et al., 2011; Kawabe et al., 2012). But in mouse ES cells, MLL2 directs H3K4me3 deposition on bivalent promoters (Hu et al., 2013; Denissov et al., 2014). Altogether, current evidence suggests that in mammalian cells, MLL3/4 complex is the major player in making the H3K4me1 histone modifications at enhancers.

Interestingly, studies on Paxip1 – a subunit of the MLL3/4 complex, provide a possible connection between cohesin and the MLL3/4 complexes. Biochemical analysis showed that PTIP interacts with SMC1 (Patel et al., 2007) and is required for SMC1 to be able to successfully repair DNA damage (Wu et al., 2009). Additionally, there is some overlap of the binding of PTIP with cohesin binding sites at the *Igh* and the *Tcrα* lymphocyte receptor loci where both cohesin and PTIP have independently been shown to be required for proper long-range interactions and transcription ensuring Immunoglobulin class switch recombination, Tcra locus rearrangement and thymocyte differentiation (Daniel et al., 2010; Degner et al., 2011; Schwab et al., 2011; Seitan et al., 2011). The PTIP subunit itself is not required for the methytransferase activity of the MLL3/4 complex (Shilatifard, 2012). Nevertheless, PTIP is required for maintenance of embryonic stem

cell pluripotency and *Paxip1* homozygous mutant mice are embryonic lethal by day 9.5 (Cho et al., 2003). Together these findings suggest that PTIP can play a significant role in transcriptional regulation. However, Daniel et al. (2010) noted that impact of *Paxip1* depletion on transcription was limited to only a few genes. They suggested that PTIP is required for the expression of specific genes either through the deployment of the MLL3/4 complex or possibly by recruiting other transcription factors. The role of PTIP in cohesin-mediated gene regulation has not been explored.

An additional intriguing observation further supports the idea that cohesin might function through the MLL3/4 complex in gene regulation. Kabuki syndrome, a rare congenital disorder, is associated with mutations in the *KMT2D* (MLL4) gene. Children with Kabuki syndrome usually have distinctive facial features and show mild to moderate mental impairment. Typical facial features include arched eyebrows, elongated and wide-set eyes, thick eyelashes, flat nasal tip, cleft palate and large or cupped ears. Other characteristic features include short stature, skeletal abnormalities and developmental delay (Cuscó et al., 2008; Bokinni, 2012). These clinical features presented in Kabuki syndrome patients are remarkably similar to those observed in patients with CdLS (Figure 1.9). Though genetic studies so far have not tested for the presence of cohesin mutations in Kabuki syndrome patients, or *KMT2D* mutations in CdLS patients, similar phenotypic characters support the idea that cohesin might regulate gene expression through the activity of effector proteins like the MLL3/4 complex.

Figure 1.9 Kabuki syndrome patients display CdLS like features. Kabuki patients typically have elongated and wide-set eyes, flat nasal tip, cleft palate, large cupped ears and show other skeletal defects along with mild to severe mental retardation (Cuscó et al., 2008).

1.4 Cohesin's role in embryonic stem cell pluripotency

Embryonic stem cells are derived from the inner cell mass (ICM) of the pre-implantation blastocyst-stage embryo. They have the remarkable property of self-renewal and are capable of differentiating into different cell types of all lineages of the developing and adult organism i.e. they are pluripotent. Consequently, these cells retain the ability to participate in normal development upon reinjection into an embryo, including the formation of functional gametes, even after indefinite propagation *in vitro*. These exceptional properties of ES cells are attributed to their unique transcriptional profile (Jaenisch and Young, 2008). Transcription factors OCT4, SOX2, NANOG form the core of the ES transcriptional regulatory network and are essential for the maintenance of a robust pluripotent state. These transcription factors function together to positively regulate their own promoters forming autoregulatory circuits. They co-occupy and activate the expression of genes necessary for maintaining the ES cell state, while contributing to repression of genes encoding lineage-specific transcription factors and thus prevent exit from the pluripotent state (Boyer et al., 2005). ES cells express uniformly high levels of OCT4 and protein levels are tightly regulated. Reduction or overexpression of OCT4 leads to differentiation of ES cells (Niwa et al., 2000). OCT4 acts synergistically with SOX2 in regulating various ES-specific genes. NANOG, on the other hand is necessary for the acquisition of pluripotency but is dispensable once pluripotency is achieved (Chambers et al., 2007). Additional ES-specific regulators required for the maintenance of pluripotency have been uncovered by several RNAi studies and examination of protein purification partners. These regulators like KLF4, ESRRB, MYC, SALL4, TCF3, DAX1, NAC1, TCF2L1, ZFP281 etc. form a part of the expanded ES pluripotency network. These factors often co-bind the same genomic regions along with the core transcriptional regulators OCT4 and SOX2 (Yeo and Ng, 2013). In addition to the unique transcriptional profile, ES cells also possess a highly dynamic and accessible 'open' chromatin landscape with a lower abundance of constitutive heterochromatin than most differentiated cells. This chromatin feature is likely to be important for attaining the transcriptionally permissible pluripotent state (Gaspar-Maia et al., 2011).

The silencing of developmental regulators in ES cells is also tightly regulated in such a way that the cells remain amenable to perturbation elicited by developmental stimuli. Most of the transcriptionally silent genes are marked by bivalent chromatin where promoters are associated with both the gene activating histone mark H3K4me3 and the repressive H3K27me3 modifications (Azuara et al., 2006; Bernstein et al., 2006). These genes are also bound by Polycomb group (PcG)

proteins which catalyse repressive H3K27me3 modifications and H2AK119 ubiquitination. It is thought that bivalency poises genes for expression and prevents their permanent silencing, so that they remain sensitive to differentiation signals. In the presence of appropriate signals, specific signal transducers are recruited which then overcome the H3K27me3-mediated repression and activate the lineage-specifying genes (Vastenhouw and Schier, 2012). Lineage commitment is also accompanied by rapid reduction in OCT4, NANOG and other ES-specific TFs by transcriptional and post-transcriptional mechanisms.

Once established, the differentiated state is stable and can be inherited through cell division. However, experimental approaches have been developed that allow the cell fate to be modified or even reverted to a pluripotent state through a process known as reprogramming. Three main strategies have so far been employed to achieve reprogramming: somatic cell nuclear transfer (SCNT), cell fusion and induced pluripotent stem (iPS) cell generation by forced expression of transcription factor combinations (Tada et al., 1997; Wilmut et al., 1997; Takahashi and Yamanaka, 2006). Reprogramming mediated by transferring the somatic cell nuclei into enucleated oocytes (SCNT) provided the first demonstration that the differentiated state is reversible (Gurdon et al., 1958). Subsequent studies based on cell fusion showed that cellular identity can be modified by trans-acting factors (Tada et al., 1997). The identity of transcriptional regulators required for reprogramming was revealed by a pioneering work from the Yamanaka lab (Takahashi and Yamanaka, 2006). Ectopic expression of OCT4, SOX2, KLF4 and MYC (OSKM) could convert fibroblasts into iPS cells. Since then, several epigenetic and transcriptional regulators have been employed to reset cellular identity with greater efficiency (Morris and Daley, 2013). It is of immense medical importance to be able to generate stem cells from somatic cells of patients. These can potentially be used for tissue repair, organ regeneration and cell replacement therapies.

1.4.1 Extrinsic signalling in ES cell transcriptional regulation

ES cells were initially derived by explanting blastocysts or ICMs on a layer of mitotically inactivated fibroblasts in medium containing fetal calf serum (Smith and Hooper, 1983). It was later discovered that the cytokine Leukemia Inhibitory Factor (LIF) can substitute for feeder cells (Smith et al., 1988). LIF promotes continual self-renewal by activating the transcription factor STAT3 (signal transducer and activator of transcription 3) (Nichols et al., 1998). Subsequently, BMP4 (bone morphogenetic protein 4) was shown to replace the need for serum, an effect which was reproduced by forced expression of *Id* (inhibitor of DNA binding) genes, allowing ES cells to

be cultured in defined conditions (Ying et al., 2003a). Thus conventional ES cell culture relies on extrinsic stimulation of STAT3 by LIF and a parallel induction of ID proteins by serum factors or BMP4.

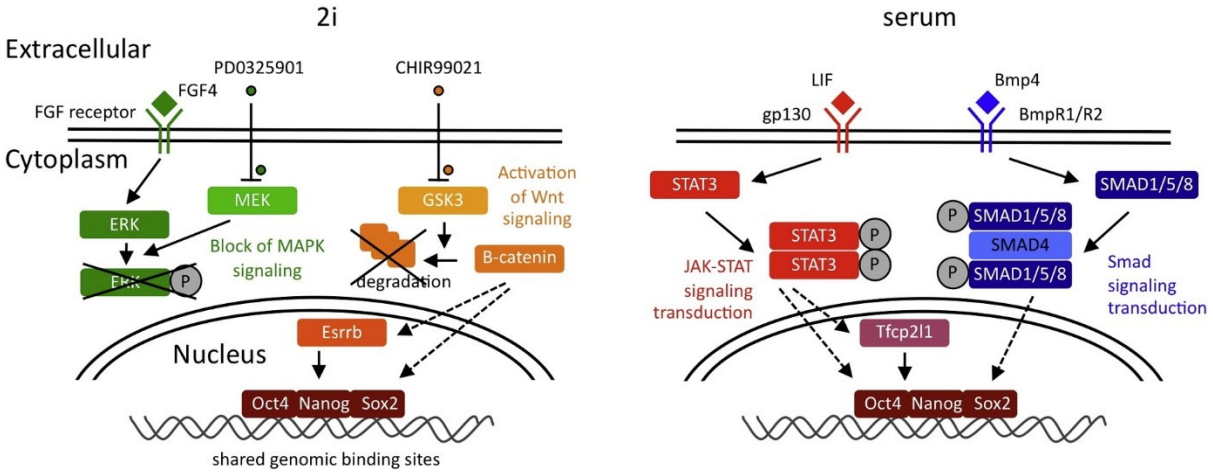


Figure 1.10 Signal transduction pathways affected by 2i and serum/BMP4 + LIF. Left panel: ES cell signalling in 2i supplemented media. PD03 inhibits the pro-differentiation MAPK/ERK signalling pathways. CHIR inhibits GSK3 and leads to the activation of canonical Wnt signalling through β-catenin stabilisation. Together these inhibitors allow the maintenance of ES cells in a ground state of pluripotency. Right panel: ES cells cultured in Serum/BMP4 + LIF media. LIF and BMP4 activate STAT3 and Smad proteins respectively, which are required for the maintenance of ES cell in pluripotent state (Adapted from Marks and Stunnenberg, 2014).

Further studies have shown that ES cells possess opposing signalling pathways that govern the balance between self-renewal and differentiation. OCT4 and SOX2, in addition to the activation of pluripotency genes, also promote the expression of the autocrine factor FGF4 (fibroblast growth factor 4), which in turn activates the Erk signalling pathway that favours differentiation (Kunath et al., 2007). These findings suggested that LIF and BMP4 act downstream of the Erk signalling pathway to block differentiation. Expanding upon this observation, Smith and colleagues showed that dual inhibition of the pro-differentiation MEK (MAP Kinase/ER Kinase) and GSK3 (Glycogen synthase kinase 3β) signalling by small molecule inhibitors, in conventional serum-cultured mouse ES cells, reinforced their capacity for self-renewal. The MEK inhibitor PDO3 (PD0325901) and the GSK3 inhibitor CHIR (CHIR99021), in combination known as “2i”, allow robust ES cell self-renewal in the absence of serum or LIF (Ying et al., 2008). Inhibition of GSK3 results in an increase in the anabolic processes and activation of canonical Wnt signalling through β-catenin stabilisation (Wray et al., 2011) (Figure 1.10). GSK3 is likely to exert β-catenin independent effects on ES cells as well. Notably, MAPK (Mitogen Activated Protein Kinase) and GSK3 signalling have been shown to have opposing effects on protein translation and on c-Myc

stability (Sears, 2004; Wu and Pan, 2010). Remarkably, the use of 2i+LIF defined media made it possible to derive ES cell lines from all strains of mice with high efficiency (Nichols et al., 2009).

ES cells growing in conventional serum and LIF culture conditions, form colonies that are composed of a mixture of undifferentiated and differentiated cells (Figure 1.11). In fact, analysis has shown that ES cells in conventional culture conditions display heterogeneous expression of transcription factors like NANOG, STELLA, REX1, ESRRB, HEX, KLF4 and TBX3. For example, NANOG is not expressed in all OCT4-expressing undifferentiated ES cells. ES cells with knockin GFP reporters for NANOG can be sorted into NANOG-low and NANOG-high expressing populations. Sorted fractions can re-establish the parental distribution of Nanog expression showing that the expression is dynamic. However, NANOG-low cells are functionally more prone to differentiation (Chambers et al., 2007). Thus heterogeneous expression reflects different cell states that coexist within the population despite their genetic homogeneity. Such stochastic changes in gene expression may provide a window of opportunity to direct lineage allocation in response to developmental cues (Torres-Padilla and Chambers, 2014).

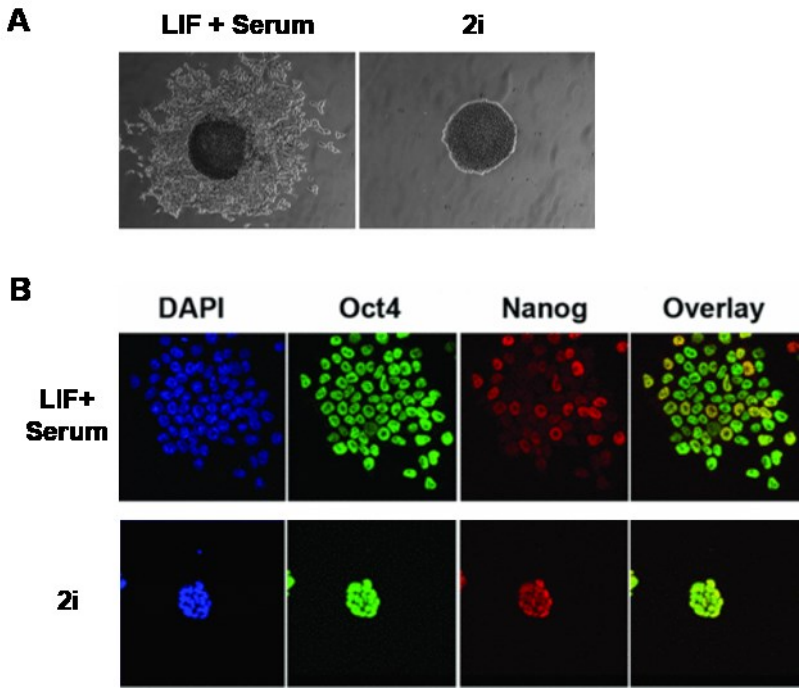


Figure 1.11 Nanog expression is heterogeneous in conventional serum ES cell cultures. A) ES cell colonies formed from single cells are morphologically different in 2i and serum conditions. **B)** Immunostaining images for Oct4 and Nanog expression in ES cells cultured in LIF+Serum or 2i. Oct4 expression is more or less uniform in both conditions. Serum+LIF ES cells show varying degrees of Nanog expression ranging from high to low while 2i ES cells are homogenously high for Nanog (Adapted from (Wray et al., 2010).

ES cells cultured in 2i medium are believed to be in a 'ground' or 'naïve' state of pluripotency equivalent to cells from the early epiblasts (Nichols and Smith, 2009). In 2i+LIF medium, mouse ES cells form colonies efficiently and these colonies are composed purely of undifferentiated cells. They have homogenously high NANOG expression (Figure 1.11) and do not show intrinsic fluctuations of gene expression as observed in conventional media. This is likely to result from the elimination of the differentiation-promoting effects of FGF4-MAPK signalling together with the consolidating effect of GSK3 inhibition. Thus, it is suggested that the pluripotency network, headed by OCT4/SOX2, NANOG and KLF4 is inherently self-sustainable, but extremely sensitive to destabilisation by exogenous signals and elimination or neutralisation of such signals captures the ES cells in a naïve state (Wray et al., 2010).

Genome-wide analysis reveals that culture conditions impose distinctive transcriptome and epigenome properties on mouse ES cells. Nearly 25% of the genes show 2-fold or greater differences in expression between 2i and serum conditions. Genes associated with metabolic processes are upregulated in 2i while those associated with differentiation are downregulated (Marks et al., 2012). 2i ES cells have much lower c-Myc expression (Ying et al., 2008) possibly resulting in the observed upregulation of cell cycle inhibitors p16, p19 and p21. Nonetheless, ES cells continue to proliferate rapidly, reflecting their freedom from cyclin checkpoint controls (Stead et al., 2002). Promoter-proximal RNA polymerase pausing is also more prominent in 2i than in serum. This may be explained by the induction of c-Myc in serum which facilitates pause release (Rahl et al., 2010). In terms of epigenetic modifications, H3K4me3 marks are similar between 2i and serum but there is a striking difference in the pattern of H3K27me3 deposition. Even though the global levels of H3K27me3 are similar between 2i and serum, the localisation of this mark at repressed genes is greatly diminished in 2i cells with no apparent increase in transcription. Overall, serum ES cells have more bivalently marked genes. Additionally, 2i ES cells have lower DNA methylation than serum ES cells (Ficz et al., 2013; Habibi et al., 2013; Hackett et al., 2013; Leitch et al., 2013; Marks and Stunnenberg, 2014).

1.4.2 Cohesin as a regulator of pluripotency

Cohesin subunits have repeatedly been identified in RNAi screens performed in ES cells, as factors required for the maintenance of pluripotency (Ding et al., 2009; Hu et al., 2009; Kagey et al., 2010; Nitzsche et al., 2011). However, it is unclear from these studies whether the requirement of cohesin is due to its necessary functions in cell division or due to its direct roles in regulating expression of pluripotency genes. Kagey et al., (2010) found that the Mediator complex

and Nipbl were also necessary for ES cell self-renewal. They showed that individual knockdown of the *Med12*, *Smc1a* and *Nipbl* each had the same effect on ES cell state as loss of Oct4 itself in terms of reduced expression of pluripotency associated transcription factors and upregulation of genes encoding developmentally important transcription factors. They further provided evidence that cohesin and NIPBL biochemically interact with the Mediator complex and co-localise on DNA at CTCF-independent cohesin binding sites associated with enhancers and promoters of active genes. To test if the enhancer-promoter associated cohesin was involved in DNA looping, they performed 3C experiments at the *Oct4*, *Nanog*, *Phc1* and *Lefty1* loci. This analysis indicated that the promoters of these pluripotency associated genes interacted with their respective enhancers only in ES cells but not in MEFs, where these genes are not expressed and show no cohesin binding. These observations suggested that cohesin is involved in enhancer-promoter interactions in a cell-type specific manner and may thus contribute to the regulation of pluripotency genes (Kagey et al., 2010; Phillips-Cremins et al., 2013).

Since then, several studies have further analysed the role of cohesin and transcription factors in organising the chromatin structure and maintenance of a transcription profile unique to pluripotent state. It is observed that regions with high density of binding sites for key pluripotency TFs OCT4, SOX2 and NANOG (OSN) cluster together, and are highly enriched for long-range DNA interactions suggesting that OSN are directly involved in higher order genome organisation in ES cells (Denholtz et al., 2013). This is supported by the demonstration that loss of either OCT4 or NANOG diminishes long-range DNA contacts (regions separated by more than 5Mb) between OSN-bound regions. Moreover, insertion of an ectopic NANOG binding site is sufficient to create long-range contacts with the endogenous OSN-bound regions. Interestingly, these long-range OSN bound contact sites are not enriched for cohesin and CTCF binding. This suggests that OSN can shape the higher order structure of the genome independent of the architectural proteins. Although OSN also anchor shorter-range *cis*-regulatory interactions that do require cohesin (de Wit et al., 2013). Oct4 expression itself is facilitated by KLF4 mediated recruitment of cohesin (Wei et al., 2013; Zhang et al., 2013a).

Comparison of chromosomal contacts and genome-wide TF binding between ES cells and differentiated progeny has revealed dynamic enhancer landscapes based on cohesin, CTCF and other TF binding which allow coordinated gene expression changes in a tissue-specific manner. During differentiation, loss of interactions is often found to be associated with loss of cohesin and TF binding and reduced expression of pluripotency genes along with the formation of new

cohesin-mediated contact sites at lineage-specifying genes (Apostolou et al., 2013; Kieffer-Kwon et al., 2013; Phillips-Cremins et al., 2013; Zhang et al., 2013b).

The process of somatic cell reprogramming to iPS cells also seems to require cohesin and Mediator subunits as their knockdown severely reduces the efficiency of iPS cell generation. This has again been related to the inability of cells to induce the expression of pluripotency genes like Oct4 and *Nanog* in the absence of enhancer-promoter interactions (Apostolou et al., 2013; Wei et al., 2013; Zhang et al., 2013a). However, it is noteworthy that generation of iPS cells requires active cellular proliferation. Thus it is not clear whether cohesin depletion impairs iPS generation directly due to transcriptional deregulation or is a consequence of the cells inability to divide properly in the absence of cohesin. In order to truly understand the role of cohesin in early steps of reprogramming, independent of its cell cycle-dependent functions, Lavagnolli et al., (2014) used the cell fusion system. They fused somatic human cells with mouse ES cells that could be conditionally and rapidly depleted of cohesin to form heterokaryons with two spatially discrete nuclei within a shared cytoplasm. The heterokaryon stage persists without cell division for up to 3 days, after which nuclear fusion occurs, giving rise to tetraploid hybrid cells. Heterokaryon formation is accompanied by the activation of pluripotency genes and the extinction of lineage-specific genes in the somatic nuclei (Pereira et al., 2008; Soza-Ried and Fisher, 2012). Therefore, cell fusion-mediated heterokaryon formation allowed the investigation of the role of cohesin in early events of somatic cell reprogramming in the absence of cell division. Interestingly, cohesin-depleted mouse ES cells consistently induced the human pluripotency-associated genes *POU5F1* (OCT4), *NANOG*, *SOX2, REX1*, and *CRIPTO* more strongly and with faster kinetics than control ES cells. These experiments indicated that cohesin is not required for the ability of ES cells to induce pluripotency gene expression by somatic nuclei in heterokaryons, and that cohesin-deficient ES cells initiate the reprogramming of somatic cell nuclei more potently than wild type ES cells (Lavagnolli, 2013). The increased reprogramming efficiency was attributed in part to *Myc* overexpression in cohesin-deficient mouse ES cells which enhanced somatic cell DNA replication (this work Section 4.1, Lavagnolli et al., 2015), previously shown to facilitate heterokaryon-mediated reprogramming (Tsubouchi et al., 2013). In contrast, cohesin-deficient somatic cells had defective DNA replication and reprogrammed poorly in heterokaryons. Nuclear transfer experiments further showed that cohesin is not required for DNA replication-independent reprogramming of cohesin-deficient somatic cells in Xenopus oocytes (Figure 1.12) (Lavagnolli et al., 2015). Overall, these studies show that the role of cohesion in somatic cell reprogramming is

therefore at least in part to facilitate DNA replication in heterokaryon-mediated reprogramming and cell division in iPS cell-mediated reprogramming.

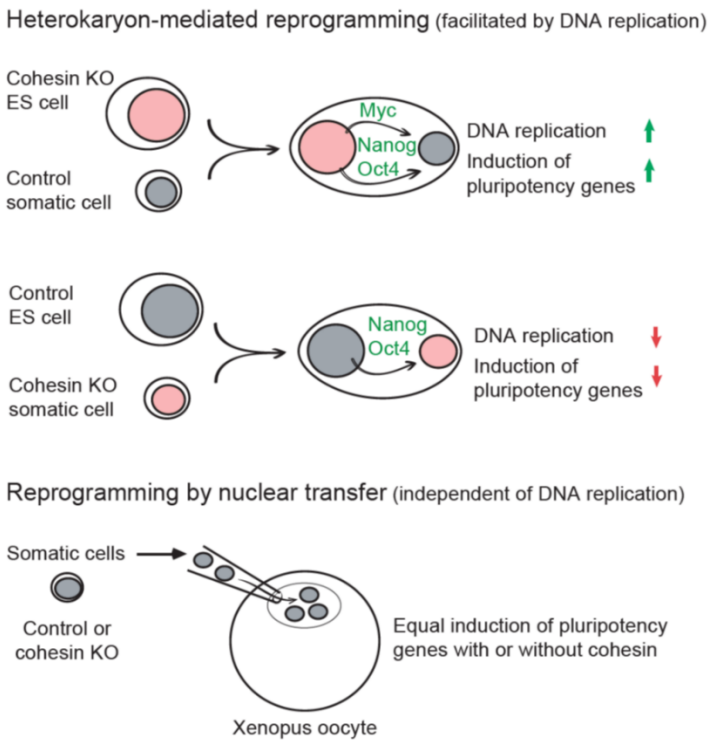


Figure 1.12 Cohesin's role in reprogramming of somatic cells towards pluripotency. (Top) Acute cohesin deletion in mouse ES cells increases their ability to reprogram somatic cells after heterokaryon formation partly due to enhanced DNA replication induced by Myc overexpression. (Middle) Cohesin depletion in somatic cells impairs DNA replication and induction of pluripotency gene expression in ES cell heterokaryons. (Bottom) Pluripotency gene expression in cohesin-deficient somatic cell nuclei is induced efficiently by nuclear transfer experiments where reprogramming occurs in the absence of DNA replication. (Adapted from Lavagnolli et al., 2015.)

Together, current evidence suggests that cohesin potentially has a role in regulating the expression of pluripotency associated genes but much of the details of its function in this process remain to be determined.

1.5 *Myc* and its regulation by cohesin

MYC regulates a multitude of biological processes ranging from cell proliferation, metabolism, growth, angiogenesis, metastasis, genomic instability to stem cell self-renewal. Deregulated *Myc* expression due to translocation, gene amplification or aberrant signalling promotes unrestrained cell proliferation and is frequently associated with human malignancies (Dang, 2012). *Myc* is also important for the maintenance of pluripotency and its induced expression along with that of Oct4, *Sox2* and *Klf4* can reprogram fibroblasts into a pluripotent stem cell state (Chappell and Dalton, 2013). Given its pivotal role in cellular processes and its potential as a therapeutic target for cancer treatment, it is of extreme importance to understand the molecular details of its various functions and modes of regulation.

1.5.1 Cellular functions of MYC

A large part of MYC's contribution to cellular processes is attributed to its role as a transcriptional regulator. MYC binds to DNA as a heterodimer with MAX (MYC-associated factor-X) (Hurlin and Huang, 2006). The canonical Myc E box DNA binding motif (5'-CAGGTG-3') occurs very frequently in the human genome (Xie et al., 2005) and MYC has been shown to extensively bind at regions of active chromatin (Dang, 2012; Nie et al., 2012). MYC is believed to increase transcription by the recruitment of histone acetyltransferases (Cowling and Cole, 2006). Additionally, it has been shown that recruitment of pTEFb by MYC stimulates the release of paused RNA polymerase partly by phosphorylation of Pol II CTD (Rahl et al., 2010). Recent studies suggest that MYC functions predominantly as a universal amplifier of active transcription (Lin et al., 2012b; Nie et al., 2012), although it can also repress certain transcriptional targets (Meyer and Penn, 2008; Sabò et al., 2014; Walz et al., 2014).

Elevated levels of c-*Myc* cause a broad spectrum of cellular effects. Myc promotes cellular growth and proliferation by boosting biosynthetic and metabolic pathways, rRNA synthesis, ribosomal protein biogenesis. It also plays a critical role in cell cycle progression. MYC is essential for G1 to S phase transition. It abrogates the transcription of cell cycle checkpoint genes, inhibits the function of cyclin-dependent kinase (CDK) inhibitors while also activating the expression of genes required for cell cycle progression (Obaya et al., 1999). Interestingly, overexpression of MYC in the absence of specific survival factors triggers apoptosis by the activation of p53-dependent pathways. In parallel, MYC induces the expression of pro-apoptotic molecules like BIM (BCL2-like 11) and represses the expression of anti-apoptotic BCL2 and BCL-XL proteins (Dang et al., 2005).

Thus MYC lies at the crossroads of several growth promoting and cellular survival signal transduction pathways. In ES cells, MYC is important for the establishment and maintenance of the pluripotency transcriptional programme (Chappell and Dalton, 2013) where it forms a stem cell module of target genes distinct from the core transcriptional network but more similar to that of cancer cells (Kim et al., 2010). MYC downregulation is required for the ES cells to exit cell cycle and undergo differentiation (Wilson et al., 2004).

MYC also has transcription-independent functions (Cole and Cowling, 2008). MYC promotes the translation of specific mRNAs by promoting the methylation of the 5' mRNA guanine or 'cap' (Cowling and Cole, 2007). MYC also plays an important role in the initiation of DNA replication. It binds to numerous components of the pre-replicative complex, including MCM proteins, ORC2, CDC6 and CDT1 and recruits them to sites of active replication origins. Studies in both *Xenopus* and mammalian cells suggest that MYC functions to control the selection of replication origins and governs the number of active replication origins. Consequently, depletion of MYC decreased DNA synthesis and addition of recombinant MYC protein rescued DNA synthesis (Dominguez-Sola et al., 2007). Thus MYC functions are multifaceted and encompass a wide variety of cellular processes.

1.5.2 Regulation of MYC expression

Physiological levels of MYC are precisely regulated by several extracellular and intracellular signals operating through an array of transcription factors, chromatin modifiers and regulatory RNAs that are brought to the *Myc* promoter and *cis*-regulatory elements. *Myc* mRNAs are spliced from primary transcripts containing three exons using multiple reading frames. The MYC protein is encoded from the exons 2 and 3 with two universally expressed forms arising from an AUG codon at the 5' end of exon 2 and a CUG initiation codon at the 3' end of exon 1, respectively (Liu and Levens, 2006). The *Myc* promoter acts as a key convergence node for a whole host of transcription factors such as CNBP, FBP, BRD4 and TCF. General *Myc* expression can be considered to be pseudo-constitutive where its expression within each cell type is controlled by a customised set of TFs and chromatin factors (Levens, 2010). Genome-wide association studies have identified common single nucleotide polymorphisms (SNPs) that are linked with multiple cancers. Detailed analysis shows that several of these SNPs lie in enhancer elements that are often involved in DNA looping with the *Myc* promoter in a context specific manner (Pomerantz et al., 2009; Tuupanen et al., 2009; Ahmadiyeh et al., 2010; Sotelo et al., 2010; Wasserman et al., 2010; Wright et al., 2010; Sur et al., 2012; Kieffer-Kwon et al., 2013; Shi et al., 2013).

However, transcription alone cannot account for the enormous differences observed in mRNA expression in response to either proliferative or anti-proliferative stimuli (Blanchard et al., 1985). In fact, *Myc* mRNA is very short lived and has a very rapid turnover rate with a half-life in the order of 15-20 minutes. *Myc* mRNA decay occurs via two mechanisms. The first is a translation-independent mechanism involving the shortening of the poly(A) tail and is regulated by AU-rich sequences in the 3' untranslated region (UTR) (Jones and Cole, 1987). The second is translation-dependent where a portion of mRNA corresponding to the CTD of the protein, known as coding region determinant, is bound by another protein, and hence, protected from endonuclease attack (Bernstein et al., 1992). In addition, *Myc* mRNA stability can be influenced by its association with ribosomes where treatment of cells with translational inhibitors prolongs its half-life (Ross, 1995). A further level of translational modulation is provided by the action of miRNAs like let-7 which inhibit *Myc* expression (Kim et al., 2009b). Precise levels of MYC protein and its activity are also regulated by a series of phosphorylation events. ERK mediated phosphorylation of the Ser62 residue on MYC increases protein stability and causes a rapid accumulation of MYC in early G1 phase. In late G1, phosphorylation of Thr58 by GSK-3β directs MYC ubiquitination and proteasomal degradation causing rapid downregulation (Sears, 2004). Therefore, MYC levels in the cell are tightly regulated at both transcriptional and post-transcriptional level.

Several reports have suggested that cohesin is also involved in the transcriptional regulation of the c-*Myc* gene. *Rad21* depleted zebrafish embryos showed a downregulation of the endogenous *Myc* gene expression. Depletion of *Rad21* or Nipped-B in *Drosophila* also reduced the expression of Myc and its target genes (Rhodes et al., 2010). Later studies in MCF7 breast cancer cell lines showed that siRNA mediated RAD21 depletion decreased *Myc* expression and blocked estradiol-mediated activation of *Myc* (McEwan et al., 2012). Proteomic profiling of cell lines from CdLS patients also revealed a deregulated network of proteins centred around MYC (Gimigliano et al., 2012). Together these experiments suggest that cohesin positively regulates *Myc* expression. However, it is noteworthy that the above studies also exhibited an upregulation of p53 and *Mdm2* expression, markers of cellular stress, upon cohesin depletion. Thus it is difficult to conclude from the above studies whether the impact on *Myc* expression is a direct or an indirect consequence of cohesin depletion. Interestingly, Wapl depletion in serum starved non-dividing MEFs, which increases the association of cohesin complex with DNA, also caused a downregulation of *Myc* expression (Tedeschi et al., 2013). These findings point towards a

potential role of cohesin in *Myc* regulation but require a more careful analysis to understand its role more definitively.

1.6 The major histocompatibility complex genes and their regulation by cohesin and CTCF

The immune system uses several mechanisms to recognise, interpret and respond to pathogens. One of the key processes is to be able to detect a pathogenic antigen and be able to discriminate it from self-antigens. Every mammalian species possesses a tightly linked cluster of genes, the major histocompatibility complex (MHC), whose products act as antigen-presenting structures for T cells which selectively recognise non-self antigens (Kindt et al., 2007).

The MHC genes were first identified by the work of Peter Gorer and George Snell as they established that the antigens encoded by these genes took part in the rejection of transplanted tumors in inbred strains of mice. Today it is known that the major histocompatibility complex is a collection of genes arrayed within a long continuous stretch of DNA on chromosome 6 in humans, known as the HLA complex and on chromosome 17 in mice, referred to as the H-2 complex (Figure 1.13). The MHC-encoded class I and class II molecules serve as the peptide-binding transport and display proteins, evoking effector response upon recognition by T cells. Other gene products, some also encoded within the MHC, participate in either the formation or translocation of peptides, or the trafficking of MHC molecules from the intracellular compartments to plasma membrane. Together they provide for antigen processing and presentation (Kindt et al., 2007).

The mouse H-2 complex

Complex	H–2				
MHC class	I			II	
Region	K	D		IA	IE
Gene products	H–2K	H–2D	H–2L	IA αβ	IE αβ

The human HLA complex

Complex	HLA					
MHC class	I			II		
Region	B	C	A	DP	DQ	DR
Gene products	HLA-B	HLA-C	HLA-A	DP αβ	DQ αβ	DR αβ

Figure 1.13 The major histocompatibility complex in mouse and human. The MHC is referred to as the H-2 complex in mice and as the HLA complex in humans. In both species the MHC is organized into a number of regions encoding class I and class II gene products. The class I and class II gene products shown in this figure are considered to be the classical MHC molecules.

The array of MHC genes, spanning a genomic region of about 4 Mb, provides an excellent prospect to study how the expression of constituent genes involved in the same pathway of antigen processing and presentation is coordinately co-regulated in a cell-type specific manner in spite of being separated by large genomic distances and investigate the role of cohesin in this process.

1.6.1 The class I and class II MHC molecules

The Class I MHC genes encode glycoproteins expressed on the surface of nearly all nucleated cells. They are designated as HLA-A, HLA-B and HLA-C in humans and H2-K, H2-D and H2-L in mice. MHC Class I molecules consist of a 45-kDa transmembrane polypeptide α chain associated with a 12-kDa light chain β_2 microglobulin molecule. Their major function is presentation of peptide antigens to $CD8^+$ cytotoxic T cells. These peptides are derived from endogenous intracellular proteins digested in the cytosol. The digested peptides are transported into the endoplasmic reticulum where they interact with the class I MHC molecules before being presented on the cell surface. The peptide-binding cleft on the top surface of the MHC class I molecules is large enough to bind a peptide of 8-10 amino acids. In normal, healthy cells, class I molecules display self-peptides resulting from normal turnover of self-proteins. In case of an infection, by a virus for example, the cell will display viral peptides as well as self-antigens associated with MHC class I molecules. The presented antigen is recognised by the T cell receptors (TCRs) and the interaction is stabilised by the co-receptor CD8. $CD8^+$ T cells recognise the MHC classI:peptide complexes and get activated to kill cells displaying foreign antigens derived from cytosolic pathogens. Different cell types express different levels of MHC class I molecules. The highest levels are expressed by lymphocytes, where they constitute approximately 1% of the total plasma membrane proteins. In contrast, fibroblasts, muscle cells, hepatocytes and neural cells express very low levels of MHC class I molecules (Kindt, 2006; Alberts, 2008).

The class II MHC genes encode the α and β glycoprotein chains of the MHC class II molecules designated as HLA-DR, DP and DQ in humans and H-2IA and –IE in mice. The class II molecules are expressed constitutively only by professional antigen presenting cells (APCs), primarily macrophages, dendritic cells and B cells. All other cell types lack constitutive expression of MHC-II molecules, but their expression can be induced under certain conditions, such as upon activation in human T cells or by exposure to cytokines, of which interferon γ (IFNγ) is the most potent. Class II MHC molecules are comprised of two different polypeptide chains, a 33-kDa α chain and a 28-kDa β chain associated by non-covalent interactions. The membrane distal portion of a class II molecule forms the antigen-binding cleft for the processed antigen of about 13-18 amino acid residues. The presented antigens are generally derived from exogenous proteins present in the microbes or viruses. The proteins are internalised by phagocytosis or by receptor-mediated endocytosis and enter the endocytic processing pathway where they are degraded. The degraded antigenic peptides bind to heterodimeric MHC class II molecules and are presented to the helper

T cells where the recognition of the antigen by TCRs is stabilised by the co-receptor CD4. CD4 T cells are specialised to activate other immune effector cells, for example B cells or macrophages, to act against the foreign antigens or pathogens that the presenting cell has taken up (Kindt et al., 2007; Alberts, 2008).

Table 1.3: Properties of class I and class II MHC proteins

	CLASS I	CLASS II
Chain structure	α chain + β_2-microglobulin	α chain + β chain
Cell distribution	most nucleated cells	dendritic cells, B cells, macrophages, thymus epithelial cells, some others
Presents antigen to	cytotoxic T cells	helper T cells, regulatory T cells
Source of peptide fragments	mainly proteins made in cytoplasm	mainly endocytosed plasma membrane and extracellular proteins
Recognition by co-receptor	CD8	CD4

The MHC locus is highly polymorphic i.e. many alternative forms of the gene, or alleles, exist at each locus among the population. As the genes within the locus lie close together, the recombination frequency within the locus is fairly low and most individuals inherit the alleles encoded by these closely linked loci as two sets, one from each parent. Each set of alleles is referred to as a haplotype. Inbred strains of mice that have identical alleles at the MHC loci, have been designated as prototype strains. For example, the C57BL/6 mice are assigned the haplotype *H-2*^{b} and the respective alleles are labelled as *H-2K*^{b}, *H-2Ab*^{b} etc. MHC polymorphism is of critical importance in antigen recognition by T cells. It increases the range of peptides bound by MHC and the direct interaction of the MHC molecule with the T-cell receptor (Doan et al., 2012).

As the MHC-I and MHC-II genes play such a critical role in immune response, their expression is tightly regulated at the transcriptional level by a variety of transcription factors that interact with conserved *cis*-regulatory promoter elements. Another level of transcriptional control is provided through chromatin histone modifications and long-range DNA interactions between regulatory elements. The activation of MHC-I genes is mediated by several conserved DNA elements formed of an enhancer A, interferon-stimulated response element (ISRE) and the SXY-module (comprising the S/W, X1, X2 and Y boxes). MHC-II promoters differ from MHC-I promoters in that they lack the typical enhancer A and ISRE. The sequence and stereo-specific alignment of the various boxes in SXY-module is highly conserved and critical for its functioning in constitutive and inducible transcriptional activation of MHC genes. The X box is bound by RFX (regulatory factor X), a trimeric complex composed of RFX5, RFXANK and RFXAP (Steimle et al., 1995; Durand et al., 1997; Masternak et al., 1998; Nagarajan et al., 1999). The X2BP complex that includes CREB, recognises the X2 box (Moreno et al., 1999). The trimeric NF-Y complex, composed

of NF-YA, NF-Yb and NF-YC, binds to the Y box (Mantovani, 1999). Together, the SXY-module is cooperatively bound by this multi-protein complex which acts as an enhanceosome. The assembly of the enhanceosome results in the formation of an enhancer surface that optimally interacts with and recruits the transcriptional apparatus (Thanos and Maniatis, 1995) driving the transactivation of the MHC genes (Figure 1.14).

1.6.2 CIITA: the master regulator of MHC class II expression

As the enhanceosome components are expressed more or less ubiquitously, they fail to account for either the cell-type specificity or the IFNγ inducibility of MHC II expression. This function can be attributed to the class II transactivator (CIITA) gene. CIITA was identified by complementation cloning using MHC II negative mutant cell lines. CIITA expression was necessary and sufficient to restore the expression of all MHC II isotypes in cells from patients with bare lymphocyte syndrome which show a total lack of MHC class II expression (Steimle et al., 1993, 1994). Indeed, CIITA exhibits a cell-specific, cytokine-inducible and differentiation-stage-specific pattern of expression that precisely parallels that of MHC class II genes. In addition, CIITA expression is controlled and induced by IFNγ. Thus CIITA acts as a master transcriptional regulator essential for MHC II transcription and contributes to the activation of MHC-I promoters (Steimle et al., 1993; Martin et al., 1997). So far, MHC II and related genes remain the major known targets of CIITA. However, there is growing evidence that CIITA may influence the expression of additional genes. (LeibundGut-Landmann et al., 2004).

CIITA belongs to the large NLR (nucleotide binding domain, leucine-rich repeat containing) family of proteins (Harton et al., 2002). The N-terminal activation domain is necessary for transcriptional activation. The centrally located NBD (nuclear binding domain) contains a GTP-binding domain which is required for the nuclear import of CIITA. The C-terminal LRRs (leucine rich repeats) are also important for nuclear import and for interactions with other proteins (Harton and Ting, 2000). As CIITA lacks a DNA binding motif, it exerts its transactivating function through protein-protein interactions with the components of the MHC-enhanceosome bound to the proximal SXY regulatory module in MHC promoters (Zhu et al., 2000; Jabrane-Ferrat, 2003). In addition, it forms a network of multiple interactions with various transcriptional cofactors. CIITA interacts with several chromatin modifying proteins including histone acetyltransferases (HATs like CBP/p300, pCAF and SRC-1), histone methyltransferases (HMTs like CARM1), ATP-dependent remodelling enzymes (BRG1) and histone deacetylases (HDAC1, weakly with HDAC2).

CIITA also recruits multiple components of the basal transcription machinery including TBP, components of the $TAF_{II}D$ complex, the kinases CDK7, CDK9 and pTEFb (Zika and Ting, 2005). Thus, CIITA acts as the central molecule modulating the chromatin structure, MHC enhanceosome assembly and transcription initiation and elongation processes. Therefore, it controls the transcription of MHC genes II genes by coordinating the temporal recruitment of distinct cofactors in response to cytokines and other stimuli (Figure 1.14).

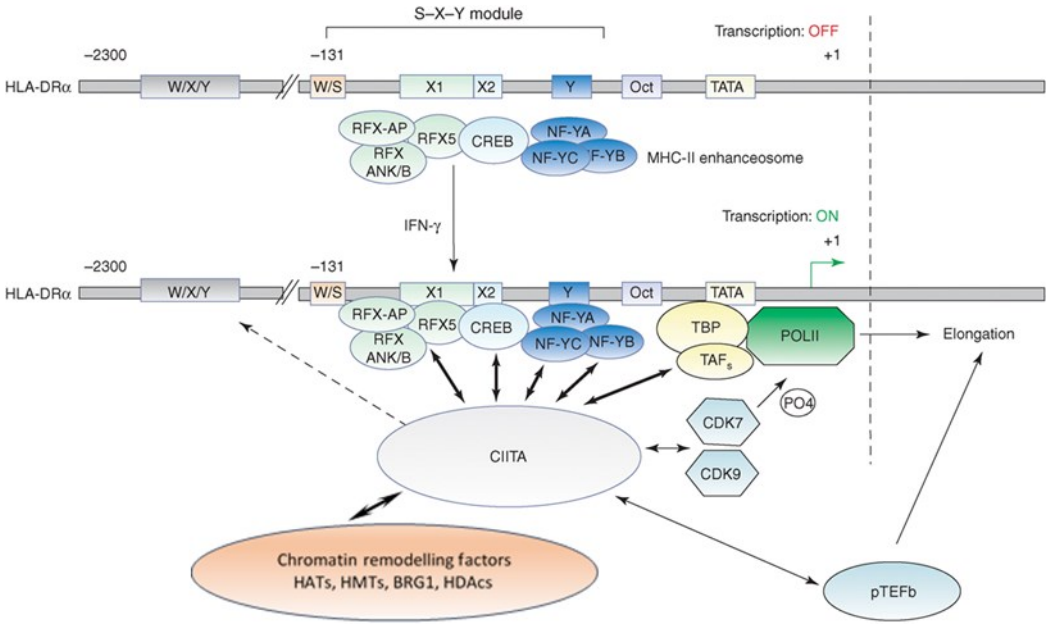


Figure 1.14 MHC II enhanceosome on the HLA-DRa promoter. MHC II promoters share a common set of cis-acting elements including the S-X-Y module which binds the heterotrimeric transcription factor NF-Y, the RFX complex and CREB. These ubiquitously expressed transcription factors bind the S-X-Y module in a stereospecific and cooperative manner and provide a surface for CIITA recruitment upon IFNγ treatment, which is essential for MHC II expression. CIITA interacts with an array of TAFs, elongation factors like pTEFb, kinases like CDK7, CDK9 and chromatin remodelling factors like HATs, HMTs, BRG1, HDACs and mediates the transcription of these genes [modified from Zika and Ting, 2005].

The regulation of CIITA expression occurs primarily at the level of transcription of the *MHC2TA* gene which involves the cell-type or stimulus-specific usage of its four different promoters (pI to pIV) (Figure 1.15). Promoters pI, pIII and pIV are highly conserved between human and mouse genes. Promoter pII has only been found in the human gene. It displays very low transcriptional activity and its significance remains unknown. The different promoters do not share any sequence homology and are not co-regulated. Each promoter precedes a distinct first exon which is alternatively spliced to the shared downstream exons. Thus transcription from

these promoters produces three types of transcripts possessing different 5' ends (Muhlethaler-Mottet et al., 1997). Experiments in knockout mouse strains lacking one or more of the promoter regions have allowed researchers to define more precisely the function of each promoter *in vivo* (Waldburger et al., 2001; Pai et al., 2002). Each of these promoters has a different function and is active only in specific cell types or in response to particular stimuli. Conventional dendritic cells and IFNγ induced macrophages (cells of myeloid origin) predominantly use the pI promoter for CIITA expression. In B cells and activated human T cells of lymphoid origin, transcription initiates from pIII (Lennon et al., 1997). Unlike conventional dendritic cells, CIITA expression in plasmacytoid dendritic cells (pDC) is also controlled by pIII. pIV is essential for CIITA expression in response to IFNγ signals in cells of non-hematopoietic origin. pIV also drives CIITA expression in cortical thymic epithelial cells which constitutively express MHC II molecules required for the positive selection of $CD4^+$ T cells (Waldburger et al., 2001; LeibundGut-Landmann et al., 2004).

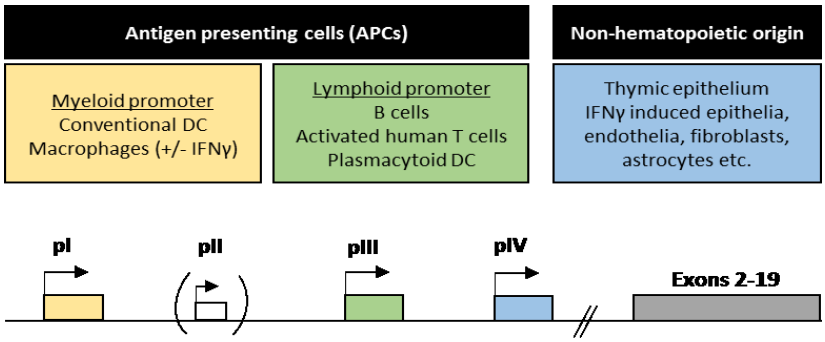


Figure 1.15 CIITA promoter usage. Expression of the CIITA gene is controlled by three independent promoters having different functions: pI is used in cells of myeloid origin (conventional dendritic cells (DC), macrophages); pIII is active in cell of lymphoid origin (B cells, activated human T cells, pDC); pIV is necessary for responsiveness to IFNγ in non-hematopoietic cells and is expressed constitutively in thymic epithelial cells. The three types of CIITA mRNA encode three different isoforms differing only at their N-terminal ends. The significance of the pII promoter present in human gene is still not known.

The activity of the CIITA promoters is further regulated by several epigenetic mechanisms and interacting factors. Activation of pIII in B cells requires CREB-1, PU.1. IRF-4 and E47 transcription factors to interact with the different regulatory motifs present in the promoter region (van der Stoep et al., 2004). Transcription mediated by CIITA-pI in immature dendritic cells requires binding of PU.1, IRF-8, NF-κB and Sp-1 to the promoter (Smith et al., 2011).

CIITA expression also needs to be repressed in order to downregulate MHC II expression under certain physiological and pathological situations. For example, immature dendritic cells

shut down *de novo* synthesis of MHC II molecules when they undergo maturation in response to an infection. This is mediated by a global repression mechanism implicating histone deacetylation spanning the CIITA regulatory region. During terminal differentiation of B cells into plasma cells, CIITA expression is actively silenced by Blimp-1 (Piskurich et al., 2000). Trophoblast cells and embryonic stem cells lack the expression of MHC II molecules both constitutively and after exposure to IFNγ which is thought to be critical in preventing rejection of the fetus by the maternal immune system. These cells are hypermethylated at CpG dinucleotides present in the pIV region rendering the cells unable to activate CIITA transcription (Morris et al., 2000; van den Elsen et al., 2000). More recently, a trophoblast derived non-coding RNA has been shown to suppress CIITA expression through an inhibitory domain (Geirsson et al., 2003). MHC molecules also play a pivotal role in presenting tumor derived antigens and hence in activating and regulating antitumor immune response. Consequently, malignant cells employ the strategy of downregulating MHC expression in order to evade elimination by immune system. Loss of constitutive MHC II expression is observed in tumor cells of hematopoietic origin while tumor cells of non-hematopoietic origin show the inability to induce MHC II expression in response to IFNγ (García-Lora et al., 2003). This inability to express MHC II has been attributed to hypermethylation of the CIITA locus (van den Elsen et al., 2003) and to the activity of histone deacetylases (Magner et al., 2000). A variety of pathogens (*cytomegalovirus* (CMV), *Mycobacterium bovis*, *Chlamydia*, *varicella-zoster* virus and Epstein Barr virus (EBV)) have also developed mechanisms to down-regulate MHC II expression in order to escape immune surveillance by the hosts (Harton and Ting, 2000).

1.6.3 IFNγ mediated CIITA induction

Interferon γ (IFNγ) is a soluble cytokine which plays important roles in innate and adaptive immunity against viral and intracellular bacterial infections and tumor control. It is produced by $CD4^+$ T helper cell type 1 (Th1) lymphocytes, $CD8^+$ cytotoxic T cells, natural killer (NK) cells, B cells, natural killer T (NKT) cells and professional APCs (Schroder and Hertzog, 2004). IFNγ production is responsive to cytokines secreted by APCs like IL-12, IL-18 and TNFα (Munder et al., 2001). IFNγ binds to the IFNγ receptor (IFNGR) and primarily signals through the Jak-Stat pathway. Ligand binding induces JAK2 autophosphorylation and activation, which allows JAK1 transphosphorylation by JAK2. The activated JAK1 phosphorylates IFNGR to form docking sites for latent STAT1. Next, STAT1 gets phosphorylated and dissociates from the receptor. The dissociated STAT1 enters the nucleus and binds to promoter elements to initiate or suppress transcription of

IFNγ regulated genes (Subramaniam et al., 2001). Exposure to IFNγ leads to upregulation of several genes in the antigen presentation pathway (like TAP1, MHC I heavy chain, β_2 microglobulin MHC II chains, CIITA), apoptosis, immunomodulation, Th development, leukocyte trafficking and production of NO intermediates (Schroder and Hertzog, 2004). *CIITA* is one of the key effector genes of the IFNγ response pathway. On exposure to IFNγ, STAT1 translocates to the nucleus and binds cooperatively with the constitutively expressed USF-1 at the GAS element and E box present at the pIV promoter of *CIITA* within 5-30 mins. This is followed by IRF1 binding. IRF1 must first be transcribed and translated. It appears at the promoter only after 30-120 mins. IRF1 binding coincides with *CIITA* transcriptional activation. This dependence on IRF1 explains the delayed kinetics of CIITA induction by IFNγ (Steimle et al., 1994). *CIITA* pIV activation is also accompanied by increased H3 and H4 acetylation (Morris et al., 2002). Inhibitors of HDAC, such as trichostatin-A (TSA), enhance *CIITA* and MHC II gene transcription, whereas the recruitment of HDACs causes the dissociation of CIITA, NF-Y and RFX from the MHC II gene promoter (Magner et al., 2000; Osborne et al., 2001; Kanaseki et al., 2003; Zika et al., 2003). The SWI/SNF ATPase BRG1 is another important factor required for transcriptional activation as the cells lacking expression of BRG1 fail to induce *CIITA* expression following IFNγ exposure (Pattenden et al., 2002). BRG1 primarily acts as a local chromatin remodeler by means of nucleosomal sliding or strand exchange which alters the winding of DNA around histones. These changes in chromatin structure increase its accessibility to DNA-binding proteins, such as activators and/or the general transcription machinery (Trotter and Archer, 2008). At *CIITA* pIV, BRG-1 recruitment facilitates the deposition of active histone marks which further aids the formation of long-range chromatin and promoter interactions through many interdependent remote enhancers. At the uninduced locus, many sites form BRG1 independent loops. The site at +59 kb loops and contacts the site at -50 kb and the element at -8 kb bridges this complex to pIV. BRG1 is recruited to the -50 kb site which triggers the appearance of an active histone mark. This change is crucial, as subsequent IFNγ-induced recruitment of transcription factors, histone modifications and additional loops, show BRG1 dependency. Thus BRG1 recruitment makes the promoter poised for induction (Ni et al., 2008).

IFNγ activated expression of CIITA can also be suppressed by a number of different stimuli including TGFβ, IL-1. IL-4 and IL-10. Suppressor of cytokine signaling (SOCS1) is induced by IFNγ which further negatively regulates the IFNγ signal transduction pathway by binding to JAK2 and inhibiting its kinase activity. Similar to SOCS1, nitric oxide – which is produced by macrophages upon IFNγ stimulation – may act as feedback inhibitor of MHC II synthesis by inhibiting *CIITA*

expression. Statins (HMG-CoA reductase inhibitors), known for their cholesterol-lowering effect, also exhibit anti-inflammatory properties due to inhibition of IFNγ induced CIITA activation (LeibundGut-Landmann et al., 2004).

1.6.4 Role of cohesin and CTCF in MHC regulation

The MHC gene cluster is interspersed with several CTCF and cohesin binding sites which are reported to be involved in the regulation of MHC expression. The intergenic region between the *HLA-DRB1* and *HLA-DQA1* genes in the human MHC contains a CTCF binding site – *XL9,* which has enhancer-blocking activity (Majumder et al., 2006). This site serves as a focus for long-range interactions with the promoters of the *HLA-DRB1* and *HLA-DQA1* genes (Majumder et al., 2008). These interactions are dependent on the presence of both CTCF and CIITA. Knockdown of CTCF resulted in decreased expression of *HLA-DRB1* and *HLA-DQA1*, reduced histone acetylation at these gene promoters and the loss of their interaction with the *XL9* site. CIITA and CTCF co-immunoprecipitate, suggesting that the two factors interact directly or indirectly through a complex of factors (Majumder et al., 2008). Further investigation identified ten other CTCF binding sites within the human MHC region and it was shown that all CIITA-regulated genes within the MHC II locus required CTCF for maximal expression. The five non-CIITA constitutively expressed genes within the locus (*TAP1, TAP2, PSMB8, PSMB9* and *BRD2*) were not affected by CTCF depletion. In the basal state, a set of self-interactions were observed between the CTCF sites which were independent of MHC II gene transcription (similar in B cells and fibroblasts) and occurred in the absence or presence of CIITA. An additional set of interactions was observed between the CTCF sites and the proximal promoter regions of MHC II genes, but only in the presence of CIITA. These interactions were also dependent on CTCF and siRNA mediated depletion of CTCF lead to loss of interactions and disrupted transcription of MHC II genes. Thus, CTCF organises the MHC II locus into a basal architecture of interacting foci and DNA loops that rearranges in the presence of CIITA. Loss of gene expression upon disruption of DNA interactions, suggests that these structures play a key role in co-regulating MHC II gene expression (Majumder and Boss, 2010).

Cohesin was also found to be associated with each of the CTCF sites identified irrespective of MHC II expression but was required for the optimal expression of the HLA-DR and HLA-DQ genes. Cohesin interacts with CTCF, RFX5 and CIITA in a DNA dependent manner and is important for DNA looping interactions between the HLA-DRA promoter region and the *C1* CTCF element.

The *C1* CTCF binding element lies 24 kb upstream of the *HLA-DRA* transcription start site and is the most 5' of the CTCF binding sites in the MHC locus. In the transcriptionally inactive state, *C1* interacts with *XL9* forming a large chromatin loop that encompasses the HLA-DR subregion. In the transcriptionally active state, *C1* interacts directly with the *HLA-DRA* gene. *C1* exhibits enhancer blocking activity and this was lost upon cohesin depletion and cohesin binding to *C1* was dependent on CTCF. Cohesin depletion also demonstrated that the interactions between *C1* and HLA-DRA promoter were dependent on cohesin subunits, but the interaction between the CTCF sites themselves were not. Thus cohesin plays an important role in MHC regulation by mediating the interactions between promoter regulatory elements and CTCF sites within the MHC locus (Majumder and Boss, 2011). Interestingly, both CTCF and cohesin depletion did not have any effect on the expression of *CIITA* (Majumder et al., 2008; Majumder and Boss, 2010, 2011).

1.7 Aims of this study

Studies in various cellular systems from different model organisms and CdLS patient cell lines show that cohesin plays an important role in the regulation of gene expression. Cohesin associates with DNA at active gene regulatory elements, particularly promoters and enhancers, in a cell type-specific manner. Current models suggest that cohesin influences gene expression by mediating long-range chromatin interactions between such gene regulatory elements. However, the rules governing this process are still unknown. Moreover, it is not clear whether the changes observed in the interactions between a gene promoter and its regulatory elements are cause or consequence of changes in gene transcription.

In this thesis, I aim to explore the mechanisms of cohesin mediated gene regulation by conditionally depleting the RAD21 subunit of the cohesin complex in different cellular systems. In doing so, I intend to analyse the gene expression changes at early time points or in non-dividing cells in order to avoid any cell-cycle stress related secondary effects. First, I will investigate the impact of cohesin depletion on the gene expression profile of pluripotent mouse ES cells. I will further examine enhancer-promoter interactions at some key pluripotency associated genes deregulated upon cohesin depletion. Next, I will focus on studying the role of cohesin in modulating the expression of *Myc* gene. Finally, I will test the hypothesis that cohesin plays a vital role in the process of gene activation by studying IFNγ-inducible MHC expression in MEFs. Using this system, I will also explore if cohesin impacts gene expression by regulating the activity of RNA polymerase and if the MLL3/4 complex is an important partner in this process.

Overall, this study will provide a better understanding of cohesin functions in gene regulation, which may help explain the etiology of CdLS and the incidence of cohesin mutations in human malignancies.

Chapter 2 : Materials and Methods

2.1 Materials

2.1.1 Antibodies

RAD21	ab992 (Rabbit polyclonal to Rad21; Abcam, Cambridge, UK). Used for Western blot at 1:1000 dilution ab154769 (Rabbit polyclonal to Rad21; Abcam, Cambridge, UK). Used for IF analysis at 1:100 dilution
CTCF	ab70303 (Rabbit polyclonal to CTCF; Abcam).
SMC3	ab9263 (Rabbit polyclonal to SMC3; Abcam) Used for Western blot at 1:1000 dilution.
OCT4	sc-8628 (Goat polyclonal to Oct3/4; Santa Cruz Biotechnology, Inc., Santa Cruz, CA, USA). Used for western blot at 1:50,000 dilution.
NANOG	REC-RCAB0001P (Cosmo Bio, USA) Used for Western blot at 1:500 dilution
MYC	N-262 sc-764 (Rabbit polyclonal to Myc; Santa Cruz Biotechnology). Used for Western blot to detect endogenous Myc at 1:1000 dilution.
Phospho c-MYC	ab32029 (Monoclonal to T58+S62 phosphorylated MYC; Abcam). Used for Western blot at 1:1000 dilution.
H3	ab1791 (Rabbit polyclonal anti-mouse histone H3; Abcam, Cambridge, UK).
H3K4me3	ab8580 (Rabbit polyclonal to anti-histone H3 (tri methyl K4); Abcam).
Tubulin	T9026 (Mouse monoclonal anti-α tubulin; Sigma). Used for western blot at 1:2000 dilution
NIPBL	A301-779A-3 (Rabbit polyclonal; Bethyl Laboratories)
RNA Pol II	H224:sc-9001 (Rabbit polyclonal to NTD of Pol II; Santa Cruz Biotechnology)

RNA Pol II CTD phospho Ser2	RNA Pol II CTD Ser2-P 3E10 (Rat polyclonal to Ser2 phosphorylated CTD of RNA Pol II; Active Motif)
RNA Pol II CTD phospho Ser5	CTD4H8 (Mouse monoclonal to Ser5 phosphorylated and unphosphorylated forms of RNA Pol II CTD)
STAT1	#9172 (Rabbit polyclonal to STAT1; Cell Signalling technology) used for Western blot at 1:1000 dilution
LAMIN B	sc-6216 (Goat polyclonal anti-mouse Lamin B; Santa Cruz Biotechnology, Inc., Santa Cruz, CA, USA). Used for western blot at 1:10,000 dilution.
IgG	DakoCytomation Z0259 (Rabbit polyclonal against mouse IgG)
I-A^k	FITC conjugated mouse anti-mouse antibody against I-A^k (Aα^k) (BD Pharmingen clone 11-5.2). Used at 1:100 dilution for FACS analysis.
I-A^b	FITC conjugated mouse anti-mouse antibody against I-A^b (Aβ^b) (BD Pharmingen clone 25-9-17). Used at 1:100 dilution for FACS analysis.
H-2D^b	FITC conjugated mouse anti-mouse antibody against H-2D^b (BD Pharmingen clone KH95). Used at 1:100 dilution for FACS analysis.
H-2D^d	FITC conjugated mouse anti-mouse antibody against H-2D^k (BD Pharmingen clone 34-2-12). Used at 1:100 dilution for FACS analysis.

2.1.2 Reagents

Doxorubicin (Cat: 11010, Serva Electrophoresis GmBH), Retinoic acid (Sigma R2625), 4'-Hydroxytamoxifen (Sigma), HDAC inhibitor PCI-34051 (Selleckchem), PhosphoStop (Phosphatase inhibitors from Roche Diagnostics GmBH), recombinant mouse interferon-gamma (mIFNγ), Myc inhibitor 10058-F4 (Sigma), Proteinase K (Sigma), Protease inhibitor cocktail (11697498001, Roche), EDTA free Protease inhibitors (11873580001, Roche), Amersham ECL plus kit (GE Healthcare) for Western blot detection, STI-571 (Imatinib), PDO325901 (Stemgent), CHIR99021 (Stemgent), Quantifast Probe PCR kit (Qiagen).

2.1.3 Primers

2.1.3.1 Genotyping primers

Gene		Primer sequence 5'-3'
Rad21	F	GAAGGTCATGGTTGGCAGAT
	R	CAATCCAAGCTTCATTTGTATTT
Smc3	F	ACCTGGACCAGAATAGCTGA
	R	CACACTGCTACACCTGAACACA
Cre	F	GCGGTCTGGCAGTAAAAACTAT
	R	CCGTCAGTACGTGAGATATCT
Trnah3	F	TGTGCCAGCATCTTTTGCC
	R	GCGTGCTGGAATTAAAAGCC

2.1.3.2 RT-qPCR primers for expression analysis

Gene		Primer sequence 5'-3'
H2-Ab1	F	TGCCTTAGAGATGGCTCTGC
	R	CCATGAACTGGTACACGAAATG
H2-Aa	F	TGGAGGTGAAGACGACATTG
	R	CTCATCACCATCAAATTCAAATG
H2-K1	F	CCGCAGAACTCAGAAGTCGC
	R	GAAATACCTCAGCGAGTGTGG
H2-D1	F	ACATGCCGTGTGTACCATGA
	R	ACACCCAGAACAGCAACGAT
Tap1	F	CTGGTCACCCTGATCAACCT
	R	TGCGTGGACTTTGCTAGAGA
Ciita **mRNA**	F	AGACAAGGGTGTGTCGAAGC
	R	CAGTGATGTTGTTTTGGGACA
Ciita **pri. Transcript**	F	GCCGACCCCCTACATCTCTA
	R	GCCTGCAGAAGTCCTGAGAA

Gene		Primer sequence 5'-3'
Rfx1	F	GCAGCTTACCAACATTCAGGT
	R	GACCTGGCCACTCTTAGTGC
Rfx5	F	TCGACTTGAAGGGATCTGAAA
	R	GCACAGGCTGCTTCTACCAG
Irf1	F	CCTGGGTCAGGACTTGGATA
	R	GAGACTGCTGCTGACGACAC
Socs1	F	GGAGACCTCATCCCACCTCT
	R	CCCAGACACAAGCTGCTACA
Tgfb	F	TCAGACATTCGGGAAGCAGT
	R	ACGCCAGGAATTGTTGCTAT
Stat1	F	CAACATGCTGGTGACAGAGC
	R	AAACTGCCAACTCAACACCTC
Prdm1	F	AGTTCCCAAGAATGCCAACA
	R	TTTCTCCTCATTAAAGCCATCAA
Irf2	F	TTGTATTGGTAGCGTGGAAAAA
	R	TGGTATCGTATTGGAATTTATCTGC
Irf8	F	CCTATGACACACACCATTCAGC
	R	AGAGACGGCAGCCTTCAAGG
Phc1	F	AACACCACTTCCCCACCTCT
	R	TCCGGTTAACCTGCAATAGG
Paxip1	F	GCAGCAGCAGCAGCTTTTTG
	R	TGCTCGGGATAGTCCGCAAT
Smc3	F	CGAAGTTACCGAGACCAAACA
	R	TCACTGAGAACAAACTGGATTGC
Pvt1	F	GGCTTGCCTGGAGATGTTAC
	R	CAGGTAGCCCGAGAGATGAC
Lefty1	F	GTTCAGCCAGAACCTTCGAG
	R	GCTCCATTCCGAACACTAGC
Klf4	F	TCCTTTCCAACTCGCTAACCC
	R	CGGATCGGATAGCTGAAGCTG

Gene		Primer sequence 5'-3'
Il2	F	AAATGTGTTGTCAGAGCCCT
	R	GTGCCTAGAAGATGAACTTGG
18SrRNA	F	GTAACCCGTTGAACCCCATT
	R	CCATCCAATCGGTAGTAGCG
bActin	F	ACCAACTGGGACGATATGGAGAAGA
	R	TACGACCAGAGGCATACAGGGACAA
Myc isoA/B	F	GCCCAAATCCTGTACCTCGTCC
	R	CTCTTCTCCACAGACACCACATCA
Myc isoA	F	TTTTTGTCTATTTGGGGACAGTG
	R	ACGGAGTCGTAGTCGAGGTC
Myc pri. ex1-in1	F	AGAGCTCCTCGAGCTGTTTG
	R	ACCCCGACTCAGATCTACCC
Myc pri. in1-ex2	F	AAGGGGAGTGGTTCAGGATT
	R	CGCAGATGAAATAGGGCTGT
Myc pri. in2-ex3	F	TTCTCACCTGTGCCCTAACC
	R	GGTTTGCCTCTTCTCCACAG
Rad21	F	AGGAAGAAGCTTTTGCGTTG
	R	CGCTAAGCTGGGCTCTAATG
Nanog	F	GCATCTTCTGCTTCCTGGCAA
	R	GAACTATTCTTGCTTACAAGGGTCTGC
Pou5f1	F	CCCAAGGTGATCCTCTTCTGCTT
	R	GAGAAGGTGGAACCAACTCCCG
Ubc	F	AGGAGGCTGATGAAGGAGCTTGA
	R	TGGTTTGAATGGATACTCTGCTGGA
Ywhaz (HAZ)	F	CGTTGTAGGAGCCCGTAGGTCAT
	R	TCTGGTTGCGGAAGCATTGGG
Nipbl	F	CCATGTCCCCATAACTACGCT
	R	AGTTCACCTCTTCTGCTATTC
Mll4	F	GCTATCACCCGTACTGTGTCAACA
	R	CACACACGATACACTCCACACAA

Gene		Primer sequence 5'-3'
Gapdh	F	TGCACCACCAACTGCTTAGC
	R	GGCATGGACTGTGGTCATGAG
p53	F	CCTGGCTGTAGGTAGCGACT
	R	ACTCCTCCATGGCAGTCATC
p21	F	GCAGACCAGCCTGACAGATT
	R	GAGGGCTAAGGCCGAAGA
Mdm2	F	TGTGTGAGCTGAGGGAGATG
	R	CACTTACGCCATCGTCAAGA
Sox1	F	GTGACATCTGCCCCCATC
	R	GAGGCCAGTCTGGTGTCAG

2.1.3.3 Primers for 3C

Primers for Nanog 3C:

TAQ PROBE:	[FAM]-TAT CAA GAA GTC AGA AGG AAG TGA GCC GC-[BHQ1]	
Distance from anchor (kb)	**3C primer**	**Reverse primer**
Anchor	**TGGCCTTCAGATAGGCTGAT**	CCAGGAAGACCCACACTCAT
Normalisation fragment	AAACGGGCTGAAGGGTTATT	CCCCGAACATATTCCAAAGA
3.6	TGAAGACTGCTTTCTCTGtcc	CGGACTAACCAAGGGCTACA
4.3	CTCGAATGTTGGGCTTAGGA	TTCTGCCACTCACACCTCAG
5.1	CCCTCCTCCCTATTCAAACC	AGTCAAGGCCACCATAGCC
6	GTGGGTGCACACAGAGAACA	AGGACATGCGTTCAGTCTCC
8.25	ATGCATTTCATCCCAGCACT	TGGGGTTGGAAAAGTCAAAG
9.1	CTGGAGAGTATTGCGCCTTC	CTGGGTTGGTGAAGATTCCA
12.7	GGTGGAGTGGCATACACCTT	CCTGTGGTCTGCTCTCCATT
13.4	CACCCCCAGACAGACTGATT	CAACCAGCCCAGGTTTCTAA

Primers for Lefty1 3C:

TAQ PROBE:	[FAM]-CTT CCC TGA GGC TAA CCA GCG ACA GTG-[BHQ]	
Distance from anchor (kb)	**3C primer**	**Reverse primer**
Anchor	**CAGGGACACACACATCCAAG**	TGTTGTAGCAGGGCCACATA
Normalisation fragment	GTCCTGGACAAGGCTGATGT	ACACCACAGGTGGAAGGAAG
2.7	AGGAGGAGCAAAGGAAGAGC	CACTGTTAACATACTTGAGAGGTGAAA
4.44	CGCAAGCAGGATGTTTTCA	TGGTTCTCTCCTCCCATCAT
6	GCACTGAGCGATACAAACCA	TTTTCTGGCATTAGCAAGCA
9.07	CCCAGTCTTTATGCCATGCT	AGTTTACCCATCCCCTCACC
10.8	AGGGGCAGAGAACATTTGAA	TCTTTAACGATGCTGCGATG
12.3	TGGTCGCCTCACTCCTAGTC	TGAGGAGAGACTGCCATGTG
14.46	GCTGTGTCCAGCCCTCTTAG	GGCCAGCATCTCAGCTTTAC
15.22	CTTTGGCTGCTCCTATCTGG	GAATCCCTCCCTGGCTACTC
16.6	ACGGATGGCAGACCTGTAAG	CCTAAGAAGCCACCTTGCAC
19.1	AGGATGACATAGCCCAGCAC	CAGCTCCAGGAACAGACCAT

Primers for Klf4 3C:

TAQ PROBE2	[FAM]-TCGTGGGAAGACAGTGTGAAAGGTTAGAAA-[BHQ1]	
Distance from anchor (kb)	**3C primer**	**Reverse primer**
Anchor	**TCCCACGTAGTGGATGTGAC**	CCATTCACAAGCTGACTTGC
Normalisation fragment	CAAAAACCGTGAGTGTGGTG	CGAGTGCCTTTCTTCAGTCC
8.4	CGGAGGCAGGAAGATTAAGA	ACTCCTGCTGATGGTTGGAC
12	ACTGTTGGCCAAAGAGAGGA	TGGCCTCGAACTCAGAAATC
13.4	AGAAATGCAAAGCCCCTAGC	GCTGGTTGACTGTGTGAGGA
15.1	GCTGGTTGACTGTGTGAGGA	AGAAATGCAAAGCCCCTAGC
24.2	AGCATCAGGACCAGAATTCAA	TATCCTGCACGGTTCACAAA
33.4	TCCTGGCCTGCTGTAGAGAT	ACAATTCTGGAGGGAGGACA
44.2	AGAAGCACAGGCAGGACCTA	GTGGAACGTTAAGGCTGGAA
48	CTCAATCCGTCTGTGCTGTG	CGGACTTCTCCACGAATCAT
49.35	TTGACCTCCATCCACATGAA	AGAGTTTGCCTGGCTGTGTT
49.7	GCAGGAGTCAGTTCCCAGAG	ACTTCCTGCCCAGCTCAGTA
51.2	CAACTTGGCAACCTCCTCAT	ACCTGTGCTTTCTGGAGTGG
51.4	GGAACACAATCAAGGTCAGGA	TGGGCAGAGAGTGGAAAGTT
53.8	TGTGGCCTGGATCCCTAATA	CTCTCCCCACGAATTAACGA
56	ATGTCCTGAAGGTTGGCTGT	TGCTGAATAGGCACTGGTTG
58.2	TCATTTGTCCTCCCTCCACT	GGCAGATCACAGGAACACCT
64	TACCATCACAAACCGGACAA	CTATGGCAGGCAGGAGACAT

Primers for Ciita 3C:

Taq probe	[FAM]-AGGTTCAGGGTGGAAGAGAAGGGGGTAG-[BHQ1]	
Distance from anchor (kb)	**3C primer**	**Reverse primer**
-40	CGACCACCTTCACCTTCTTT	GATCCCAGGATCCACACAGT
-32	GACCAGCAGGCTCCTACAAA	AAGCTCCTCTCTGGGCTTATG
-24	CCCTACCATCTGGGGAAACT	AAAATCTGGGGTGTCCCATT
-21	TTGCCCATTCTAAGCCTGAC	GGAGGGTCCCACCAGTTAGT
-19	CTGTCATTGCTTGGATCGTC	CCTCTAGCTCCTCGAAAGCA
-6	ACTACCCCGGCTTCTTCTGT	CAGCCAGCCAGTCTATGTGA
Neighbouring fragment	CTGTTGGTGGGACTCAGCTT	AGTCTGTTGACAGGCTgctg
Anchor	**GAACCTCCAGCTGTGAGTAAGA**	AACCACTGTCATCCGAGGTC
8	CAACCTAGCCACCTGCCTTC	TTCCCGTGGGAGCTAGATAA
12	AAGATTTCCCTGACCCTGCT	GGACACCTTAGGGTTGGTCA
22	TGTGGTCTCAAAAGCCACTG	CACAGAGAAGTGGGGACAGA
33	AGATTCGGGGGACTTGATCT	TAAGGATGGACTGGCAGCAT
37	CAGGGAGGAAACTCGAACTG	GGGGAATGAACCATCTGCTA
40	GATGCAACCCTACCTGGAGA	AGGAGCTTCGACCAGATGAA
50	TTTCCTACAGCAAACTGCAAA	GGGGTTGGTCTGGGAACTAT
60	ACATGGGCAGGGTCATTTTA	GCGATACCGTCCATCTTAGG
70	GCCAATAAAACCGACTGGAA	CGGATCCCTAAAAGGCAAGT
74	ATCCTGGAAGGACCTCTGGT	GCTCTCTCCCATGAGGATCA
80	AAATTGGAATTCCCTCAcca	AACCCAGAGCACAGAACTGG

2.1.3.4 ChIP primers

For ChIP in ES cells

Position/Approximate distance from TSS	**Forward primer**	**Reverse primer**
Nanog promoter	CCCTTTAAATCTATCGCCTTGA	AAGGTTTTAGGCAACAACCAAA
Nanog enahncer	ACTCCAAGGCTAGCGATTCA	CTTATCCAGGGAAGCGGTTT
Lefty1 promoter	ACTGGTCTCGAGCCAAGAAA	AAGACTCGTCCCTGGTGTGT
Lefty1 enhancer	TTGCACAATGGGCTTGATTA	GCAGGGTGACAAACTTGGTT
Klf4 p1 (TSS)	CGCCTCTTGCTTAATCTTGG	TTAGCAAAGGAAGCCCAGAC
Klf4 p2 (+15kb)	TTCCAGTCCAGTCCCAAGTC	CCTGGATGGTCTACGTGCTT
Klf4 p3 (+50kb)	CTTGGACACGGTTTTGGTTT	ACTGTGATGTGGCTCTGTCG
Klf4 p4 (+53kb)	CAACTTGGCAACCTCCTCAT	ACCTGTGCTTTCTGGAGTGG
Klf4 p5 (+55kb)	TGTGGCCTGGATCCCTAATA	CTCTCCCCACGAATTAACGA
Klf4 p6 (+66kb)	TCCCTTGCTAGGCGATAATG	GGAGCAAGGAACTTGGCTTA

For ChIP in fibroblasts:

Gene		Primer sequence 5'-3'
Ciita pI	F	TCCTTCATCCTGGGTCTCAC
	R	GGCTGGAGAGCAGTGTCTGT
Ciita pIII	F	AAGCTGGAGCTCACCATGTC
	R	AATTGGGTGACCACAGAAGC
Ciita pIV	F	AGCGCCACAGATACTCCCTA
	R	CAAAGGGGATCTTGGAGACA
Ciita enh2	F	GGCAAAGTCCAGTTCTGAGG
	R	GTGCATTTCCCGTAAGTGGT
Tap1 TSS	F	AGCTTCCAGAACAGCCTGAG
	R	AGTGCTGGCGTTTAGAGGAA
H-2K1 TSS	F	TTCGCGACTTCTGAGTTCTG
	R	CTGGGTCAGGTCCTTCTGTC
CTCF1	F	CAGCGCTATCCCTAGTGAGC
	R	CAGAAGGTGGCAGTGTGAGA
CTCF2	F	GTTTTTGAGGGGCTGTGGAG
	R	TGGATGGCTGTGCTTCATAG
CTCF3	F	AGATGTCCCTCTGCTGCACT
	R	GGAATTTGCCTCATGCTCTC
Oct4 TSS	F	GGGTGAGAAGGCGAAGTCTGAA
	R	GTGAGCCGTCTTTCCACCAGG
IL4-6	F	TCCAGTCCAGTTCCAAAAGGAGC
	R	CAGGATGAGGATATGACTAGCTGTGG
17:16298	F	AGGACATAGTCGCTTGAGTGATGG
	R	TTGGGCCAGGCTGGTACTTT
Xist7Na	F	TGGCTTGTACTTCCAGATCAT
	R	AATGTATAAGCAAGCTAGTACGCA
Igf2r-2	F	GGTCTCGCCAGCTTGCTATTTC
	R	TGGCTAGATGTCATTGTGGTGG

2.1.4 Genome wide datasets used

The ES cell ChIP-sequencing datasets: Mediator, Nipbl, Smc1, Rad21 and CTCF from (Kagey et al., 2010); p300, H3K4me1, H3K27ac and H3K4me3 from (Shen et al., 2012). ChIP-seq datasets for H3K4me1, H3K4me3, Pol II, CTCF in MEFs are from ENCODE (Bing Ren lab). Macrophage STAT1, H3K27ac, H3K4me1 and H3K4me3 ChIP-seq datasets are from (Ostuni et al., 2013). preB Rad21 and CTCF datasets were generated in our lab. Nipbl and p300 ChIP-seq datasets in activated B cells are from (Yamane et al., 2011). ES cell and B cell ChIA-PET datasets are from (Kieffer-Kwon et al., 2013).

2.2 Methods

2.2.1 Cell culture

2.2.1.1 Mouse ES cells in 2i media

ERT2Cre-*Rad21*$^{WT/WT}$ (A7 cell line) and ERT2Cre-*Rad21*$^{lox/lox}$ (A4 cell line) mouse ES cells were maintained undifferentiated in N2B27 medium containing 2i and LIF. The composition of N2B27 + 2i is as follows: Neurobasal media supplemented with 50% v/v DMEM:F12 media, 0.5% v/v N2 supplement. 1% v/v B27 supplement, 2mM L-glutamine, antibiotics (100U/ml penicillin/ 100µg/ml streptomycin) and 100µM β-mercaptoethanol (all from Life Technologies, NY, USA). Chemical inhibitors (Ying et al., 2008) were then added at a final concentration of 1µM (PD032590) and 3µM (CHIR99021) (Stemgent, MA, USA). Cells were cultured on vessels pre-coated with 10ng/ml Laminin (Invitrogen) for 2-4 hours. Approximately 0.5x10^6 cells were plated in a well of a 6-well dish. Fresh media was added every day and cells were split as required every two days.

For *Rad21* deletion experiments, cells were trypsinised and 0.5x10^6 cells were replated in a laminin coated 6-well plate. The trypsinised cells were used as 0 hour sample. The activity of the Cre recombinase was induced by the addition of 100nM 4'-Hydroxytamoxifen (4'-OHT) to culture media. Fresh media supplemented with 4'-OHT was added after 36 hours.

For differentiation, 0.4x10^6 ES cells were plated on laminin coated plates and cultured in N2B27-2i neural differentiation media having the same composition as N2B27 media except with twice as much N2 and B27 supplements in the absence of 2i and LIF.

2.2.1.2 Mouse ES cells in serum media

ERT2Cre-*Rad21*$^{WT/WT}$ (mA7 cell line) and ERT2Cre-*Rad21*$^{lox/lox}$ (mA4 cell line) mouse ES cells were maintained in Knockout Dulbecco's Modified Eagle Medium (KO DMEM) (Life Technologies, NY, USA) supplemented with 10% v/v FCS (Life Technologies, NY, USA), 2mM L-glutamine, antibiotics (100U/ml penicillin/ 100µg/ml streptomycin), 50µM β-mercaptoethanol, 1mM non-essential amino acids (NEAA) (Life Technologies, NY, USA) and 1000U/ml leukaemia inhibitory factor (LIF). Cells were cultured on vessels coated with 0.1% gelatin (Sigma-Alrich, MA, USA).

2.2.1.3 Mouse preB cells

The Abelson transformed mouse preB cell lines expressing ERT2Cre recombinase and *Bcl2* transgene (preB ERT2Cre-Rad21$^{lox/WT}$ and preB ERT2Cre-Rad21$^{lox/lox}$) were grown in IMDM (Iscove's modified Dulbecco's medium), 10% v/v FCS (Biosera), 2mM L-Glutamine, 50µM β-mercaptoethanol and antibiotics (100U/ml penicillin/ 100µg/ml streptomycin) (all reagents from GIBCO, Invitrogen).

2.2.1.4 Preadipocytes

Transformed Mll3$^{-/-}$Mll4$^{fl/fl}$ preadipocyte cell lines stably transfected with either GFP or Cre expressing adenoviral vector (Mll3$^{-/-}$Mll4$^{fl/fl}$ + Ad-GFP and Mll3$^{-/-}$Mll4$^{fl/fl}$ + Ad-Cre preadipocytes) were cultured in DMEM supplemented with 10% FCS, 2mM L-Glutamine, 50µM β-mercaptoethanol and antibiotics (100U/ml penicillin/ 100µg/ml streptomycin) (all reagents from GIBCO, Invitrogen). Cells were cultured in 6-well plates and did not require gelatin coating.

2.2.2 Extraction and culture of MEFs

Pregnant mice were sacrificed at 13 or 14 d.p.c. (day post-coitum) by cervical dislocation. The uterine horns were dissected out, briefly rinsed in 70% (v/v) ethanol and place into a 50 ml Falcon tube containing PBS without Ca2+Mg2+ (Gibco, Invitrogen). The following steps were carried out in a tissue culture hood under aseptic conditions and using sterile instruments. Uterine horns were placed into a Petri dish and each embryo was separated from its placenta and embryonic sac. Head and red organs were removed from each embryo. All embryos were washed in PBS and placed in a clean Petri dish on ice. The tissue was finely minced using a sterile razor blade until they became possible to pipette. The minced tissue was suspended in fresh MEF medium [450 ml of DMEM, 50 ml of FCS (10% (v/v)), 5 ml of 200 mM L-glutamine (1/100 (v/v)), 5 ml of Penicillin-streptomycin (1/100 (v/v)]. 1 ml of 0.05% trypsin/EDTA (Gibco, Invitrogen) was added to the suspension, along with 100K units of DNase I (USB), per embryo. The tissue was transferred into a 50 ml Falcon tube and incubated for 15 min at 37 °C. After each 5 min of incubation, cells were dissociated by pipetting up and down thoroughly. The trypsin was inactivated by adding about 1 volume of freshly prepared MEF medium. The cells were centrifuged at low-speed (1200 rpm or 210g), 5 min, and the supernatant was carefully removed. The cell pellet was resuspended in warm MEF medium and plated on a 10cm dish. Fresh media was added the next day and confluent cells were split and frozen for future use (passage P0). These MEFs were cultured in serum containing MEF media (DMEM+ 10% FCS) in a low oxygen

incubator (5% CO_2, 3% O_2). These primary MEFs can be expanded in culture for upto 5-6 passages after which they start senescing and eventually stop dividing. For serum starvation, the MEFs were cultured in DMEM supplemented with 1% FCS and 5 ml of Penicillin-streptomycin (1/100 (v/v).

2.2.3 RNA extraction and cDNA synthesis

RNA extraction was performed using the QIAshredder and RNeasy Mini kits (Qiagen, Valencia, CA), and residual DNA was eliminated using DNA-free kit (Ambion) according to manufacturer's instructions.

RNA was then reverse transcribed using Superscript^{TM} First-Strand Synthesis system (Invitrogen). 1ug of total RNA was diluted in RNase free water to a final volume of 11uL and supplemented with 1uL of 10mM dNTP mix (Invitrogen) and 1uL of oligo (dT) (Invitrogen). The mixture was incubated at 65°C for 5 min and put on ice for 1min, when 1 uL of 0.1M DTT, 4uL of 5X first strand buffer, 1uL of RnaseOUT (Invitrogen) and 1uL of 200 U/uL Superscript III were added. A reaction mixture without the enzyme was also set up as a control (designated "-RT"). The mixture was incubated at 25°C for 5 min, at 50°C for 1 h and at 75°C for 15 min. cDNAs of interest were then detected by real-time quantitative PCR.

2.2.4 Real-time quantitative PCR analysis (RT-qPCR)

Real-time quantitative PCR (RT-qPCR) of cDNA was carried out on an OpticonTM DNA engine using Opticon Monitor 3 software (MJ Research Inc.) under the following cycling conditions: an initial denaturating step at 95°C for 15 min, 40 cycles of denaturation at 94°C for 15 sec, annealing at 60°C for 30 sec, elongation at 72°C for 30sec at which point the fluorescence was read at 72°C, 75°C, 78°C and 83°C. The melting curve was determined from 70°C to 90°C, at 0.2°C intervals. PCR reactions included 2X Syber-Green PCR Mastermix (Qiagen), 1uM primers and 1uL of template in a 10uL reaction volume. A reaction without DNA was included to control for the formation of primer dimmers and each measurement was performed in duplicate. The RT-qPCR data analysis was performed with the Opticon Monitor 3 software and the relative abundance of sequences was calculated using the ΔΔC(T) method. The primer amplification efficiencies of all primers were tested and confirmed on serial RNA dilutions to be equivalent and close to 2. When the amplification efficiency is close to 2, the relative amount of PCR products between reactions 1 and 2 can be calculated as $2^{-\Delta C(T)1}/2^{-\Delta C(T)2}$, C(T) being the threshold cycle at which fluorescence due to PCR products becomes detectable above background. *Ywhaz* and *Ubc* were generally used for gene expression data normalisation.

RT-qPCR of ChIP samples were carried out as the cDNA RT-qPCR except that 0.5uL of DNA, 2x SYBR-Green PCR Mastermix (BioRad, Bio Rad Laboratories Inc., Waltham, MA) and a 3min enzyme activation period were used instead.

All experiments were performed atleast as 'n' independent biological replicates with each sample being analysed on the PCR in duplicates.

2.2.5 Genomic DNA extraction and genomic PCR

One million cells or mouse tails for genotyping were lysed by resuspension in 100µl of lysis buffer (100mM Tris-Cl pH 8.5, 5mM EDTA, 0.2% SDS, 200mM NaCl, 100µg/ml Proteinase K), followed by overnight incubation at 65°C. DNA was then precipitated by an equal volume of isopropanol. Genomic DNA was amplified using 1U of HotStarTaq™ DNA Polymerase (Qiagen) in a final reaction volume of 10µl, with 1 X HotStarTaq™ buffer (Qiagen), 0.2mM dNTPs (Life Technologies, Invitrogen), 0.2mM primers, and 1µl DNA as prepared above. PCR amplification was performed using the following cycle: 15 minutes at 95°C, followed by 40 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 ethidium bromide staining.

2.2.6 Western Blot

Whole cell extracts were prepared by direct lysis of cells (5min at 95°C and subsequent vortexing) in protein sample buffer (50mM Tris-HCl pH6.8, 1% SDS, 10% glycerol, 0.001% Bromophenol Blue, 5% β-mercaptoethanol) (Laemmli, 1970). $1x10^6$ cells were resuspended in 100µl 1x sample buffer. Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out with the Bio-Rad minigel system. 15-25µl of protein sample and the benchmark pre-stained protein ladder (Invitrogen) were loaded on an acrylamide (BioRad) stacking gel [4% (w/v) acrylamide, 0.125M Tris (pH6.8), 0.1% (w/v) SDS, 0.1% (w/v) ammonium persulphate, and 0.1% (v/v) N,N,N',N'-tetramethylethylenediamine and separated in a 8% or 10% acrylamide resolving gel [8% or 10% (w/v) acrylamide, 0.4M Tris (pH8.8), 0.1% (w/v) SDS, 0.1% (w/v) ammonium persulphate, and 0.1% (v/v) N,N,N',N'-tetramethylethylenediamine using Tris-glycine electrophoresis buffer [1.5% (w/v) Tris, 7.2% glycine, 0.5% (w/v) SDS]. Resolved gels were blotted onto a Protran nitrocellulose transfer membrane (Schleicher & Schuell Bioscience) in transfer buffer (48mM Trizma base(pH 9.2), 39mM glycine, 0.037% (w/v) SDS and 20% (v/v) methanol) using the trans-blot semi-dry electrophoretic transfer apparatus (BioRad). The membranes were incubated for 30min with

blocking buffer [5% (w/v) milk powder (Marvel), 1.2g/L Tris pH7.4, 8.75g/L NaCl] followed by primary antibody incubation diluted in blocking buffer for 2h at room temperature, with agitation. After washing 3 times in wash buffer [1.2g/L Tris pH7.4, 8.75g/L NaCl] for 10min, blots were incubated with horseradish peroxidise-coupled secondary antibodies in blocking buffer for 1h at room temperature. The secondary antibodies used were anti-mouse HRP (1:2000, GE Healthcare), anti-rabbit HRP (1:5,000, GE Healthcare). Detection was done with the ECL-Plus western blotting detection kit (Amersham, Amersham Pharmacia Biotech, Little Chalfont, UK) following manufacturer's instructions and using Kodak X-Omat photographic films.

For quantitative Western blot, resolved gels were blotted onto PVDF membranes instead of nitrocellulose membranes. Before use, the PVDF membranes were preincubated in methanol for 30 secs and subsequently soaked in MiliQ water for 1 min and transferred into transfer buffer for 5 min. Blocking was done using Rocklands blocking buffer MB-070 (proprietary protein formulation in TRIS buffered saline at pH 7.6). Rest of the procedure was similar as described above. Fluorescently labelled secondary antibodies were used for detection. Signals were quantified by the Odyssey Imaging system (LI-COR biosciences).

2.2.7 Immunofluorescence (IF)

Sterile glass coverslips were placed in 6-well plates and cells were cultured onto these coverslips. At appropriate time points samples were fixed with 4% paraformaldehyde (PFA) in PBS for 15 minutes at room temperature. Fixed samples were washed in PBS and, for intracellular staining, samples were permeabilised with 0.5% Triton X-100 for 5 min. Samples were incubated sequentially in blocking solution [3% Normal Goat Serum (Vector), 0.1% Triton X-100 in PBS] for 30 min at room temperature in a humid chamber. Primary antibody was diluted in blocking solution at specific dilutions (see section 2.1.1) and added to the samples for 2h at room temperature or incubated overnight at 4°C in a humid chamber. Coverslips were washed 3 times in PBS and incubated with secondary antibodies coupled with appropriate fluorophores (Molecular Probes) that were diluted in blocking solution for 1h at room temperature in a dark humid chamber. Cells were then washed twice in PBS and mounted in Vecatshield (Vector Laboratories, Peterborough, UK) with DAPI (0.1µg/ml). Samples were visualised using a TCS SP5 Leica laser scanning confocal microscope. Images were processed using Leica Confocal software and ImageJ. Microscope settings and laser power were kept constant between controls and samples.

2.2.8 Inter-species heterokaryon formation - Cell Fusion

Cell fusion experiments were performed as described previously (Pereira et al., 2008). Inter-species heterokaryons can be generated by cell fusion of adherent ES cells and lymphocytes (non-adherent cells). The resulting heterokaryons will attach to gelatin coated dishes. Heterokaryons (cells in which parental nuclei share the same cytoplasm but remain discrete) were cultured under conditions that promote mES self-renewal. Briefly, mES and hB cells were mixed at a ratio 1:1 and washed twice in PBS. The supernatant was completely removed and 1mL of pre-warmed polyethylene glycol (PEG) (Sigma) was added to the pellet of cells over 60 sec. The mixture was incubated at 37°C for 90 sec with constant stirring. Then 4mL of serum-free medium (DMEM; Invitrogen) were carefully added over a period of 3 min, followed by 10mL of DMEM and incubation at 37°C for 3 min. After centrifugation (1350 rpm or 275g, 5 min), the pellet was allowed to swell in complete medium for 3 min. Cell mixtures were then resuspended and cultured under conditions promoting the maintenance of undifferentiated mES cells at $0.5x10^6$ cells/cm^2. Non-fused mES cells were eliminated by the addition of puromycin (1.5 µg/ml puromycin; Sigma) 6-12 hours after fusion onwards.

2.2.9 EdU labelling

One day after fusion, heterokaryons were pulse labelled with 10µM EdU for 45 minutes. Samples were fixed in 4% PFA in PBS for 10min and permeabilised in 0.5% Triton-X100 (Sigma) in PBS. Washing steps were performed in 3% BSA in PBS. EdU was detected with the Click-iT™ EdU Alexa Fluor 647 HCS Assay kit (MolecularProbes, Invitrogen) according to the manufacturer's instructions.

2.2.10 Cell cycle analysis – Propidium Iodide staining

Cells were collected, counted, and up to $1x10^6$ cells were fixed by resuspension in 1ml of 70% EtOH. For propidium iodide (PI) staining, fixed cells were resuspended in PI buffer (PBS-/- supplemented with 50µg/ml propidium iodide (Sigma), 1mg/ml RNase A (Life Technologies, Invitrogen), and 0.05% v/v NP40 (Calbiochem, Merck Millipore) and incubated on ice for 20 minutes. Cells were analysed using a FACSCalibur™ flow cytometer. Data were analysed using FlowJo software.

2.2.11 Fluorescence activated cell sorting (FACS) analysis

2.2.11.1 Cell surface marker analysis

Each antibody was resuspended in wash buffer at a 1:100 dilution. Up to 1 x 10_6 cells were resuspended in 50µl of the wash buffer + antibody, and incubated for 20 minutes at room temperature in the dark. Cells were then washed once in PBS -/- by centrifuging the cells for 5 minutes at 12000 rpm or 13000g, removing the antibody solution, resuspending in wash buffer, and centrifuging again. Cells were then resuspended in wash buffer for further analysis. Cells were analysed using a FACSCalibur™ flow cytometer. Data were analysed using FlowJo software.

2.2.11.2 Sorting GFP expressing cells

To sort GFP positive fibroblasts transfected with MSCV-IRES-GFP (MIG) or MSCV-IRES-Cre-GFP vector, cells were resuspended in FACS buffer (10% FCS in PBS-/-) and subjected to a FACS purification using the FACSAria cell sorter (BD Biosciences).

2.2.12 Bacterial transformation

BAC DNA or plasmid vector DNA 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 H_2O. 100µl of competent DHS5α bacteria (Invitrogen) was added to the mixture and incubated on ice for 20 minutes, and then at room temperature for 10 minutes. Transformed bacteria were plated into Lysogeny Broth (LB) agar supplemented with 50µg/µl of ampicilin and incubated overnight at 37°C. Individual colonies were picked and transferred to 3mL of LB broth with ampicilin and left overnight at 37°C with agitation. DNA was extracted from exponentially growing cultures with Maxipreps kits (Qiagen) according to manufacturer's instructions.

2.2.13 293T transfection and virus production

The pMIG or the Cre-GFP retroviral vector were co-transfected with pCL-Eco (packing vector) into 293T cells using a calcium phosphate protocol. Briefly, DNA-containing precipitates were formed by slowly adding 500µL of 2x HEBS buffer (280nM NaCl, 10mM KCl, 1.5mM Na2HPO4.2H2O, 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 packing vector DNA. 1mL of DNA precipitate was slowly added to a 40-60% confluent 293T cells cultured in 9mL medium in a 10cm dish. Cells were fed with fresh media 24h post-transfection and supernatant containing retrovirus particles were harvested 48, 60 and 72 hours post-transfection and pooled.

2.2.14 Retroviral transduction of fibroblasts

Confluent fibroblast-like cells were trypsinised and about $0.5x10^6$ cells were plated in each well of a 6-well plate. 1ml virus supernatant supplemented with 4μg/mL polybrene and 10mM HEPES pH 7.6 (both from Sigma) was added to each well and cells were centrifuged at 900g for 1h and 30minutes at 37°C. Cells were then incubated for 4 hours at 37°C when the virus supernatant was replaced by fresh MEF media.

2.2.15 Cell fractionation

To prepare total cell extracts, tissue-cultured cells were harvested by centrifugation, washed in PBS, and directly resuspended in protein sample buffer (or Laemmli buffer). To isolate chromatin, cells were resuspended (4×10^7cells/ml) in buffer A (10 mM HEPES, [pH 7.9], 10 mM KCl, 1.5 mM MgCl2, 0.34 M sucrose, 10% glycerol, 1 mM DTT, 5 μg of aprotinin per ml, 5 μg of leupeptin per ml, 0.5 μg of pepstatin A per ml 0.1 mM phenylmethylsulfonyl fluoride). Triton X-100 (0.1%) was added, and the cells were incubated for 5 min on ice. Nuclei were collected in pellet 1 (P1) by low-speed centrifugation (4 min, 1,300 × g, 4°C). The supernatant (S1) was further clarified by high-speed centrifugation (15 min, 20,000 × g, 4°C) to remove cell debris and insoluble aggregates. Nuclei were washed once in buffer A, and then lysed in buffer B (3 mM EDTA, 0.2 mM EGTA, 1 mM DTT, protease inhibitors as described above). Insoluble chromatin was collected by centrifugation (4 min, 1,700 × g, 4°C), washed once in buffer B, and centrifuged again under the same conditions. The final chromatin pellet (P3) was resuspended in Laemmli buffer and sonicated for 15sec in a Diagenode sonicator at high intensity.

2.2.16 Alkaline Phosphatase staining

Cells were stained for Alkaline phosphatase using the 86-R kit from Sigma following manufacturer's specifications with the following modifications. Briefly, cells were fixed in fixing buffer (4.5mM citric acid, 2.25mM sodium citrate, 3mM NaCl, 65% methanol, 4% paraformaldehyde) for 15 mins at RT. Fixed cells were rinsed in water and sufficient amount of Alkaline-dye mixture (1ml Sodium Nitrite solution, 1ml FRV-Alkaline solution, 45ml dH_20, 1ml Naphthol As-B1 alkaline solution) was added to cover the cells. Cells were incubated at RT for 15min, away from direct light. Dye mixture was discarded and cells were rinsed with dH_2O.

2.2.17 Chromatin immunoprecipitation (ChIP)

2.2.17.1 ChIP in ES cells

50-100 million cells were harvested by trypsinisation, resuspended in 40ml Serum ES media and crosslinked with 1% formaldehyde (0.33M) for 10min at 37°C. After quenching of formaldehyde with glycine (both from Sigma) at a final concentration of 125mM for 5min at room temperature, cells were washed twice with icecold PBS. Cells were pelleted at 2000rpm for 5min at 4°C. Pellets were frozen in dry ice and stored at -80°C for later use.

Upto 50 million cells were resuspended in Sonication buffer (20mM Tris-HCl pH8, 150mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton X-100) and lysed on ice for 15 mins. All buffers are supplemented with protease inhibitor cocktail (Roche, West Sussex, UK) The suspension was transferred to polypropylene 15ml Falcon tubes for sonication. Using a Bioruptor (Diagenode, Denville, NJ, USA), the chromatin was sonicated for 20min at 4°C to achieve an average size of 500-1000bp running the settings: output-high, 30sec ON and 30 sec OFF. Chromatin fragment size was assessed on 1% agarose gel by running 4uL of sonicated chromatin (supplemented with 2uL 5% N-lauroylsacrosine, 2uL methyl-orange DNA loading buffer and 2 uL of PBS). Insoluble proteins were discarded after centrifugation of the lysate at 14,000 rpm for 30min, at 4°C and DNA concentration was quantified by spectrophotometry (NanoDropR ND-1000).

100µl of protein G magnetic Dynabeads (Sigma) were rinsed twice in 1mL of ice cold Sonication buffer. Washed beads were incubated with required amounts of antibody for 4hrs at 4°C on a rotator. Beads coupled with antibody were again washed twice with sonication buffer before an incubation with the chromatin sample.

Fragmented chromatin (150µg DNA per IP) was diluted to a final volume of 500µl in Sonication buffer. 1% (v/v) input sample was collected from the diluted chromatin sample. The chromatin samples were then added to Dynabeads coupled with primary antibody and incubated at 4°C overnight. Mouse IgG antibody was used as negative control. The antibody/bead/chromatin ratio used for different ChIPs is as follows: an equivalent of 50µl beads coupled to Rad21 (20µg ab992 antibody), Nipbl (10µg A301-779A antibody) were incubated with 150µg chromatin in 500µl IP volume. For CTCF ChIP, an equivalent of 25µl beads coupled to 5µg CTCF antibody (ab70303) were incubated with 150µg chromatin in 500µl IP volume.

On the next day, beads were washed 1X with the sonication buffer, 1X with 20mM Tris-HCl pH8, 500mM NaCl, 2mM EDTA, 0.1% SDS, 1%Triton X-100, 1X with 10mM Tris-HCl pH8, 250nM

LiCl, 2mM EDTA, 1% NP40 and 1X with TE containing 50 mM NaCl. Finally, the samples were reverse-crosslinked in 450 µl Elution Buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA and 1% SDS, 220 µg proteinase K and 50 µg RNaseA) for 2 hours at 37°C and overnight at 65°C, shaking at 1000rpm for 30sec intervals every minute. DNA was sequentially extracted with phenol/chloroform/isoamylalcohol and precipitated in EtOH, NaAc and 150µg/ml of glycogen carrier (Glycoblue; Ambion, Austin,TX) for 1hr at -80°C. After centrifugation at 14000rpm (13750g) for 30min at 4°C, pellets were washed with 70% EtOH and resuspended in 100µl H_2O. 1µl of ChIP DNA was used per qPCR reaction.

2.2.17.2 ChIP in MEFs

About 5-10 million fibroblasts were fixed in a 15cm dish. Media in the dishes was adjusted to a known volume and formaldehyde was added directly to the media to a final concentration of 1%. The dishes were incubated at 37°C for 10 mins. After this, glycine was added to a final concentration of 125mM to quench the formaldehyde fixation at RT for 5 min. Subsequent procedures were performed on ice. The cells were washed 3x with ice-cold PBS directly in the flasks/dishes they were grown in. The PBS was removed and sufficient amount of Swelling buffer (25mM HEPES pH7.9, 1.5mM $MgCl_2$, 10mM KCl, 0.1% NP40) was added to cover the cells and incubated at 4°C for 10 min. Cells were collected by scraping and dounce homogenised with a 'tight' pestle (50 strokes). Nuclei were collected by centrifugation at 3000xg, 4°C, 5min. The pelleted nuclei were stored at -80°C for later use. The pellet containing 5-10 million nuclei was resuspended in 500µl RIPA buffer (50mM HEPES pH 7.9, 140mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS). Swelling and RIPA buffers were supplemented with phosphatase and protease inhibitors. The chromatin was sonicated as described in 2.2.17.1. The beads were coupled to antibody as described in 2.2.17.1 but using RIPA buffer.

The rest of the steps were performed similarly to 2.2.17.1 with the following modifications. For histone ChIP, 1µg each of H3 and H3K4me3 antibody was coupled to an equivalent of 10µl beads which were then incubated with 10µg chromatin in a total IP volume of 200µl. For Polymerase ChIP, antibodies against total Pol II (10µg sc-H224), Ser5-Pol II (2.5µg CTD4H8) and Ser2-Pol II (10µg 3E10) were coupled to an equivalent of 10µl beads which were then incubated with 100µg chromatin in a total IP volume of 200µl. On the next day, beads were washed 1X with the RIPA buffer, 1X with high-salt RIPA buffer (50mM HEPES pH8, 500mM NaCl, 1mM EDTA, 0.1% SDS, 1%Triton X-100, 0.1% Na-deoxycholate), 1X with LiCl-RIPA (20mM Tris-HCl pH8, 250mM LiCl, 1mM EDTA, 0.5% NP40, 0.1% Na-deoxycholate), and 1X with TE containing 50 mM NaCl.

Rad21 and CTCF ChIP in fibroblasts was performed as described in 2.2.17.1 except 5µg ab154769 was used instead of ab992.

2.2.18 Chromosome Conformation Capture

2.2.18.1 Testing chromatin digestion efficiency

Forward and reverse 20-24mer primers with a Tm range of 60-65°C amplifying a region of 100-300bp were designed spanning the restriction digestion sites of interest. The primers were tested on 2-fold serial dilutions of gDNA (starting conc. 20ng).

The chromatin samples were digested following the 3C protocol until day 1 and then samples were decrosslinked overnight. Digestion efficiency was calculated by PCR amplifying DNA from digested and undigested chromatin samples using these primers. SDS and enzyme concentration were optimised to yield uniformly high digestion efficiencies across all the tested sites at a particular locus.

2.2.18.2 Designing and testing 3C primers

BACs spanning the genomic region of interest were purified and 15µg of BAC DNA was digested overnight 37°C with the restriction enzyme of interest. The digested BAC fragments were isolated by phenol/chloroform extraction. These fragments were resuspended in 10µl water and quantified using Nanodrop. 2.5µg BAC DNA was religated at RT for 2 hrs with 400U T4 DNA ligase. The religated BAC was purified and resuspended in 20µl volume. BAC concentration was measured using Picogreen. 10 fold dilutions of this BAC product were prepared for testing the amplification efficiency of each pair of 3C 'bait' primer and 'anchor' primer.

The Taqman Probe is designed to be about 30 nucleotides long with a Tm ranging from 68-75°C. The FAM fluorophore is attached at the 5'end while the 3' end is covalently attached to the quencher BHQ. The probe hybridises to the opposite strand of the PCR products.

2.2.18.3 3C experiment

Cells were fixed in 10% FCS, 1% formaldehyde for 10 minutes at room temperature and fixation was stopped with glycine (0.125M). 10^7 cells per sample were lysed in 10mM Tris, pH8, 10mM NaCl, 5mM MgCl2, 0.2% NP-40 for 30 min on ice. The nuclei were pelleted and re-suspended in 0.5ml 1.2x digestion buffer (NEB2, New England Biolabs) and permeabilised with SDS (0.3% final concentration) for 1 hr at 37°C, shaking at 800rpm and 3.3% Triton X-100 were added for an additional 1 hour at 37°C. Required amount of enzyme (New England Biolabs) was

added before incubation over night at (37°C, 800rpm) and inactivated with SDS (1.5%, 65°C, 30 min). The reaction was diluted in 6.2ml 1.1x T4 ligase buffer (New England Biolabs) and incubated at 37°C for 1h after addition of 1% Triton X-100. 800U T4 DNA ligase (New England Biolabs) was added for 4 hrs at 16°C, crosslinking was reversed by 300µg proteinase K (65°C, 16h). 300µg RNase A was added for 1h at 37°C. DNA was isolated by phenol/chloroform extraction and ethanol precipitation, quantified using Quant-iT PicoGreen (Invitrogen) and 200ng DNA were used per TaqMan PCR reaction (QuantiFast, Qiagen). Data were normalised to the cross-linking frequency between the anchor and the neighbouring fragment. The distance of the fragment of interest was calculated as the distance between the centre of the anchor fragment and the centre of the digested fragment corresponding to the respective reverse primer used for PCR amplification.

The following digestion setup was used for the different genes analysed: Nanog, Lefty1 locus - 1500U DpnII enzyme (overnight digestion at 37°C). Klf4 locus - 2500U PvuII enzyme (750U for 2hrs + 750U overnight + 1000U for 4hrs). 0.3% SDS was used for permeabilisation in ES cells. For the Ciita 3C, fibroblasts were permeabilised with 0.5% SDS and digested with 1000U NcoI enzyme (250U for 2hr + 250U O/N + 500U for 4hr).

Chapter 3 : Role of cohesin in regulating the expression of pluripotency associated genes

Embryonic stem cells possess the remarkable properties of self-renewal and pluripotency. Thier unique gene expression profile enables them to continually propagate in the pluripotent state and be poised to execute a broad range of developmental programmes. As discussed in Section 1.4.2, cohesin subunits have been identified in RNAi screens performed in ES cells, as factors required for the maintenance of pluripotency and self-renewal (Ding et al., 2009; Hu et al., 2009; Kagey et al., 2010; Nitzsche et al., 2011). However, it is not clear from these studies whether the requirement for cohesin stems from its direct involvement in regulation of pluripotency genes, or is an indirect effect of the loss of its functions in cell division.

RNAi mediated knockdown approach requires the culture of ES cells for relatively long periods of time to obtain a substantial reduction in cohesin proteins. Consequently, Kagey et al., (2010) evaluated gene expression changes in ES cells after 5 days of Smc1 depletion. This approach is problematic because ES cells are rapidly dividing cells, and depletion of cohesin for such an extended period of time will significantly affect cell cycle progression. Therefore, prolonged depletion of cohesin in ES cells will cause activation of stress response pathways and result in indirect effects on gene expression. For example, p53 activation in ES cells, in response to upstream stress signals, supresses *Nanog* expression and triggers ES cell differentiation (Lin et al., 2005; Maimets et al., 2008; Li et al., 2012). Thus, in order to discern the role of cohesin in regulating pluripotency, it is critical to avoid stress induced secondary effects on gene expression. Moreover, as discussed previously in Section 1.4.1, conventional ES cell cultures in Serum + LIF media display heterogeneous expression of some pluripotency genes and contain a sub-population of ES cells more prone to differentiation. Hence, it is likely that by the fifth day of cohesin knockdown, many of the fast dividing ES cells have died and there is an experimental bias to enrich for the slow dividing differentiated cell population. This would further falsely accentuate the impact of cohesin depletion on pluripotency. These concerns have been substantiated by Gene Ontology (GO) analysis of the microarray dataset from Kagey et al. (2010) by our lab, which revealed a significant enrichment for DNA repair and cell stress related pathways (Figure 3.5B, left panel) (Mira-Bontenbal, 2011). This indicates that the knockdown methodology is not the ideal strategy to investigate the direct role of cohesin in regulation of gene expression in ES cells.

In order to overcome these limitations and to truly understand how cohesin regulates gene expression, our lab developed an ERT2Cre-lox conditional knockout system for the Rad21 subunit of the cohesin complex. LoxP sites flank exons 5 and 6 of the *Rad21*^{lox} allele while Cre recombinase, targeted to the ubiquitously expressed *Rosa26* locus, is expressed as a fusion partner to the ligand binding domain of a mutated tamoxifen-specific estrogen receptor 2 (ERt2). The ERt2-Cre fusion protein remains in the cytoplasm, but translocates to the nucleus after addition of 4'hydroxy-tamoxifen (4'OHT). Upon nuclear translocation, Cre excises exons 5 and 6 of *Rad21* gene, which is sufficient to completely inactivate the Rad21 protein. Using this strategy, our lab derived ERT2Cre-*Rad21*$^{lox/lox}$ and ERT2Cre-*Rad21*$^{WT/WT}$ mouse ES cells in defined N2B27 media supplemented with LIF and 2i. The use of 2i allows ES cells to be maintained in a ground state of pluripotency (Ying et al., 2008; Wray et al., 2010) and enables us to study the impact of cohesin depletion in a more uniform ES cell population.

Using these ES cells, which can be conditionally and rapidly depleted of Rad21, I aim to ascertain the role of cohesin in regulating the gene expression programme of pluripotent cells independently of its essential functions in cell division. I further intend to gain a mechanistic insight into cohesin's mode of action as a gene regulator by assessing its role in mediating long-range enhancer-promoter interactions at some key pluripotency associated genes.

3.1 The conditional knockout system allows rapid and acute depletion of RAD21 within 24 hours

In order to identify a suitable time for the investigation of the impact of cohesin depletion on pluripotency in the absence of any secondary effects due to activation of stress, an in depth analysis and characterisation of cohesin depletion was carried out in the conditional ERT2Cre-*Rad21* ES cells by using the following experimental setup (Figure 3.1). ERT2Cre-*Rad21*$^{lox/lox}$ and ERT2Cre-*Rad21*$^{WT/WT}$ mouse ES cells were treated with either 100 nM 4'OHT to induce deletion of the loxed *Rad1* allele or with an equivalent volume of 100% ethanol (carrier control) after plating the cells and samples were collected for analysis after 24, 36 and 48 hrs. In parallel, ERT2Cre-*Rad21*$^{lox/lox}$ ES cells were subjected to differentiation by growing them in N2B27 media without LIF and 2i. In the absence of LIF and 2i, retinoic acid present in the N2B27 media induces rapid differentiation of ES cells (Ying et al., 2003b). These differentiated cells were used to compare the extent of differentiation caused by cohesin depletion during the same time span. ERT2Cre-*Rad21*$^{lox/lox}$ ES cells treated with the DNA damage inducing agent doxorubicin (0.5 µM for 6 hours) were used as a positive control for the effect of stress on changes in gene expression.

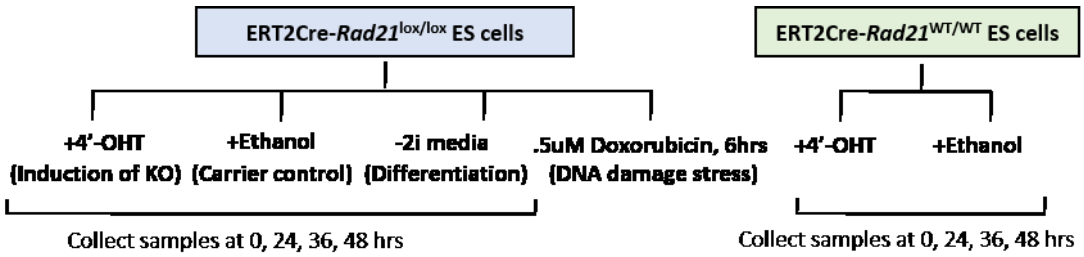


Figure 3.1 Schematic outline of the experimental setup. ERT2Cre-Rad21$^{lox/lox}$ ES cells were treated with 100nM 4'OHT to induce Cre mediated knockout of the Rad21 allele or ethanol as the carrier control. In parallel, cells were plated in N2B27 media lacking 2i to induce differentiation. To control for any discrete effects of tamoxifen, ERT2Cre-Rad21$^{WT/WT}$ ES cells were also treated with tamoxifen and ethanol. As a positive control for the effect of stress on gene expression, cells were treated with .5µM Doxorubicin for 6 hours. Cells were trypsinised and collected for experiments at 0, 24, 36 and 48 hour time points.

The kinetics of cohesin depletion post 4'OHT treatment were monitored by analysing the deletion of the *Rad21* allele and the loss of *Rad21* mRNA transcript and protein levels in the ERT2Cre-*Rad21* ES cells. qPCR on genomic DNA samples showed that the loxed *Rad21* allele is completely deleted by 24 hours (Figure 3.2B). RT-qPCR analysis showed that negligible amounts of *Rad21* mRNA were left after 24 hours of 4'OHT-induced *Rad21* knockout (Figure 3.2C). Quantitative Western analysis showed that 12 hours post 4'OHT treatment, RAD21 protein levels

were already reduced by 80% and a 90% reduction was observed by 24 hours (Figure 3.2D). To ensure that the loss of cohesin was functionally relevant, 4'OHT treated ERT2-Cre *Rad21*$^{lox/lox}$ ES cells were fractionated into the chromatin bound and cytoplasmic fractions and RAD21 protein levels in each fraction were analysed by Western blot. Results showed that cellular depletion of RAD21 corresponded with a depletion in the chromatin associated RAD21 protein (Figure 3.2E). Together, these experiments show that the ERT2Cre conditional knockout system allows rapid depletion of RAD21 protein levels to less than 15% within 24 hrs in ES cells.

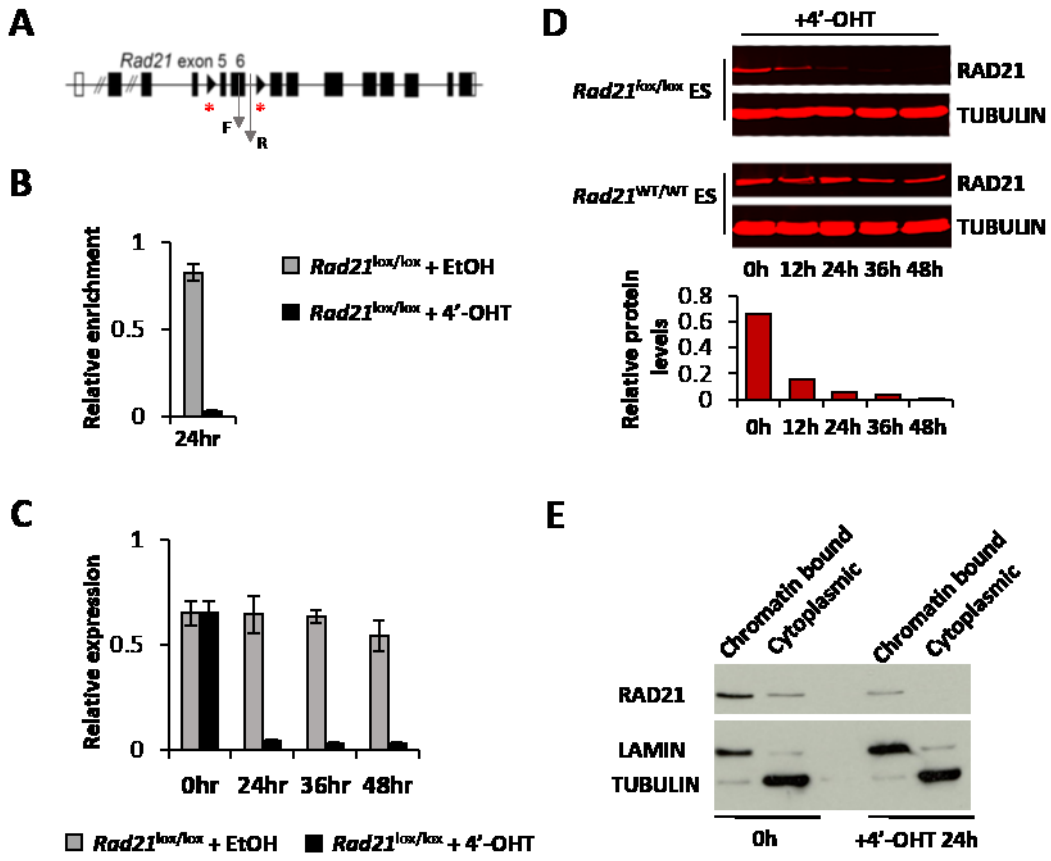


Figure 3.2 **The conditional knockout system allows rapid and acute depletion of cohesin in ES cells within 24 hours A)** Schematic of the conditional *Rad21* allele with LoxP sites (marked in red asterix) flanking exons 5 and 6. Arrows indicate the position of primers F and R used for estimation of genomic deletion of the *Rad21* allele. **B)** Genomic deletion of the loxed allele is complete by 24 hours. Genomic DNA from cells was PCR amplified using primers flanking the LoxP sites and normalised to primers binding outside the *Rad21* allele. **C)** *Rad21* mRNA levels reduced by more than 90% within 24hours in the conditional knockout cells. Expression is normalised to that of *UBC* and *HAZ* genes. **D)** Rad21 protein levels reduced to 15% in 24hours. Protein samples were Western Blotted onto PVDF membranes and probed with appropriate fluorescent secondary antibodies. Fluorescence was measured and normalised to Tubulin loading control. **E)** Cohesin depletion is functionally relevant as indicated by the reduction in chromatin bound Rad21. Cells were fractionated into the cytoplasmic and chromatin bound components and protein samples from each were Western blotted.

3.2 Depletion of cohesin predates cell cycle defects and activation of stress response

As cohesin is essential for proper cell cycle progression and DNA repair, cohesin-depleted ES cells will eventually arrest during cell cycle and activate stress response pathways. Nonetheless, it may still be possible to capture a window of opportunity where the ES cells have been substantially depleted of cohesin before the onset of mitotic arrest and activation of stress response. With this aim, the cell cycle profile of the ES cells was analysed by propidium iodide (PI) staining post tamoxifen treatment. After 24hrs of 4'OHT treatment, the *Rad21*$^{lox/lox}$ ES cells showed the same cell cycle distribution of G1, S and G2 phases as the *Rad21*$^{WT/WT}$ ES cells. However, by 48hrs, majority of the cohesin-depleted ES cells accumulated in G2/M phase (Figure 3.3).

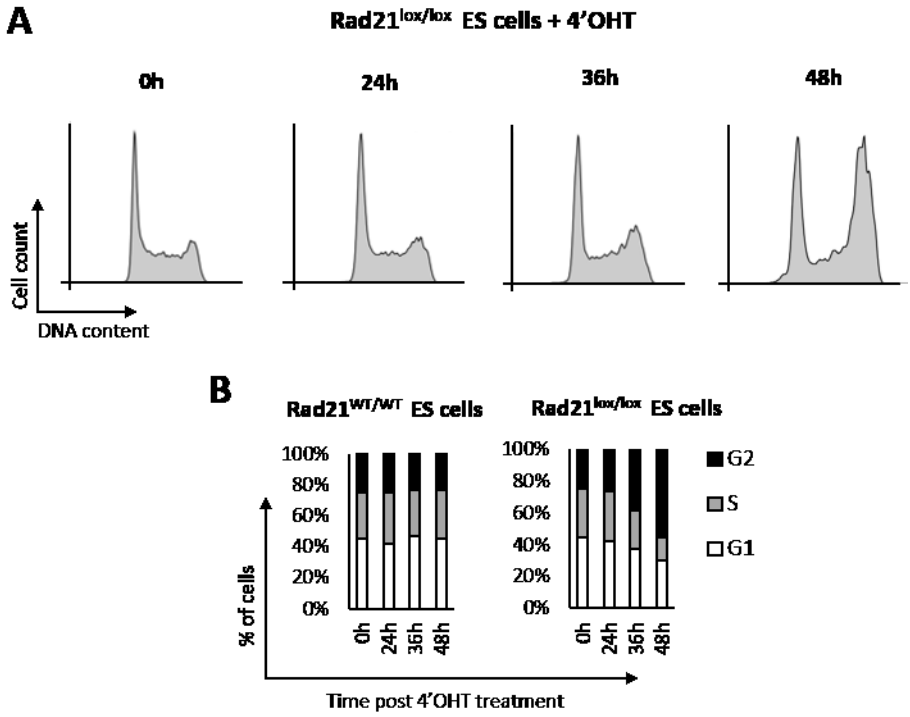


Figure 3.3 Cohesin deplete ES cells show normal cycle profile until 24 hours of 4'OHT treatment and arrest in G2 phase thereafter. A) Cells were stained with propidium iodide and analysed by flow cytometry. After 24 hours of tamoxifen treatment, cohesin deplete ES cells have normal cell cycle profile while majority of them are arrested in G2 phase by 48 hours. B) Proportion of cells in G1, S and G2 phases of cell cycle in wildtype and cohesin-depleted ES cells.

Additionally, to evaluate cellular stress levels upon cohesin depletion, I checked the expression of *Trp53* (p53) and its downstream target genes – *Cdkn1a* (p21) and *Mdm2* as markers which are upregulated upon activation of stress response pathways. Cells were treated with

0.5µM Doxorubicin, for 6hrs to provide a positive control for the impact of stress on changes in gene expression. The results showed that *Mdm2* mRNA levels at 24hrs were identical to untreated cells but increased markedly by 36 and 48hrs of cohesin depletion. *Trp53* mRNA levels did not change while *Cdkn1a* was downregulated in response to cohesin loss (Figure 3.4). These experiments demonstrate that prolonged periods of cohesin depletion in ES cells activates the cellular stress response pathways and renders them unsuitable for gene expression analysis. However, the 24hr time point provides the perfect opportunity to study gene expression changes where the cohesin levels are already reduced by 90% but the ES cells do not yet show signs of cell cycle defects or activation of stress response pathways.

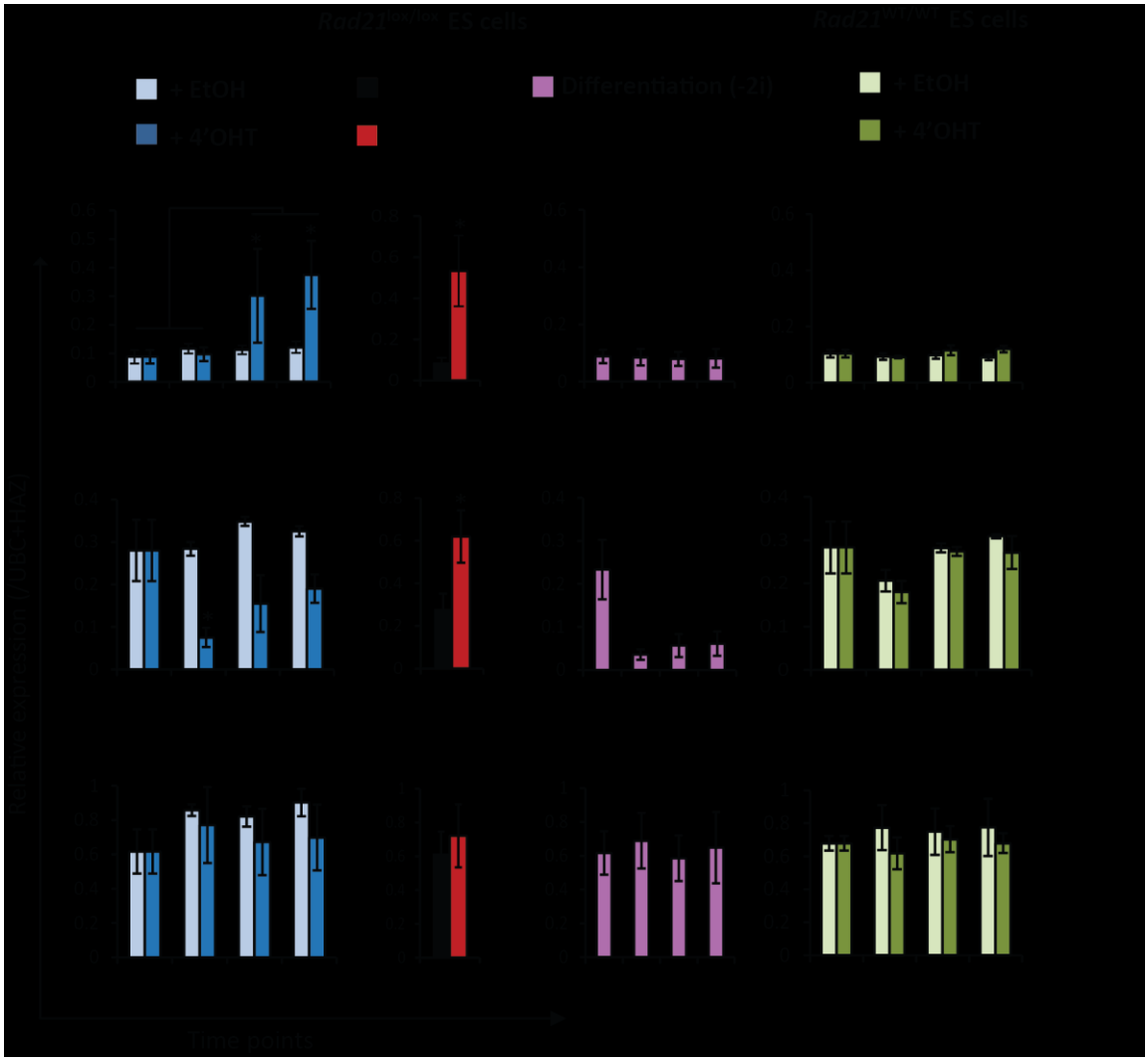


Figure 3.4 Cohesin depletion predates activation of stress response. qRT-PCR analysis of mRNA expression of stress response genes *Mdm2*, *Cdkn1a* (p21), *Trp53* (p53). *Mdm2* is significantly upregulated only after 36 hours of cohesin depletion. Cdkn1a is downregulated upon cohesin depletion (Normalised to *UBC* and *Ywhaz*; mean ± SE; n=3; * p-value ≤ 0.05 from T-test).

3.3 Cohesin depletion impacts the gene expression programme of ES cells in a locus specific manner, but does not lead to a general collapse of pluripotency

To measure the impact of cohesin depletion on the gene expression programme of ES cells at 24 hours post 4'OHT treatment on a genome-wide scale, a colleague in the lab performed microarray experiments (Mira-Bontenbal, 2011). Transcriptional profiling showed that 598 genes were differentially expressed (p-value < 0.05 with Benjamini-Hochberg correction) of which 257 were upregulated and 341 were downregulated. It is noteworthy that much lesser number of genes were deregulated in our dataset as compared to that published in Kagey et al., 2010 where more than 10,000 genes were deregulated (Figure 3.5A, B).

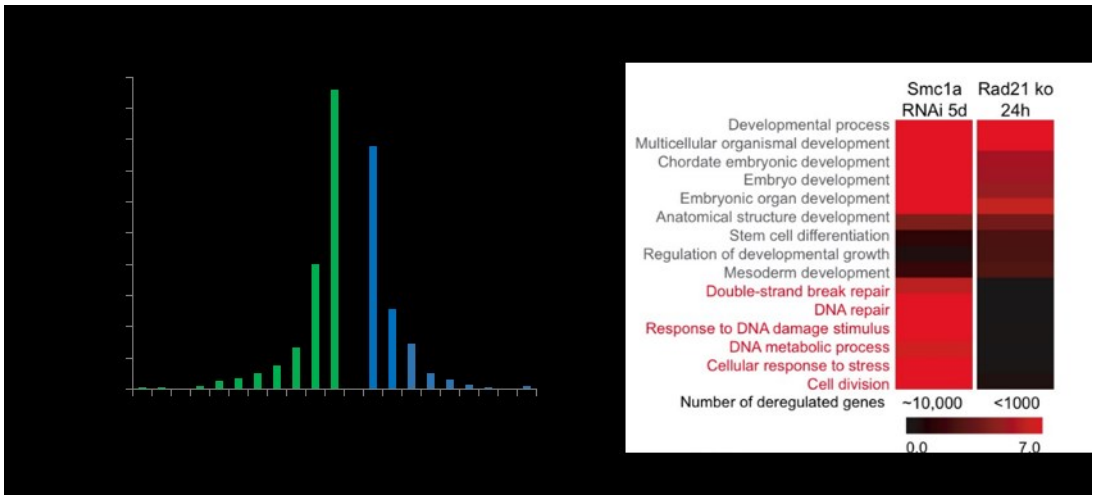


Figure 3.5 Microarray data analysis of RAD21-depleted ES cells at 24 hours shows deregulation of specific genes. A) Log_2 fold change distribution of the 598 significantly differentially expressed genes upon Rad21 depletion in ES cells for 24h. Downregulated genes are in green and upregulated genes are in blue. Only a limited number of genes show high magnitudes of fold changes. **B)** Heat map analysis for the Smc1a 5day knockdown expression data from Kagey et al., 2010 (left panel) and Rad21 24h KO data (right panel) showing $-Log_{10}$ of the p-values for each GO Term. Genes related to DNA damage and chromosome missegregation are preferentially induced by prolonged knockdown of cohesin subunits, but not by acute cohesin deletion (black: no enrichment, red: enrichment).

Moreover, the magnitude of the impact in terms of fold changes in expression was limited as majority of the differentially expressed genes were moderately effected (maximal log_2 fold change from -0.75 to +0.75) and only a few genes showed higher changes in expression (maximal log_2 fold change less than -1 and greater than +1) (Figure 3.5A). We next performed Gene Ontology (GO) analysis on our microarray dataset which revealed that deregulated genes were enriched for developmental functions (GO terms in grey type-face in Figure 3.5B). In contrast to

prolonged knockdown of *Smc1a* (Figure 3.5B, left), we found no enrichment of cell cycle and DNA damage-related genes (GO terms in red type-face in Figure 3.5B). Integration with published ChIP-seq data (Kagey et al., 2010), showed that the binding of cohesin, the cohesin loading factor NIPBL, Mediator12 subunit and RNA Polymerase II was highly correlated with genes showing deregulated expression upon acute cohesin depletion (4.44x10-50, Wilcoxon gene set test) (Mira-Bontenbal, 2011). Overall, this analysis indicates that rapid and acute depletion of cohesin in ES cells within 24 hours is a suitable way to measure direct effects of cohesin depletion on gene expression. This strategy avoids secondary effects due to the activation of stress response. Also noteworthy is the limited yet specific range of changes observed in gene expression upon cohesin depletion.

I next sought to determine the impact of cohesin depletion on some key pluripotency associated genes. Data presented here shows the mRNA expression levels for *Nanog*, *Pou5f1* (Oct4), *Klf4* and *Rex1* (Figure 3.6A). At 24 hours, *Nanog* expression was only marginally reduced while *Klf4* expression was substantially decreased as compared to control cells. On the contrary, *Pou5f1* and *Rex1* did not show any significant decrease. Even though *Nanog* expression was reduced in *Rad21* depleted ES cells at 24 hours, the decrease was much less than that observed in differentiating cells in the same time span. The changes in mRNA expression observed for *Nanog* and Oct4 were also reiterated at the protein level. NANOG protein levels showed a progressive reduction upon cohesin depletion while OCT4 protein levels remained unchanged (Figure 3.6B). Importantly, doxorubicin induced stress led to significantly reduced expression of all the pluripotency genes tested within 6 hours. Also noteworthy is the observation that by 48 hours of 4'OHT treatment, when the cells start to get stressed due to lack of cohesin, all the genes tested showed a more significant decrease in expression as compared to that at 24 hours. Altogether, this data indicates that expression of pluripotency genes is strongly affected by the activation of stress response pathway in ES cells. This also explains the differences in gene expression observed 24 and 48 hours after cohesin depletion. By 48 hours, cells are under stress and consequently display a more pronounced downregulation of most pluripotency genes tested, as also observed in prolonged Smc1a knockdown studies (Kagey et al., 2010). On the other hand, gene expression changes at 24 hours, assessed to be a direct impact of cohesin depletion, are restricted to only a few specific pluripotency genes.

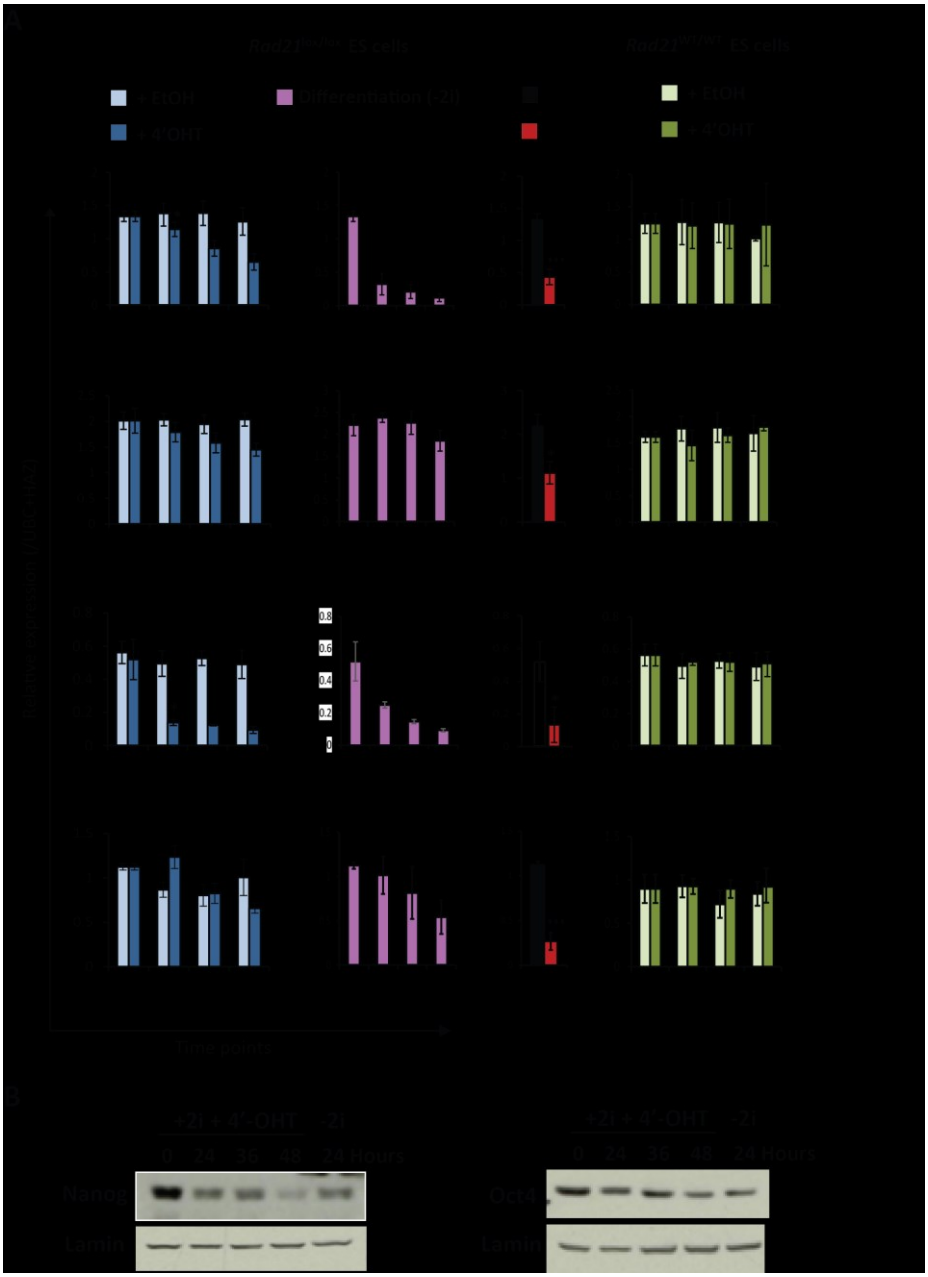


Figure 3.6 Specific pluripotency associated genes are downregulated upon cohesin depletion. A) qRT-PCR analysis of pluripotency associated genes *Nanog, Pou5f1* (Oct4)*, Klf4* and *Rex1*. Only *Nanog* and *Klf4* expression is reduced significantly upon cohesin depletion at 24 hours. It is noteworthy that all these genes are downregulated in response to Doxorubicin treatment (normalised to *Ubc* and *Ywhaz*; mean ± SE; n=3; p-value *≤ 0.05, **≤ 0.005, ***≤ 0.0005) **B)** Western blot showing Nanog and Oct4 protein levels in ERT2Cre-*Rad21*$^{lox/lox}$ ES cells treated with tamoxifen. Levels are compared to differentiating cells in -2i media at 24hours.

To further extend the scope of this analysis I assessed the impact of acute cohesin depletion on the expression of genes assigned to an extended pluripotency network (Kim et al., 2008b). A curated list of genes whose promoters are targeted by at least 5 of the pluripotency factors Nanog, Dax1, Sox2, Nac1, Oct4, Klf4, Myc, Rex1 and Zfp281 was compiled. Of these 354 genes, only 40 (11.3 expected, enrichment significance 8.4 x 10-13) were found to be significantly deregulated upon cohesin depletion at 24 hours. Among these 40 genes, 16 were upregulated and 24 downregulated (Figure 3.7) with a log_2fold change of 0.45 to 1.80 and -0.42 to -1.55, respectively.

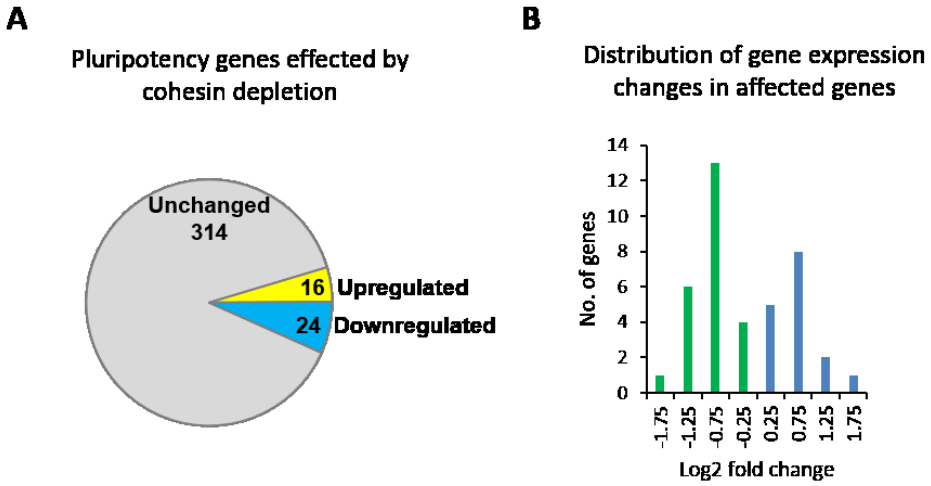


Figure 3.7 Cohesin depletion has limited impact on the expression of pluripotency associated genes. **A)** Within a set of 354 pluripotency associated genes, only 16 were upregulated and 24 downregulated in the microarray dataset 24 hours after cohesin depletion **B)** The distribution of the log_2 fold changes in gene expression of the 40 pluripotency genes affected by cohesin depletion.

Given the limited number of pluripotency genes being affected, combined with their direction (up or down) and extent (magnitude of fold change) of deregulation, this analysis argues for locus-specific effects rather than a global collapse of pluripotency gene expression in acutely cohesin-depleted ES cells.

Next, I assessed if differentiation associated genes were preferentially deregulated upon cohesin depletion. First, I checked the expression of *Phc1* and *Lefty1* genes which are expressed early during embryonic development and play a role in pattern formation and left-right axis determination respectively. *Lefty1* was found to be significantly upregulated while *Phc1* showed no significant difference at 24hours post tamoxifen treatment. The impact on both of these genes was opposite to that observed in differentiating ES cells. It is also to be noted that Kagey et al.

(2010), described *Lefty1* and *Phc1* to be significantly downregulated upon Smc1a knockdown, which can be explained by the decreased expression seen upon doxorubicin induced stress. Another gene analysed was *Sox1*, a neural differentiation marker, which also did not show any significant upregulation upon cohesin depletion (Figure 3.8).

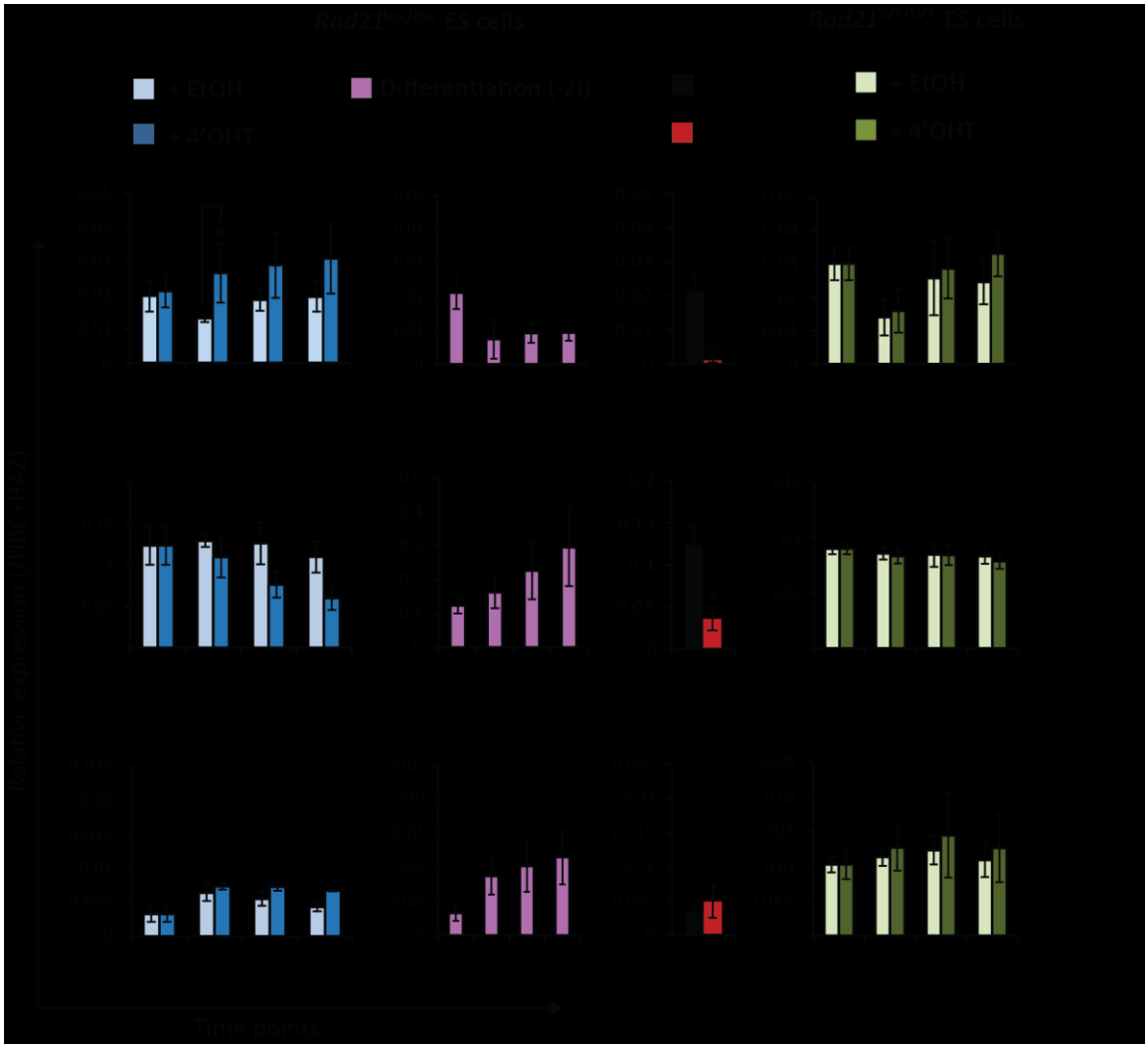
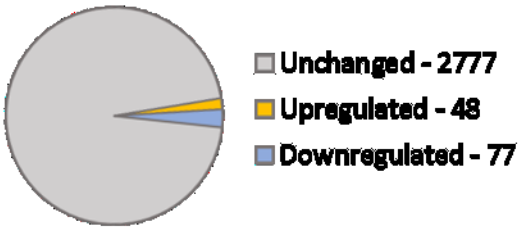


Figure 3.8 Cohesin depletion does not cause an overall upregulation of differentiation associated genes. qRT-PCR analysis of embryonically expressed genes important in various stages during differentiation. *Lefty1* is significantly upregulated 24 hours after cohesin depletion, while *Phc1* is downregulated and *Sox1* remains unchanged (normalised to *UBC* and *Ywhaz*; mean ± SE; n=3; p-value * ≤ 0.05, ** ≤ 0.005).

Key developmental genes are marked by bivalent chromatin marks in ES cells (Azuara et al., 2006; Bernstein et al., 2006). These genes are maintained in a transcriptionally repressed state in ES cells but can be rapidly activated at different stages of differentiation. Therefore, I used bivalently marked genes to test if cohesin-depleted ES cells showed a globally increased propensity for upregulation of differentiation associated genes. The chromatin state of all the genes in our microarray dataset was assessed based on published datasets and the proportion of deregulated bivalent genes was calculated (Figure 3.9). Only 125 of a total of 2902 bivalent genes were found to be differentially expressed upon cohesin depletion at 24 hours. Of these, 48 were upregulated and 77 downregulated. Together, bivalent genes made a very small percentage of total deregulated genes and show no indication of a preferential upregulation upon cohesin depletion.

A

Change in Rad21 KO	All genes	Bivalent genes	Overlap significance (p-value)	Percentage
Up	257	48	0.01557889	18.4
Down	341	77	1.84E-05	22.8
No Change	20336	2777	0.9999948	13.6
Up+Down	598	125	3.14E-06	20.9



B

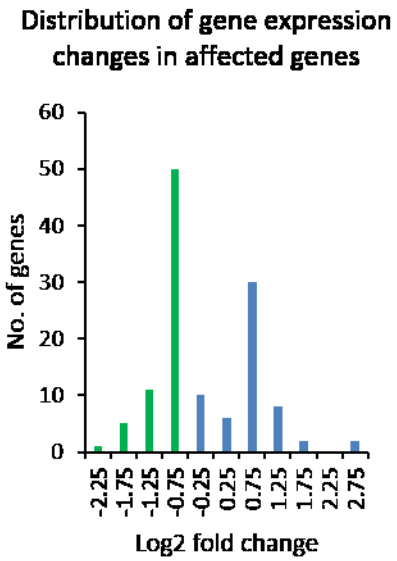


Figure 3.9 Cohesin depletion has no preferential impact on upregulation of bivalent genes. A) Bivalent genes were identified in the microarray dataset and their proportion in up- and downregulated genes was calculated. Even though the overlap between bivalent genes and all deregulated genes was significant, upregulation of bivalent genes was not favoured. **B)** The distribution of the $\log_2$ fold changes in gene expression of the 40 pluripotency genes affected by cohesin depletion.

Wild-type, cohesin-depleted and differentiating ES cells were further analysed for the cell surface expression of Alkaline Phosphatase (AP), a pluripotency marker. AP staining was markedly reduced in differentiating ES cells while there was no notable difference in wild-type and cohesin depleted ES cells at 24 hrs (Figure 3.10). Normal AP expression in cohesin-depleted ES cells further contradicts a total loss in pluripotency.

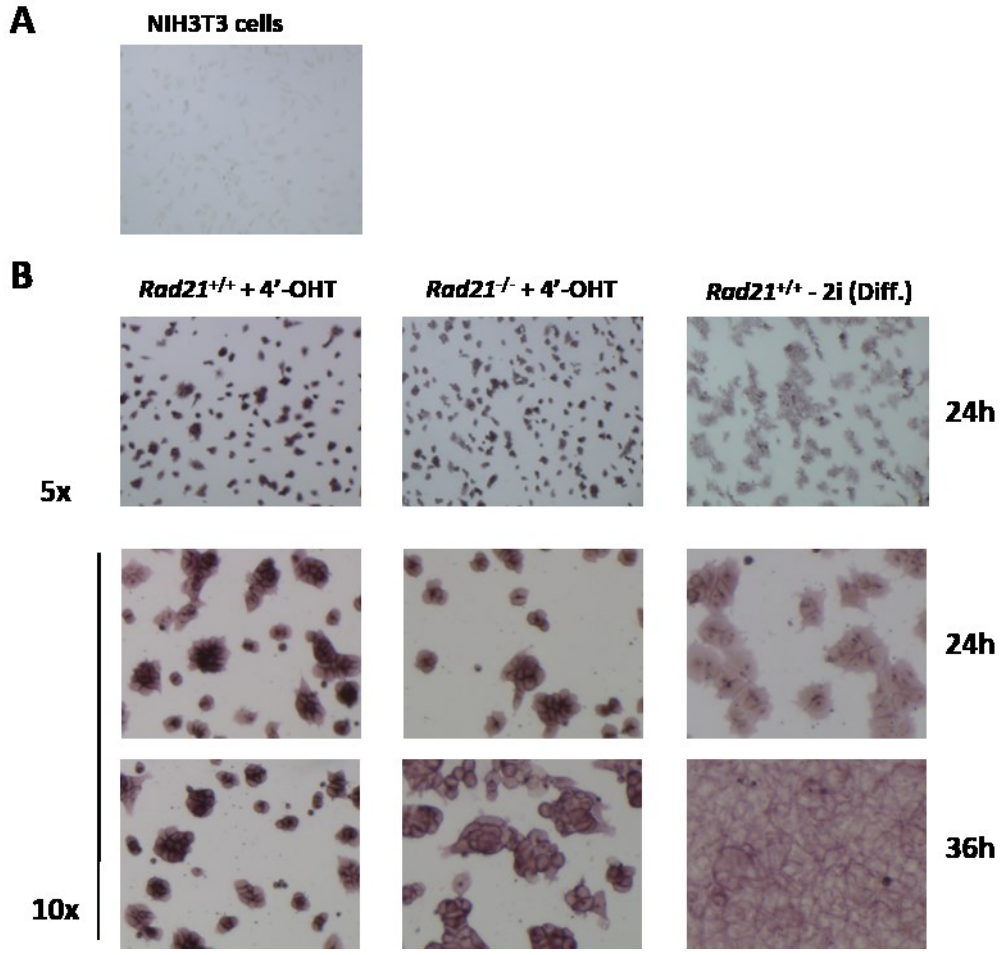


Figure 3.10 Cohesin depleted ES cells are positive for Alkaline phosphatase staining. The relative levels of Alkaline Phosphatase staining were compared in wild-type, cohesin-depleted and differentiating ES cells. NIH3T3 cells were used as negative control (**A**). Differentiated ES cells showed a significantly reduced expression of AP, while the levels of expression in cohesin-depleted cells were comparable to that of wild-type cells at 24 hours **(B)**. Presence of Alkaline phosphatase stain in Rad21 depleted ES cells argues against loss of pluripotency.

Overall, this analysis reveals that acutely cohesin-depleted ES cells do not show a global collapse in pluripotency or an increased propensity for differentiation. In fact, the impact of cohesin depletion seems to be limited to certain genes, and the effect is highly locus-specific.

It has previously been reported that cohesin depletion in non-cycling double positive thymocytes (DPTs) compresses the dynamic range of gene expression. Genes with low expression levels are more often upregulated, whereas highly expressed genes are more often downregulated (Seitan et al., 2013). In order to check how cohesin depletion affected the overall range of gene expression in ES cells, the list of deregulated genes was subjected to a similar analysis and genes were stratified into bins according to the mean of their expression in control and cohesin depleted cells. The stratified mean expression was plotted against log2 fold changes in gene expression (Figure 3.11). Unlike cohesin depleted DPTs, results did not indicate any preferential upregulation of lowly expressed genes or downregulation of highly expressed genes. In fact, the proportion of upregulated and downregulated genes was mostly similar across the whole range of expression (Figure 3.11B). Hence, cohesin depletion in ES cells did not lead to more homogenous levels of gene expression. Once more, this observation argues against a universal reduction in expression of pluripotency genes and upregulation of lowly/un-expressed differentiation genes. Instead it points towards specific gene expression changes across the whole range of genes.

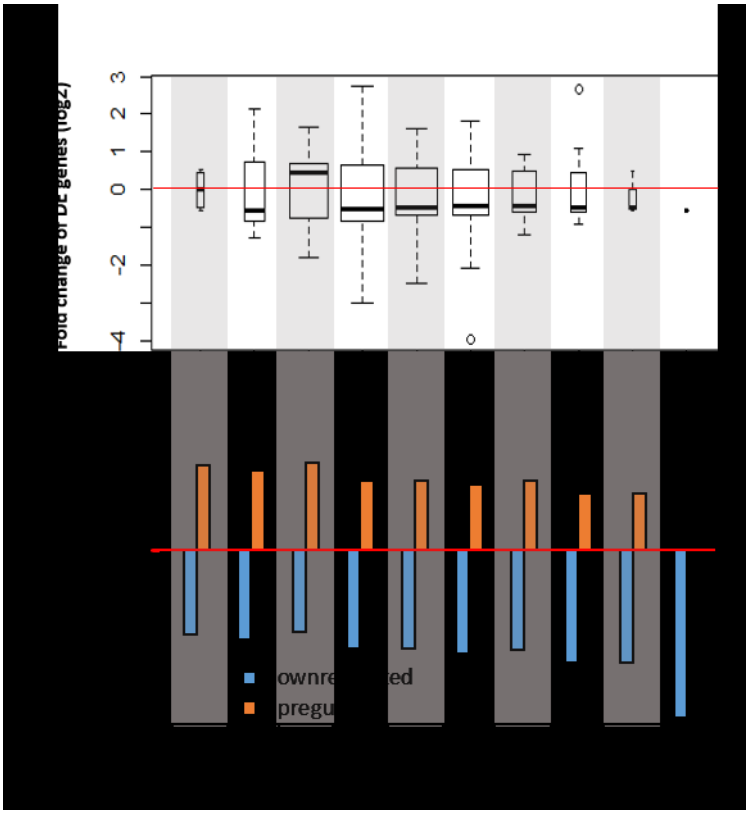


Figure 3.11 Cohesin depletion in ES cells does not compress the dynamic range of gene expression. **A)** Genes were stratified into 12 equally sized log intervals from low (0-1) to high (11-12) based on the average gene expression of control and cohesin deficient ES cells. Boxplot indicates the distribution of gene expression fold changes. Width of the box plot is representative of number of genes in each bin. Lowly expressed genes are not preferentially upregulated and vice versa. **B)** Bar plot indicating the proportion of up- and down-regulated genes in each gene expression interval.

Taken together, the data presented here suggests that cohesin depletion does not lead to a more random gene expression programme resulting in a global collapse of pluripotency, but rather has locus-specific effects.

3.4 Long-range regulatory DNA interactions can be maintained after loss of cohesin and are not sufficient to explain the observed changes in gene expression

To gain insight into the possible mechanisms by which cohesin regulates gene expression, I tested the impact of cohesin loss on long-range enhancer-promoter DNA interactions which have previously been shown to play an important role in cohesin-mediated gene regulation. For this purpose, I selected a subset of genes deregulated upon cohesin depletion based on the positioning of their regulatory elements and relevance to the pluripotent state. At these loci, chromosome conformation capture (3C) analysis was combined with information regarding the binding of Rad21, CTCF and Nipbl using ChIP analysis, to get a more comprehensive view of gene expression changes and the role played by cohesin.

First, I studied the *Nanog* gene which is downregulated upon cohesin depletion. *Nanog* expression is regulated by an enhancer located 5 kb upstream of the transcription start site (TSS) (Wu et al., 2006). This enhancer region is co-occupied by Rad21, Smc1, Nipbl, Mediator, p300 and is marked by histone H3K27 acetylation as assessed from published ChIP-sequencing datasets (Figure 3.12A). A 3C experiment was designed to map DNA interactions between the TSS (anchor) and the upstream enhancer in control, 24h cohesin-depleted and 24h differentiating ES cell populations. In accordance with previous observations, enhancer-promoter interactions were weakened upon cohesin depletion and differentiation (Figure 3.12A bottom panel). Additionally, the extent of loss of interaction was comparable to the reduction in expression (Figure 3.12B). However, it is to be noted that even though cohesin binding was reduced by 80% at the *Nanog* enhancer and promoter in cohesin depleted ES cells, the reduction in interaction was not proportional. On the other hand, decreased enhancer-promoter interactions upon differentiation were accompanied by loss of cohesin binding (Figure 3.12C). No significant binding of CTCF was observed at either *Nanog* enhancer or promoter in the three states of ES cells described here (not shown). The binding of cohesin loading factor Nipbl remained unchanged after cohesin depletion, but was significantly reduced in response to differentiation (Figure 3.12D). Thus, at the *Nanog* locus, the reduction in gene expression upon cohesin depletion can partly be explained by a decrease in enhancer-promoter interactions. It is also highlighted here that Rad21 and Nipbl binding is reduced at the Nanog locus in response to differentiation signals which may consequently result in decreased enhancer-promoter interactions.

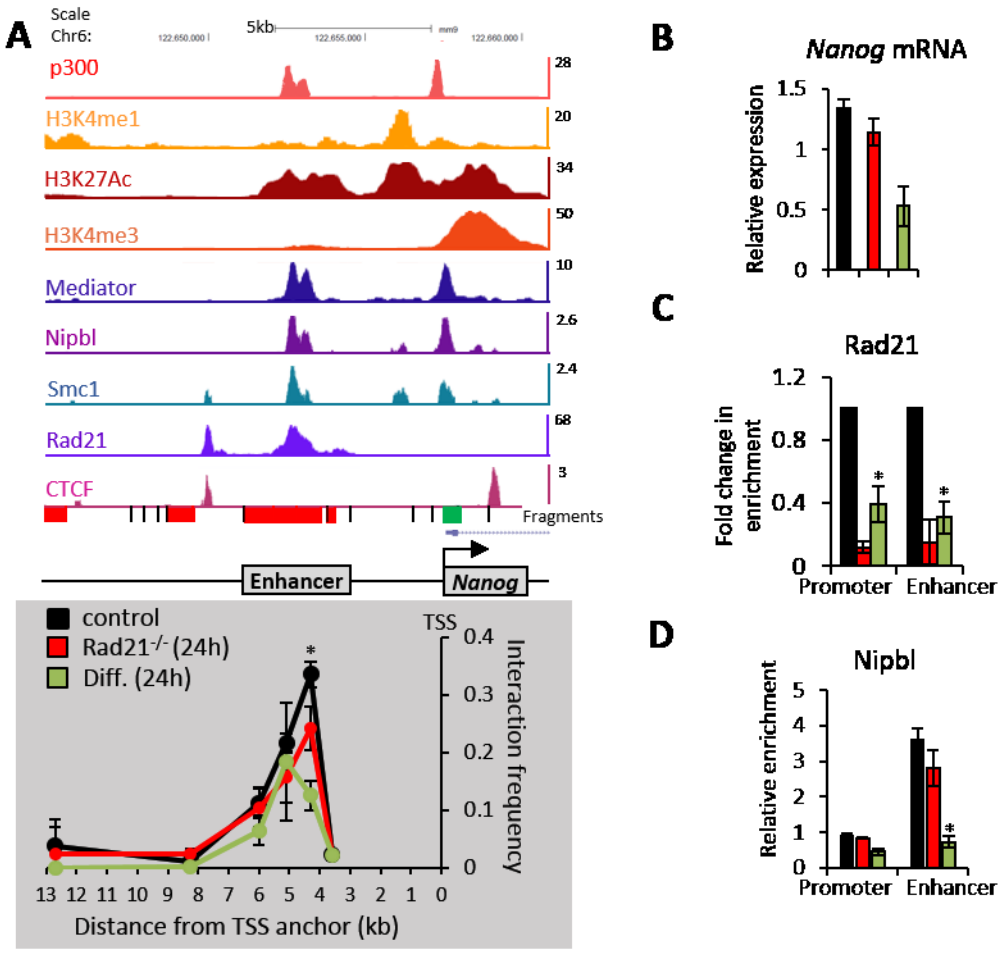


Figure 3.12 Cohesin depletion leads to weakened enhancer-promoter interactions at the *Nanog* locus. **A)** Top panel: Binding of mediator, cohesin subunits and histone modifications at the *Nanog* locus. Restriction digested fragments are marked – Anchor in green and analysed fragments in red. Bottom panel: 3C at the *Nanog* locus. Chromatin was fixed and digested with DpnII enzyme. The religated products were PCR amplified and detected using a Taq Probe binding in the TSS anchor fragment. Values for interaction frequency were normalised to BAC templates and control regions. DNA interaction between the upstream enhancer at 5 kb and TSS was reduced upon cohesin depletion. Differentiated ES cells showed more pronounced reduction in this interaction. **B)** *Nanog* mRNA expression (normalised to *UBC* and *Ywhaz*). **C)** Rad21 ChIP at the *Nanog* enhancer and promoter. Data is plotted as fold change in Rad21 enrichment relative to that in control cells. Rad21 binding is reduced by 80% in cohesin-depleted ES cells. Rad21 binding also decreases in response to differentiation. **D)** Nipbl ChIP at the *Nanog* enhancer and promoter. Nipbl enrichment is plotted relative to 1% input chromatin. Nipbl binding is significantly reduced upon differentiation but remains unaffected upon cohesin depletion. (Mean ± SE; n = 3; Color code: Black - control ES cells, Red - *Rad21*$^{lox/lox}$ + 4'OHT 24h ES cells, Green – *Rad21*$^{lox/lox}$ – 2i 24h ES cells)

Next, I chose to investigate the impact of cohesin loss on enhancer-promoter interactions at genes that are upregulated in cohesin-depleted ES cells. *Lefty1* was selected for this analysis. *Lefty1* expression is tightly regulated during development by a combination of *cis*-regulatory elements comprising of the lateral plate mesoderm specific enhancer (LPE), the neural plate enhancer (NPE), located 9.6 kb and 3 kb upstream of TSS respectively, and a promoter-proximal right side-specific silencer (RSS) (Saijoh et al., 1999). In ES cells, the *Lefty1* enhancer was recognised by the presence of H3K4me1 and H3K27ac marks along with the binding of cohesin, Nipbl, mediator and p300 approximately 10 kb upstream of *Lefty1* TSS (Figure 3.13A). Using *Lefty1* TSS as the anchor, interactions were mapped between the promoter and enhancer using 3C. Interestingly, these interactions remained unchanged between wild-type and cohesin-depleted ES cells (Figure 3.13A), in spite of a substantial loss in Rad21 binding as confirmed by ChIP (Figure 3.13C). By contrast, in differentiating ES cells, where *Lefty1* expression was found to be reduced (Figure 3.13B), the enhancer-promoter interactions were evidently diminished. This demonstrated the validity of the 3C assay while also suggesting that the interaction of the promoter with this enhancer is functionally relevant for *Lefty1* expression. From these experiments it can be concluded that *Lefty1* enhancer-promoter interactions are maintained at wildtype levels despite efficient depletion of cohesin.

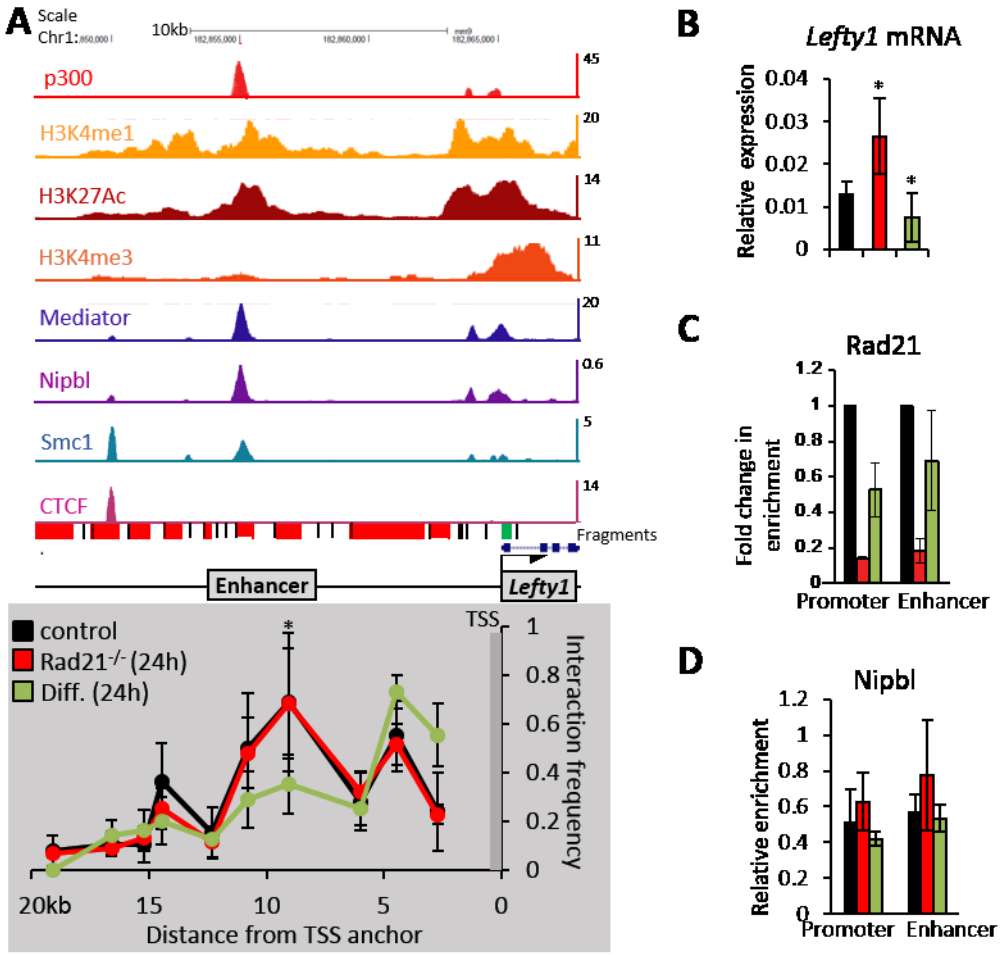


Figure 3.13 *Lefty1* expression and enhancer-promoter interactions can be maintained in the absence of cohesin. A) Top panel: Binding of mediator, cohesin subunits and histone modifications at the *Lefty1* locus. Restriction digested fragments are marked – Anchor in green and analysed fragments in red. Bottom panel: 3C at the *Lefty1* locus. Chromatin was fixed and digested with DpnII enzyme. Religated products were PCR amplified and detected using a Taq Probe binding in the TSS anchor fragment. Values for interaction frequency were normalised to BAC templates and control regions. The panel above shows the DNA interaction between the upstream enhancer and TSS remained unchanged upon cohesin depletion and was decreased upon differentiation. **B)** *Lefty1* mRNA expression (normalised to *UBC* and *Ywhaz)*. **C)** Rad21 ChIP at the *Lefty1* enhancer and promoter. Results are plotted as fold change in Rad21 enrichment relative to that in control cells. Rad21 binding is reduced by 80% in KO ES cells. Rad21 binding decreases partially in response to differentiation. **D)** Nipbl ChIP at the *Lefty1* enhancer and promoter. Nipbl enrichment is plotted relative to 1% input chromatin. (Mean ± SE; n = 3; Color code: Black - control ES cells, Red - *Rad21*$^{lox/lox}$ + 4'OHT 24h ES cells, Green – *Rad21*$^{lox/lox}$ – 2i 24h ES cells)

I next analysed *Klf4*, a gene which is substantially downregulated upon cohesin depletion. *Klf4* enhancers have not been previously functionally tested but putative *Klf4* enhancers were identified based on ChIP-sequencing data for cohesin, Nipbl, mediator, p300, H3K4me1 and H3K27ac marks at around 50 kb downstream of TSS (Figure 3.14A top panel). This region has also been recently termed as a super-enhancer (Whyte et al., 2013). 3C experiments at *Klf4* locus showed that indeed *Klf4* promoter (anchor) interacted with this putative enhancer region in wildtype ES cells and that the interaction was abolished upon differentiation when *Klf4* is downregulated. However, despite efficient cohesin depletion and substantial decrease in expression in cohesin-depleted ES cells, the enhancer-promoter interactions remained intact (Figure 3.14A). Thus, *Klf4* provides another example that dissociates cohesin binding and gene expression changes from enhancer-promoter interactions. A more detailed analysis highlighted the dynamic binding of Rad21, Nipbl and CTCF at the +53kb and +55kb enhancer regions upon differentiation. CTCF and Nipbl binding at these two sites was unaffected by cohesin depletion. However differentiating ES cells showed significant reduction in Rad21, CTCF and Nipbl binding, specifically in the +53, +55kb region (Figure 3.14 C, D, and E). These observations suggest that in response to physiological cues, decreased transcription was associated with a loss of cohesin binding and a concomitant reduction in enhancer-promoter interactions. However, the reverse is not true, i.e. loss of cohesin binding is not sufficient to abrogate the enhancer-promoter interactions even though it results in downregulation of *Klf4*.

Overall, these experiments indicate that cohesin binding is dynamically and specifically regulated in response to physiological cues at enhancers and promoters and is very well correlated with the long-range interactions. However, it is also highlighted here that enhancer-promoter interactions can be maintained upon removal of cohesin as exemplified by the *Lefty1* and *Klf4* loci.

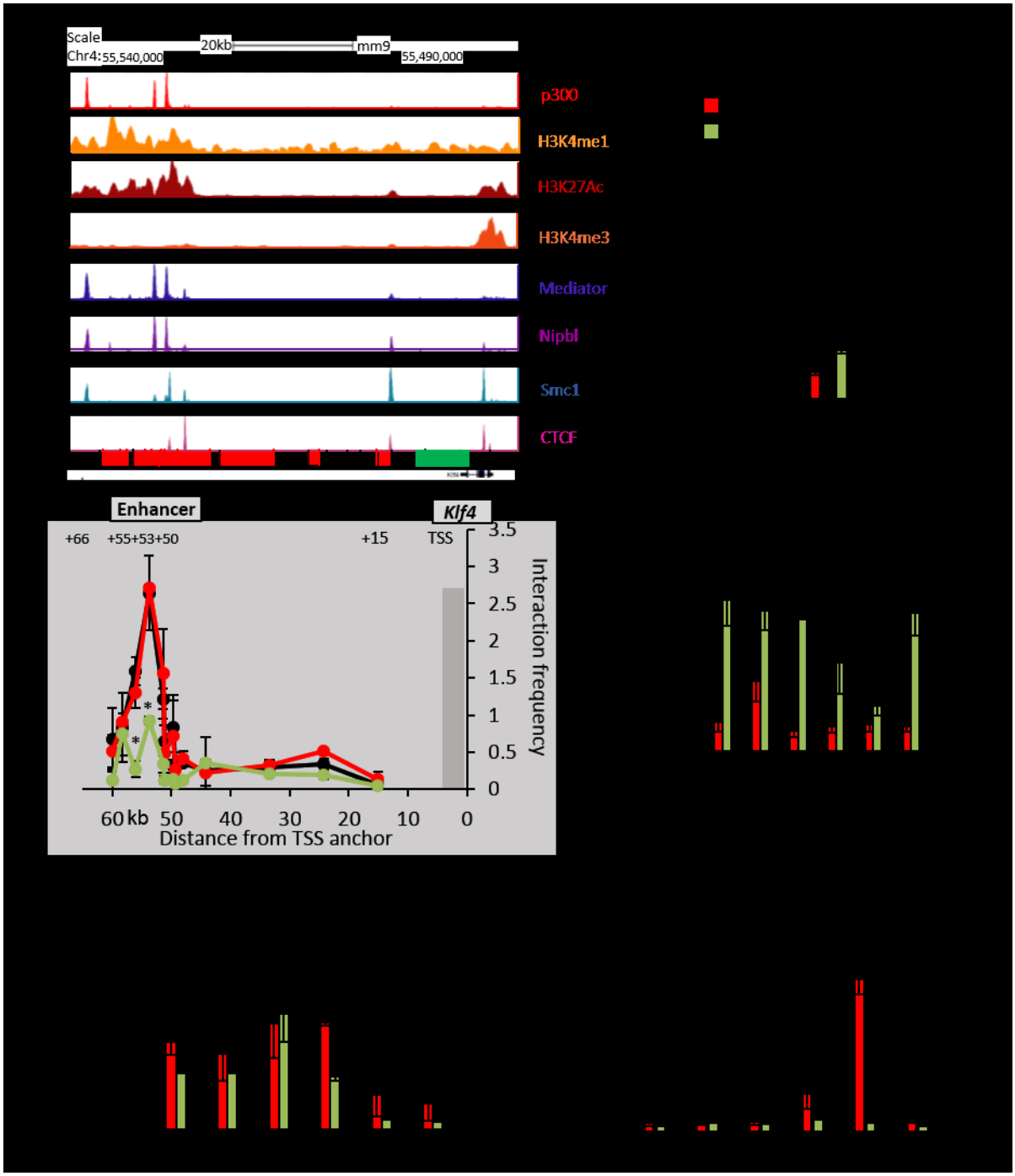


Figure 3.14 Enhancer-promoter interaction is maintained in spite of Klf4 downregulation upon cohesin depletion. A) Top panel: Binding of enhancer associated marks at the Klf4 locus. Digested fragments are marked – Anchor in green and analysed fragments in red. Bottom panel: 3C at the Klf4 locus. Fixed chromatin was digested with PvuII enzyme. The religated products were PCR amplified using TSS as the anchor fragment. Values for interaction frequency were normalised to BAC templates and control regions. The panel above shows the DNA interactions between the putative downstream enhancer and TSS remain unchanged upon cohesin depletion and are reduced upon cohesin depletion. B) Klf4 mRNA expression normalised to UBC and Ywhaz. C) Rad21 ChIP at the Klf4 enhancer and promoter. Results are plotted as fold change in Rad21 enrichment relative to that in control cells. Rad21 binding is reduced by 80% in KO ES cells and specifically decreases at the +53 and +55 kb sites upon differentiation. D) and E) CTCF and Nipbl ChIP at the Klf4 locus. Enrichment is plotted relative to 1% input chromatin. CTCF binding reduced at +53kb site while Nipbl reduced at both +53 and +55kb sites upon differentiation. (Mean ± SE; n = 3; Color code: Black - control ES cells, Red - $Rad21^{lox/lox}$ + 4'OHT 24h ES cells, Green – $Rad21^{lox/lox}$ – 2i 24h ES cells)

3.5 Discussion and future perspectives

Embryonic stem cells possess a unique transcriptional programme which allows them to perpetually self-renew while still being capable of giving rise to progenitors of the various different cell types in an organism. Cohesin has previously been reported to play an important role in these process (Ding et al., 2009; Hu et al., 2009; Kagey et al., 2010; Nitzsche et al., 2011). However, prolonged RNAi mediated knockdown of cohesin fails to dissociate essential cohesin functions in mitosis from its role as a regulator of transcription. In order to overcome such limitations, a conditional knockout system was used in this study to rapidly deplete the Rad21 subunit. I combined the use of ERT2Cre-*Rad21*^{lox} conditional allele with the culture of ES cells in 2i media. This provided a suitable system to study the direct impact of cohesin depletion on pluripotency. Additionally, as ES cells in 2i media show a more homogeneous transcriptional profile akin to a 'naïve' state of pluripotency (Ying et al., 2008; Silva et al., 2009), any bias for the selection of differentiated cells was avoided. Data presented here shows that cohesin protein levels were reduced to less than 10% within 24 hours of inducing *Rad21* knockout, before the onset of mitotic arrest and activation of stress response pathways. Thus in essence, any gene expression changes observed at 24 hours post *Rad21* knockout should be independent of any secondary effects due to cohesin depletion. Genome-wide transcriptional analysis of cohesin-depleted ES cells by microarrays at this time point revealed modest yet significant changes in the expression of a specific set of genes. I did not observe a global collapse in pluripotency nor a prominent upregulation of differentiation associated genes. On the contrary, the limited scale (598 genes), direction (up- or down-regulated) and magnitude of the gene expression changes observed, suggest that cohesin has locus-specific impacts on transcriptional regulation. Deregulated genes are enriched for developmental functions and are found to be bound by cohesin or Nipbl at the promoters and distal regulatory elements, again supporting a direct role of cohesin in the regulation of these genes. These observations are in agreement with studies in CdLS animal models where depletion of Nipbl and cohesin subunits also leads to modest yet specific gene expression changes (Kawauchi et al., 2009; Muto et al., 2011).

These results are in disagreement with previously published work by Kagey et al., (2010) who have done a similar analysis but using the knockdown strategy in serum ES cultures. They reported a massive downregulation of pluripotency genes. However, this observed discrepancy can be attributed to a bias for selection of differentiated ES cells due to prolonged cohesin depletion and the consequent stress induced secondary effects on gene expression. This

speculation is supported by the fact that their dataset is highly enriched for cell cycle and DNA damage related genes. Activation of cell cycle and DNA damage checkpoints has already been shown to supress the expression of key pluripotency genes like *Nanog* and trigger ES cell differentiation (Lin et al., 2005; Maimets et al., 2008; Li et al., 2012). Thus, the gene expression changes observed by Kagey et al, (2010) might in large part be a secondary effect of stress induced differentiation. The same effect is observed when ES cells were forcefully stressed by the use of doxorubicin leading to the downregulation of all the pluripotency associated genes tested in this study. In fact, closer examination of gene expression changes at 48 hours post cohesin depletion, when ES cells are arrested in G2 phase and show upregulation of stress markers like *Mdm2*, also reveals a more pronounced downregulation of pluripotency genes as compared to that at 24 hours. Thus, in addition to delineating cohesin's role in pluripotent gene expression programme, this study also points to critical shortcomings in existing knockdown studies. It provides further evidence that activation of stress responses in ES cells can lead to differentiation, and emphasises on the avoidance of such stress while analysing cohesin's role in regulating gene expression in cycling cells.

Having ascertained that cohesin plays a direct role in the transcriptional regulation of specific genes expressed in ES cells, I selected a few key pluripotency associated genes to understand the mechanisms of cohesin action. Given the previously reported roles of cohesin in mediating long-range DNA interactions between *cis*-regulatory elements (Kagey et al., 2010; Seitan et al., 2011; Merkenschlager and Odom, 2013), I focused on analysing how enhancer-promoter interactions are affected by the loss of cohesin and how does this relate to observed changes in gene expression. At the *Nanog* locus, reduction in expression upon cohesin depletion and differentiation at the 24 hour time point was associated with a similar decrease in enhancer-promoter interactions. At the *Klf4* locus, however, enhancer-promoter interactions were maintained in the absence of cohesin even though *Klf4* expression was significantly reduced. Another interesting contrast was observed in case of *Lefty1* where the gene displayed intact enhancer-promoter contacts and was in fact upregulated upon cohesin depletion. In contrast to the impact of cohesin depletion, enhancer-promoter interactions at both *Klf4* and *Lefty1* locus were diminished in response to differentiation and were associated with reduced Rad21 and Nipbl binding. These observations highlight several aspects of cohesin's role in mediating enhancer-promoter interactions and the resultant impact on gene expression.

The first intriguing observation is that cohesin depletion does not necessarily lead to a loss in enhancer-promoter interactions at deregulated genes. This is particularly evident at the *Lefty1* and *Klf4* loci. Even at the *Nanog* locus, it is striking that >80% depletion in cohesin binding is accompanied by only a small reduction in enhancer-promoter interactions. These findings suggest that even though cohesin associated active enhancers are engaged with their respective target promoters via long-range DNA interactions, some of these established interactions can be maintained in the absence of cohesin. This is in agreement with the observation that cohesin's impact on pluripotency is locus-specific because if cohesin was required to maintain all enhancer-promoter interactions, cohesin depletion would be accompanied by downregulation of most active genes. Rather, it is more probable that some genes and enhancer-promoter interactions are more sensitive to perturbations upon cohesin depletion. In addition, the actual maintenance of DNA looping between enhancers and promoters in the absence of cohesin may be attributed to interactions between transcription factors like Oct4, Nanog and components of the mediator complex that are bound to enhancers and promoters. This hypothesis is supported by previous studies which show that Oct4, Sox2 and Nanog cluster together at regions involved in long-range DNA interactions and that loss of either Oct4 or Nanog diminishes these contacts. In fact, insertion of an ectopic Nanog binding site is sufficient for the formation of DNA contacts with other endogenous OSN bound regions (de Wit et al., 2013; Denholtz et al., 2013). Thus, it is possible that preformed enhancer and promoter DNA interactions, probably mediated by cohesin, are supported by protein-protein interactions which are strong enough to maintain these DNA contacts upon later removal of cohesin.

Another important characteristic of cohesin mediated gene regulation illustrated in this study is the dynamic nature of cohesin binding in response to cellular cues. This is clearly evident at the *Klf4* locus where there are several cohesin and CTCF binding sites across the 70kb region analysed. The +55kb region is a CTCF independent cohesin and Nipbl binding site with a high interaction frequency with *Klf4* TSS in ES cells. Upon differentiation, cohesin binding at the CTCF sites remains invariant but is specifically reduced at the +55kb enhancer region. Similarly, Nipbl binding is also specifically decreased at the +55kb region. Importantly, the reduced cohesin and Nipbl binding is associated with decreased enhancer-promoter interactions. The same was observed at the *Nanog* locus, where cohesin and Nipbl binding and DNA contacts were reduced at the enhancer and promoter regions upon differentiation. These results show that cohesin loading at *cis*-regulatory elements is highly responsive to cellular signals like differentiation. As

cohesin binding at enhancer regions was reduced in both cohesin-depleted ES cells and differentiating ES cells, but reduced DNA interactions were observed only in differentiating ES cells, cohesin binding alone cannot be the sole determinant of the formation of long-range DNA interactions. It can be surmised that cohesin unloading in response to differentiation signals accompanied by reduced binding of downregulated pluripotency associated TFs might together result in decreased enhancer-promoter interactions.

An additional feature that emerges from this study is that enhancer-promoter interactions alone cannot explain the observed gene expression changes upon cohesin depletion. For example, *Lefty1* upregulation could not be ascribed to increase enhancer-promoter interactions. It is possible that *Lefty1* overexpression is caused due to the loss of an interaction with a repressor element or a gain in interaction with additional enhancer elements both of which could be made possible due to the breakdown of the existing chromatin framework in the absence of cohesin. However, these hypotheses could not be tested in the present study which relies on the use of 3C technique. 3C is dependent on pre-existing knowledge of DNA elements whose interactions are to be measured. Even the *Nanog* gene has been reported to be involved in multiple long-range DNA interactions apart from the 5kb upstream enhancer. However, the functional significance of any of these other interactions for transcriptional regulation has not been tested. Nevertheless, it is possible that the impact of cohesin depletion is a combined result of the effect on each of these individual interactions and not a single regulatory element. As such, it would be useful to extend the scope of this analysis to more genes and study the associated DNA interactions on a more global scale by using 4C or Hi-C. However, increasing coverage with these technologies would be limited by the loss of resolution and it might not be possible to pick up small differences in interactions at shorter ranges as assessed here by 3C. The *Klf4* locus also provides a rather contrasting observation where gene transcription is significantly reduced even when enhancer-promoter contacts are maintained in the absence of cohesin. This suggests a dissociation of the mere presence of enhancer-promoter contacts from their role in activating transcription i.e. bringing of an enhancer element in close proximity of a gene promoter is not necessarily sufficient for the activation of transcription, and signifies the role of other TFs and regulatory factors in making this association a productive one. At the same time, downregulation of *Klf4* in the absence of cohesin implicates cohesin's role in transcription beyond its role in mediating enhancer-promoter interactions.

Altogether, this work shows that cohesin plays a direct role in regulating specific genes in ES cells. Furthermore, the data presented here, even though limited, provides strong evidence that enhancer-promoter interactions can be maintained upon cohesin removal.

Chapter 4 : Cohesin functions in regulation of *Myc* expression

Myc is one of the genes that is significantly upregulated in the microarray dataset of cohesin deficient ES cells. *Myc* is involved in several cellular processes including stem cell self-renewal and pluripotency. Its expression is precisely regulated in response to several extracellular and intracellular signals and deregulated *Myc* expression can lead to malignant transformation of cells (Meyer and Penn, 2008; Dang, 2012). Interestingly, previous reports suggest that depletion of cohesin and its regulatory subunits results in *Myc* downregulation (Rhodes et al., 2010; McEwan et al., 2012). However, these studies also reported an accompanied upregulation of p53 and *Mdm2* expression. Thus it is difficult to conclude whether *Myc* deregulation was a direct effect of cohesin depletion, or an indirect impact of activated stress response. Nevertheless, the observed contrast between the upregulation of Myc in ES cells, and the previously reported downregulation of *Myc* in the absence of cohesin, is extremely intriguing. Furthermore, cohesin mutations have been identified in several types of tumors but the underlying cause is not yet understood. It is possible that the cancerous growth of cells results from deregulated *Myc* expression due to cohesin mutations. Therefore, it is of tremendous value to understand if and how cohesin regulates *Myc* expression.

The work in this chapter explores the upregulation of Myc expression in cohesin deficient ES cells and further attempts to examine the impact of cohesin depletion on *Myc* expression in various different cell types.

4.1 Myc expression is post-transcriptionally upregulated in cohesin deficient ES cells growing in 2i media

To study the impact of cohesin depletion on *Myc* expression in ES cells, I first used the previously described ERT2Cre-*Rad21*$^{WT/WT}$ and ERT2Cre-*Rad21*$^{lox/lox}$ ES cells growing in 2i supplemented media. The cells were treated with 4'OHT and RNA samples were collected at different time points. *Myc* mRNA levels were assessed by qRT-PCR analysis using primers spanning the exon2-exon3 junction, which is common to both isoforms A and B of *Myc* transcripts. Results clearly showed that *Myc* mRNA levels were markedly upregulated upon 24 hours of cohesin depletion and the increased levels were maintained even after 48 hours (Figure 4.1B left). Primers designed to amplify the exon1-exon2 junction exclusively present in isoform A also confirmed the observed upregulation of *Myc* mRNA (Figure 4.1B right). Western blot analysis revealed that even MYC protein levels were significantly increased upon cohesin depletion. The presence of Ser62 and Thr58 phosphorylated forms of MYC protein indicates that the upregulated MYC is functionally active (Figure 4.1C).

I further analysed the transcriptional status of the *Myc* gene by designing primers which bind to an intron-exon pair and thus amplify the primary transcript. As the RNA samples were subjected to DNAse treatment prior to cDNA preparation, amplification of gDNA was avoided and intron-exon spanning primers specifically detected primary transcripts. Results showed that the untreated ERT2Cre-*Rad21*$^{lox/lox}$ ES cells at 0 hour already displayed slightly higher levels of *Myc* transcripts as compared to ERT2Cre-*Rad21*$^{WT/WT}$ ES cells. The *Myc* gene also demonstrated changes in gene expression with changing cellular density in the culture dish with increasing time (Appendix 1). This accounts for the observed increase in *Myc* transcripts at 24 hours and the subsequent reduction in both ERT2Cre-*Rad21*$^{WT/WT}$ and ERT2Cre-*Rad21*$^{lox/lox}$ ES cells. After accounting for the cell-variant and density-dependent differences, the qRT-PCR analysis surprisingly revealed that the primary transcript levels did not show an equivalent increase in levels as was observed for *Myc* mRNA upon cohesin depletion (Figure 4.1D). While *Myc* mRNA showed more than 5-fold increase in the absence of cohesin, the impact on primary transcript levels was minor at best. This indicates that cohesin depletion did not affect *Myc* transcription and the impact on increased *Myc* mRNA and protein is post-transcriptional. As the *Myc* mRNA is short-lived with a half-life of only 15-20 minutes, the sustained increase in mRNA and protein levels without increased transcription can be attributed to an increase in mRNA stability in the absence of cohesin.

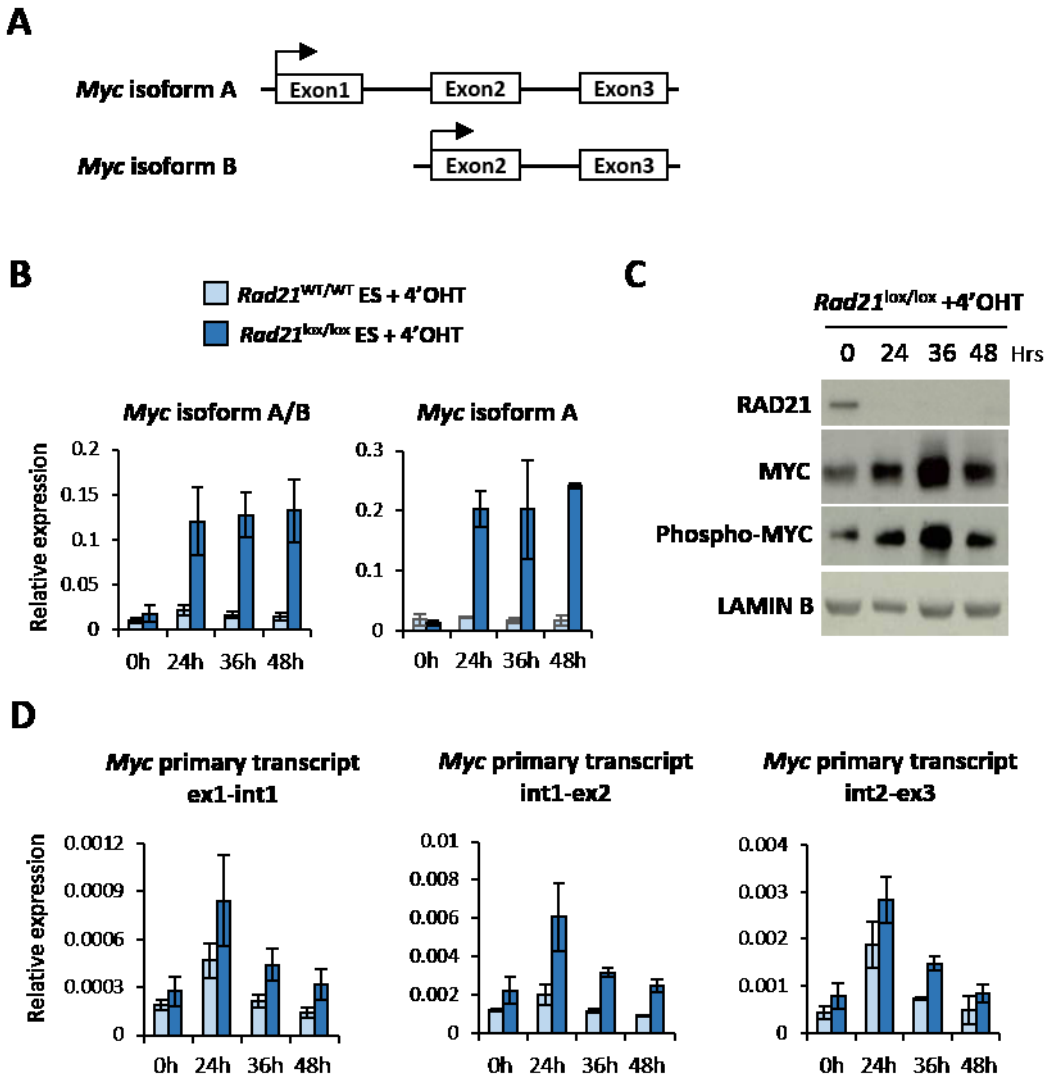


Figure 4.1 Myc expression is post-transcriptionally upregulated upon cohesin depletion in ES cells growing in 2i media. A) Schematic representing the two isoforms of *Myc* transcripts. **B)** qRT-PCR analysis of *Myc* mRNA isoforms A and B (left) or A only (right). *Myc* mRNA is highly upregulated upon cohesin depletion. **C)** Western blot analysis shows an increase in the total and phosphorylated forms of MYC protein. **D)** *Myc* primary transcript analysis by qRT-PCR with primers binding in the indicated exons and introns. Transcription of *Myc* gene was not markedly upregulated upon cohesin depletion. (Normalised to *Ubc* and *Ywhaz*; mean ± SE; n≥3)

Although further experiments are required to prove that the observed *Myc* upregulation is due to an increase in mRNA and protein stability, these experiments provide the first instance where cohesin's involvement in gene regulation is post-transcriptional. Since cohesin is predominantly a nuclear DNA associated protein in its functional form, it is highly unlikely that cohesin directly interacts with *Myc* mRNA or protein to affect its stability. On the other hand, it is quite possible that increased Myc stability is a consequence of the direct impact of cohesin on the transcription of another regulatory factor, like RNA binding proteins or non-coding RNAs. A

cursory look at the list of deregulated genes from the microarray dataset did not reveal any obvious possible regulators of protein stability. However, I noted that *Pvt1* (plasmacytoma variant translocation), a long non-coding RNA (lncRNA) encoded within the 8q24 gene desert containing the *Myc* gene, is significantly upregulated upon cohesin depletion. The upregulation of *Pvt1* was also confirmed by qRT-PCR analysis (Figure 4.2). Until now, the function of *Pvt1* was not known. However, a recent study reported that *PVT1* interacts with and stabilises MYC protein by preventing MYC phosphorylation and degradation, and depletion of *PVT1* reduced the half-life of MYC protein (Tseng et al., 2014). Therefore, it is possible that increased *Myc* levels are a consequence of *Pvt1* upregulation due to cohesin depletion. Altogether, these experiments provide an intriguing novel mechanism of cohesin-mediated gene regulation, where cohesin regulates the expression of a lncRNA which in turn post-transcriptionally regulates the expression of an important gene like *Myc*.

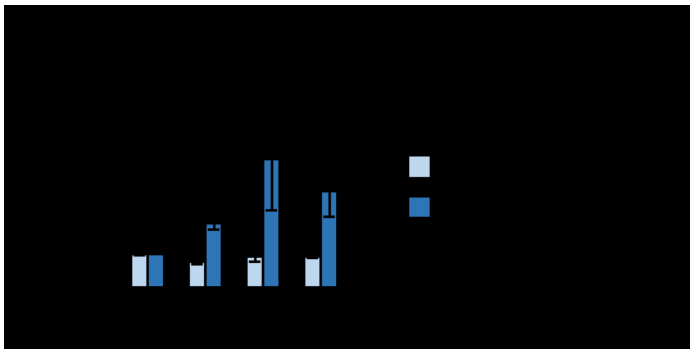


Figure 4.2 The lncRNA *Pvt1* is upregulated upon cohesin depletion in ES cells growing in 2i media. qRT-PCR analysis shows that *Pvt1* expression is significantly upregulated in cohesin depleted ES cells. (Normalised to *Ubc* and *Ywhaz*; mean ± SE; n=2)

The stability of MYC is determined in part by the combined activity of ERK and GSK3β signalling (Sears, 2004). Consequently, ES cells growing in 2i media have lower levels of *Myc* mRNA and protein due to ERK and GSK3β inhibition, as compared to ES cells growing in Serum + LIF culture conditions (Ying et al., 2008). Since the impact of cohesin depletion on ES cells may vary depending on extracellular and intracellular signalling in different culture conditions, I decided to check if *Myc* upregulation in the absence of cohesin was a feature of ES cells in general, or a culture-dependent phenomenon. For this, the ERT2Cre-*Rad21*$^{WT/WT}$ and ERT2Cre-*Rad21*$^{lox/lox}$ ES cells originally derived and cultured in 2i media were adapted to serum + LIF culture conditions. The cohesin depletion experiments were performed as previously described (Figure 3.1). Interestingly, cohesin depletion in serum adapted ES cells did not cause *Myc* upregulation. In fact,

both Myc mRNA and primary transcript levels were unaffected in cohesin depleted serum ES cells (Figure 4.3A). MYC protein levels were unchanged at 24 hours post cohesin depletion, but were drastically reduced thereafter (Figure 4.3B). The reduced MYC levels observed in spite of proper *Myc* transcription indicate that MYC protein is degraded, possibly due to the activation of stress response pathway after 24 hours.

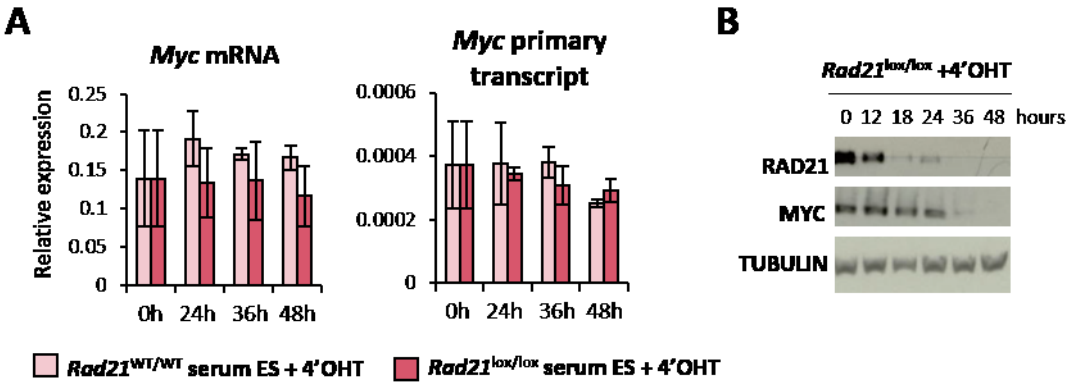


Figure 4.3 Cohesin depletion does not affect *Myc* expression in serum ES cells. A) qRT-PCR analysis for *Myc* mRNA and primary transcript levels shows no significant difference between widltype and cohesin depleted ES cells growing in serum + LIF media (Normalised to *Ubc* and *Ywhaz*; mean ± SE; n=3). **B)** Western blot analysis shows that MYC protein levels are not affected until 24 hours of cohesin depletion but are significantly reduced afterwards.

The striking difference between the observed impact of cohesin depletion on *Myc* expression in 2i and serum ES cells shows that cohesin functions in gene regulation are sensitive to extracellular and intracellular signalling. Thus, the stabilisation of MYC in cohesin depleted 2i ES cells could possibly be a combined effect of *Pvt1* overexpression and alleviation of the inhibitory effect of the 2i inhibitors. The impact of cohesin depletion on *Pvt1* expression in serum ES cells remains to be determined in order to be able to dissect the molecular differences between *Myc* expression in 2i and serum ES cells.

4.2 Serum stimulated activation of *Myc* in serum starved MEFs is independent of cohesin

Given the observed disparity between the impact of cohesin depletion on *Myc* expression in ES cells growing in 2i media and those growing in serum + LIF media, I decided to extend the analysis to several different cell types to further understand the role of cohesin in *Myc* regulation. MEFs can be arrested in a G1/G0 phase by serum starvation. Serum starved resting MEFs have very low levels of *Myc* mRNA which are dramatically increased when cells are stimulated with serum (Blanchard et al., 1985). I decided to check if cohesin is required for *Myc* expression in response to serum stimulation. For this purpose, I used MEFs derived from ERT2Cre-*Rad21*$^{lox/lox}$ embryos which can be conditionally depleted of *Rad21* upon 4'-OHT treatment. Briefly, ERT2Cre-*Rad21*$^{lox/lox}$ MEFs were treated with 100nM 4'-OHT or 100% ethanol for 2 days in serum supplemented media (DMEM + 10% FCS), after which they were serum starved (DMEM + 0.1% FCS) for 6 days. This allowed for proper cohesin depletion and as the serum starved MEFs are arrested in G1/G0 phase, it provided a system where gene expression changes can be analysed independently of cohesin's role in cell division (refer to Section 5.1 for detailed characterisation of the system). At the end of 6 days, cells were re-introduced into serum supplemented media and *Myc* expression was analysed 6 and 12 hours post serum stimulation.

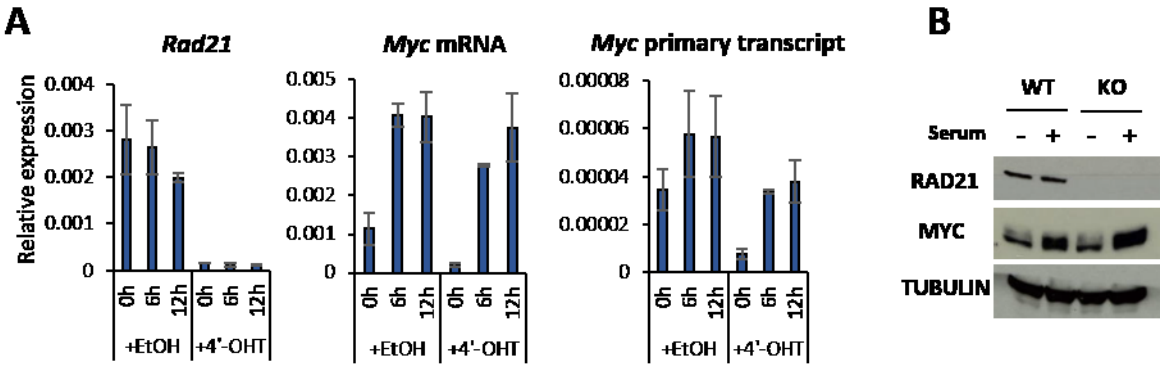


Figure 4.4 *Myc* activation in MEFs upon serum stimulation does not require cohesin. A) qRT-PCR analysis in ERT2Cre-*Rad21*$^{lox/lox}$ MEFs shows *Myc* transcription is appropriately activated upon 6 and 12 hours of serum stimulation in the absence of cohesin (Normalised to *18S rRNA*; mean ± SE; n=3) **B)** Western blot confirms MYC induction in both wildtype and *Rad21*$^{-/-}$ MEFs 6 hours post serum treatment.

Results showed that even though both *Myc* mRNA and primary transcript levels were reduced in serum starved cohesin deficient MEFs, *Myc* activation upon serum stimulation was not affected (Figure 4.7A) and MYC protein levels were equivalently increased in both wildtype and *Rad21*$^{-/-}$ MEFs (Figure 4.7B). These experiments indicate that cohesin is not necessary for the

increased transcription and expression of MYC protein when serum starved MEFs are stimulated by serum growth factors.

4.3 Cohesin depletion does not affect *Myc* expression in G1 arrested preB cells.

Next I decided to analyse the dependence of *Myc* expression on cohesin in Abelson transformed preB cells, which can be arrested in G1 phase by the addition of imatinib (STI-571), an inhibitor of the Abelson protein tyrosine kinase (Schindler et al., 2000). Mouse ERT2Cre-*Rad21*$^{lox/WT}$ and ERT2Cre-*Rad21*$^{lox/lox}$ preB cells containing a *Bcl2* transgene were derived in our laboratory. These cells were transformed by the Abelson murine leukaemia virus allowing for the prolonged culture of these cells *in vitro*. By preventing apoptosis, the Bcl2 transgene ensured survival of these cells upon STI treatment. The ERT2Cre-*Rad21*$^{lox/WT}$ and ERT2Cre-*Rad21*$^{lox/lox}$ preB cells were treated with 400nM 4'-OHT to induce cohesin deletion and then arrested in G1 by the addition of 2µM STI (Figure 4.8A, C). Rad21 mRNA (Figure 4.8D) and protein (Figure 4.8B) were efficiently depleted in ERT2Cre-*Rad21*$^{lox/lox}$ preB cells 4 days after 4'-OHT addition. Therefore, G1 arrested preB cells with the conditional cohesin allele also provide a valuable system where the effects of cohesin depletion on *Myc* gene expression can be studied in non-dividing cells.

qRT-PCR analysis revealed that *Gapdh* and *Myc* mRNA levels were significantly reduced upon STI treatment in both heterozygous and homozygous cohesin depleted preB cells. Expression of other housekeeping genes like *Ubc*, *β-Actin* and 18SrRNA, however, remained unaffected. *Myc* transcription was also effectively reduced in the presence of STI (Figure 4.8D) and MYC protein was undetectable in Western blot (not shown). The repression of *Gapdh* and *Myc* could likely be a consequence of the inhibition of Abl kinase by STI, which influences the growth and survival of the cells (Roskoski, 2003). Thus, the cell cycle arrest of preB cells appears to be accompanied by a decline in growth and metabolic activities as well. *Myc* levels were similarly reduced in both heterozygous and homozygous cohesin deficient cells treated with STI. However, this system is not ideal to evaluate cohesin function in Myc regulation because of the unavoidable inhibitory effect of STI treatment.

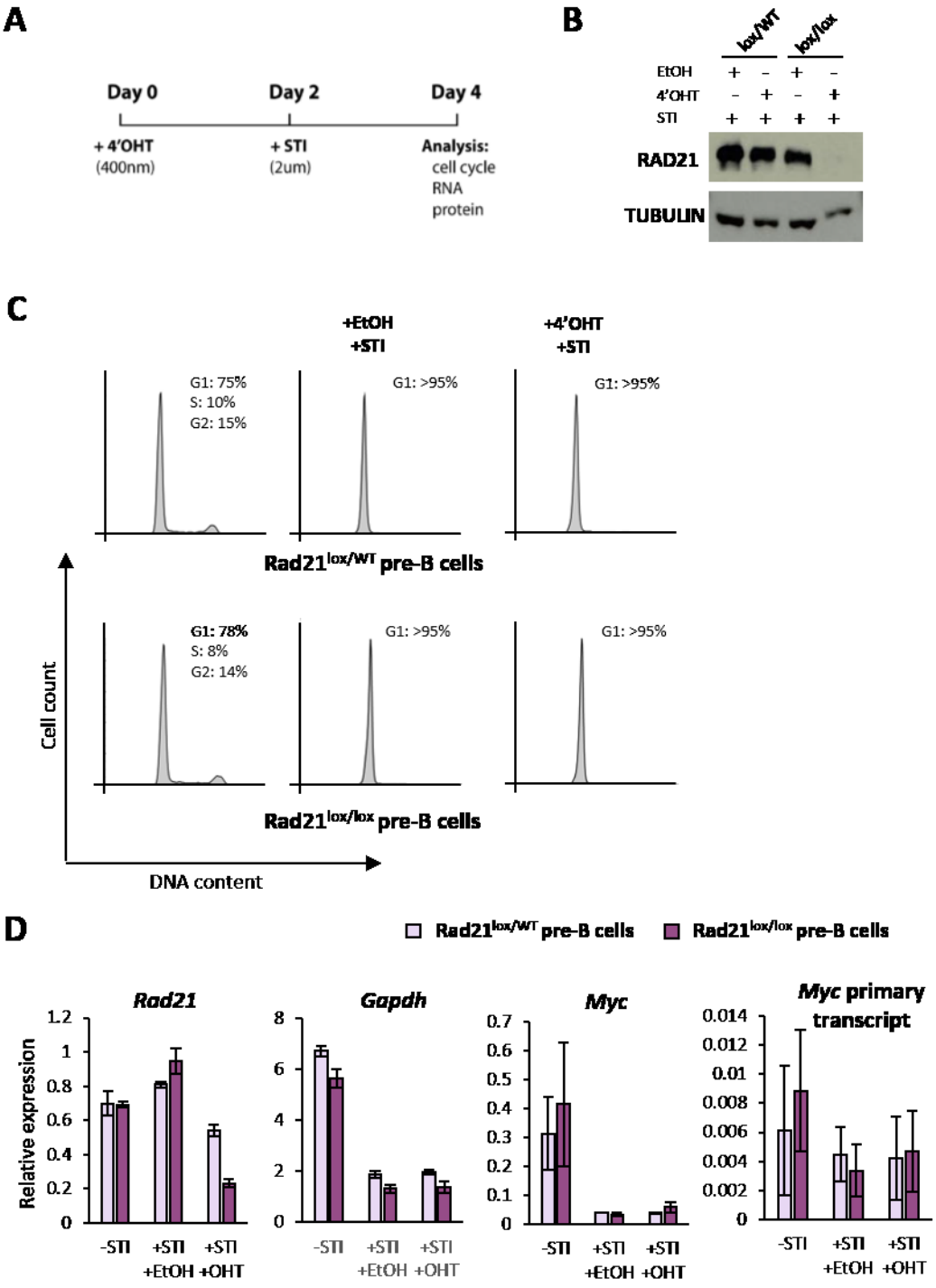


Figure 4.5 Cohesin depletion does not affect Myc downregulation in G1 arrested preB cells. A) Schematic outline of the experimental setup. ERT2Cre-*Rad21*$^{lox/WT}$ and ERT2Cre-*Rad21*$^{lox/lox}$ preB cells were treated with 400nM 4'-OHT or ethanol for 2 days before the addition of 2μM STI. Cells were collected for analysis at the end of 4 days. **B)** Western blot analysis shows that RAD21 protein levels are efficiently reduced in4'-OHT treated ERT2Cre-*Rad21*$^{lox/lox}$ preB cells. C) Cell cycle analysis shows that majority of the STI treated cells are arrested in G1 phase. D) qRT-PCr analysis of RNA samples shows similar reduction in *Myc* and *Gapdh* expression upon STI treatment (Normalised to *Ubc* expression; mean ± SE; n=3).

4.4 Cohesin depletion does not affect *Myc* expression in resting thymocytes

Developing thymocytes go through several stages of maturation that are associated with the expression of specific cell surface markers. Late stage $CD4^-CD8^-$ double negative (DN) thymocytes go through a phase of proliferation that extends to the early $CD4^+$ $CD8^+$ double-positive (DP) stage. Shortly after the acquisition of CD4 and CD8, DP thymocytes lose the expression of the transferrin receptor CD71 and become small, non-proliferating $CD71^-$ DP cells, comprising the majority of cells in the thymus. At this stage, successful TCRα gene rearrangement, surface expression and the engagement of TCRαβ heterodimer selects a minority (3–5%) of DP thymocytes for differentiation towards long-lived CD4 or CD8 single-positive (SP) cells, again with minimal proliferation (Figure 4.9A) (Seitan et al., 2011). These non-proliferating thymocytes provide another opportunity to study the role of cohesin in gene regulation.

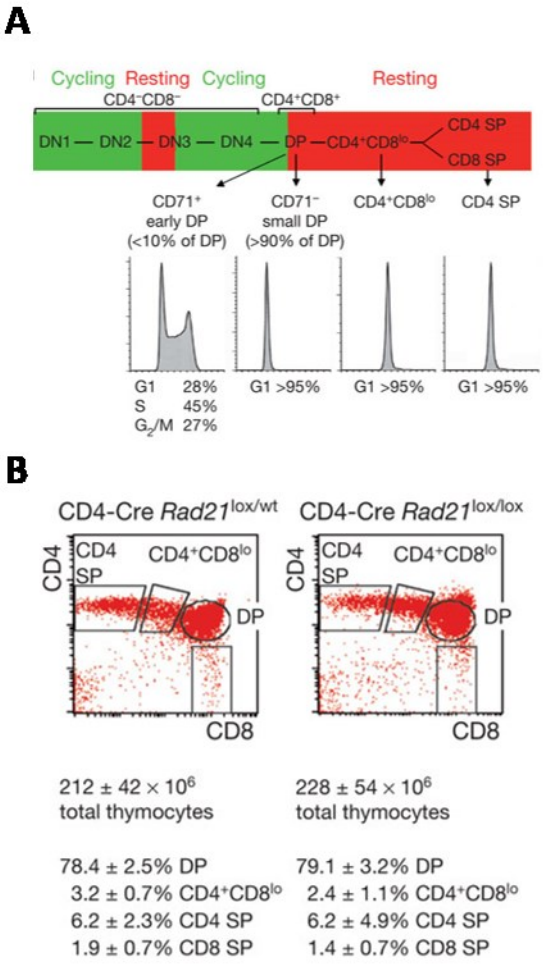


Figure 4.6 Cellular profile of $CD4^+$ non-cycling thymocytes.

A) Schematic of thymocyte differentiation and cell cycle profiles. From left to right: $CD4^-$ $CD8^-$ double negative (DN) stages 1 to 4; $CD4^+$ $CD8^+$ double positive (DP), $CD4^+$ $CD8^{lo}$ DP transitional stage; $CD4^+$ or $CD8^+$ single positive (SP) cells. Proliferative stages are represented in green, cell cycle arrest in red. Histograms show DNA content.
B) Cell numbers and flow cytometric analysis of thymocyte subsets in 6-week-old CD4-Cre *Rad21*$^{lox/lox}$ and CD4-Cre *Rad21*$^{lox/wt}$ mice (mean ± standard deviation, *n* = 12).

In order to conditionally deplete RAD21 in resting thymocytes, *Rad21*$^{lox/lox}$ mice were crossed with mice expressing a Cre transgene under the control of *Cd4* regulatory elements. The CD4Cre transgene is activated just before the transition from the DN to DP stage, when developing thymocytes initiate *Cd4* expression and enter the non-proliferative stage. As a result, CD4Cre selectively drives the loss of cohesin in non-dividing thymocytes. *Rad21* genomic deletion is essentially complete and *Rad21* mRNA and protein levels are substantially reduced in CD4^{+} CD8^{+} DP and CD4^{+} SP thymocytes from CD4Cre-*Rad21*$^{lox/lox}$ mice (Figure 4.10A, B). Importantly, CD4Cre-*Rad21*$^{lox/lox}$ DP thymocyte numbers were normal. Intermediate CD4^{+} CD8^{lo} and mature CD4 SP and CD8 SP thymocytes accumulated slowly in CD4Cre-*Rad21*$^{lox/lox}$ mice due to reduced efficiency of differentiation but were present in normal numbers by 6 weeks of age (Figure 4.9B) (Seitan et al., 2011). RNA-seq showed that *Myc* was expressed at very low levels in both CD4Cre-*Rad21*$^{lox/lox}$ and control (CD4Cre-*Rad21*$^{lox/WT}$) DP thymocytes (Figure 4.10C). Furthermore, cohesin depletion also did not have a significant effect on *Myc* expression in CD4 SP thymocytes (Figure 4.10D).

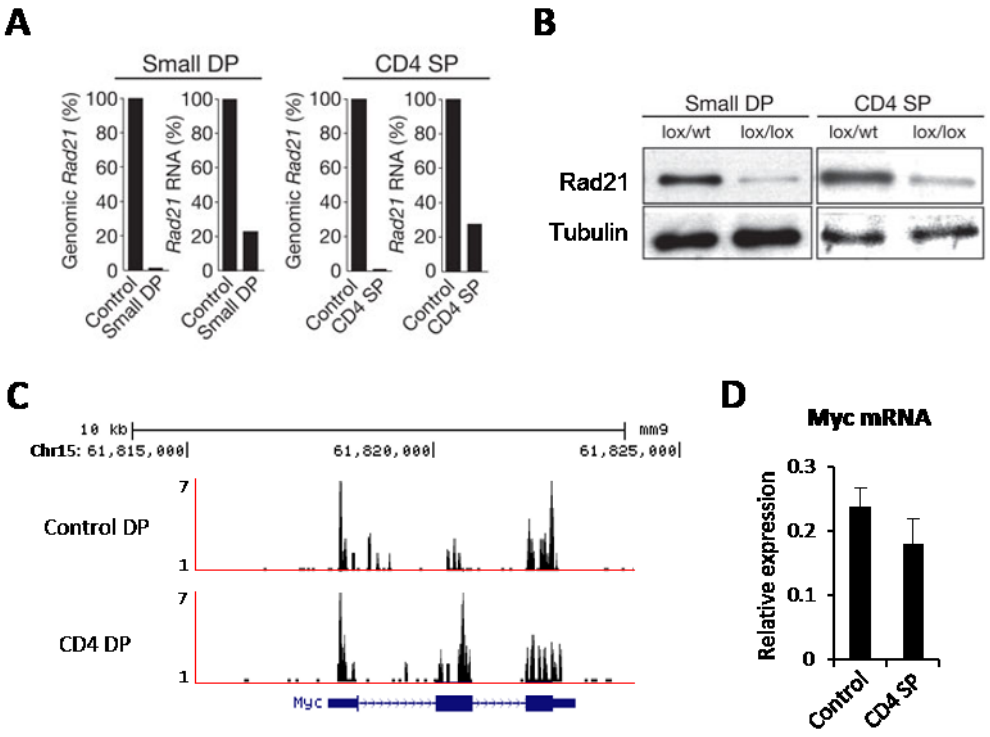


Figure 4.7 Conditional cohesin depletion in resting DP and CD4 SP thymocytes does not affect *Myc* expression. A) & B) Genomic deletion of the Rad21 allele and mRNA as assessed by RT-PCR, Rad21 Western blot shows efficient cohesin depletion in CD4Cre-*Rad21*$^{lox/lox}$ and control (CD4Cre-*Rad21*$^{lox/lox}$) DP and CD4 SP thymocytes. **C)** RNA-seq reads at the *Myc* locus in control and CD4 DP thymocytes. **D)** qRT-PCR analysis for *Myc* mRNA in CD4 SP thymocytes (Normalised to *Ubc, Ywhaz* expression; mean±SE; n=3). *Myc* expression remains unaffected upon cohesin depletion in both DP and CD4 SP thymocytes.

4.6 Discussion and future perspectives

4.6.1 Upregulation of *Myc* upon cohesin depletion in 2i cultured ES cells

Myc plays an important role in the maintenance of ES cell pluripotency and *Myc* expression is tightly regulated during developmental processes. Interestingly, cohesin depletion in ES cells growing in 2i media lead to increased *Myc* mRNA and protein levels. Interestingly, the loss of cohesin did not affect the transcriptional status of *Myc* but the cause of *Myc* upregulation was post-transcriptional. I speculate that cohesin depletion resulted in an increased *Myc* mRNA and protein stability but further experiments like mRNA half-life analysis are required to confirm this. So far, cohesin's role in the regulation of gene expression has essentially been attributed to its ability to associate with DNA, and there is no evidence for a direct interaction of cohesin with mRNAs and soluble proteins which may influence their stability. Therefore, it is conceivable that cohesin depletion directly deregulates the expression of another factor which in turn is an important determinant of *Myc* stability.

Both *Myc* mRNA and protein have a very rapid turnover. Degradation of *Myc* mRNA is regulated by several mechanisms, including the activity of RNA binding proteins and microRNAs (Ross, 1995; Jackstadt and Hermeking, 2014). In addition, translation-coupled mechanisms have also been proposed where association with ribosomes protects *Myc* mRNA from degradation (Bernstein et al., 1992). On the other hand, MYC protein stability is determined mainly by post-translational modifications. While phosphorylation of the Ser62 residue on MYC increases protein stability and causes rapid accumulation of MYC, phosphorylation of Thr58 directs MYC ubiquitination and proteasomal degradation causing rapid downregulation (Sears, 2004). Therefore, it will be useful to study if cohesin depletion disrupts the precise balance between the Ser62 and Thr58 phosphorylations by using specific antibodies against both of these modifications to observe the changes at protein level. Investigation into whether MYC protein is stabilised due to increased Ser62 phosphorylation or due to decreased Thr58 phosphorylation or a combination of both, will further help us deduce other effector molecules which may be involved in the cohesin-dependent upregulation of MYC. It would also be interesting to explore whether the absence of cohesin stabilises both mRNA and protein independently of each other or increased amounts of protein are a consequence of increased mRNA levels in a process that is translationally-coupled. Experiments comparing the half-life of *Myc* mRNA and protein in cohesin-depleted ES cells treated with a transcriptional inhibitor (like Actinomycin D) or an inhibitor of

translation elongation (like Cycloheximide) or both, will help us further understand the role of cohesin in this process.

In order to make a mechanistic link between cohesin and *Myc* upregulation, the list of deregulated genes in cohesin-depleted ES cells was scrutinised for the indication of any prominent RNA binding protein genes known to be involved in *Myc* regulation. Although, this basic analysis did not reveal any definitive targets, it is interesting that the expression of *Pvt1*, a long non-coding RNA gene located adjacent to the *Myc* gene within the 8q24 gene desert region, was significantly upregulated upon cohesin depletion. A recent study shows that *PVT1* interacts with and stabilises MYC protein by preventing its degradation. Depletion of *PVT1* reduces the half-life of MYC protein (Tseng et al., 2014). Further experiments will address whether increased *Myc* levels are a consequence of *Pvt1* overexpression. As it stands, the work presented here shows that cohesin prevents *Pvt1* overexpression in ES cells growing in 2i media and *Pvt1* upregulation in the absence of cohesin might be responsible for the observed increase in *Myc* levels. It remains to be seen if cohesin regulates *Pvt1* expression in other cell types as well, or if this function is specific to ES cells growing in 2i media. Cohesin-mediated regulation of *Pvt1* may be of significance for evaluating the role of cohesin in the development of cancer.

In addition to its role in the maintenance of pluripotency, *Myc* also plays an important role in somatic cell reprogramming. Myc is one of the original Yamanaka factors that mediate the formation of iPS cells (Takahashi and Yamanaka, 2006). Ectopic c-*Myc* is dispensable for the creation of iPS cells, but acts as an enhancer of kinetics and efficiency of reprogramming (Nakagawa et al., 2008; Wernig et al., 2008). It has been proposed that *Myc* facilitates reprogramming by increasing the proliferation of iPS cells (Hanna et al., 2009) and by cooperating with other reprogramming factors in the transcriptional regulation of target genes (Soufi et al., 2012). In the context of cell-fusion mediated heterokaryon formation, where the early steps of reprogramming occur without cell division, a recent study (Lavagnolli et al., 2015) provides evidence that *Myc* enhances reprogramming efficiency by promoting DNA replication. The observed upregulation of *Myc* in cohesin-depleted ES cells was shown to cause the unexpected increase in their potential to reprogram somatic cells. Together, these experiments show that unlike previous reports claiming that cohesin is required for iPS reprogramming (Apostolou et al., 2013; Zhang et al., 2013a), cohesin is not necessary for the ability of ES cells to induce the expression of pluripotency genes in somatic cells during cell-fusion mediated reprogramming. In fact, cohesin depletion increased the reprogramming potential of ES cells in part by promoting

Myc-dependent DNA replication. As somatic cell reprogramming occurs in the absence of cell division in interspecies heterokaryons, they provide an ideal system to study the role of cohesin in initiation of somatic cell reprogramming. Since iPS cell formation typically requires cellular proliferation, it is likely that the requirement of cohesin for iPS reprogramming stems from its essential functions in cell division (Lavagnolli et al., 2015).

4.6.2 *Myc* expression is not dependent on the availability of cohesin

The upregulation of Myc in cohesin-deficient ES cells growing in 2i media was in striking contrast with the previously reported positive correlation between cohesin and *Myc* gene expression (Rhodes et al., 2010; McEwan et al., 2012). Consequently, the role of cohesin in regulation of *Myc* expression was further analysed in various different cellular systems. Interestingly, in ES cells the impact of cohesin depletion on *Myc* expression varied depending upon culture conditions. The upregulation of *Myc* in cohesin-deficient ES cells was seen only when they were cultured in 2i media. In serum + LIF culture conditions, cohesin depletion in ES cells of the same genetic background did not have any effect on *Myc* mRNA levels. These experiments show that the impact of cohesin depletion is not only cell-type specific but is also sensitive to extracellular signalling.

The analysis was extended to study the impact of cohesin depletion in non-dividing somatic cells. Very low amounts of *Myc* transcripts were present in resting preB cells, $CD4^+CD8^+$ double positive and CD4 single positive thymocytes. These levels of *Myc* remained unaffected by cohesin depletion. Experiments in serum starved MEFs showed that although steady state levels of Myc mRNA were reduced upon cohesin depletion, cohesin was essentially dispensable for activation of *Myc* transcription in response to stimulation by serum growth factors.

The above experiments in non-proliferating cells avoid secondary effects due to disruption of cell cycle functions in cohesin-deficient cells and collectively show that cohesin is not necessary for appropriate *Myc* transcription. It was also highlighted that impact of cohesin depletion on gene expression was determined in part by extracellular signals. And finally, this work suggests that cohesin might play an important role in the regulation of the lncRNA Pvt1 and may help explain the high incidence of cohesin mutations in human malignancies.

Chapter 5 : IFNγ mediated activation of MHC class II genes requires cohesin

Results from 3C experiments indicated that long-range enhancer-promoter interactions at the tested deregulated genes can be maintained even after cohesin depletion and are not enough to explain the observed changes in gene expression. This questions the prevailing model of cohesin mediated gene regulation and calls for alternative explanations to how loss of cohesin leads to changes in expression of genes associated with cohesinopathies. An insight into this problem can be gained by careful observation of patients with Cornelia de Lange syndrome who regularly show developmental defects in limb formation, a process which requires spatio-temporally coordinated activation of Hox genes (Mallo and Alonso, 2013). One might surmise that cohesin is important for the limb development programme. But in the larger context, one can hypothesise that cohesin is required for the establishment of a new gene expression programme. Hence, this chapter focuses on testing the hypothesis that cohesin is necessary for the process of gene activation and in exploring the various components of the transcriptional machinery associated with this process.

I decided to use the IFNγ inducible expression of MHC genes in fibroblasts as the model system for this analysis. As discussed earlier, MHC class I molecules are expressed on the surface of most vertebrate cells and are over-expressed in response to IFNγ exposure. MHC II gene expression, on the other hand, is restricted to professional antigen presenting cells. However, most cells can be stimulated to express MHC II genes when exposed to IFNγ. Therefore, using IFNγ to induce MHC expression would potentially allow me to delineate the requirement of cohesin for induction of previously expressed genes (MHC class I) as opposed to that for the initiation of new gene expression (MHC class II).

Mouse embryonic fibroblasts (MEFs) isolated from *Rosa26*$^{ERT2Cre/ERT2Cre}$*Rad21*$^{WT/WT}$ (ERT2Cre-*Rad21*$^{WT/WT}$) or *Rosa26*$^{ERT2Cre/ERT2Cre}$*Rad21*$^{lox/lox}$ (ERT2Cre-*Rad21*$^{lox/lox}$) embryos were used as the cellular model for these studies as they do not express MHC II molecules on cell surface unless stimulated by IFNγ. Moreover, MEFs can be arrested in a quiescent state (G0) by serum starvation and thus provide an ideal system where effects of cohesin depletion on gene expression can be studied in non-dividing cells.

5.1 IFNγ stimulated activation of MHC class II genes in fibroblasts requires cohesin

Using MEFs extracted from ERT2Cre-*Rad21*$^{WT/WT}$ or ERT2Cre-*Rad21*$^{lox/lox}$ embryos, a cohesin depletion system was setup as depicted in Figure 5.1A, which allowed for gene expression analysis upon IFNγ induction in non-dividing cells. Briefly, early passage MEFs were grown in serum containing media (DMEM + 10% FCS) and treated with 100nM 4'OHT for 24 hours. Cells were washed with PBS and seeded with fresh serum media for another 24 hours after which the cells were serum starved (DMEM + 0.1% FCS) for 6 days. At this point (d0 or 0 hour) cells were induced with 100U/ml IFNγ and samples were collected after 6 hours, 24 hrs (d1) and 48 hrs (d2).

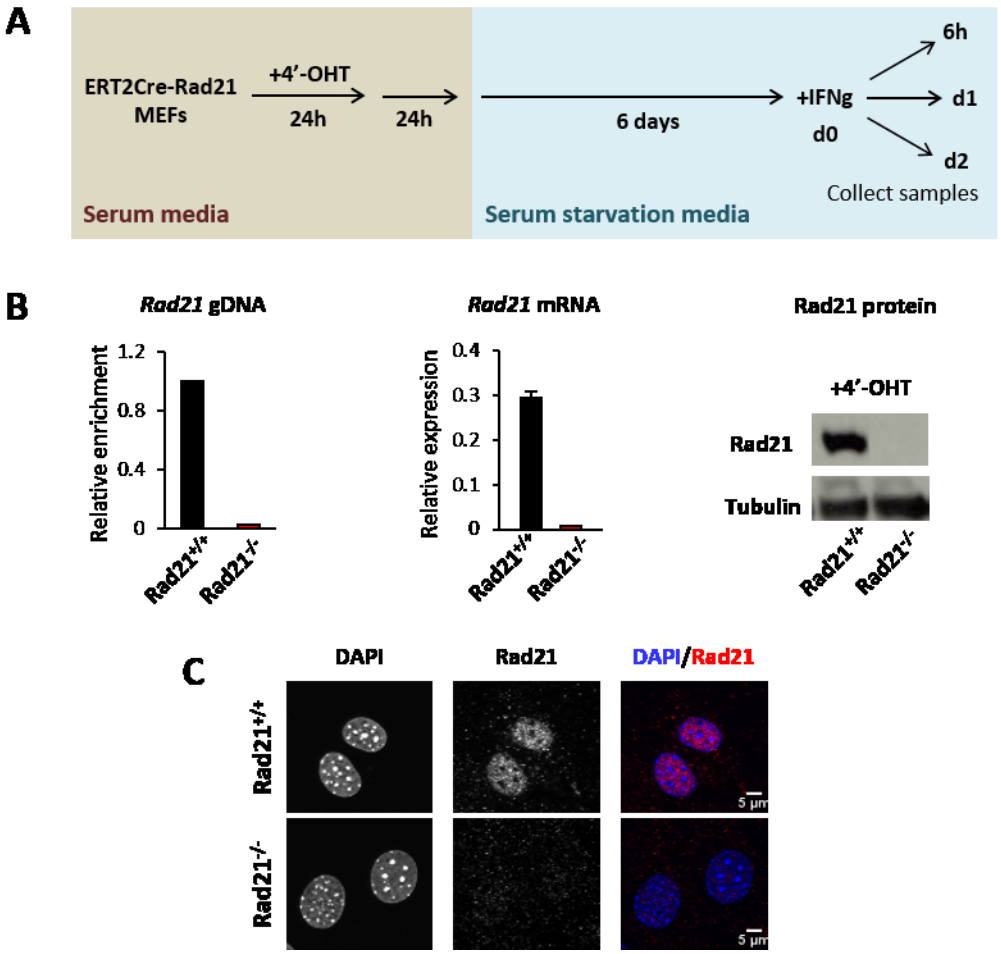


Figure 5.1 Serum starved ERT2Cre-*Rad21* MEFs can be efficiently depleted of cohesin. A) Schematic outline of the experimental setup - MEFs were treated with 100nM 4'-OHT while growing in serum media for 24h. Cells were then supplemented with fresh serum media for additional 24h before being left in serum starvation media for 6 days. These cells were then induced with 100U/ml IFNγ. **B)** Analysis of genomic DNA depletion of the *Rad21* allele, *Rad21* mRNA (normalised to *Ubc, Ywhaz*) and Rad21 protein in tamoxifen treated MEFs. **C)** Rad21 protein depletion was confirmed by IF.

Tamoxifen-treated serum starved ERT2Cre-*Rad21*$^{lox/lox}$ MEFs at day0 (*Rad21*$^{-/-}$ MEF) showed complete loss of the loxed *Rad21* allele, *Rad21* mRNA as compared to tamoxifen-treated ERT2Cre-*Rad21*$^{WT/WT}$ MEFs (*Rad21*$^{+/+}$ MEF) (Figure 5.1B). RAD21 protein was not detectable in Rad21$^{-/-}$ MEFs by Western and IF (Figure 5.1B, C). Importantly, both cohesin wildtype and cohesin knockout MEFs were arrested in G1 (or G0) phase upon serum starvation as compared to cycling MEFs growing in serum media as assessed by propidium iodide staining (Figure 5.2A). Also, there was no indication of cellular stress as determined by p53, p21 and *Mdm2* mRNA expression. Specifically, *Mdm2* expression was not elevated upon cohesin depletion and was similar to wildtype levels (Figure 5.2B). Thus this setup allows us to efficiently deplete cohesin and study its requirement for gene activation in non-dividing cells while avoiding any secondary effects due to activation of stress response in the absence of cohesin.

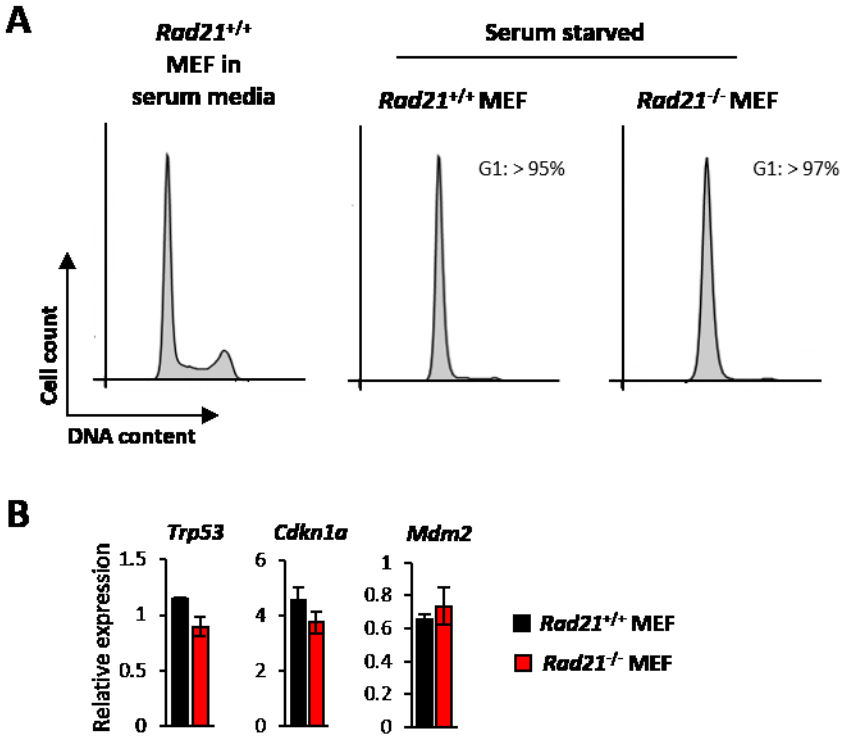


Figure 5.2 Serum starved cohesin deficient MEFs are arrested in G1. A) Propidium iodide staining. MEFs growing in serum media are actively cycling but become arrested in G1 (or G0) phase upon serum starvation and stop dividing. **B)** qRT-PCR analysis of wildtype and cohesin deficient serum starved MEFs for the expression of stress response genes. *Trp53* (p53), *Cdkn1a* (p21) and *Mdm2* expression is not elevated upon cohesin depletion and is maintained at wildtype levels.

Using the above setup, I next treated wildtype and cohesin deficient MEFs with IFNγ and checked the expression of MHC genes by qRT-PCR analysis at different time points. The induction of MHC class I gene *H2-K* was moderately affected in cohesin deficient MEFs as compared to wildtype MEFs while the expression of *H2-D* was similar to that observed in wildtype MEFs. However, the MHC class II genes – *H2-Aα* and *H2-Aβ,* failed to be induced in the absence of cohesin (Figure 5.3).

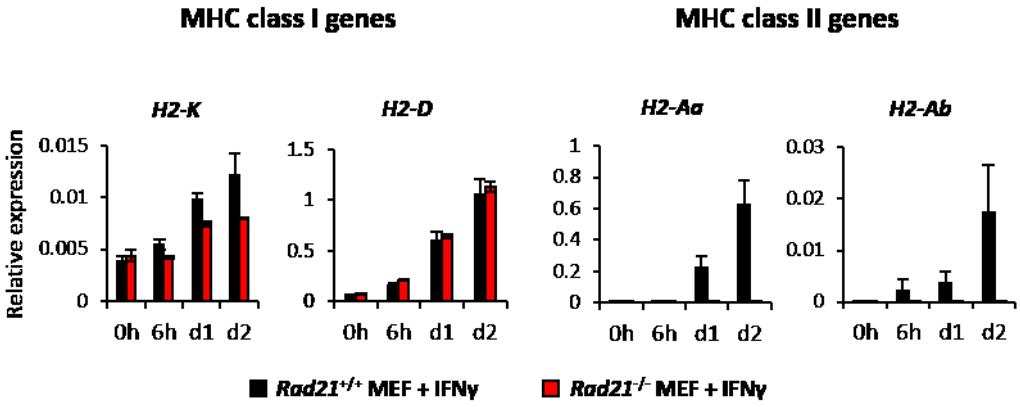


Figure 5.3 MHC class II genes fail to be induced by IFNγ in the absence of cohesin. qRT-PCR analysis of for MHC genes in wildtype and cohesin deficient serum starved MEFs at 0h and 6h, 24h (d1), 48h (d2) post IFNγ treatment. MHC class II genes failed to express in the absence of cohesin but MHC class I genes remained unaffected (Normalised to *Ubc, Ywhaz*; mean ± SE; n=3).

This observation was further confirmed by analysing the cell surface presentation of MHC molecules in MEFs upon IFNγ treatment using FACS analysis (Figure 5.4). Uninduced serum starved MEFs were used as negative controls. As the ERT2Cre-*Rad21* mice were C57BL/6 background with haplotype b, antibodies against haplotype d and k were used as isotype controls. H-2D^{b} MHC class I molecules were efficiently presented on the cell surface in both wildtype and cohesin deficient MEFs. However, MHC class II presentation of I-A^{b} molecules was severely reduced in cohesin deficient cells where the levels were close to that observed in untreated cells.

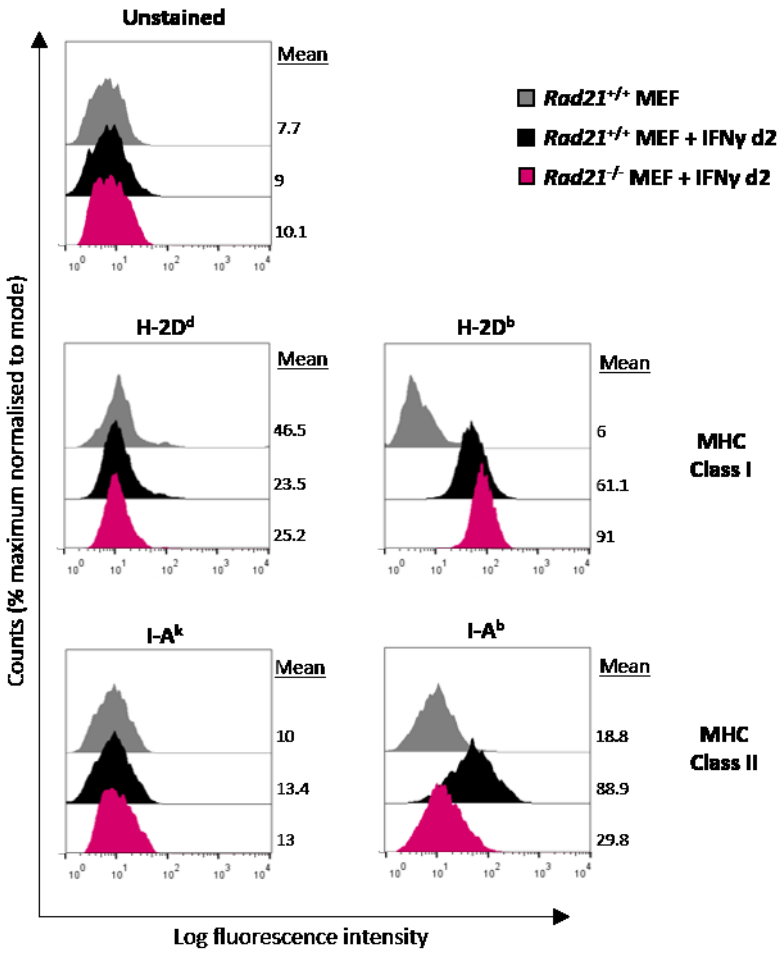


Figure 5.4 FACS analysis for MHC presentation on MEF cell surface upon IFNγ induction. Uninduced or 48h IFNγ induced wildtype and cohesin deficient MEFs were stained with isotype specific antibodies against MHC class I and class II molecules and subjected to FACs analysis. Number of cells and the observed fluorescence intensity are plotted. Mean fluorescence intensity is indicated by the side of each sample. Middle panel: Staining for MHC class I H-2D molecules. H-2D^{d} was used as isoptype control. H-2D^{b} expression is similar in *Rad21*$^{+/+}$ and *Rad21*$^{-/-}$ MEFs. Lower panel: Staining for MHC class II molecules – I-A^{k} (or *H-2A*^{k}) (isotype control), I-A^{b} (or *H-2A*^{b}). Rad21 KO MEFs displayed negligible amounts of MHC class II molecules.

The observed MHC II activation defect could be the result of a direct requirement of cohesin for the expression of these genes, or could even result from the lack of another downstream factor involved in IFNγ induction. So, I decided to check the expression of the factors known to be required in the IFNγ induction pathway. Almost all these accessory factors, *Stat1, Tap1, Rfx1, Rfx5, Irf1 and Irf2* were induced at wildtype levels and were not affected by cohesin depletion (Figure5.5).

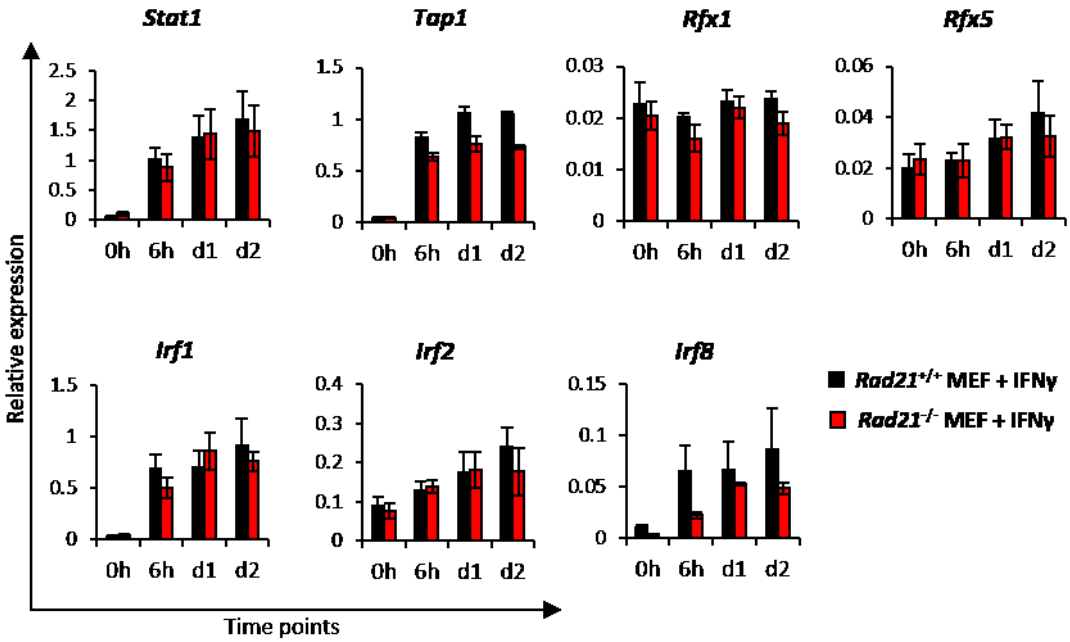


Figure 5.5 Expression of accessory factors remains unaffected by cohesin depletion. qRT-PCR analysis of WT and Rad21 KO MEFs induced with IFNγ for the shown genes (Normalised to *Ubc, Ywhaz*; mean±SE; n=3)

The expression of MHC class II genes, additionally requires the master regulatory transcription factor CIITA. qRT-PCR analysis revealed that *Ciita* specifically failed to be activated in the absence of cohesin. The reduced expression of *Ciita* was a transcriptional defect as assessed by quantification of primary *Ciita* transcripts (Figure 5.6)

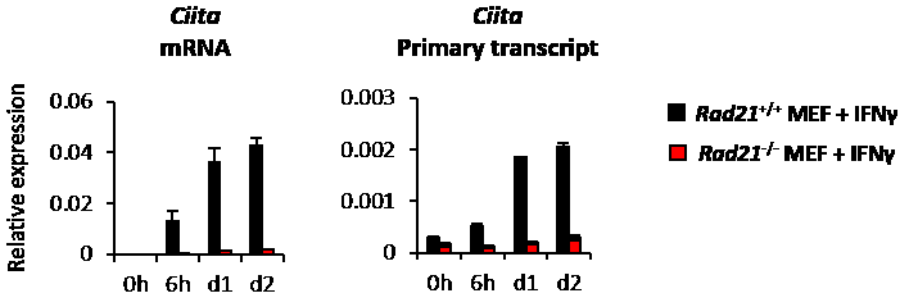


Figure 5.6 *Ciita* expression is abrogated in the absence of cohesin. qRT-PCR analysis of WT and Rad21 KO MEFs induced with IFNγ. mRNA levels were assessed by designing primers binding in two successive exons while primary transcript primers spanned an exon and the subsequent intron (Normalised to *Ubc, Ywhaz*; mean±SE; n=3).

As *Ciita* expression can be repressed by factors like Prdm1 (Blimp-1), TGFβ and SOCS1, it was tested if the observed defect in *Ciita* induction was due to the upregulation of these inhibitory factors. qPCR analysis did not indicate any increase in the expression of these factors in cohesin depleted cells above those observed in wildtype cells (Figure 5.7).

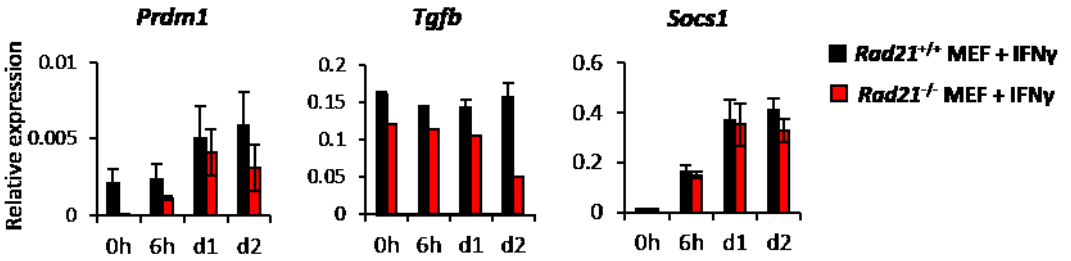


Figure 5.7 *Ciita* inhibitory factors are not overexpressed in cohesin deficient MEFs. qRT-PCR analysis of wildtype and cohesin depleted MEFs induced with IFNγ (Normalised to *Ubc, Ywhaz*; mean ± SE; n=3).

Since, all the factors regulating *Ciita* expression are normally expressed (Figure 5.5) and none of the known negative regulators of *Ciita* are overexpressed (Figure 5.7) in cohesin depleted cells, it can be concluded that cohesin is required for the activation of *Ciita* expression.

To verify if depletion of other subunits of the cohesin complex had the same effect on MHC II activation, I made use of MEFs that can be conditionally depleted of *Smc3*. Experiments were performed on the ERT2Cre-*Smc3* MEFs as described previously in Figure 5.1A. Tamoxifen-treated ERT2Cre-*Smc3*$^{lox/lox}$ MEFs showed efficient depletion of both *Smc3* mRNA and protein as compared to tamoxifen-treated wildtype MEFs (Figure 5.8B, C). The absence of Smc3 had the same effect as that of Rad21 loss and MEFs failed to induce *Ciita* and MHC class II genes while the MHC class I genes and accessory factors were induced at normal levels (Figure 5.7D, E, F).

Altogether, these experiments provide compelling evidence that MHC class I genes and the accessory factors which are already expressed at basal level before IFNγ induction, do not require cohesin for increased expression. However, cohesin is essential for the activation of previously silent master regulator *Ciita* and MHC class II genes in response to IFNγ treatment.

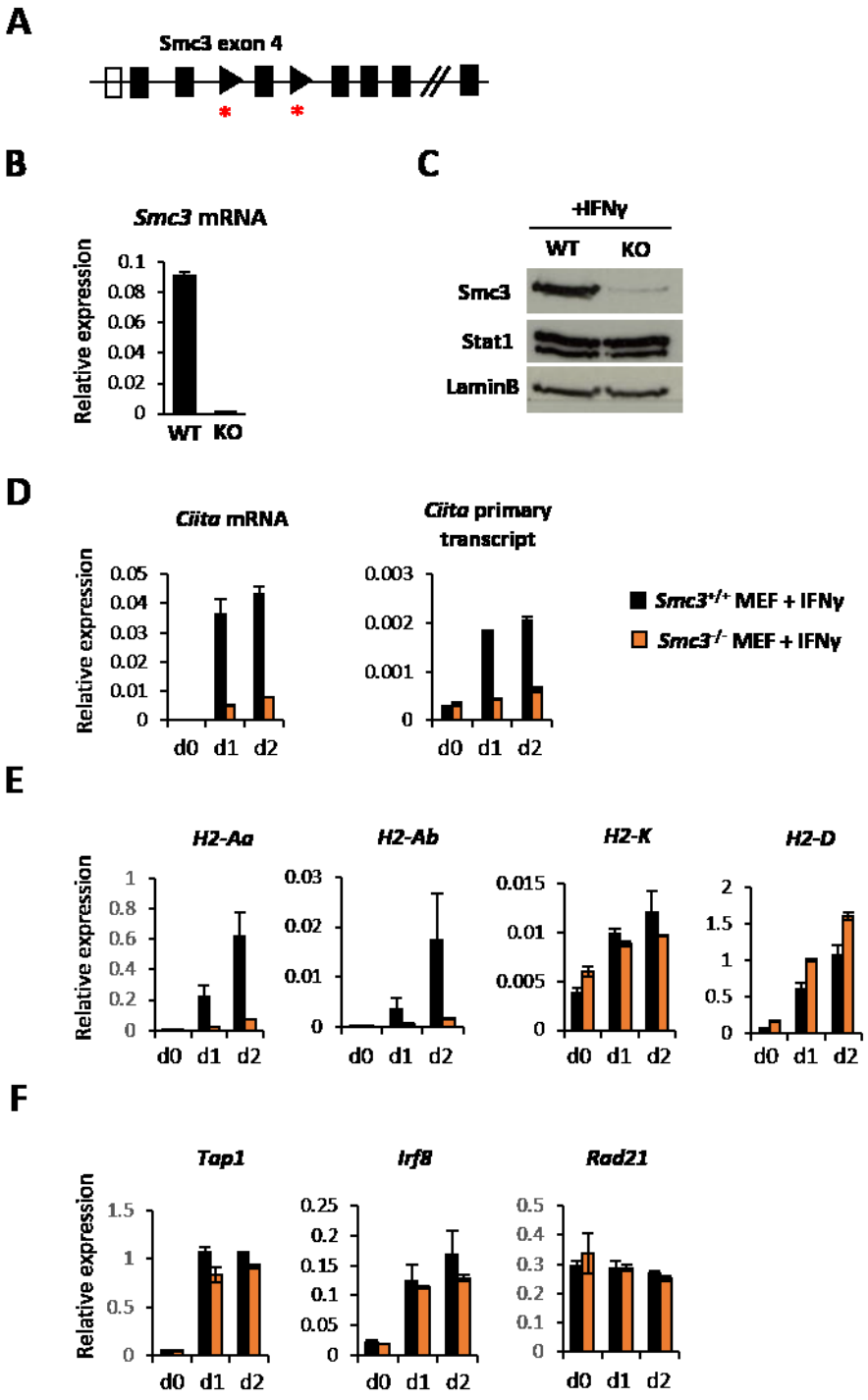


Figure 5.8 Cohesin subunit Smc3 is essential for the activation of *Ciita* and MHC class II genes upon IFNγ induction. A) Schematic of the conditional Smc3 allele with loxP sites flanking exon 4. **B) and C)** Efficient *Smc3* mRNA and protein depletion in serum starved MEFs respectively. **D), E) and F**) qRT-PCR analysis for the expression of *Ciita*, MHC genes and associated factors respectively in response to IFNγ treatment (Normalised to *Ubc, Ywhaz*; mean ± SE; n=2). Similar to Rad21 depletion, activation of *Ciita* and MHC class II genes was severely affected.

5.2 HDAC8 inhibition mimics the effect of cohesin deficiency on *Ciita* induction

Smc3 acetylation promotes stable loading of cohesin complexes onto DNA by counteracting the anti-establishment action of Pds5-Wapl (Unal et al., 2008; Zhang et al., 2008). Once cohesin is released from chromatin, Smc3 is deacetylated by HDAC8. This is a critical step in the cohesin cycle (Figure 1.2) and is required for reusing the pool of cohesin unloaded from chromatin by the prophase pathway (Beckouët et al., 2010; Borges et al., 2010). If HDAC8 activity is inhibited, acetylated levels of Smc3 increase in the cell (Deardorff et al., 2012a). This would essentially diminish the pool of 'usable' cohesin available for loading onto chromatin and in effect create a cellular state depleted of cohesin. This phenomenon can potentially explain how mutations in HDAC8 lead to a CdLS phenotype (Deardorff et al., 2012a). However, the role of HDAC8 and Smc3 acetylation in gene regulation remains to be explored further. Here, I hypothesize that using a specific inhibitor for HDAC8 would have the same effect on IFNγ mediated MHC class II induction, as would cohesin depletion.

In order to test this hypothesis, I treated serum starved MEFs with 25µM PCI (selective inhibitor for HDAC8) for 8 hours before inducing them with IFNγ for 24 hours (Figure 5.8A). Cells treated with a low (20nM) and a high (100nM) concentration of Trichostatin A (TSA), which inhibits class I and class II mammalian HDACs, were used as a control to assess the general impact of increased acetylation on gene expression. Cells treated with the HDAC8 inhibitor selectively showed reduced induction of *Ciita* and MHC class II genes – *H2-Aα* and *H2-Aβ.* PCI treated MEFs also showed a slight increase in the activation of MHC class I genes – *H2-K* and *H2-D*, and the accessory factors *Stat1, Tap1, Irf1*. In contrast, TSA treatment lead to an overall decrease in the induction of *Ciita* and both MHC class I and MHC class II genes along with reduction in *Stat1* and *Tap1* expression (Figure 5.9).

These results show that HDAC8 inhibition specifically impaired *Ciita* and MHC class II gene activation while only moderately enhancing the expression of MHC class I genes and accessory factors. On the other hand, TSA treatment resulted in a ubiquitously reduced expression of IFNγ induced genes. Further experiments are required to verify these findings and to understand the mechanistic details. One possible explanation is that inhibition of HDAC8 causes increased Smc3 acetylation and a consequent reduction of the 'usable' pool of Smc3 required for initiating the expression of *Ciita* and MHC class II genes. At active genes where Smc3 is already present, the

inhibition of the HDAC8 deactylation activity favours cohesin association with DNA and supports the expression of pre-existing gene transcription. Overall these experiments support the previous results that cohesin is required for appropriate induction of *Ciita* and MHC class II genes in response to IFNγ stimulation.

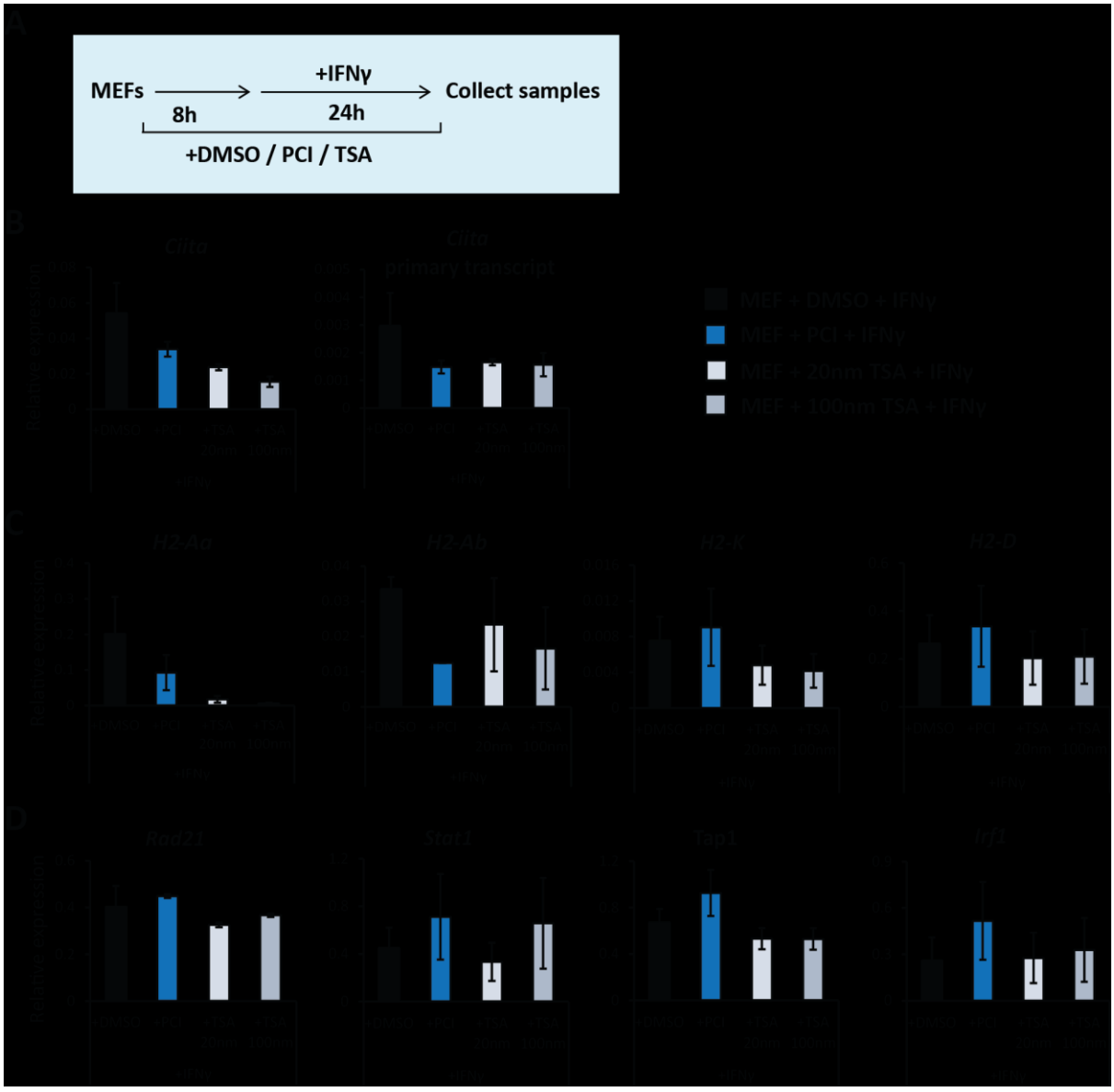


Figure 5.9 HDAC8 inhibition specifically impairs *Ciita* and MHC class II gene expression. A) Experimental setup: MEFs were treated with either DMSO, 25μM PCI (HDAC8 inhibitor) or with 20nM and 100nM TSA and induced with IFNγ 8h afterwards, for 24h when samples were collected for RNA extraction. **B), C) and D)** qRT-PCR analysis of Ciita, MHC and associated gene expression respectively. PCI treatment specifically impaired Ciita and MHC II expression. TSA treatment, on the other hand, lead to an overall decrease in the expression of IFNγ induced genes (Normalised to *Ubc, Ywhaz*; mean ± SE; n=3).

5.3 Cohesin depletion impairs the progression but not recruitment of RNA polymerase during activation of gene transcription

Having ascertained that cohesin is required for IFNγ mediated activation of *Ciita*, I next sought to elucidate the mechanisms by which cohesin controls *Ciita* activation. Given the role of cohesin in mediating looping between cis-regulatory elements, I first explored the chromatin landscape around the *Ciita* locus. I analysed published ChIP-sequencing datasets to identify possible enhancers for *Ciita* expression. Because little information was available for chromatin signatures on IFNγ activated non-classical MHC class II presentation, information was collected from uninduced MEFs and mES cells – both of which do not express MHC class II genes. This was compared to datasets from cells of the B cell lineage and IFNγ induced macrophages, which do express *Ciita* and MHC class II molecules. In addition, data on cohesin and CTCF binding, enhancer-associated chromatin modifications – H3K4me1, H3K27ac, Nipbl, p300 binding and specific transcription factors like Stat1 was compiled and assessed. ChIA-PET experiments which combined Polii ChIP with 3C technology were also used to evaluate DNA interactions in ES cells and B cells at the *Ciita* locus (Figure 5.10).

This helped to identify conserved CTCF binding sites which were found to be associated with CTCF in ES cells, MEFs and pre-B cells and will henceforth be called CTCF1, CTCF2 and CTCF3 peaks. As expected, *Ciita* promoters were not associated with any Polii and H3K4me3 modifications in MEFs nor did they show any association with CTCF. Macrophages showed STAT1 (a pioneer transcription factor) binding associated with increased H3K27ac and H3K4me1 marks in response to IFNγ activation at the regions demarcated as enhancer 2 (Enh2) and enhancer 3 (Enh3). These modifications were specifically enriched in activated macrophages at the putative enhancer regions and were absent or diminished in both MEFs and ES cells, indicating their possible role in *Ciita* expression in macrophages. B cells also did not show enhancer associated modifications at the Enh2 and Enh3 regions but Nipbl and p300 binding was instead observed at the region demarcated Enh4. Closer inspection revealed that Enh4 was not enriched for enhancer-associated marks in macrophages suggesting that Enh4 might act as a B cell-specific enhancer. This idea is also supported by the presence of DNA interactions between Enh4 and Ciita gene body as judged from the presence of ChIA-PET signals in B cells and their absence in ES cells. Thus this analysis revealed that *Ciita* expression is not only controlled by cell-type specific promoters but also probably by previously under-appreciated action of cell-type specific enhancers.

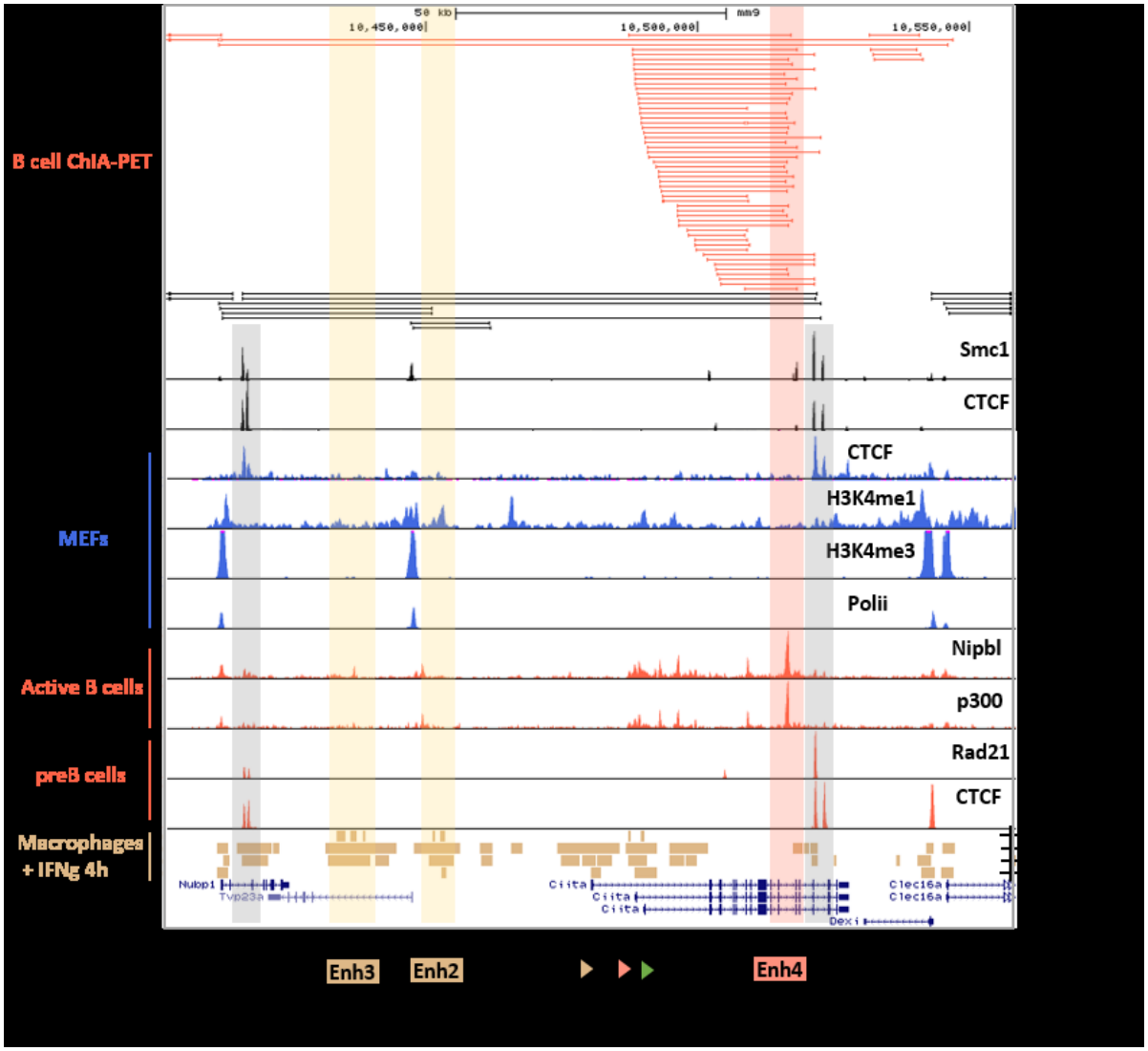


Figure 5.10 Localisation of relevant proteins and histone modifications along with DNA interactions at the *Ciita* locus. Top - Polii ChIA-PET DNA interactions *Ciita* expressing B cells and non-expressing ES cells. Below – illustration for the binding of Smc1, Rad21, CTCF, Polii, H3K4me3 and enhancer associated marks – H3K4me1, p300, Nipbl, H3K27ac in ES cells, MEFs, B cell lineage and IFNγ treated macrophages. The grey regions highlight the conserved cohesin, CTCF binding sites across all cell types called as CTCF1, CTCF2 and CTCF3 peaks. In yellow are highlighted the putative enhancers specific to macrophages – Enh2 and Enh3. Region highlighted in red, demarcated as Enh4 has B cell specific enhancer features.

Using this information, I next tested if any of these regulatory elements were involved in IFNγ mediated induction of *Ciita* expression in MEFs. First, I analysed cohesin and CTCF binding at the *Ciita* locus in MEFs both pre- and post-induction (Figure 5.11). Low levels of cohesin binding were detected at *Tap1* and *H2-K1* TSS which remain unchanged upon induction. No cohesin was detected at the IFNγ responsive promoter IV of *Ciita* (similar to negative controls) in uninduced MEFs. However, cohesin binding was significantly increased at *Ciita* promoter IV upon IFNγ treatment. Cohesin was also found to be associated with the CTCF sites studied but more

experimental replicates are required to check if cohesin binding at these sites was significantly increased upon IFNγ treatment. Likewise, cohesin binding at Enh2 site did not show a significant increase upon induction. Enh3 (not shown) and Enh4, however, were not found to be associated with cohesin at all. Also, Rad21 ChIP in cohesin-deficient MEFs showed a uniformly abolished cohesin loading across all sites studied (Figure 5.11A).

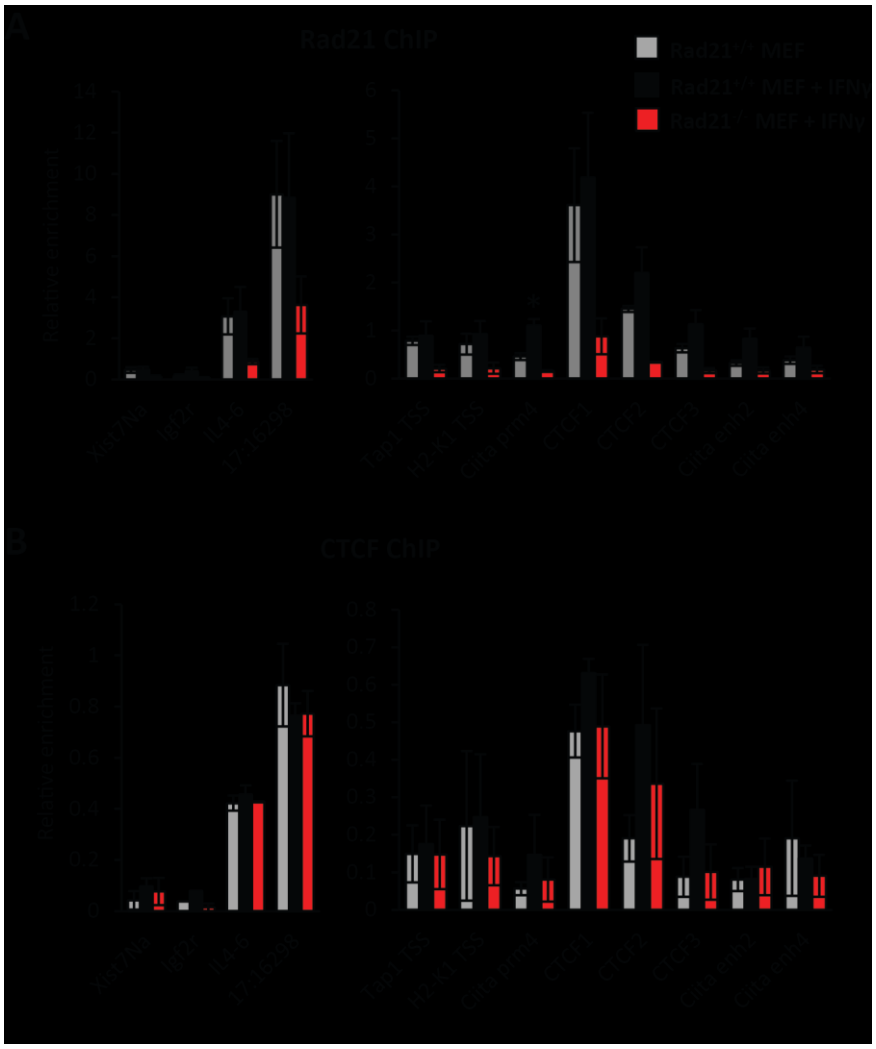


Figure 5.11 Cohesin is recruited to *Ciita* promoter IV in response to IFNγ treatment. A) and B) Rad21 and CTCF ChIP in uninduced, Rad21$^{+/+}$ IFNγ induced, Rad21$^{-/-}$ IFNγ induced MEFs. Left panels show enrichment in negative and positive controls. Right panel shows binding at *Tap1*, *H2-K1* TSS and *Ciita* locus. Enrichment is plotted relative to 1% input DNA. Rad21 binding is specifically increased at *Ciita* promoter IV (*Ciita* prm4) region.

Binding of CTCF was similar to that of Rad21 at the above tested sites. CTCF binding was also increased at *Ciita* promoter IV and the CTCF1,2,3 sites and more replicates are required to confirm this. In contrast to Rad21 binding in cohesin-deficient MEFs, CTCF binding largely remained unaffected by reduced cohesin levels (Figure 5.11B).

Overall, this analysis showed that cohesin binding was significantly increased at the *Ciita* promoter IV in response to IFNγ stimulation, and revealed other cohesin and CTCF associated sites with likely important roles in the regulation of *Ciita* expression.

One of the possibilities of how cohesin helps initiate expression of a gene is by mediating intra-chromosomal interactions between enhancers and promoter. To explore this prospect, I performed 3C experiments using *Ciita* promoter IV as the anchor and measured its interactions with the above described putative cis-regulatory sites and CTCF binding sites around the *Ciita* locus.

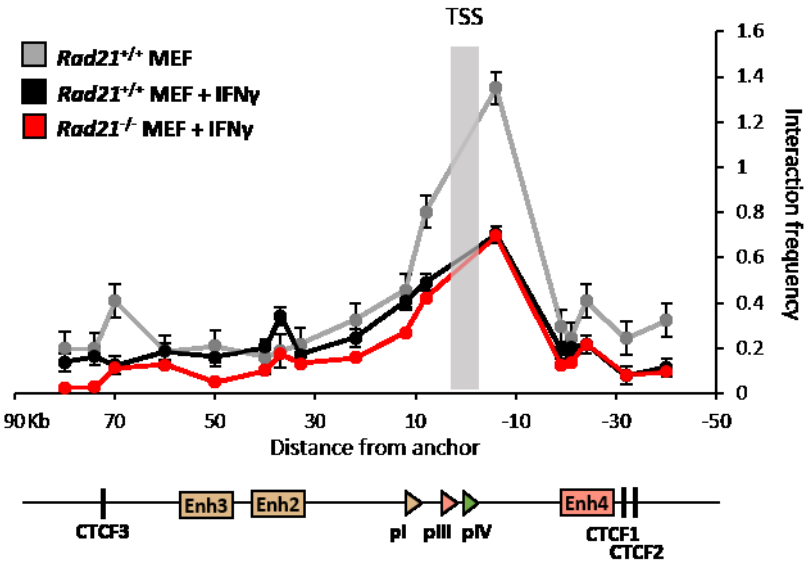


Figure 5.12 DNA interactions at the *Ciita* locus. 3C performed at the *Ciita* locus with pIV as the anchor in uninduced, Rad21$^{+/+}$ induced and Rad21$^{-/-}$ induced MEFs. Formaldehyde fixed chromatin was digested with NcoI and religated. Interactions were quantified by PCR after normalising to BAC templates. Results indicate a possible reduction in interaction with CTCF3 peak upon induction. Also WT induced MEF *Ciita* pIV possibly interacts with Enh2 (Mean ± SE; n=2) which is lost in Rad21 KO MEFs (n=1).

Preliminary results showed that the chromatin conformation around the *Ciita* promoter changes dramatically upon IFNγ induction in both wildtype and cohesin-deficient MEFs. It is likely that in the basal state, *Ciita* promoter IV interacts with the CTCF3 site. Possibly, this interaction is

lost after IFNγ induction and instead the Ciita promoter IV interacts with the Enh2 region. No evident interactions with the Enh3 and Enh4 regions were observed upon induction. IFNγ induced cohesin depleted cells also showed the loss of interaction with the CTCF3 site but no interactions at the Enh2 were observed (Figure 5.12). The differences in Ciita promoter IV interaction between wildtype and cohesin deficient cells could potentially help explain the cohesin-dependent induction defect. However, a more detailed analysis is required to be able to assess the importance of these interactions

Results from 3C experiments so far have repeatedly underscored the need to evaluate alternate roles of cohesin in mediating gene regulation. As cohesin is often found at the promoters of active genes (Kagey et al., 2010; Faure et al., 2012) and has been shown to be associated with subunits of the Mediator complex (Kagey et al., 2010), it is possible that cohesin may have a direct role in influencing RNA polymerase activity. Keeping this in mind, I next studied the effects of cohesin depletion on the steps of RNA polymerase recruitment and its licensing to successful initiation and elongation phases.

First, I assessed the H3K4me3 chromatin modification which marks the TSS of active genes. H3K4me3 mark was completely absent from *Ciita* promoters in uninduced MEFs where Ciita is silent. Upon IFNγ induction, H3K4 was selectively trimethylated at the promoter IV. The gain of H3K4me3 remained unaffected by cohesin depletion (Figure 5.13A), indicating that cohesin deficient cells received proper signals to initiate transcription and were appropriately primed to do so.

The next step was to test if RNA polymerase could be efficiently recruited to the primed promoter in the absence of cohesin. For this, I performed a ChIP with an antibody recognising the N-terminus of RNA Pol II. Results showed that Pol II recruitment was increased to similar levels in both wildtype and cohesin depleted MEFs induced with IFNγ at Tap1, H2-K1 TSS and specifically at *Ciita* promoter IV sites. Therefore, even though *Ciita* transcription was severely abrogated in cohesin-deficient cells, Pol II recruitment at promoter IV was not affected (Figure 5.13B).

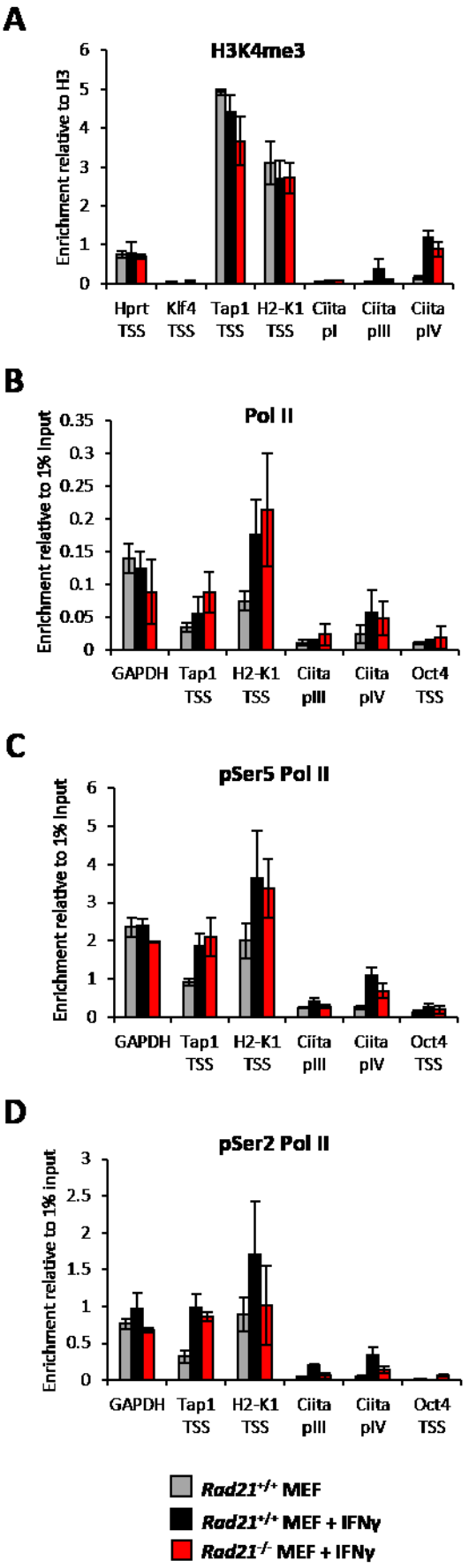


Figure 5.13 RNA polymerase is recruited to *Ciita* pIV in cohesin-depleted MEFs upon induction but elongation is impaired.

A) H3K4me3 ChIP. Enrichment is plotted relative to H3 binding. H3K4me3 modification is specifically gained at Ciita pIV upon IFNγ stimulation and remains unaffected by cohesin depletion.

B) Total Pol II ChIP with an antibody against the N-terminal. Enrichment is plotted relative to 1% input DNA. Pol II is also specifically binds to Ciita pIV upon IFNγ stimulation in both WT and Rad21 KO MEFs. Cohesin depletion does not affect Pol II recruitment.

C) and D) ChIP against the Ser5 and Ser2 modifications of the Pol II CTD representative of initiation and elongation state of Pol II. Enrichment is plotted relative to 1% input DNA. While Ser5 and Ser2 modifications at most sites tested remained unaffected by cohesin depletion, their incidence at Ciita pIV was reduced in cohesin deficient IFNγ induced MEFs. This indicates that the recruited Pol II could not initiate and elongate effectively.

(Mean ± SE; n=3).

But the presence of Pol II at the promoter is not enough to ensure successful transcription of a gene. The entry of Pol II into the initiation and elongation phase is also carefully regulated. The CTD of Pol II is phosphorylated at Ser5 and Ser2 during the initiation and elongation phase respectively (Saunders et al., 2006; Zhou et al., 2012). ChIP experiments were performed using antibodies against the Ser5 or Ser2 modified CTD of Pol II. Results showed that appropriate levels of Ser5 and Ser2 phosphorylated forms of Pol II were present at *Gapdh, Tap1* and *H2-K1* TSS, in both wildtype and cohesin deficient IFNγ treated MEFs. Specifically at the *Ciita* promoter IV, however, a decrease was observed in the binding of the Ser5, Ser2 modified Pol II in the absence of cohesin (Figure 5.13C, D).

Thus, these experiments indicate that cohesin depletion did not affect the recruitment of Pol II in response to IFNγ treatment but prevented the progression of Pol II into the initiation and elongation phase which may account for the observed abrogation of *Ciita* expression.

5.4 *Paxip1* deficient fibroblast-like cells also show reduced Ciita induction upon IFNγ stimulation

It is possible that cohesin impacts enhancer function by more than being a mediator of enhancer-promoter interactions. Perhaps, it is also important for the designation and activation of enhancer activity, a process which is associated with histone modifications like H3K4me1 and H3K27ac (Creyghton et al., 2010; Rada-Iglesias et al., 2011). The deposition of H3K4me1 marks is attributed to the methyltransferase activity of the MLL3/4 complexes (Lee et al., 2013). Given the phenotypic overlap between the impact of cohesin depletion with that of the depletion of MLL3/4 subunit - PTIP at the *Igh* locus (Daniel et al., 2010), along with the similarity between CdLS and Kabuki phenotype, as discussed earlier (Section 1.3.4), it can be speculated that cohesin functions through modulating the activity of the MLL3/4 complex. One can hypothesize that if cohesin functioned through the MLL3/4 complex, the absence of MLL3/4 complex would have the same effect on IFNγ induction as that of cohesin depletion.

In order to test this, I obtained ES cells with a floxed *Paxip1* allele (as described in Kim et al., 2009) where *Paxip1* – a component of the MLL3/4 complex, can be deleted by transfecting with Cre recombinase. As ES cells failed to induce *Ciita* expression upon IFNγ treatment (not shown) as previously reported (Morris et al., 2000; van den Elsen et al., 2000), these ES cells were differentiated into fibroblast-like cells. Several strategies for differentiation were tested. Cells remained immune to IFNγ induction even 10 days after embryoid body formation post LIF removal from the media. Differentiating these ES cells as a monolayer on gelatin in the absence of LIF for extended periods of time also did not alleviate this immunity. Treatment of these monolayer differentiated ES cells with factors like Activin A or Fgf2, known to promote differentiation into ectodermal lineages, also did not yield cells that could be induced to express *Ciita* by IFNγ treatment (not shown). Eventually, treatment of monolayer differentiated ES cells with high concentrations of Retinoic acid for more than 4 days gave rise to population of fibroblast-like cells which could be induced with IFNγ to express *Ciita* and MHC class II genes. These fibroblast-like cells were then used for further analysis.

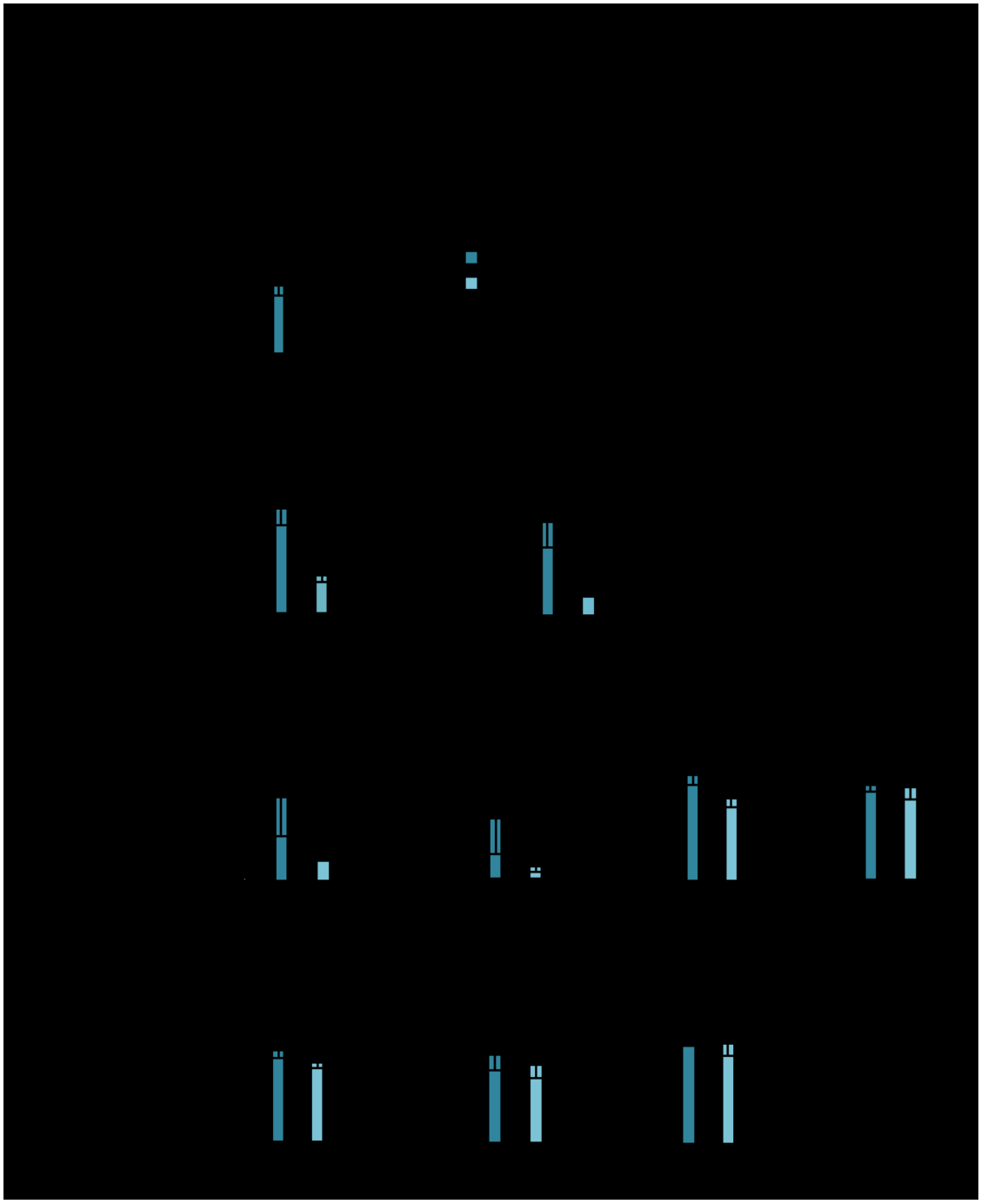


Figure 5.14 *Paxip1* is required for optimal activation of Ciita and MHC class II genes in response to IFNγ. **A)** Schematic experimental setup: ES cells were differentiated into fibroblast-like cells by LIF withdrawal and subsequent retinoic acid treatment. These cells were then transfected with either empty vector-GFP (mig) or Cre-GFP vector (cre). Sorted cells were IFNγ induced for 48h. **B)** qRT-PCR analysis showed efficient Paxip1 mRNA depletion. Ciita and MHC class II gene activation was impaired upon Paxip1 depletion but MHC class I and other genes remained unaffected (Normalised to *Ubc, Ywhaz*; mean ± SE; n=3).

Briefly, the *Paxip1*$^{fl/fl}$ ES cells growing in serum media were allowed to differentiate by LIF withdrawal into a monolayer on gelatin for 4 days. These cells were trypsinised and replated sparsely onto gelatin for 10 days in ES serum media supplemented with 1µM Retinoic acid without LIF. The cells were then transfected with either the retroviral MSCV-IRES-Cre-GFP vector or the empty MSCV-IRES-GFP (MIG) vector as control. Cells were sorted 5 days post transfection based on GFP expression. These cells were replated and induced with 100U/ml of IFNγ for 48 hours before collecting for analysis (Figure 5.14A).

qRT-PCR analysis on the collected samples revealed that *Paxip1* mRNA was efficiently depleted upon Cre transfection. These *Paxip1* depleted fibroblast-like cells also showed a significantly reduced induction of *Ciita* and MHC class II genes while MHC class I gene activation remained unaffected (Figure 5.14B). Thus, the effects of *Paxip1* depletion on IFNγ stimulation mirrored those observed upon cohesin depletion and HDAC8 inhibition in being selectively restrictive to *Ciita* and MHC class II gene expression. This shows that PTIP subunit is required for the proper activation of these genes and suggests that cohesin might function through the action of the PTIP subunit of the MLL3/4 complex.

5.5 Knockout of MLL3 and MLL4 subunits in preadipocytes does not impair *Ciita* induction

To better understand the role of MLL3/4 complex in IFNγ mediated gene activation, I used SV40 immortalised preadipocytes from *Mll3*$^{-/-}$*Mll4*$^{flox/flox}$ mouse. Preadipocytes are not classical antigen presenting cells, thus their response to IFNγ stimulation should be similar to that observed in fibroblasts. *Mll3/Mll4* double knockout was achieved by stable transduction with adenoviral Cre (Ad-Cre) vector. Adenoviral GFP (Ad-GFP) transfected preadipocytes served as a control (as described in Lee et al., 2013). The Ad-GFP or Ad-Cre transfected *Mll3*$^{-/-}$*Mll4*$^{flox/flox}$ preadipocytes were induced with IFNγ for 48 hours and gene expression changes were analysed by qRT-PCR. Interestingly, and in contrast to deletion of the PTIP subunit of the MLL3/4 complex, *Mll3/Mll4* double knockout itself did not impair the activation of *Ciita* and MHC genes (Figure 5.15).

These results suggest that cohesin mediated gene activation requires the activity of specific subunits of the MLL3/MLL4 complex like PTIP. Other components like the MLL3 and MLL4 subunits, are not essentially required.

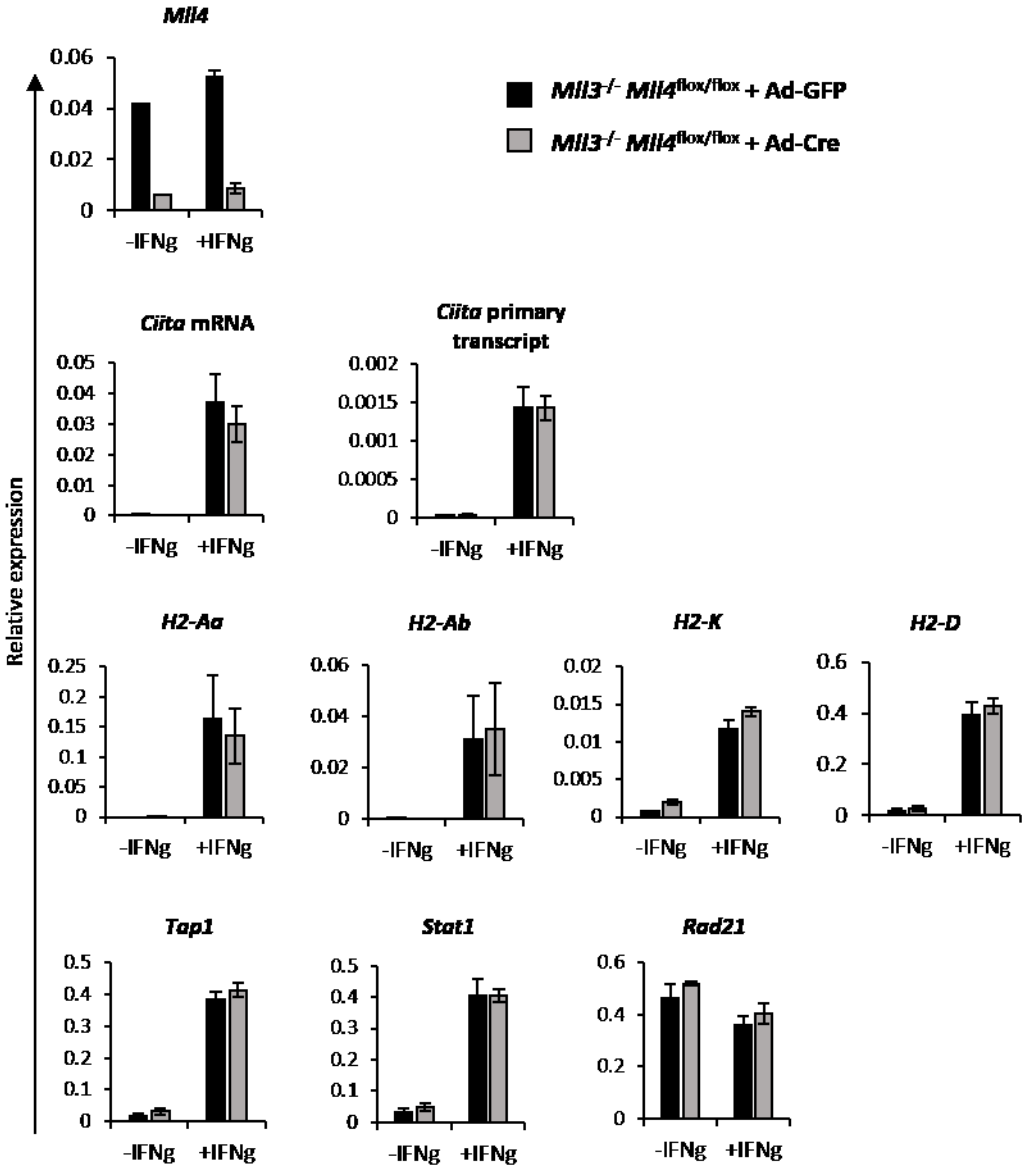


Figure 5.15 *Mll3/Mll4* double knockout does not impair IFNγ induction. Ad-Cre or Ad-GFP transfected *Mll3*$^{-/-}$*Mll4*$^{fl/fl}$ preadipocytes were induced with IFNγ for 48h and samples were collected for RNA analysis. qRT-PCR results show that *Mll3/Mll4* double knockout did not affect the induction of *Ciita*, MHC and other genes tested (Normalised to *Ubc, Ywhaz*; mean ± SE; n=3).

5.6 *Ciita* and MHC class II gene expression can be rescued by TSA treatment in cohesin deficient fibrobalsts

Activation of *Ciita* expression from pIV is accompanied by increased histone acetylation (Morris et al., 2002), and loss of *Ciita* expression is associated with decreased acetylation (Majumder et al., 2008). Thus, it is possible that absence of cohesin impairs the critical process of histone acetylation required for *Ciita* activation. If that is the case, an increased histone acetylation should help to recover from the effect of cohesin deficiency.

To test this hypothesis, I pharmacologically inhibited class I and class II HDACs by using TSA which increases the global acetylation levels. First, serum starved MEFs were treated with 20nM and 100nM TSA alone for 24h to control for any effects of increased acetylation in the absence of IFNγ stimulation (Figure 5.16A top). Notably, TSA treatment alone did not activate the expression of *Ciita* and MHC class II genes. For the rescue experiment, ERT2Cre-*Rad21* MEFs were treated with tamoxifen and serum starved as described previously in Figure 5.1A. Wildtype and cohesin depleted MEFs thus obtained were induced with 100U/ml of IFNγ. 6h post IFNγ induction, cells were treated with DMSO or 20nM TSA or 100nM TSA and samples were collected at the end of 24h (Figure 5.16A bottom). As observed from the qRT-PCR analysis, TSA treatment post IFNγ induction did in fact increase the transcription of *Ciita* and MHC class II genes in cohesin deficient cells (Figure 5.16B).The higher the TSA concentration, the greater was the increase in expression of these genes. Even though TSA treatment did not affect the transcription of most genes in WT MEFs, it did reduce the expression of *Ciita* and *H2-Aa*. This can be explained by the observation that Ciita recruits HDACs to gene promoters to promote gene expression and therefore optimal expression is a result of a balance of acetylation and deacetylation activity of HATs and HDACs, respectively. Thus while TSA treatment increased *Ciita* and MHC class II gene expression in cohesin deficient cells, it inevitably lead to a decrease in their expression in WT cells.

Altogether, these experiments indicate that cohesin might also be playing an important role in modulating histone acetylation facilitating activation of transcription. In addition, they point towards the potential use of HDAC inhibitors to partially rescue the activation defect of genes in the absence of cohesin.

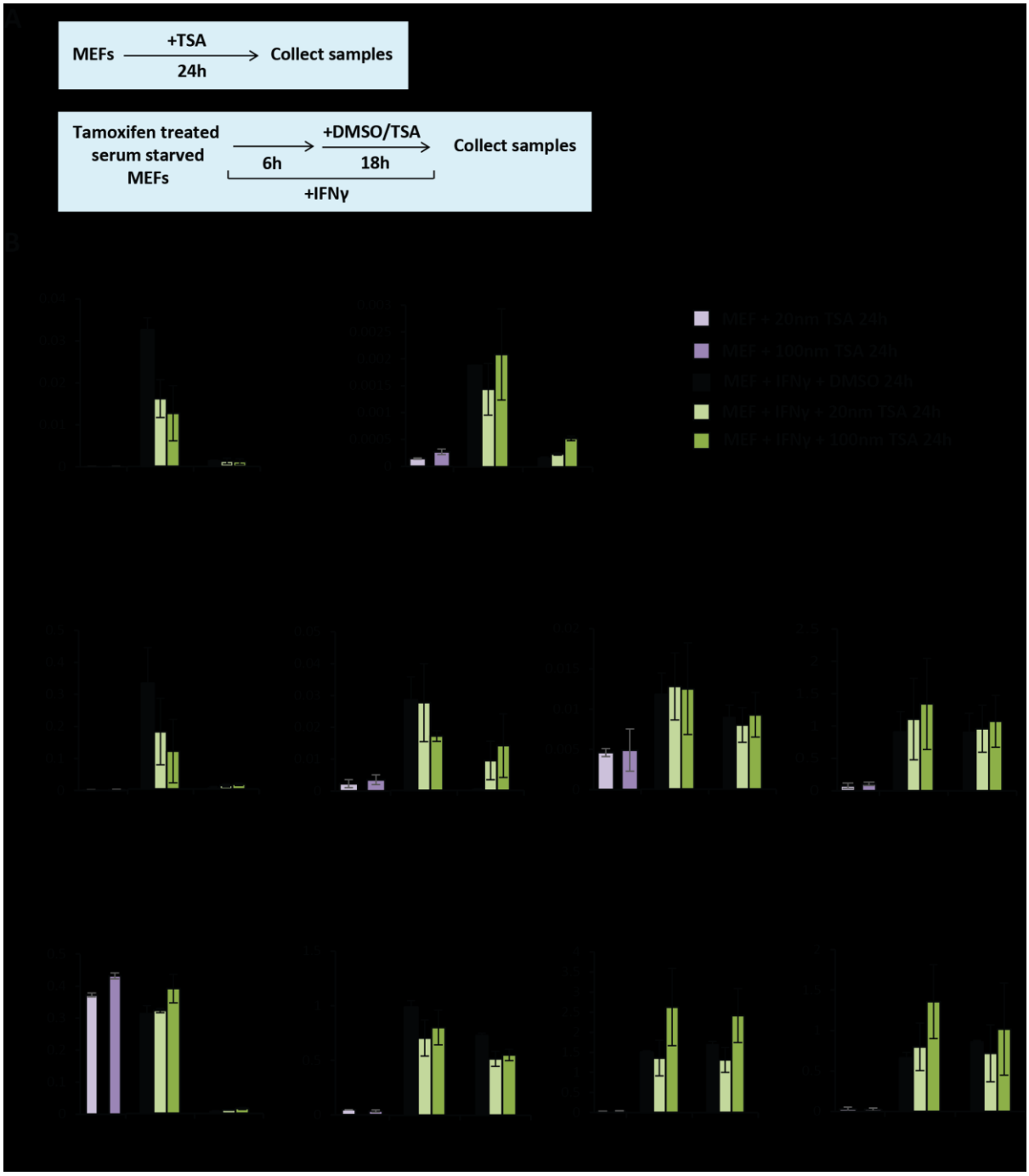


Figure 5.16 TSA can rescue cohesin-dependent *Ciita* induction defect. A) Experimental setup. Top: Serum starved MEFs treated with 20nM or 100nM TSA for 24h. Bottom: WT and Rad21 KO serum starved MEFs were treated with TSA 6h post IFNγ induction and RNA samples collected at the end of 24h. **B)** qRT-PCR analysis for the expression of Ciita, MHC and associated genes. TSA treatment increases transcription of *Ciita* and MHC class II genes in the absence of cohesin (Normalised to *Ubc, Ywhaz*; mean ± SE; n=2).

5.7 Discussion and future perspectives

The work presented in this chapter was aimed at investigating the role of cohesin in the activation of gene expression and focused on the study of MHC induction by IFNγ stimulation in serum starved MEFs. This system has two major advantages. Firstly, serum starvation arrests the MEFs in a G1 (or a quiescent G0) phase. This allowed me to analyse the impact of cohesin depletion on gene expression in non-dividing cells, thus avoiding secondary effects due to cell cycle stress. Secondly, MEFs express very low basal levels of MHC class I molecules and do not express MHC class II genes at all. But they can be stimulated by IFNγ to increase the expression of MHC class I genes as well as to induce the expression of MHC class II genes which is otherwise restricted to antigen-presenting cells only. Therefore, this system allowed me to study if cohesin was required for the activation of MHC genes in response to IFNγ in non-dividing cells.

The results clearly showed that MHC class II genes failed to be induced normally in the absence of cohesin. In contrast, the upregulation of MHC class I genes remained mostly unaffected by cohesin depletion. Detailed examination revealed that cohesin depletion abolished the induction of *Ciita*, the master regulator essential for the expression of MHC class II genes. Since all the regulatory factors known to be important for *Ciita* expression were appropriately expressed and none of the factors known to negatively impact *Ciita* expression were upregulated, it is likely that cohesin is directly involved in *Ciita* activation. At this stage, it is not clear whether the MHC class II induction defect is a direct effect of cohesin depletion or a consequence of the lack of Ciita. Further experiments based on ectopic expression of Ciita in cohesin deficient cells can help determine if cohesin is directly required for the activation of MHC class II genes in response to IFNγ stimulation. It is noteworthy that cohesin depletion did not affect the induction of all MHC genes. In fact, the genes which were already expressed at basal levels prior to IFNγ stimulation, like MHC class I genes and the associated regulatory factors, did not require cohesin for upregulation. On the contrary, MHC class II genes and *Ciita* are silent in MEFs and their promoters are not associated with RNA polymerase prior to IFNγ treatment. To test if the absence of pre-bound RNA polymerase was the determining factor for cohesin sensitivity, the expression of *Irf8* was analysed. IRF8 is another regulatory factor upregulated upon IFNγ stimulation, but like *Ciita* it also does not have bound RNA polymerase as seen from Pol II ChIP-sequencing data in MEFs. Unlike *Ciita*, *Irf8* expression could be induced in the absence of cohesin, albeit with a slower kinetics (Figure 5.5). This suggests that not all inactive genes require cohesin for activation and it remains to be seen what characterises the activation of a gene to be cohesin-dependent.

Nevertheless, these experiments show that cohesin plays a crucial role in the initiation of the expression of specific genes.

To elucidate the specific role of cohesin in the induction of *Ciita*, I did ChIP and 3C experiments. ChIP experiments showed that Rad21 binding was significantly increased at the activated *Ciita* promoter IV upon IFNγ induction. Given the role of cohesin in mediating long-range DNA interactions, it is possible that cohesin facilitates activation of transcription by establishing contacts between the promoter and the enhancer. This was tested by 3C experiments. In the absence of information regarding enhancer elements required for IFNγ mediated *Ciita* activation in non-hematopoietic cells, preliminary results did not reveal a prominent role of cohesin in bridging distant regulatory elements to *Ciita* promoter IV. Still, small differences in the genomic architecture around CTCF binding sites and putative enhancer regions were observed. The functional significance of these changes remains to be validated. As the genomic landscape plays a significant role in determining the potential for gene activation, it is of great interest to explore this further. A 4C or a 5C experiment centred around the *Ciita* locus would extend the scope of this analysis. Coupled with ChIP data for enhancer-associated chromatin marks like H3K27ac, H3K4me1 and for Nipbl, Mediator and STAT1 binding, this analysis would help us understand the role of cohesin in mediating DNA looping between regulatory DNA elements as required for gene activation.

It is possible that cohesin can directly impact transcription initiation independently of its role in mediating enhancer-promoter interactions. Some of the early steps of transcriptional initiation involve the deposition of the H3K4me3 mark and recruitment of RNA polymerase at the promoter. ChIP results showed that appropriate amounts of H3K4me3 and Pol II were present at *Ciita* promoter IV after IFNγ induction in cohesin deficient cells. However, ChIP for the Ser5 and Ser2 modified CTD of Pol II indicated a reduction in the initiating and elongating forms of Pol II specifically at *Ciita* promoter in the absence of cohesin. This suggests that even though Pol II is recruited to the *Ciita* promoter upon IFNγ induction, it may not proceed into productive transcription. Cohesin may therefore, play an important role in licensing RNA polymerase into successful initiation and elongation, either directly or through the recruitment of other factors.

This work made use of different antibodies directed against NTD and the modified CTD of Pol II to assess the state of transcription and the analysis was limited to the promoters of *Ciita* gene. As the presence of Pol II does not indicate its transcriptional status, it would be interesting

to further analyse the role of cohesin in modulating polymerase activity with greater refinement using GRO-seq (Global Run ON sequencing). GRO-seq allows profiling of nascent transcripts produced by Pol II activity by the incorporation of biotinylated nucleotides thus providing a genome-wide snapshot of transcriptionally engaged polymerase. Thus the GRO-seq profile across the *Ciita* gene body would essentially be able to clarify if indeed cohesin depletion does not allow the progression of recruited polymerase into active transcription. Further to this aim, it would also be interesting to asses if cohesin depletion affects the recruitment of elongation factors like pTEFb or NELF, thus causing polymerase pausing.

In addition to cohesin's role in DNA looping and RNA polymerase processivity, this work also explored the role of MLL3/4 complex subunits in the regulation of IFNγ mediated gene activation. Interestingly it was observed that while the depletion of MLL3 and MLL4 subunits had no effect on *Ciita* induction, depletion of the PTIP subunit of the same complex specifically impaired the activation of *Ciita* and MHC class II genes, but not the other IFNγ responsive genes tested. Moreover, since the deposition of H3K4me3 at *Ciita* promoter IV upon IFNγ induction is not affected in cohesin-deficient cells, it is unlikely that cohesin functions in gene activation by modulating the methyltransferase activity of MLL3/MLL4 at promoters. It is possible that PTIP plays a role in *Ciita* activation independently of cohesin. However, given the similarity in the impact of deletion of both *Paxip1* and *Rad21* selectively on *Ciita* and MHC class II genes, and the observation that the effect of *Paxip1* deletion is milder than that of cohesin depletion, it is possible that PTIP is one of the effectors of cohesin function in gene activation.

The apparent disparity in the impact of the absence of different subunits of the same complex may be due to a role of PTIP in gene regulation independent of its role as part of the MLL3/4 complex or due to a possible redundancy of the MLL3, MLL4 subunits with the activity of MLL1 and MLL2 subunits. It has previously been shown that PTIP itself is not required for the methyltransfearse activity of the MLL3/4 complex and its absence does not affect the integrity of the complex (Cho et al., 2007). In addition, studies at the *Igh* locus reveal that PTIP is required for the association of RNA polymerase at regions which show PTIP dependent H3K4me3 marks. The authors also suggest an alternate model where PTIP regulates gene expression via direct interaction with DNA binding transcription factors (Daniel et al., 2010). My results show that PTIP depletion, but not MLL3/MLL4 depletion, specifically impacts induction of *Ciita* and MHC class II genes upon IFNγ stimulation. This is consistent with the idea of a direct role of PTIP in gene regulation. As PTIP and SMC1 have previously been shown to interact (Patel et al., 2007), it may

be that cohesin directs PTIP to gene promoters in response to gene activation signals. Further experiments, including PTIP ChIP at the *Ciita* promoter in the presence or absence of cohesin, are required to validate this hypothesis. Nevertheless, the overlapping impact of cohesin and PTIP depletion on gene expression can possibly explain the similarities in the phenotypic features associated with CdLS and Kabuki syndrome. The observation that mutations in cohesin subunits lead to similar developmental defects as mutations in other regulatory proteins, should be explored in greater detail. This would help us explain the occurrence of CdLS in patients which show no mutations in cohesin and its regulatory subunits. It also emphasises on the need to check for the incidence of cohesin mutations in other broad spectrum complex genetic disorders like Kabuki syndrome.

So far, there is no remedial treatment to alleviate the various problems faced by CdLS patients. Medical care is mostly focused at disease management rather than treatment. Here, I showed that treatment of cohesin deficient cells with the HDAC inhibitor TSA, could partly rescue the *Ciita* induction defect. Even though TSA would not be a feasible remedy due to its systemic toxicity, these preliminary results point towards the exciting potential for the development of drugs based on targeting epigenetic modifiers involved in the process of cohesin mediated gene regulation.

Therefore, the work presented here explores the potential role of cohesin in gene activation and provides strong experimental evidence that cohesin is required for the expression of *Ciita* and MHC class II genes in response to IFNγ stimulation. This cohesin-dependent *Ciita* induction defect can possibly contribute to the increased incidence of recurrent infections in CdLS patients, where the patients display impaired immune responses. Preliminary results further implicate cohesin functions beyond DNA looping potentially via regulating the transition of paused RNA polymerase into active transcription and through the action of DNA binding factors like the PTIP subunit of the MLL3/4 complex. Altogether, this study underscores the role of cohesin during transcriptional activation, and uncovers previously unexplored possibilities of cohesin mediated gene regulation.

Chapter 6 : General discussion

The cohesin complex was initially recognised for its essential functions in chromosome segregation during cell division and DNA repair. During the last decade, several studies have shown that cohesin also plays an important role in regulating gene expression. Moreover, defects in cohesin functions in gene regulation are thought to be the major cause of human genetic disorders like CdLS and are suspected to be involved in the development of several cancers. However, mechanistic details of how cohesin contributes to the characteristic features associated with CdLS patients and occurrence of human malignancies remain unknown. The work presented here highlights some interesting aspects of cohesin mediated gene regulation.

6.1 Cohesin depletion in mouse ES cells does not cause a global collapse in pluripotency but deregulates specific genes

The use of ERT2Cre mediated conditional deletion of *Rad21* in mouse ES cells proved to be a better strategy for analysing the impact of cohesin depletion, than previously published knockdown studies. The conditional Rad21 knockout strategy was particularly useful as it allowed us to rapidly deplete chromatin bound cohesin in cycling ES cells within 24 hours. This permitted me to analyse gene expression changes prior to arrest in G2 and before the onset of stress responses due to the loss of cohesin functions in cell division. Therefore, secondary effects on gene expression were avoided. Transcriptional analysis post 24 hours of cohesin depletion in ES cells revealed about 600 deregulated genes. However, not all pluripotency associated genes were downregulated nor was there a preferential upregulation of differentiation associated genes. In fact, the impact of cohesin depletion was found to be highly locus specific. This observation contrasts with previously published reports, which suggest that cohesin is required for the maintenance of pluripotency (Ding et al., 2009; Hu et al., 2009; Kagey et al., 2010; Nitzsche et al., 2011). The observed differences may be attributed to the activation of stress response due to prolonged culture of ES cells (5 days) with reduced cohesin levels using the RNAi knockdown strategy (Figure 3.5B). This speculation is supported by the evidence that stress signals trigger the differentiation of ES cells (Lin et al., 2005; Maimets et al., 2008; Li et al., 2012). In fact, gene expression changes observed in my conditionally cohesin depleted ES cells at 48 hours and upon doxorubicin treatment, showed a more drastic impact on pluripotency and resembled the changes observed in the knockdown studies. Therefore, the loss of pluripotency-associated transcriptional profile most likely results from indirect effects of stress response due to prolonged

cohesin loss and is not a true indication of cohesin functions in ES cell gene regulation. These results illustrate the importance of avoiding the activation of stress response due to cell cycle defects while studying cohesin functions in gene regulation. This study also establishes that limited number of genes are affected by perturbations in cohesin levels in the cell and the impact is context dependent.

6.2 Cohesin is not essential for the maintenance of enhancer-promoter interactions

In this study I compared the enhancer-promoter interactions at some key pluripotency associated genes in ES cells with or without cohesin and with ES cells which had been differentiated for 24 hours. The chromatin interactions were assessed by 3C while the presence of cohesin on DNA was quantified by ChIP. These experiments unexpectedly revealed that enhancer-promoter interactions at certain loci in cohesin deficient ES cells were not affected even though loss of cohesin binding was associated with deregulation of the expression of those genes. This was true for the *Lefty1* and *Klf4* loci. Even at the *Nanog* locus, where both gene expression and enhancer-promoter interactions were reduced in response to cohesin loss, the loss in interactions was much smaller than the loss in the amount of DNA bound cohesin at the enhancer. Incidentally, the enhancer-promoter interactions at the same loci were more significantly reduced upon differentiation and showed an associated loss in cohesin binding specifically at the enhancer region. Taken together, it can be concluded that enhancer-promoter interactions can be maintained in the absence of cohesin at the genes tested. Additionally, it also suggests that changes in enhancer-promoter interactions alone cannot explain the observed changes in gene expression and cohesin may impact on other aspects of transcriptional regulation. Although, it is noteworthy that in this study only interactions between the promoter and the corresponding functional enhancer region were measured. It is possible that gene promoters are involved in additional chromatin interactions which may also play a part in transcriptional regulation.

Mechanistically, it is possible that the cohesin ring structure stabilises the DNA loop formed between enhancers and promoters but the loop is actually maintained by multiple protein-protein interactions between the several proteins associated at enhancer and promoter regions. This hypothesis is supported by the observation that transcription factors like OCT4, SOX2 and NANOG are directly involved in both higher order genome organisation and local chromatin conformation. Moreover, loss of either OCT4 or NANOG diminishes long-range interactions

between OSN bound region (de Wit et al., 2013; Denholtz et al., 2013). Additionally, many other factors like SATB1 and LDB1 have also been shown to contribute to genome organisation and transcriptional regulation (Cai et al., 2006; Soler et al., 2010). In cohesin-deficient ES cells, OCT4 is still expressed appropriately and NANOG only shows a small reduction. Therefore, enhancer-promoter interactions may in part be maintained by protein-protein interactions centred around the OSN cluster in the absence of cohesin. And upon differentiation, OCT4 and NANOG expression is significantly reduced.

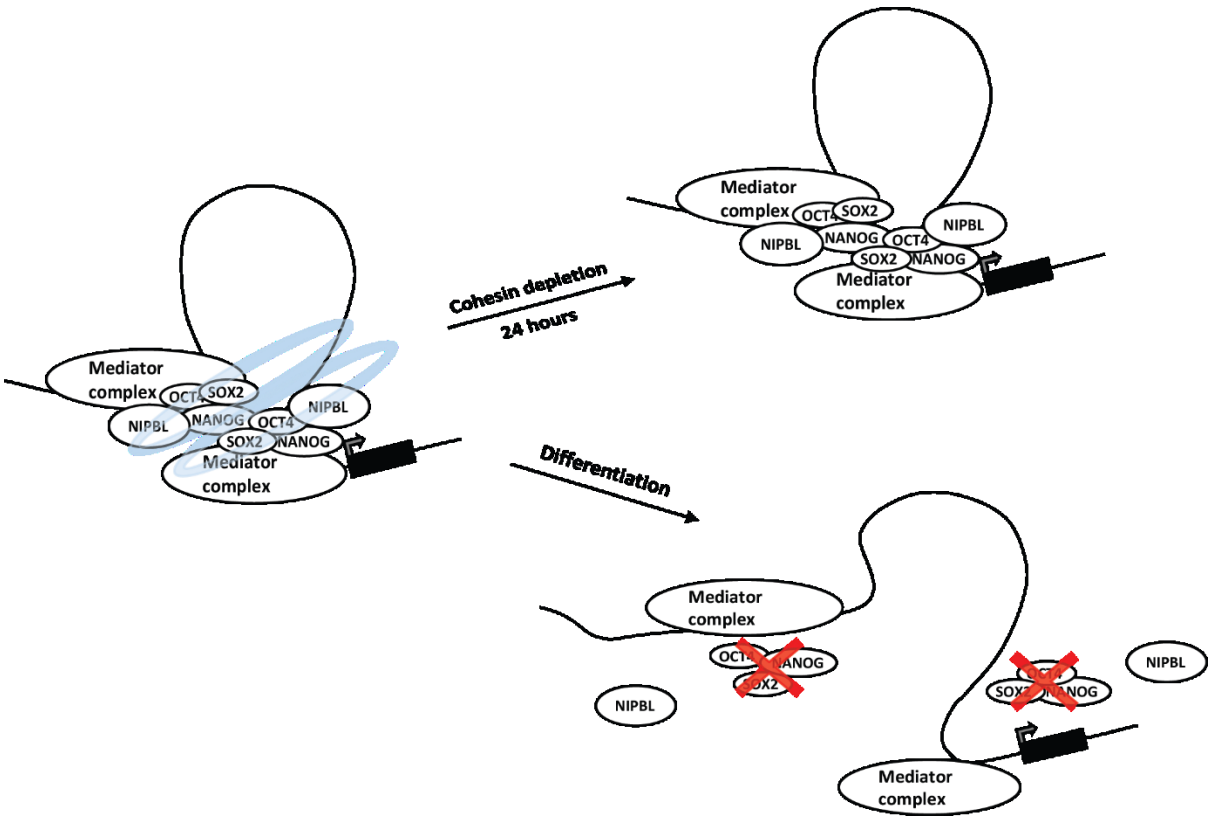


Figure 6.1 A schematic depiction of possible chromatin landscape changes in ES cells associated with cohesin depletion and with differentiation. Cohesin rings are depicted in light blue and the target gene as small black rectangle. Enhancer-promoter interactions can still be maintained at certain loci after cohesin depletion, possibly through interactions between associated transcription factors. Differentiation, however, is associated with reduced expression of pluripotency associated transcription factors. It is also accompanied by loss of Nipbl binding and consequently cohesin loading. In the absence of a physical trapping of chromatin by cohesin as well as loss of transcription factor based interactions, chromatin looping is no longer maintained.

Thus, together with loss of cohesin binding at enhancer, reduced protein-protein interactions between transcription factors, might together be responsible for reduced enhancer-promoter interactions in differentiating ES cells. If this is the case, these experiments indicate that

mere contact between enhancer and promoter does not specify the transcriptional status of the gene. It also suggests that the changes observed in long-range cis-regulatory interactions upon cohesin depletion can possibly be a consequence rather than being the cause of the associated transcriptional changes.

6.3 Cohesin plays essential roles in the activation of specific gene transcription

While most studies so far have relied on the analysis of the impact of cohesin depletion on the steady state of the cellular system, I have here explored the possibility of cohesin being involved in the process of gene activation. Experiments in IFNγ induced MEFs showed that cohesin is required for the activation of *Ciita* gene expression. *Ciita* and MHC class II genes, which are not expressed in MEFs unless treated with IFNγ, failed to be induced in the absence of cohesin. On the other hand, the induction of MHC class I genes, which are expressed at basal levels in all cell types, was not affected by cohesin depletion. These observations show that cohesin is essential for the establishment of a new gene expression programme upon IFNγ treatment and regulates the activation of master regulators like *Ciita.* Contrarily, upregulation of previously expressed genes does not require cohesin. This provides a new paradigm in the field of cohesin mediated gene regulation whereupon cohesin is necessary for the successful activation of key transcription factors in response to cellular signals. It opens up several directions of future investigation like the impact of cohesin mutations in the process of limb and brain development, cognition and mounting of immune response – some typical processes affected in CdLS patients. In fact, reduced MHC class II expression might in part be responsible for the increased incidence of infections in CdLS patients (Jyonouchi et al., 2013) and needs to be investigated further in patient cell lines.

This activity of cohesin is in some way similar to that of pioneer transcription factors which play critical roles in gene activation (Zaret and Carroll, 2011). However, unlike pioneer factors, cohesin functions are limited to only select genes within the same pathway. It will be of utmost importance to determine what features demarcate genes that show cohesin-dependent gene activation. Not all genes which were silent in uninduced MEFs showed cohesin-dependent activation. Therefore, a genome-wide analysis of several chromatin features like histone marks, RNA polymerase binding and binding of chromatin remodelling factors might be useful in discovering what makes some genes cohesin-sensitive and others not. In fact, it might be more interesting to find what makes the other genes, which also show increased cohesin binding upon induction, to not be dependent on it. Additionally, further investigation is required to figure out

what brings or signals cohesin loading at activated genes. In this respect, an examination into the requirement of CTCF or NIPBL in this process might give some perspective.

6.4 Cohesin can potentially operate by regulating the activity of RNA polymerase and recruitment of PTIP

The data presented here shows that cohesin can potentially regulate transcription by modulating the activity and processivity of RNA polymerase but further experiments are required to clarify this role. It will be useful to compare the transcriptional profile of *Ciita* gene in the absence of cohesin with that of cells treated with specific inhibitors which block the initiation or elongation phase of Polymerase activity.

Another interesting observation made here is that the phenotype or the impact of *Paxip1* deletion on *Ciita* induction is similar to that of cohesin depletion but this is not the case for *Mll3/Mll4* deletion. These experiments were done contemplating cohesin functions in depositing histone methylation marks at enhancers and promoters through the activity of MLL3/MLL4 complex. Given the striking similarity between the clinical features of Kabuki syndrome and CdLS patients, it is still promising to explore this possibility further. Consequently, it will be useful to map H3K4me1 marks around the *Ciita* locus. The impact of PTIP depletion further argues for the role of MLL complexes in cohesin mediated gene regulation. It is probable that at the *Ciita* locus, MLL3/MLL4 subunit function is redundant with that of the MLL1/MLL2 subunits. It would be interesting to check the impact of depleting other individual subunits of the MLL complexes on *Ciita* induction. Additionally, it is also possible that PTIP functions independently of the MLL complexes to regulate gene expression.

Given the role of cohesin in mediating long-range enhancer promoter interactions, it will be valuable to evaluate the connection between enhancer looping and polymerase activity. Recent reports have shown that preformed enhancer-promoter interactions exist prior to activation of gene transcription (Fanucchi et al., 2013; Jin et al., 2013) and are frequently associated with paused Polymerase (Ghavi-Helm et al., 2014). It remains to be seen if cohesin is bound at the sites of preformed chromatin interactions or is loaded in response to gene activating stimuli. In another scenario, cohesin initiated looping of an enhancer close to a promoter region might trigger the elongation of polymerase into successful transcription.

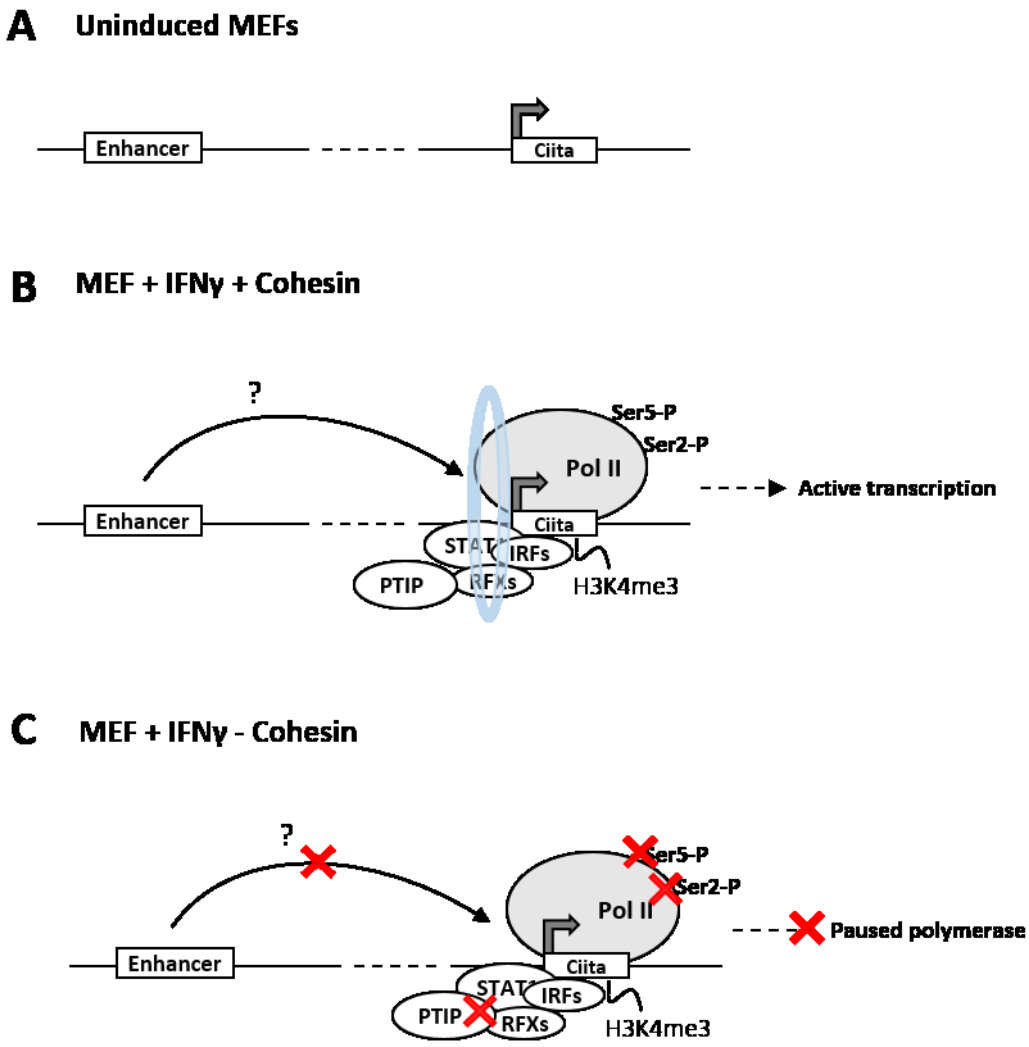


Figure 6.2 Possible mechanism of cohesin-mediated *Ciita* activation. A) The Ciita gene locus is not transcribed in uninduced MEFS. **B)** Stimulation of MEFs with IFNγ causes STAT1 binding along with the associated regulatory factors like RFX1, RFX5, IRF1, IRF2, to the *Ciita* promoter. This is accompanied by cohesin loading which is known to occur at active regulatory elements. The histones are modified to H3K4me3 at the TSS and Ser5/Ser2 phosphorylated polymerase actively transcribes *Ciita* gene. **C)** Stimulation of MEFs in the absence of cohesin results in the failure to induce *Ciita* transcription. The absence of cohesin could potentially affect any of the steps mentioned in B). Based on the data presented in this study, I speculate that in the absence of cohesin, the enhancer cannot form stable contacts with the *Ciita* promoter. Absence of cohesin also prevents the release of polymerase into active elongation due to reduced Ser5 and Ser2 phosphorylation. A role for PTIP in the induction of *Ciita* transcription is suggested by the failure of Paxip1-deficient MEFs to express Ciita in response to IFNγ. Together these events prevent the successful transcription of the *Ciita* gene.

6.5 The impact of cohesin depletion on *Myc* expression is context dependent

Experiments in fibroblasts showed that cohesin is not essential for appropriate induction of *Myc* expression. Steady state levels of *Myc* were reduced in G1 arrested cohesin deficient preB cells but not in resting CD4 DP and SP thymocytes. More interestingly, *Myc* was post transcriptionally upregulated in cohesin depleted ES cells cultured in 2i media but not in those cultured in serum + LIF media. These observations show that cohesin does play a role in fine tuning *Myc* expression but is not essential for its expression. Moreover, the post-transcriptional regulation of *Myc* provides yet another novel mechanism of cohesin mediated gene regulation. It remains to be confirmed if the upregulation is mediated through stabilization of *Myc* by the action of *Pvt1,* which is also upregulated in the absence of cohesin. If confirmed, the deregulation of *Pvt1* expression in human malignancies, which show mutations in cohesin and its regulatory subunits, provides a promising avenue for exploring cohesin functions in cancerous cells.

Taken together, this work shows that cohesin impacts gene expression not only by regulating chromatin interactions but also influences transcription more directly by modulating the activity of components of the transcriptional machinery. This work also provides evidence that cohesin plays an essential role in the process of *Ciita* gene activation.

References

Ahmadiyeh, N., Pomerantz, M.M., Grisanzio, C., Herman, P., Jia, L., Almendro, V., He, H.H., Brown, M., Liu, X.S., Davis, M., et al. (2010). 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. Proc. Natl. Acad. Sci. U. S. A. *107*, 9742–9746.

Alberts, B. (2008). Molecular Biology of the Cell (Garland Science; 5 edition).

Apostolou, E., Ferrari, F., Walsh, R.M., Bar-Nur, O., Stadtfeld, M., Cheloufi, S., Stuart, H.T., Polo, J.M., Ohsumi, T.K., Borowsky, M.L., et al. (2013). Genome-wide chromatin interactions of the Nanog locus in pluripotency, differentiation, and reprogramming. Cell Stem Cell *12*, 699–712.

Ardehali, M.B., Mei, A., Zobeck, K.L., Caron, M., Lis, J.T., and Kusch, T. (2011). Drosophila Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription. EMBO J. *30*, 2817–2828.

Arumugam, P., Gruber, S., Tanaka, K., Haering, C.H., Mechtler, K., and Nasmyth, K. (2003). ATP hydrolysis is required for cohesin's association with chromosomes. Curr. Biol. *13*, 1941–1953.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., et al. (2006). Chromatin signatures of pluripotent cell lines. Nat. Cell Biol. *8*, 532–538.

Balbás-Martínez, C., Sagrera, A., Carrillo-de-Santa-Pau, E., Earl, J., Márquez, M., Vazquez, M., Lapi, E., Castro-Giner, F., Beltran, S., Bayés, M., et al. (2013). Recurrent inactivation of STAG2 in bladder cancer is not associated with aneuploidy. Nat. Genet. *45*, 1464–1469.

Ball, A.R., Chen, Y.-Y., and Yokomori, K. (2014). Mechanisms of cohesin-mediated gene regulation and lessons learned from cohesinopathies. Biochim. Biophys. Acta *1839*, 191–202.

Barber, T.D., McManus, K., Yuen, K.W.Y., Reis, M., Parmigiani, G., Shen, D., Barrett, I., Nouhi, Y., Spencer, F., Markowitz, S., et al. (2008). Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers. Proc. Natl. Acad. Sci. U. S. A. *105*, 3443–3448.

Beckouët, F., Hu, B., Roig, M.B., Sutani, T., Komata, M., Uluocak, P., Katis, V.L., Shirahige, K., and Nasmyth, K. (2010). An Smc3 acetylation cycle is essential for establishment of sister chromatid cohesion. Mol. Cell *39*, 689–699.

Bénard, C.Y., Kébir, H., Takagi, S., and Hekimi, S. (2004). mau-2 acts cell-autonomously to guide axonal migrations in Caenorhabditis elegans. Development *131*, 5947–5958.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315–326.

Bernstein, P.L., Herrick, D.J., Prokipcak, R.D., and Ross, J. (1992). Control of c-myc mRNA half-life in vitro by a protein capable of binding to a coding region stability determinant. Genes Dev. *6*, 642–654.

Blanchard, J.M., Piechaczyk, M., Dani, C., Chambard, J.C., Franchi, A., Pouyssegur, J., and Jeanteur, P. (1985). c-myc gene is transcribed at high rate in G0-arrested fibroblasts and is post-transcriptionally regulated in response to growth factors. Nature *317*, 443–445.

Bokinni, Y. (2012). Kabuki syndrome revisited. J. Hum. Genet. *57*, 223–227.

Borges, V., Lehane, C., Lopez-Serra, L., Flynn, H., Skehel, M., Rolef Ben-Shahar, T., and Uhlmann, F. (2010). Hos1 deacetylates Smc3 to close the cohesin acetylation cycle. Mol. Cell *39*, 677–688.

Bose, T., Lee, K.K., Lu, S., Xu, B., Harris, B., Slaughter, B., Unruh, J., Garrett, A., McDowell, W., Box, A., et al. (2012). Cohesin proteins promote ribosomal RNA production and protein translation in yeast and human cells. PLoS Genet. *8*, e1002749.

Boyer, L. a, Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947–956.

Buheitel, J., and Stemmann, O. (2013). Prophase pathway-dependent removal of cohesin from human chromosomes requires opening of the Smc3-Scc1 gate. EMBO J. *32*, 666–676.

Cai, S., Lee, C.C., and Kohwi-Shigematsu, T. (2006). SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. Nat. Genet. *38*, 1278–1288.

Calo, E., and Wysocka, J. (2013). Modification of Enhancer Chromatin: What, How, and Why? Mol. Cell *49*, 825–837.

Carretero, M., Ruiz-Torres, M., Rodríguez-Corsino, M., Barthelemy, I., and Losada, A. (2013). Pds5B is required for cohesion establishment and Aurora B accumulation at centromeres. EMBO J. *32*, 2938–2949.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230–1234.

Chan, K.-L., Roig, M.B., Hu, B., Beckouët, F., Metson, J., and Nasmyth, K. (2012). Cohesin's DNA Exit Gate Is Distinct from Its Entrance Gate and Is Regulated by Acetylation. Cell 1–14.

Chappell, J., and Dalton, S. (2013). Roles for MYC in the establishment and maintenance of pluripotency. Cold Spring Harb. Perspect. Med. *3*, a014381.

Chatterjee, A., Zakian, S., Hu, X.-W., and Singleton, M.R. (2013). Structural insights into the regulation of cohesion establishment by Wpl1. EMBO J. *32*, 677–687.

Chen, H.-S., Wikramasinghe, P., Showe, L., and Lieberman, P.M. (2012). Cohesins repress Kaposi's sarcoma-associated herpesvirus immediate early gene transcription during latency. J. Virol. *86*, 9454–9464.

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., et al. (2011). Cohesin mediates chromatin interactions that regulate mammalian β-globin expression. J. Biol. Chem. *286*, 17870–17878.

Cho, E.A., Prindle, M.J., and Dressler, G.R. (2003). BRCT Domain-Containing Protein PTIP Is Essential for Progression through Mitosis. Mol. Cell. Biol. *23*, 1666–1673.

Cho, Y.-W., Hong, T., Hong, S., Guo, H., Yu, H., Kim, D., Guszczynski, T., Dressler, G.R., Copeland, T.D., Kalkum, M., et al. (2007). PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. J. Biol. Chem. *282*, 20395–20406.

Cho, Y.-W., Hong, S., Jin, Q., Wang, L., Lee, J.-E., Gavrilova, O., and Ge, K. (2009). Histone methylation regulator PTIP is required for PPARgamma and C/EBPalpha expression and adipogenesis. Cell Metab. *10*, 27–39.

Choi, H.-K., Kim, B.-J., Seo, J.-H., Kang, J.-S., Cho, H., and Kim, S.-T. (2010). Cohesion establishment factor, Eco1 represses transcription via association with histone demethylase, LSD1. Biochem. Biophys. Res. Commun. *394*, 1063–1068.

Ciosk, R., Shirayama, M., Shevchenko, A., Tanaka, T., Toth, A., and Nasmyth, K. (2000). Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. Mol. Cell *5*, 243–254.

Cole, M.D., and Cowling, V.H. (2008). Transcription-independent functions of MYC: regulation of translation and DNA replication. Nat. Rev. Mol. Cell Biol. *9*, 810–815.

Cowling, V.H., and Cole, M.D. (2006). Mechanism of transcriptional activation by the Myc oncoproteins. Semin. Cancer Biol. *16*, 242–252.

Cowling, V.H., and Cole, M.D. (2007). The Myc transactivation domain promotes global phosphorylation of the RNA polymerase II carboxy-terminal domain independently of direct DNA binding. Mol. Cell. Biol. *27*, 2059–2073.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M. a, Frampton, G.M., Sharp, P. a, et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U. S. A. *107*, 21931–21936.

Cuscó, I., del Campo, M., Vilardell, M., González, E., Gener, B., Galán, E., Toledo, L., and Pérez-Jurado, L.A. (2008). Array-CGH in patients with Kabuki-like phenotype: identification of two patients with complex rearrangements including 2q37 deletions and no other recurrent aberration. BMC Med. Genet. *9*, 27.

Dang, C. V (2012). MYC on the path to cancer. Cell *149*, 22–35.

Dang, C. V, O'donnell, K.A., and Juopperi, T. (2005). The great MYC escape in tumorigenesis. Cancer Cell *8*, 177–178.

Daniel, J.A., Santos, M.A., Wang, Z., Zang, C., Schwab, K.R., Jankovic, M., Filsuf, D., Chen, H.-T., Gazumyan, A., Yamane, A., et al. (2010). PTIP promotes chromatin changes critical for immunoglobulin class switch recombination. Science *329*, 917–923.

Deardorff, M. a, Bando, M., Nakato, R., Watrin, E., Itoh, T., Minamino, M., Saitoh, K., Komata, M., Katou, Y., Clark, D., et al. (2012a). HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. Nature *489*, 313–317.

Deardorff, M.A., Kaur, M., Yaeger, D., Rampuria, A., Korolev, S., Pie, J., Gil-Rodríguez, C., Arnedo, M., Loeys, B., Kline, A.D., et al. (2007). Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of cornelia de Lange syndrome with predominant mental retardation. Am. J. Hum. Genet. *80*, 485–494.

Deardorff, M.A., Wilde, J.J., Albrecht, M., Dickinson, E., Tennstedt, S., Braunholz, D., Mönnich, M., Yan, Y., Xu, W., Gil-Rodríguez, M.C., et al. (2012b). RAD21 mutations cause a human cohesinopathy. Am. J. Hum. Genet. *90*, 1014–1027.

Degner, S.C., Verma-Gaur, J., Wong, T.P., Bossen, C., Iverson, G.M., Torkamani, A., Vettermann, C., Lin, Y.C., Ju, Z., Schulz, D., et al. (2011). CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. Proc. Natl. Acad. Sci. U. S. A. *108*, 9566–9571.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306–1311.

Dekker, J., Marti-Renom, M. a, and Mirny, L. a (2013). Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat. Rev. Genet. *14*, 390–403.

DeMare, L.E., Leng, J., Cotney, J., Reilly, S.K., Yin, J., Sarro, R., and Noonan, J.P. (2013). The genomic landscape of cohesin-associated chromatin interactions. Genome Res. *23*, 1224–1234.

Denholtz, M., Bonora, G., Chronis, C., Splinter, E., de Laat, W., Ernst, J., Pellegrini, M., and 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*, 602–616.

Denissov, S., Hofemeister, H., Marks, H., Kranz, A., Ciotta, G., Singh, S., Anastassiadis, K., Stunnenberg, H.G., and Stewart, A.F. (2014). Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. Development *141*, 526–537.

Ding, L., Paszkowski-Rogacz, M., Nitzsche, A., Slabicki, M.M., Heninger, A.-K., de Vries, I., Kittler, R., Junqueira, M., Shevchenko, A., Schulz, H., et al. (2009). A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. Cell Stem Cell *4*, 403–415.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 1–5.

Doan, T., Melvold, R., Viselli, S., and Waltenbaugh, C. (2012). Lippincott's Illustrated Reviews: Immunology (LWW; Second, North American Edition edition).

Dominguez-Sola, D., Ying, C.Y., Grandori, C., Ruggiero, L., Chen, B., Li, M., Galloway, D. a, Gu, W., Gautier, J., and Dalla-Favera, R. (2007). Non-transcriptional control of DNA replication by c-Myc. Nature *448*, 445–451.

Donze, D., Adams, C.R., Rine, J., and Kamakaka, R.T. (1999). The boundaries of the silenced HMR domain in Saccharomyces cerevisiae. Genes Dev. *13*, 698–708.

Dorsett, D., and Ström, L. (2012). The ancient and evolving roles of cohesin in gene expression and DNA repair. Curr. Biol. *22*, R240–50.

Durand, B., Sperisen, P., Emery, P., Barras, E., Zufferey, M., Mach, B., and Reith, W. (1997). RFXAP, a novel subunit of the RFX DNA binding complex is mutated in MHC class II deficiency. EMBO J. *16*, 1045–1055.

Eichinger, C.S., Kurze, A., Oliveira, R.A., and Nasmyth, K. (2013). Disengaging the Smc3/kleisin interface releases cohesin from Drosophila chromosomes during interphase and mitosis. EMBO J. *32*, 656–665.

Van den Elsen, P.J., van der Stoep, N., Viëtor, H.E., Wilson, L., van Zutphen, M., and Gobin, S.J. (2000). Lack of CIITA expression is central to the absence of antigen presentation functions of trophoblast cells and is caused by methylation of the IFN-gamma inducible promoter (PIV) of CIITA. Hum. Immunol. *61*, 850–862.

Van den Elsen, P.J., Holling, T.M., van der Stoep, N., and Boss, J.M. (2003). DNA methylation and expression of major histocompatibility complex class I and class II transactivator genes in human developmental tumor cells and in T cell malignancies. Clin. Immunol. *109*, 46–52.

Fanucchi, S., Shibayama, Y., Burd, S., Weinberg, M.S., and Mhlanga, M.M. (2013). Chromosomal Contact Permits Transcription between Coregulated Genes. Cell *155*, 606–620.

Faure, A.J., Schmidt, D., Watt, S., Schwalie, P.C., Wilson, M.D., Xu, H., Ramsay, R.G., Odom, D.T., and Flicek, P. (2012). Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules. Genome Res. *22*, 2163–2175.

Fay, A., Misulovin, Z., Li, J., Schaaf, C. a, Gause, M., Gilmour, D.S., and Dorsett, D. (2011). Cohesin selectively binds and regulates genes with paused RNA polymerase. Curr. Biol. *21*, 1624–1634.

Ficz, G., Hore, T.A., Santos, F., Lee, H.J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.-L., Walter, J., et al. (2013). FGF Signaling Inhibition in ESCs Drives Rapid Genome-wide Demethylation to the Epigenetic Ground State of Pluripotency. Cell Stem Cell *null*.

García-Lora, A., Algarra, I., Collado, A., and Garrido, F. (2003). Tumour immunology, vaccination and escape strategies. Eur. J. Immunogenet. *30*, 177–183.

Gaspar-Maia, A., Alajem, A., Meshorer, E., and Ramalho-Santos, M. (2011). Open chromatin in pluripotency and reprogramming. Nat. Rev. Mol. Cell Biol. *12*, 36–47.

Geirsson, A., Paliwal, I., Lynch, R.J., Bothwell, A.L.M., and Hammond, G.L. (2003). Class II transactivator promoter activity is suppressed through regulation by a trophoblast noncoding RNA. Transplantation *76*, 387–394.

Gerlich, D., Koch, B., Dupeux, F., Peters, J.-M., and Ellenberg, J. (2006). Live-cell imaging reveals a stable cohesin-chromatin interaction after but not before DNA replication. Curr. Biol. *16*, 1571–1578.

Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E.E.M. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. Nature *512*, 96–100.

Gillis, L.A., McCallum, J., Kaur, M., DeScipio, C., Yaeger, D., Mariani, A., Kline, A.D., Li, H., Devoto, M., Jackson, L.G., et al. (2004). NIPBL mutational analysis in 120 individuals with Cornelia de Lange syndrome and evaluation of genotype-phenotype correlations. Am. J. Hum. Genet. *75*, 610–623.

Gimigliano, A., Mannini, L., Bianchi, L., Puglia, M., Deardorff, M. a, Menga, S., Krantz, I.D., Musio, A., and Bini, L. (2012). Proteomic profile identifies dysregulated pathways in Cornelia de Lange syndrome cells with distinct mutations in SMC1A and SMC3 genes. J. Proteome Res. *11*, 6111–6123.

Gomes, N.P., and Espinosa, J.M. (2010). Gene-specific repression of the p53 target gene PUMA via intragenic CTCF-Cohesin binding. Genes Dev. *24*, 1022–1034.

Gorkin, D.U., Leung, D., and Ren, B. (2014). The 3D Genome in Transcriptional Regulation and Pluripotency. Cell Stem Cell *14*, 762–775.

Gruber, S., Haering, C.H., and Nasmyth, K. (2003). Chromosomal cohesin forms a ring. Cell *112*, 765–777.

Gruber, S., Arumugam, P., Katou, Y., Kuglitsch, D., Helmhart, W., Shirahige, K., and Nasmyth, K. (2006). Evidence that loading of cohesin onto chromosomes involves opening of its SMC hinge. Cell *127*, 523–537.

Guacci, V., Koshland, D., and Strunnikov, A. (1997). A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of MCD1 in S. cerevisiae. Cell *91*, 47–57.

Gullerova, M., and Proudfoot, N.J. (2008). Cohesin complex promotes transcriptional termination between convergent genes in S. pombe. Cell *132*, 983–995.

Guo, C., Yoon, H.S., Franklin, A., Jain, S., Ebert, A., Cheng, H.-L., Hansen, E., Despo, O., Bossen, C., Vettermann, C., et al. (2011). CTCF-binding elements mediate control of V(D)J recombination. Nature *477*, 424–430.

Gurdon, J.B., Elsdale, T.R., and Fischberg, M. (1958). Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei. Nature *182*, 64–65.

Gutiérrez-Caballero, C., Cebollero, L.R., and Pendás, A.M. (2012). Shugoshins: from protectors of cohesion to versatile adaptors at the centromere. Trends Genet. *28*, 351–360.

Haarhuis, J.H.I., Elbatsh, A.M.O., van den Broek, B., Camps, D., Erkan, H., Jalink, K., Medema, R.H., and Rowland, B.D. (2013). WAPL-mediated removal of cohesin protects against segregation errors and aneuploidy. Curr. Biol. *23*, 2071–2077.

Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H.D., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C., et al. (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. Cell Stem Cell *13*, 360–369.

Hackett, J.A., Dietmann, S., Murakami, K., Down, T.A., Leitch, H.G., and Surani, M.A. (2013). Synergistic Mechanisms of DNA Demethylation during Transition to Ground-State Pluripotency. Stem Cell Reports *1*, 518–531.

Hadjur, S., Williams, L.M., Ryan, N.K., Cobb, B.S., Sexton, T., Fraser, P., Fisher, A.G., and Merkenschlager, M. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature *460*, 410–413.

Haering, C.H., Löwe, J., Hochwagen, A., and Nasmyth, K. (2002). Molecular architecture of SMC proteins and the yeast cohesin complex. Mol. Cell *9*, 773–788.

Hanna, J., Saha, K., Pando, B., van Zon, J., Lengner, C.J., Creyghton, M.P., van Oudenaarden, A., and Jaenisch, R. (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. Nature *462*, 595–601.

Harton, J.A., and Ting, J.P.-Y. (2000). Class II Transactivator: Mastering the Art of Major Histocompatibility Complex Expression. Mol. Cell. Biol. *20*, 6185–6194.

Harton, J.A., Linhoff, M.W., Zhang, J., and Ting, J.P.-Y. (2002). Cutting Edge: CATERPILLER: A Large Family of Mammalian Genes Containing CARD, Pyrin, Nucleotide-Binding, and Leucine-Rich Repeat Domains. J. Immunol. *169*, 4088–4093.

Heidinger-Pauli, J.M., Mert, O., Davenport, C., Guacci, V., and Koshland, D. (2010). Systematic reduction of cohesin differentially affects chromosome segregation, condensation, and DNA repair. Curr. Biol. *20*, 957–963.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.

Herz, H.-M., Mohan, M., Garruss, A.S., Liang, K., Takahashi, Y.-H., Mickey, K., Voets, O., Verrijzer, C.P., and Shilatifard, A. (2012). Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. Genes Dev. *26*, 2604–2620.

Higashi, T.L., Ikeda, M., Tanaka, H., Nakagawa, T., Bando, M., Shirahige, K., Kubota, Y., Takisawa, H., Masukata, H., and Takahashi, T.S. (2012). The prereplication complex recruits XEco2 to chromatin to promote cohesin acetylation in Xenopus egg extracts. Curr. Biol. *22*, 977–988.

Hirano, T. (2006). At the heart of the chromosome: SMC proteins in action. Nat. Rev. Mol. Cell Biol. *7*, 311–322.

Hirano, M., and Hirano, T. (2002). Hinge-mediated dimerization of SMC protein is essential for its dynamic interaction with DNA. EMBO J. *21*, 5733–5744.

Hirayama, T., Tarusawa, E., Yoshimura, Y., Galjart, N., and Yagi, T. (2012). CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons. Cell Rep. *2*, 345–357.

Horsfield, J. a, Anagnostou, S.H., Hu, J.K.-H., Cho, K.H.Y., Geisler, R., Lieschke, G., Crosier, K.E., and Crosier, P.S. (2007). Cohesin-dependent regulation of Runx genes. Development *134*, 2639–2649.

Hou, F., and Zou, H. (2005). Two human orthologues of Eco1/Ctf7 acetyltransferases are both required for proper sister-chromatid cohesion. Mol. Biol. Cell *16*, 3908–3918.

Hu, D., Garruss, A.S., Gao, X., Morgan, M.A., Cook, M., Smith, E.R., and Shilatifard, A. (2013). The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1093–1098.

Hu, G., Kim, J., Xu, Q., Leng, Y., Orkin, S.H., and Elledge, S.J. (2009). A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. Genes Dev. 837–848.

Hurlin, P.J., and Huang, J. (2006). The MAX-interacting transcription factor network. Semin. Cancer Biol. *16*, 265–274.

Jabrane-Ferrat, N. (2003). MHC class II enhanceosome: how is the class II transactivator recruited to DNA-bound activators? Int. Immunol. *15*, 467–475.

Jackstadt, R., and Hermeking, H. (2014). MicroRNAs as regulators and mediators of c-MYC function. Biochim. Biophys. Acta.

Jaenisch, R., and Young, R. (2008). Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell *132*, 567–582.

Jeong, K.W., Kim, K., Situ, A.J., Ulmer, T.S., An, W., and Stallcup, M.R. (2011). Recognition of enhancer element-specific histone methylation by TIP60 in transcriptional activation. Nat. Struct. Mol. Biol. *18*, 1358–1365.

Jeppsson, K., Kanno, T., Shirahige, K., and Sjögren, C. (2014). The maintenance of chromosome structure: positioning and functioning of SMC complexes. Nat. Rev. Mol. Cell Biol. *15*, 601–614.

Jessberger, R. (2012). Age-related aneuploidy through cohesion exhaustion. EMBO Rep. *13*, 539–546.

Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.-A., Schmitt, A.D., Espinoza, C. a., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature *503*, 290–294.

Jones, T.R., and Cole, M.D. (1987). Rapid cytoplasmic turnover of c-myc mRNA: requirement of the 3' untranslated sequences. Mol. Cell. Biol. *7*, 4513–4521.

Jyonouchi, S., Orange, J., Sullivan, K.E., Krantz, I., and Deardorff, M. (2013). Immunologic Features of Cornelia de Lange Syndrome. Pediatrics *132(2)*, e484–e489.

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., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435.

Kanaseki, T., Ikeda, H., Takamura, Y., Toyota, M., Hirohashi, Y., Tokino, T., Himi, T., and Sato, N. (2003). Histone Deacetylation, But Not Hypermethylation, Modifies Class II Transactivator and MHC Class II Gene Expression in Squamous Cell Carcinomas. J. Immunol. *170*, 4980–4985.

Kang, H., and Lieberman, P.M. (2011). Mechanism of glycyrrhizic acid inhibition of Kaposi's sarcoma-associated herpesvirus: disruption of CTCF-cohesin-mediated RNA polymerase II pausing and sister chromatid cohesion. J. Virol. *85*, 11159–11169.

Kawabe, Y.-I., Wang, Y.X., McKinnell, I.W., Bedford, M.T., and Rudnicki, M.A. (2012). Carm1 regulates Pax7 transcriptional activity through MLL1/2 recruitment during asymmetric satellite stem cell divisions. Cell Stem Cell *11*, 333–345.

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., et al. (2009). Multiple organ system defects and transcriptional dysregulation in the Nipbl(+/-) mouse, a model of Cornelia de Lange Syndrome. PLoS Genet. *5*, e1000650.

Kieffer-Kwon, K.-R., Tang, Z., Mathe, E., Qian, J., Sung, M.-H., Li, G., Resch, W., Baek, S., Pruett, N., Grøntved, L., et al. (2013). Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. Cell *155*, 1507–1520.

Kim, B.-J., Kang, K.-M., Jung, S.Y., Choi, H.-K., Seo, J.-H., Chae, J.-H., Cho, E.-J., Youn, H.-D., Qin, J., and Kim, S.-T. (2008a). Esco2 is a novel corepressor that associates with various chromatin modifying enzymes. Biochem. Biophys. Res. Commun. *372*, 298–304.

Kim, D., Patel, S.R., Xiao, H., and Dressler, G.R. (2009a). The role of PTIP in maintaining embryonic stem cell pluripotency. Stem Cells *27*, 1516–1523.

Kim, H.H., Kuwano, Y., Srikantan, S., Lee, E.K., Martindale, J.L., and Gorospe, M. (2009b). HuR recruits let-7/RISC to repress c-Myc expression. Genes Dev. *23*, 1743–1748.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008b). An extended transcriptional network for pluripotency of embryonic stem cells. Cell *132*, 1049–1061.

Kim, J., Woo, A.J., Chu, J., Snow, J.W., Fujiwara, Y., Kim, C.G., Cantor, A.B., and Orkin, S.H. (2010). A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. Cell *143*, 313–324.

Kim, Y.J., Cecchini, K.R., and Kim, T.H. (2011). Conserved, developmentally regulated mechanism couples chromosomal looping and heterochromatin barrier activity at the homeobox gene A locus. Proc. Natl. Acad. Sci. U. S. A. *108*, 7391–7396.

Kindt, T.J.. R.A.G.B.A.O. (2006). Kuby Immunology (W.H.Freeman & Co Ltd; 6Rev Ed edition).

Kindt, T.J., Goldsby, R.A., Osborne, B.A., and Kuby, J. (2007). Kuby Immunology (W. H. Freeman).

Klein, F., Mahr, P., Galova, M., Buonomo, S.B., Michaelis, C., Nairz, K., and Nasmyth, K. (1999). A central role for cohesins in sister chromatid cohesion, formation of axial elements, and recombination during yeast meiosis. Cell *98*, 91–103.

Kon, A., Shih, L.-Y., Minamino, M., Sanada, M., Shiraishi, Y., Nagata, Y., Yoshida, K., Okuno, Y., Bando, M., Nakato, R., et al. (2013). Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. Nat. Genet. *45*, 1232–1237.

Kunath, T., Saba-El-Leil, M.K., Almousailleakh, M., Wray, J., Meloche, S., and Smith, A. (2007). FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. Development *134*, 2895–2902.

Laemmli, U.K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. Nature *227*, 680–685.

Lavagnolli, T. (2013). The role of cohesin in regulating early steps of cellular differentiation and reprogramming. Imperial College London.

Lavagnolli, T., Gupta, P., Hörmanseder, E., Mira-Bontenbal, H., Dharmalingam, G., Carroll, T., Gurdon, J.B., Fisher, A.G., and Merkenschlager, M. (2015). Initiation and maintenance of pluripotency gene expression in the absence of cohesin. Genes Dev. *29*, 23–38.

Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature *505*, 495–501.

Lee, J.-E., Wang, C., Xu, S., Cho, Y.-W., Wang, L., Feng, X., Baldridge, A., Sartorelli, V., Zhuang, L., Peng, W., et al. (2013). H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. Elife *2*, e01503.

Lefevre, P., Witham, J., Lacroix, C.E., Cockerill, P.N., and Bonifer, C. (2008). The LPS-induced transcriptional upregulation of the chicken lysozyme locus involves CTCF eviction and noncoding RNA transcription. Mol. Cell *32*, 129–139.

LeibundGut-Landmann, S., Waldburger, J.-M., Krawczyk, M., Otten, L. a, Suter, T., Fontana, A., Acha-Orbea, H., and Reith, W. (2004). Mini-review: Specificity and expression of CIITA, the master regulator of MHC class II genes. Eur. J. Immunol. *34*, 1513–1525.

Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., et al. (2013). Naive pluripotency is associated with global DNA hypomethylation. Nat. Struct. Mol. Biol. *20(3)*, 311–316.

Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G.P., Itoh, T., Watanabe, Y., Shirahige, K., and Uhlmann, F. (2004). Cohesin relocation from sites of chromosomal loading to places of convergent transcription. Nature *430*, 573–578.

Lennon, A.-M., Ottone, C., Rigaud, G., Deaven, L.L., Longmire, J., Fellous, M., Bono, R., and Alcaïde-Loridan, C. (1997). Isolation of a B-cell-specific promoter for the human class II transactivator. Immunogenetics *45*, 266–273.

Levens, D. (2010). You Don't Muck with MYC. Genes Cancer *1*, 547–554.

Li, M., He, Y., Dubois, W., Wu, X., Shi, J., and Huang, J. (2012). Distinct regulatory mechanisms and functions for p53-activated and p53-repressed DNA damage response genes in embryonic stem cells. Mol. Cell *46*, 30–42.

Lin, C., Garruss, A.S., Luo, Z., Guo, F., and Shilatifard, A. (2012a). The RNA Pol II Elongation Factor Ell3 Marks Enhancers in ES Cells and Primes Future Gene Activation. Cell *152*, 144–156.

Lin, C.Y., Lovén, J., Rahl, P.B., Paranal, R.M., Burge, C.B., Bradner, J.E., Lee, T.I., and Young, R.A. (2012b). Transcriptional amplification in tumor cells with elevated c-Myc. Cell *151*, 56–67.

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

Liu, J., and Levens, D. (2006). Making Myc. Curr. Top. Microbiol. Immunol. *302*, 1–32.

Liu, J., Zhang, Z., Bando, M., Itoh, T., Deardorff, M. a, Clark, D., Kaur, M., Tandy, S., Kondoh, T., Rappaport, E., et al. (2009). Transcriptional dysregulation in NIPBL and cohesin mutant human cells. PLoS Biol. *7*, e1000119.

Losada, A. (2014). Cohesin in cancer: chromosome segregation and beyond. Nat. Rev. Cancer *14*, 389–393.

Losada, A., Hirano, M., and Hirano, T. (1998). Identification of Xenopus SMC protein complexes required for sister chromatid cohesion. Genes Dev. *12*, 1986–1997.

Losada, A., Yokochi, T., Kobayashi, R., and Hirano, T. (2000). Identification and characterization of SA/Scc3p subunits in the Xenopus and human cohesin complexes. J. Cell Biol. *150*, 405–416.

Lyons, N.A., and Morgan, D.O. (2011). Cdk1-dependent destruction of Eco1 prevents cohesion establishment after S phase. Mol. Cell *42*, 378–389.

Magner, W.J., Kazim, A.L., Stewart, C., Romano, M.A., Catalano, G., Grande, C., Keiser, N., Santaniello, F., and Tomasi, T.B. (2000). Activation of MHC Class I, II, and CD40 Gene Expression by Histone Deacetylase Inhibitors. J. Immunol. *165*, 7017–7024.

Maimets, T., Neganova, I., Armstrong, L., and Lako, M. (2008). Activation of p53 by nutlin leads to rapid differentiation of human embryonic stem cells. Oncogene *27*, 5277–5287.

Majumder, P., and Boss, J.M. (2010). CTCF controls expression and chromatin architecture of the human major histocompatibility complex class II locus. Mol. Cell. Biol. *30*, 4211–4223.

Majumder, P., and Boss, J.M. (2011). Cohesin Regulates MHC Class II Genes through Interactions with MHC Class II Insulators. J. Immunol. *187*, 4236–4244.

Majumder, P., Gomez, J.A., and Boss, J.M. (2006). The human major histocompatibility complex class II HLA-DRB1 and HLA-DQA1 genes are separated by a CTCF-binding enhancer-blocking element. J. Biol. Chem. *281*, 18435–18443.

Majumder, P., Gomez, J.A., Chadwick, B.P., and Boss, J.M. (2008). The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. J. Exp. Med. *205*, 785–798.

Mallo, M., and Alonso, C.R. (2013). The regulation of Hox gene expression during animal development. Development *140*, 3951–3963.

Mannini, L., Cucco, F., Quarantotti, V., Krantz, I.D., and Musio, A. (2013). Mutation spectrum and genotype-phenotype correlation in Cornelia de Lange syndrome. Hum. Mutat. *34*, 1589–1596.

Mantovani, R. (1999). The molecular biology of the CCAAT-binding factor NF-Y. Gene *239*, 15–27.

Marks, H., and Stunnenberg, H.G. (2014). Transcription regulation and chromatin structure in the pluripotent ground state. Biochim. Biophys. Acta *1839*, 129–137.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Francis Stewart, a., Smith, A., et al. (2012). The Transcriptional and Epigenomic Foundations of Ground State Pluripotency. Cell *149*, 590–604.

Martin, B.K., Chin, K.C., Olsen, J.C., Skinner, C.A., Dey, A., Ozato, K., and Ting, J.P. (1997). Induction of MHC class I expression by the MHC class II transactivator CIITA. Immunity *6*, 591–600.

Masternak, K., Barras, E., Zufferey, M., Conrad, B., Corthals, G., Aebersold, R., Sanchez, J.C., Hochstrasser, D.F., Mach, B., and Reith, W. (1998). A gene encoding a novel RFX-associated transactivator is mutated in the majority of MHC class II deficiency patients. Nat. Genet. *20*, 273–277.

McEwan, M. V, Eccles, M.R., and Horsfield, J. a (2012). Cohesin is required for activation of MYC by estradiol. PLoS One *7*, e49160.

Merkenschlager, M. (2010). Cohesin: a global player in chromosome biology with local ties to gene regulation. Curr. Opin. Genet. Dev. *20*, 555–561.

Merkenschlager, M., and Odom, D.T. (2013). CTCF and cohesin: linking gene regulatory elements with their targets. Cell *152*, 1285–1297.

Meyer, N., and Penn, L.Z. (2008). Reflecting on 25 years with MYC. Nat. Rev. Cancer *8*, 976–990.

Michaelis, C., Ciosk, R., and Nasmyth, K. (1997). Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. Cell *91*, 35–45.

Mira-Bontenbal, H. (2011). Developmental regulation of cohesin positioning on mammalian chromosome arms. Imperial College London.

Mishiro, T., Ishihara, K., Hino, S., Tsutsumi, S., Aburatani, H., Shirahige, K., Kinoshita, Y., and Nakao, M. (2009). Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. EMBO J. *28*, 1234–1245.

Misulovin, Z., Schwartz, Y.B., Li, X.-Y., Kahn, T.G., Gause, M., MacArthur, S., Fay, J.C., Eisen, M.B., Pirrotta, V., Biggin, M.D., et al. (2008). Association of cohesin and Nipped-B with transcriptionally active regions of the Drosophila melanogaster genome. Chromosoma *117*, 89–102.

Monahan, K., Rudnick, N.D., Kehayova, P.D., Pauli, F., Newberry, K.M., Myers, R.M., and Maniatis, T. (2012). Role of CCCTC binding factor (CTCF) and cohesin in the generation of single-cell diversity of Protocadherin-α gene expression. Proc. Natl. Acad. Sci. U. S. A. *109*, 9125–9130.

Moreno, C.S., Beresford, G.W., Louis-Plence, P., Morris, A.C., and Boss, J.M. (1999). CREB regulates MHC class II expression in a CIITA-dependent manner. Immunity *10*, 143–151.

Morris, S.A., and Daley, G.Q. (2013). A blueprint for engineering cell fate: current technologies to reprogram cell identity. Cell Res. *23*, 33–48.

Morris, A.C., Spangler, W.E., and Boss, J.M. (2000). Methylation of Class II trans-Activator Promoter IV: A Novel Mechanism of MHC Class II Gene Control. J. Immunol. *164*, 4143–4149.

Morris, A.C., Beresford, G.W., Mooney, M.R., and Boss, J.M. (2002). Kinetics of a Gamma Interferon Response: Expression and Assembly of CIITA Promoter IV and Inhibition by Methylation. Mol. Cell. Biol. *22*, 4781–4791.

Muhlethaler-Mottet, A., Otten, L.A., Steimle, V., and Mach, B. (1997). Expression of MHC class II molecules in different cellular and functional compartments is controlled by differential usage of multiple promoters of the transactivator CIITA. EMBO J. *16*, 2851–2860.

Munder, M., Mallo, M., Eichmann, K., and Modolell, M. (2001). Direct stimulation of macrophages by IL-12 and IL-18 — a bridge built on solid ground. Immunol. Lett. *75*, 159–160.

Murayama, Y., and Uhlmann, F. (2014). Biochemical reconstitution of topological DNA binding by the cohesin ring. Nature *505*, 367–371.

Murrell, A., Heeson, S., and Reik, W. (2004). Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. Nat. Genet. *36*, 889–893.

Musio, A., Selicorni, A., Focarelli, M.L., Gervasini, C., Milani, D., Russo, S., Vezzoni, P., and Larizza, L. (2006). X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. Nat. Genet. *38*, 528–530.

Muto, A., Calof, A.L., Lander, A.D., and Schilling, T.F. (2011). Multifactorial Origins of Heart and Gut Defects in nipbl-Deficient Zebrafish, a Model of Cornelia de Lange Syndrome. PLoS Biol. *9*, e1001181.

Nagarajan, U.M., Louis-Plence, P., DeSandro, A., Nilsen, R., Bushey, A., and Boss, J.M. (1999). RFX-B is the gene responsible for the most common cause of the bare lymphocyte syndrome, an MHC class II immunodeficiency. Immunity *10*, 153–162.

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat. Biotechnol. *26*, 101–106.

Nasmyth, K., and Haering, C.H. (2009). Cohesin: its roles and mechanisms. Annu. Rev. Genet. *43*, 525–558.

Nativio, R., Wendt, K.S., Ito, Y., Huddleston, J.E., Uribe-Lewis, S., Woodfine, K., Krueger, C., Reik, W., Peters, J.-M., and Murrell, A. (2009). Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. PLoS Genet. *5*, e1000739.

Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B.R., Mirny, L.A., and Dekker, J. (2013). Organization of the mitotic chromosome. Science *342*, 948–953.

Ni, Z., Abou El Hassan, M., Xu, Z., Yu, T., and Bremner, R. (2008). The chromatin-remodeling enzyme BRG1 coordinates CIITA induction through many interdependent distal enhancers. Nat. Immunol. *9*, 785–793.

Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. Cell Stem Cell *4*, 487–492.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H., and Smith, A. (1998). Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4. Cell *95*, 379–391.

Nichols, J., Jones, K., Phillips, J.M., Newland, S.A., Roode, M., Mansfield, W., Smith, A., and Cooke, A. (2009). Validated germline-competent embryonic stem cell lines from nonobese diabetic mice. Nat. Med. *15*, 814–818.

Nie, Z., Hu, G., Wei, G., Cui, K., Yamane, A., Resch, W., Wang, R., Green, D.R., Tessarollo, L., Casellas, R., et al. (2012). c-Myc Is a Universal Amplifier of Expressed Genes in Lymphocytes and Embryonic Stem Cells. Cell *151*, 68–79.

Nishiyama, T., Ladurner, R., Schmitz, J., Kreidl, E., Schleiffer, A., Bhaskara, V., Bando, M., Shirahige, K., Hyman, A. a, Mechtler, K., et al. (2010). Sororin mediates sister chromatid cohesion by antagonizing Wapl. Cell *143*, 737–749.

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., et al. (2011). RAD21 Cooperates with

Pluripotency Transcription Factors in the Maintenance of Embryonic Stem Cell Identity. PLoS One *6*, e19470.

Niwa, H., Miyazaki, J., and Smith, a G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat. Genet. *24*, 372–376.

Van de Nobelen, S., Rosa-Garrido, M., Leers, J., Heath, H., Soochit, W., Joosen, L., Jonkers, I., Demmers, J., van der Reijden, M., Torrano, V., et al. (2010). CTCF regulates the local epigenetic state of ribosomal DNA repeats. Epigenetics Chromatin *3*, 19.

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., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Obaya, A.J., Mateyak, M.K., and Sedivy, J.M. (1999). Mysterious liaisons: the relationship between c-Myc and the cell cycle. Oncogene *18*, 2934–2941.

Ong, C.-T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. Nat. Rev. Genet. *15*, 234–246.

Osborne, A., Zhang, H., Yang, W.M., Seto, E., and Blanck, G. (2001). Histone deacetylase activity represses gamma interferon-inducible HLA-DR gene expression following the establishment of a DNase I-hypersensitive chromatin conformation. Mol. Cell. Biol. *21*, 6495–6506.

Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., Curina, A., Prosperini, E., Ghisletti, S., and Natoli, G. (2013). Latent enhancers activated by stimulation in differentiated cells. Cell *152*, 157–171.

Pai, R.K., Askew, D., Boom, W.H., and Harding, C. V (2002). Regulation of class II MHC expression in APCs: roles of types I, III, and IV class II transactivator. J. Immunol. *169*, 1326–1333.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., et al. (2008a). Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell *132*, 422–433.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., et al. (2008b). Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell *132*, 422–433.

Patel, S.R., Kim, D., Levitan, I., and Dressler, G.R. (2007). The BRCT-domain containing protein PTIP links PAX2 to a histone H3, lysine 4 methyltransferase complex. Dev. Cell *13*, 580–592.

Pattenden, S.G., Klose, R., Karaskov, E., and Bremner, R. (2002). Interferon-gamma-induced chromatin remodeling at the CIITA locus is BRG1 dependent. EMBO J. *21*, 1978–1986.

Pauli, A., Althoff, F., Oliveira, R. a, Heidmann, S., Schuldiner, O., Lehner, C.F., Dickson, B.J., and Nasmyth, K. (2008). Cell-type-specific TEV protease cleavage reveals cohesin functions in Drosophila neurons. Dev. Cell *14*, 239–251.

Pauli, A., van Bemmel, J.G., Oliveira, R.A., Itoh, T., Shirahige, K., van Steensel, B., and Nasmyth, K. (2010). A direct role for cohesin in gene regulation and ecdysone response in Drosophila salivary glands. Curr. Biol. *20*, 1787–1798.

Pereira, C.F., Terranova, R., Ryan, N.K., Santos, J., Morris, K.J., Cui, W., Merkenschlager, M., and Fisher, A.G. (2008). Heterokaryon-based reprogramming of human B lymphocytes for pluripotency requires Oct4 but not Sox2. PLoS Genet. *4*, e1000170.

Peters, J.-M., Tedeschi, A., and Schmitz, J. (2008). The cohesin complex and its roles in chromosome biology. Genes Dev. *22*, 3089–3114.

Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.-T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. Cell *153*, 1281–1299.

Piskurich, J.F., Lin, K.I., Lin, Y., Wang, Y., Ting, J.P., Calame, K., and Ii, C. (2000). BLIMP-I mediates extinction of major histocompatibility class II transactivator expression in plasma cells. Nat. Immunol. *1*, 526–532.

Pomerantz, M.M., Ahmadiyeh, N., Jia, L., Herman, P., Verzi, M.P., Doddapaneni, H., Beckwith, C.A., Chan, J.A., Hills, A., Davis, M., et al. (2009). The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. Nat. Genet. *41*, 882–884.

Prickett, A.R., Barkas, N., McCole, R.B., Hughes, S., Amante, S.M., Schulz, R., and Oakey, R.J. (2013). Genome-wide and parental allele-specific analysis of CTCF and cohesin DNA binding in mouse brain reveals a tissue-specific binding pattern and an association with imprinted differentially methylated regions. Genome Res. *23*, 1624–1635.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. a, Flynn, R. a, and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R. a, McCuine, S., Burge, C.B., Sharp, P. a P.A., and Young, R. a (2010). c-Myc regulates transcriptional pause release. Cell *141*, 432–445.

Remeseiro, S., and Losada, A. (2013). Cohesin, a chromatin engagement ring. Curr. Opin. Cell Biol. *25*, 63–71.

Remeseiro, S., Cuadrado, A., Gómez-López, G., Pisano, D.G., and Losada, A. (2012). A unique role of cohesin-SA1 in gene regulation and development. EMBO J. *31(9)*, 2090–2102.

Rhodes, J.M., Bentley, F.K., Print, C.G., Dorsett, D., Misulovin, Z., Dickinson, E.J., Crosier, K.E., Crosier, P.S., and Horsfield, J. a (2010). Positive regulation of c-Myc by cohesin is direct, and evolutionarily conserved. Dev. Biol. *344*, 637–649.

Ribeiro de Almeida, C., Stadhouders, R., de Bruijn, M.J.W., Bergen, I.M., Thongjuea, S., Lenhard, B., van Ijcken, W., Grosveld, F., Galjart, N., Soler, E., et al. (2011). The DNA-binding protein CTCF limits proximal Vκ recombination and restricts κ enhancer interactions to the immunoglobulin κ light chain locus. Immunity *35*, 501–513.

Rolef Ben-Shahar, T., Heeger, S., Lehane, C., East, P., Flynn, H., Skehel, M., and Uhlmann, F. (2008). Eco1-dependent cohesin acetylation during establishment of sister chromatid cohesion. Science *321*, 563–566.

Rollins, R.A., Morcillo, P., and Dorsett, D. (1999). Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. Genetics *152*, 577–593.

Roskoski, R. (2003). STI-571: an anticancer protein-tyrosine kinase inhibitor. Biochem. Biophys. Res. Commun. *309*, 709–717.

Ross, J. (1995). mRNA stability in mammalian cells. Microbiol. Rev. *59*, 423–450.

Rubio, E.D., Reiss, D.J., Welcsh, P.L., Disteche, C.M., Filippova, G.N., Baliga, N.S., Aebersold, R., Ranish, J. a, and Krumm, A. (2008). CTCF physically links cohesin to chromatin. Proc. Natl. Acad. Sci. U. S. A. *105*, 8309–8314.

Sabò, A., Kress, T.R., Pelizzola, M., de Pretis, S., Gorski, M.M., Tesi, A., Morelli, M.J., Bora, P., Doni, M., Verrecchia, A., et al. (2014). Selective transcriptional regulation by Myc in cellular growth control and lymphomagenesis. Nature *511*, 488–492.

Saijoh, Y., Adachi, H., Mochida, K., Ohishi, S., Hirao, A., and Hamada, H. (1999). Distinct transcriptional regulatory mechanisms underlie left-right asymmetric expression of lefty-1 and lefty-2. Genes Dev. *13*, 259–269.

Saunders, A., Core, L.J., and Lis, J.T. (2006). Breaking barriers to transcription elongation. Nat. Rev. Mol. Cell Biol. *7*, 557–567.

Schaaf, C. a, Misulovin, Z., Sahota, G., Siddiqui, A.M., Schwartz, Y.B., Kahn, T.G., Pirrotta, V., Gause, M., and Dorsett, D. (2009). Regulation of the Drosophila Enhancer of split and invected-engrailed gene complexes by sister chromatid cohesion proteins. PLoS One *4*, e6202.

Schaaf, C. a., Kwak, H., Koenig, A., Misulovin, Z., Gohara, D.W., Watson, A., Zhou, Y., Lis, J.T., and Dorsett, D. (2013). Genome-Wide Control of RNA Polymerase II Activity by Cohesin. PLoS Genet. *9*, e1003382.

Schindler, T., Bornmann, W., Pellicena, P., Miller, W.T., Clarkson, B., and Kuriyan, J. (2000). Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. Science *289*, 1938–1942.

Schleiffer, A., Kaitna, S., Maurer-Stroh, S., Glotzer, M., Nasmyth, K., and Eisenhaber, F. (2003). Kleisins: a superfamily of bacterial and eukaryotic SMC protein partners. Mol. Cell *11*, 571–575.

Schmidt, D., Schwalie, P.C., Ross-innes, C.S., Hurtado, A., Brown, G.D., Carroll, J.S., Flicek, P., and Odom, D.T. (2010). A CTCF-independent role for cohesin in tissue-specific transcription. Genome Res. 578–588.

Schroder, K., and Hertzog, P. (2004). Interferon-γ: an overview of signals, mechanisms and functions. J. Leukoc. … *75*, 163–189.

Schuldiner, O., Berdnik, D., Levy, J.M., Wu, J.S., Luginbuhl, D., Gontang, A.C., and Luo, L. (2008). piggyBac-based mosaic screen identifies a postmitotic function for cohesin in regulating developmental axon pruning. Dev. Cell *14*, 227–238.

Schwab, K.R., Patel, S.R., and Dressler, G.R. (2011). Role of PTIP in class switch recombination and long-range chromatin interactions at the immunoglobulin heavy chain locus. Mol. Cell. Biol. *31*, 1503–1511.

Sears, R.C. (2004). The Life Cycle of C-Myc: from synthesis to degradation. Cell Cycle *3*, 1133–1137.

Seitan, V.C., Banks, P., Laval, S., Majid, N.A., Dorsett, D., Rana, A., Smith, J., Bateman, A., Krpic, S., Hostert, A., et al. (2006). Metazoan Scc4 homologs link sister chromatid cohesion to cell and axon migration guidance. PLoS Biol. *4*, e242.

Seitan, V.C., Hao, B., Tachibana-Konwalski, K., Lavagnolli, T., Mira-Bontenbal, H., Brown, K.E., Teng, G., Carroll, T., Terry, A., Horan, K., et al. (2011). A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. Nature *476*, 467–471.

Seitan, V.C., Krangel, M.S., and Merkenschlager, M. (2012). Cohesin, CTCF and lymphocyte antigen receptor locus rearrangement. Trends Immunol. *33*, 153–159.

Seitan, V.C., Faure, A.J., Zhan, Y., Mccord, R.P., Lajoie, B.R., Ing-simmons, E., Lenhard, B., Giorgetti, L., Heard, E., Fisher, A.G., et al. (2013). Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. Genome Res. *23*, 2066–2077.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V. V., et al. (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116–120.

Shi, J., Whyte, W.A., Zepeda-mendoza, C.J., Milazzo, J.P., Shen, C., Roe, J., Minder, J.L., Mercan, F., Wang, E., Eckersley-Maslin, M.A., et al. (2013). Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. Genes Dev. 2648–2662.

Shilatifard, A. (2012). The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annu. Rev. Biochem. *81*, 65–95.

Shintomi, K., and Hirano, T. (2010). Sister chromatid resolution: a cohesin releasing network and beyond. Chromosoma *119*, 459–467.

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature *479*, 74–79.

Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. Cell *138*, 722–737.

Sjögren, C., and Nasmyth, K. (2001). Sister chromatid cohesion is required for postreplicative double-strand break repair in Saccharomyces cerevisiae. Curr. Biol. *11*, 991–995.

Skibbens, R. V, Colquhoun, J.M., Green, M.J., Molnar, C.A., Sin, D.N., Sullivan, B.J., and Tanzosh, E.E. (2013). Cohesinopathies of a feather flock together. PLoS Genet. *9*, e1004036.

Smith, T.A., and Hooper, M.L. (1983). Medium conditioned by feeder cells inhibits the differentiation of embryonal carcinoma cultures. Exp. Cell Res. *145*, 458–462.

Smith, A.G., Heath, J.K., Donaldson, D.D., Wong, G.G., Moreau, J., Stahl, M., and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature *336*, 688–690.

Smith, M.A., Wright, G., Wu, J., Tailor, P., Ozato, K., Chen, X., Wei, S., Piskurich, J.F., Ting, J.P.-Y., and Wright, K.L. (2011). Positive regulatory domain I (PRDM1) and IRF8/PU.1 counter-regulate MHC class II transactivator (CIITA) expression during dendritic cell maturation. J. Biol. Chem. *286*, 7893–7904.

Smith, T.G., Laval, S., Chen, F., Rock, M.J., Strachan, T., and Peters, H. (2014). Neural crest cell-specific inactivation of Nipbl or Mau2 during mouse development results in a late onset of craniofacial defects. Genesis *52*, 687–694.

Sofueva, S., Yaffe, E., Chan, W.-C., Georgopoulou, D., Vietri Rudan, M., Mira-Bontenbal, H., Pollard, S.M., Schroth, G.P., Tanay, A., and Hadjur, S. (2013). Cohesin-mediated interactions organize chromosomal domain architecture. EMBO J. *32*, 1–11.

Soler, E., Andrieu-Soler, C., de Boer, E., Bryne, J.C., Thongjuea, S., Stadhouders, R., Palstra, R.-J., Stevens, M., Kockx, C., van Ijcken, W., et al. (2010). The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. Genes Dev. *24*, 277–289.

Solomon, D.A., Kim, T., Diaz-Martinez, L.A., Fair, J., Elkahloun, A.G., Harris, B.T., Toretsky, J.A., Rosenberg, S.A., Shukla, N., Ladanyi, M., et al. (2011). Mutational inactivation of STAG2 causes aneuploidy in human cancer. Science *333*, 1039–1043.

Solomon, D.A., Kim, J.-S., Bondaruk, J., Shariat, S.F., Wang, Z.-F., Elkahloun, A.G., Ozawa, T., Gerard, J., Zhuang, D., Zhang, S., et al. (2013). Frequent truncating mutations of STAG2 in bladder cancer. Nat. Genet. *45*, 1428–1430.

Sopher, B.L., Ladd, P.D., Pineda, V. V, Libby, R.T., Sunkin, S.M., Hurley, J.B., Thienes, C.P., Gaasterland, T., Filippova, G.N., and La Spada, A.R. (2011). CTCF Regulates Ataxin-7 Expression through Promotion of a Convergently Transcribed, Antisense Noncoding RNA. Neuron *70*, 1071–1084.

Sotelo, J., Esposito, D., Duhagon, M.A., Banfield, K., Mehalko, J., Liao, H., Stephens, R.M., Harris, T.J.R., Munroe, D.J., and Wu, X. (2010). Long-range enhancers on 8q24 regulate c-Myc. Proc. Natl. Acad. Sci. U. S. A. *107*, 3001–3005.

Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell *151*, 994–1004.

Soza-Ried, J., and Fisher, A.G. (2012). Reprogramming somatic cells towards pluripotency by cellular fusion. Curr. Opin. Genet. Dev. *22*, 459–465.

Spencer, R.J., del Rosario, B.C., Pinter, S.F., Lessing, D., Sadreyev, R.I., and Lee, J.T. (2011). A boundary element between Tsix and Xist binds the chromatin insulator Ctcf and contributes to initiation of X-chromosome inactivation. Genetics *189*, 441–454.

Splinter, E., Heath, H., Kooren, J., Palstra, R.-J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev. *20*, 2349–2354.

Stead, E., White, J., Faast, R., Conn, S., Goldstone, S., Rathjen, J., Dhingra, U., Rathjen, P., Walker, D., and Dalton, S. (2002). Pluripotent cell division cycles are driven by ectopic Cdk2, cyclin A/E and E2F activities. Oncogene *21*, 8320–8333.

Stedman, W., Kang, H., Lin, S., Kissil, J.L., Bartolomei, M.S., and Lieberman, P.M. (2008). Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. EMBO J. *27*, 654–666.

Steimle, V., Otten, L.A., Zufferey, M., and Mach, B. (1993). Complementation cloning of an MHC class II transactivator mutated in hereditary MHC class II deficiency (or bare lymphocyte syndrome). Cell *75*, 135–146.

Steimle, V., Siegrist, C., Mottet, A., Lisowska-grospierre, B., and Mach, B. (1994). Regulation of MHC Class 11 Expression by Interferon-y Mediated by the Transactivator Gene CIITA. Science (80-.). *265*, 1–4.

Steimle, V., Durand, B., Barras, E., Zufferey, M., Hadam, M.R., Mach, B., and Reith, W. (1995). A novel DNA-binding regulatory factor is mutated in primary MHC class II deficiency (bare lymphocyte syndrome). Genes Dev. *9*, 1021–1032.

Van der Stoep, N., Quinten, E., Marcondes Rezende, M., and van den Elsen, P.J. (2004). E47, IRF-4, and PU.1 synergize to induce B-cell-specific activation of the class II transactivator promoter III (CIITA-PIII). Blood *104*, 2849–2857.

Subramaniam, P.S., Torres, B.A., and Johnson, H.M. (2001). So many ligands, so few transcription factors: a new paradigm for signaling through the STAT transcription factors. Cytokine *15*, 175–187.

Sumara, I., Vorlaufer, E., Gieffers, C., Peters, B.H., and Peters, J.-M.M. (2000). Characterization of vertebrate cohesin complexes and their regulation in prophase. J. Cell Biol. *151*, 749–762.

Sur, I.K., Hallikas, O., Vähärautio, A., Yan, J., Turunen, M., Enge, M., Taipale, M., Karhu, A., Aaltonen, L. a, and Taipale, J. (2012). Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. Science *338*, 1360–1363.

Tada, M., Tada, T., Lefebvre, L., Barton, S.C., and Surani, M.A. (1997). Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells. EMBO J. *16*, 6510–6520.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Takahashi, T.S., Yiu, P., Chou, M.F., Gygi, S., and Walter, J.C. (2004). Recruitment of Xenopus Scc2 and cohesin to chromatin requires the pre-replication complex. Nat. Cell Biol. *6*, 991–996.

Tedeschi, A., Wutz, G., Huet, S., Jaritz, M., Wuensche, A., Schirghuber, E., Davidson, I.F., Tang, W., Cisneros, D.A., Bhaskara, V., et al. (2013). Wapl is an essential regulator of chromatin structure and chromosome segregation. Nature *501*, 564–568.

Terret, M.-E., Sherwood, R., Rahman, S., Qin, J., and Jallepalli, P. V (2009). Cohesin acetylation speeds the replication fork. Nature *462*, 231–234.

Thanos, D., and Maniatis, T. (1995). Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. Cell *83*, 1091–1100.

Tomkins, D., Hunter, A., and Roberts, M. (1979). Cytogenetic findings in Roberts-SC phocomelia syndrome(s). Am. J. Med. Genet. *4*, 17–26.

Torres-Padilla, M.-E., and Chambers, I. (2014). Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage. Development *141*, 2173–2181.

Trotter, K.W., and Archer, T.K. (2008). The BRG1 transcriptional coregulator. Nucl. Recept. Signal. *6*, e004.

Tseng, Y.-Y., Moriarity, B.S., Gong, W., Akiyama, R., Tiwari, A., Kawakami, H., Ronning, P., Reuland, B., Guenther, K., Beadnell, T.C., et al. (2014). PVT1 dependence in cancer with MYC copy-number increase. Nature *512*, 82–86.

Tsubouchi, T., Soza-Ried, J., Brown, K., Piccolo, F.M.M., Cantone, I., Landeira, D., Bagci, H., Hochegger, H., Merkenschlager, M., and Fisher, A.G.G. (2013). DNA synthesis is required for reprogramming mediated by stem cell fusion. Cell *152*, 873–883.

Tuupanen, S., Turunen, M., Lehtonen, R., Hallikas, O., Vanharanta, S., Kivioja, T., Björklund, M., Wei, G., Yan, J., Niittymäki, I., et al. (2009). The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. Nat. Genet. *41*, 885–890.

Uhlmann, F., and Nasmyth, K. (1998). Cohesion between sister chromatids must be established during DNA replication. Curr. Biol. *8*, 1095–1101.

Unal, E., Heidinger-Pauli, J.M., Kim, W., Guacci, V., Onn, I., Gygi, S.P., and Koshland, D.E. (2008). A molecular determinant for the establishment of sister chromatid cohesion. Science *321*, 566–569.

Vastenhouw, N.L., and Schier, A.F. (2012). Bivalent histone modifications in early embryogenesis. Curr. Opin. Cell Biol. *24*, 374–386.

Vega, H., Waisfisz, Q., Gordillo, M., Sakai, N., Yanagihara, I., Yamada, M., van Gosliga, D., Kayserili, H., Xu, C., Ozono, K., et al. (2005). Roberts syndrome is caused by mutations in ESCO2, a human homolog of yeast ECO1 that is essential for the establishment of sister chromatid cohesion. Nat. Genet. *37*, 468–470.

Vernì, F., Gandhi, R., Goldberg, M.L., and Gatti, M. (2000). Genetic and molecular analysis of wings apart-like (wapl), a gene controlling heterochromatin organization in Drosophila melanogaster. Genetics *154*, 1693–1710.

Waldburger, J.M., Suter, T., Fontana, a, Acha-Orbea, H., and Reith, W. (2001). Selective abrogation of major histocompatibility complex class II expression on extrahematopoietic cells in mice lacking promoter IV of the class II transactivator gene. J. Exp. Med. *194*, 393–406.

Walz, S., Lorenzin, F., Morton, J., Wiese, K.E., von Eyss, B., Herold, S., Rycak, L., Dumay-Odelot, H., Karim, S., Bartkuhn, M., et al. (2014). Activation and repression by oncogenic MYC shape tumour-specific gene expression profiles. Nature *511*, 483–487.

Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M.U., Ohgi, K. a, et al. (2011a). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature *474*, 390–394.

Wang, R., Dillon, C.P., Shi, L.Z., Milasta, S., Carter, R., Finkelstein, D., McCormick, L.L., Fitzgerald, P., Chi, H., Munger, J., et al. (2011b). The transcription factor Myc controls metabolic reprogramming upon T lymphocyte activation. Immunity *35*, 871–882.

Wasserman, N.F., Aneas, I., and Nobrega, M. a (2010). An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. Genome Res. *20*, 1191–1197.

Watrin, E., Schleiffer, A., Tanaka, K., Eisenhaber, F., Nasmyth, K., and Peters, J.-M. (2006). Human Scc4 is required for cohesin binding to chromatin, sister-chromatid cohesion, and mitotic progression. Curr. Biol. *16*, 863–874.

Wei, Z., Gao, F., Kim, S., Yang, H., Lyu, J., An, W., Wang, K., and Lu, W. (2013). Klf4 organizes long-range chromosomal interactions with the oct4 locus in reprogramming and pluripotency. Cell Stem Cell *13*, 36–47.

Weitzer, S., Lehane, C., and Uhlmann, F. (2003). A model for ATP hydrolysis-dependent binding of cohesin to DNA. Curr. Biol. *13*, 1930–1940.

Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature *451*, 796–801.

Wernig, M., Meissner, A., Cassady, J.P., and Jaenisch, R. (2008). c-Myc is dispensable for direct reprogramming of mouse fibroblasts. Cell Stem Cell *2*, 10–12.

Whelan, G., Kreidl, E., Wutz, G., Egner, A., Peters, J.-M., and Eichele, G. (2012). Cohesin acetyltransferase Esco2 is a cell viability factor and is required for cohesion in pericentric heterochromatin. EMBO J. *31*, 71–82.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. Cell *153*, 307–319.

Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., and Campbell, K.H. (1997). Viable offspring derived from fetal and adult mammalian cells. Nature *385*, 810–813.

Wilson, A., Murphy, M.J., Oskarsson, T., Kaloulis, K., Bettess, M.D., Oser, G.M., Pasche, A.-C., Knabenhans, C., Macdonald, H.R., and Trumpp, A. (2004). c-Myc controls the balance between hematopoietic stem cell self-renewal and differentiation. Genes Dev. *18*, 2747–2763.

De Wit, E., Bouwman, B.A.M., Zhu, Y., Klous, P., Splinter, E., Verstegen, M.J.A.M., Krijger, P.H.L., Festuccia, N., Nora, E.P., Welling, M., et al. (2013). The pluripotent genome in three dimensions is shaped around pluripotency factors. Nature *501*, 227–231.

Wray, J., Kalkan, T., and Smith, A.G. (2010). The ground state of pluripotency. Biochem. Soc. Trans. *38*, 1027–1032.

Wray, J., Kalkan, T., Gomez-Lopez, S., Eckardt, D., Cook, A., Kemler, R., and Smith, A. (2011). Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. Nat. Cell Biol. *13*, 838–845.

Wright, J.B., Brown, S.J., and Cole, M.D. (2010). Upregulation of c-MYC in cis through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells. Mol. Cell. Biol. *30*, 1411–1420.

Wu, D., and Pan, W. (2010). GSK3: a multifaceted kinase in Wnt signaling. Trends Biochem. Sci. *35*, 161–168.

Wu, N., and Yu, H. (2012). The Smc complexes in DNA damage response. Cell Biosci. *2*, 5.

Wu, J., Prindle, M.J., Dressler, G.R., and Yu, X. (2009). PTIP regulates 53BP1 and SMC1 at the DNA damage sites. J. Biol. Chem. *284*, 18078–18084.

Wu, Q., Chen, X., Zhang, J., Loh, Y.-H., Low, T.-Y., Zhang, W., Zhang, W., Sze, S.-K., Lim, B., and Ng, H.-H. (2006). Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. J. Biol. Chem. *281*, 24090–24094.

Xie, X., Lu, J., Kulbokas, E.J., Golub, T.R., Mootha, V., Lindblad-Toh, K., Lander, E.S., and Kellis, M. (2005). Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. Nature *434*, 338–345.

Yamane, A., Resch, W., Kuo, N., Kuchen, S., Li, Z., Sun, H., Robbiani, D.F., McBride, K., Nussenzweig, M.C., and Casellas, R. (2011). Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. Nat. Immunol. *12*, 62–69.

Yeo, J.-C., and Ng, H.-H. (2013). The transcriptional regulation of pluripotency. Cell Res. *23*, 20–32.

Ying, Q.-L., Nichols, J., Chambers, I., and Smith, A. (2003a). BMP Induction of Id Proteins Suppresses Differentiation and Sustains Embryonic Stem Cell Self-Renewal in Collaboration with STAT3. Cell *115*, 281–292.

Ying, Q.-L., Stavridis, M., Griffiths, D., Li, M., and Smith, A. (2003b). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat. Biotechnol. *21*, 183–186.

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519–523.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev. *25*, 2227–2241.

Zhang, B., Jain, S., Song, H., Fu, M., Heuckeroth, R.O., Erlich, J.M., Jay, P.Y., and Milbrandt, J. (2007). Mice lacking sister chromatid cohesion protein PDS5B exhibit developmental abnormalities reminiscent of Cornelia de Lange syndrome. Development *134*, 3191–3201.

Zhang, H., Jiao, W., Sun, L., Fan, J., Chen, M., Wang, H., Xu, X., Shen, A., Li, T., Niu, B., et al. (2013a). Intrachromosomal looping is required for activation of endogenous pluripotency genes during reprogramming. Cell Stem Cell *13*, 30–35.

Zhang, J., Shi, X., Li, Y., Kim, B.-J., Jia, J., Huang, Z., Yang, T., Fu, X., Jung, S.Y., Wang, Y., et al. (2008). Acetylation of Smc3 by Eco1 is required for S phase sister chromatid cohesion in both human and yeast. Mol. Cell *31*, 143–151.

Zhang, Y., Wong, C.-H., Birnbaum, R.Y., Li, G., Favaro, R., Ngan, C.Y., Lim, J., Tai, E., Poh, H.M., Wong, E., et al. (2013b). Chromatin connectivity maps reveal dynamic promoter–enhancer long-range associations. Nature *504*, 306–310.

Zhou, Q., Li, T., and Price, D.H. (2012). RNA polymerase II elongation control. Annu. Rev. Biochem. *81*, 119–143.

Zhu, X.-S., Linhoff, M.W., Li, G., Chin, K.-C., Maity, S.N., and Ting, J.P.-Y. (2000). Transcriptional Scaffold: CIITA Interacts with NF-Y, RFX, and CREB To Cause Stereospecific Regulation of the Class II Major Histocompatibility Complex Promoter. Mol. Cell. Biol. *20*, 6051–6061.

Zika, E., and Ting, J.P.-Y. (2005). Epigenetic control of MHC-II: interplay between CIITA and histone-modifying enzymes. Curr. Opin. Immunol. *17*, 58–64.

Zika, E., Greer, S.F., Zhu, X.-S., and Ting, J.P.-Y. (2003). Histone deacetylase 1/mSin3A disrupts gamma interferon-induced CIITA function and major histocompatibility complex class II enhanceosome formation. Mol. Cell. Biol. *23*, 3091–3102.

Zuin, J., Dixon, J.R., van der Reijden, M.I.J. a., Ye, Z., Kolovos, P., Brouwer, R.W.W., van de Corput, M.P.C., van de Werken, H.J.G., Knoch, T. a., van IJcken, W.F.J., et al. (2013). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc. Natl. Acad. Sci. *111*, 996–1001.

Appendix 1: Gene expression in ES cells changes with cellular density in the culture dish over time

A confluent well of a 6-well plate of ES cells growing in 2i media was trypsinised (0h sample) and $0.5x10^6$ cells were replated onto new wells and seeded with fresh media. Samples were collected at the indicated time points below and gene expression was analysed by RT-qPCR. Results showed that Lefty1, Lefty2 and Myc expression varied over time after the initial replating of the cells while oct4 was more uniform. At lower density at 12h, Lefty genes showed reduced expression while Myc expression was increased. The gene expression profile of confluent cells was regained after 48 hours. Therefore, the gene expression changes were inherently dynamic and not a permanent effect.

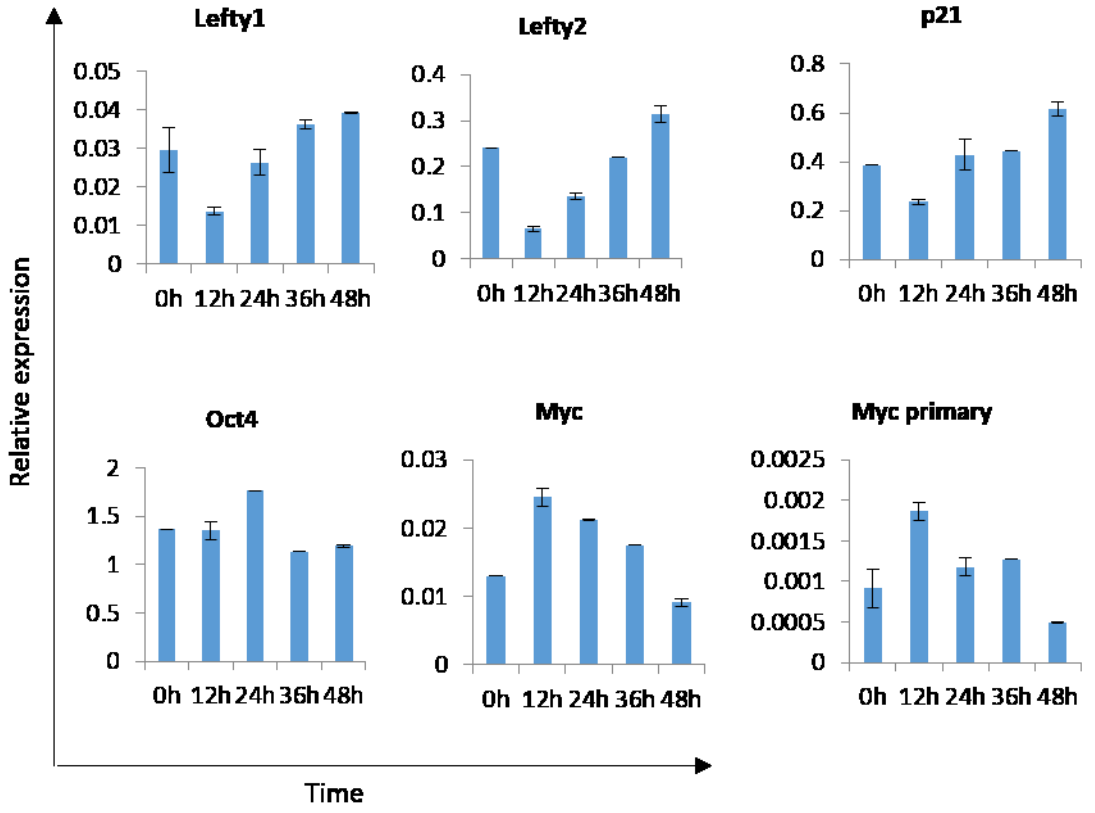


Figure: Relative gene expression profiles of pluripotency associated genes after different times of plating the ES cells on culture dish. Expression is normalised to *Ubc* and *Ywhaz*.

Appendix 2: *Nipbl* depletion does not affect the upregulation of *Myc* expression upon activation of naïve CD4$^+$CD25$^-$ T cells

Naïve CD4$^+$ T cells are quiescent non-dividing cells. These cells can be activated by simultaneous engagement of the TCR-CD3 complex and the co-stimulatory molecule CD28. Upon activation, CD4$^+$ T cells undergo a growth phase of ~24 hr, followed by massive clonal expansion and differentiation phases that are essential for appropriate immune defence and regulation. This process of CD4$^+$ T cell activation is accompanied by the induction of *Myc* expression (Wang et al., 2011b). Therefore, the activation of CD4$^+$ T cells provides another opportunity to investigate the role of cohesin in regulating *Myc* expression in response to extracellular signals. However, the number of peripheral T cells in CD4Cre-*Rad21*$^{lox/lox}$ mice is greatly reduced, probably due to the requirement of RAD21 in proper TCRα rearrangement (Seitan et al., 2011). Instead, I sought to determine if the cohesin loading factor NIPBL is required for the induction of *Myc* expression.

For this purpose, I used CD4$^+$ T cells from the lymph nodes of CD4Cre-*Nipbl* $^{lox/WT}$ and CD4Cre-*Nipbl* $^{lox/lox}$ mice. The CD4Cre-*Nipbl*^{lox} mice were generated by crossing the mice carrying Cre recombinase transgene under the control of *Cd4* regulatory elements with those containing the floxed *Nipbl* allele where two *loxP* sites flank exon 2. Cre-mediated recombination results in a *Nipbl* null allele lacking exon 2 and the start codon (Smith et al., 2014). CD4$^+$CD25$^-$ naïve T cells were isolated by FACS (Figure 4.11A) and activated using anti-CD3 plus anti-CD8 coated Dynabeads *in vitro*. RNA was extracted from the activated CD4$^+$ T cells obtained from CD4Cre-*Nipbl* $^{lox/WT}$ and CD4Cre-*Nipbl* $^{lox/lox}$ mice. Gene expression was analysed by qRT-PCR. Experimental results showed an effective depletion of Nipbl mRNA in CD4Cre-*Nipbl* $^{lox/lox}$ CD4$^+$ T cells (Figure 4.11B). The reduced Nipbl mRNA levels observed in activated CD4Cre-*Nipbl* $^{lox/WT}$ T cells is an artefact of PCR normalisation (T cell activation leads to an overall increased gene expression including that of housekeeping genes but not *Nipbl*. Therefore, standard PCR normalisation using housekeeping genes shows an apparent decrease in *Nipbl* expression. mRNA copy number analysis should instead be used ideally to compare changes in gene expression – unpublished work by Luke Williams). The expression of *Il2* was analysed as an indicator of CD4$^+$ T cell activation. The Il2 expression was reduced in activated CD4Cre-*Nipbl* $^{lox/lox}$ T cells as compared to the *Nipbl* heterozygous (CD4Cre-*Nipbl* $^{lox/WT}$) T cells. *Myc* mRNA levels, on the other hand, were appropriately upregulated in both CD4Cre-*Nipbl* $^{lox/lox}$ and CD4Cre-*Nipbl* $^{lox/WT}$ T cells. Interestingly, *Myc* primary transcript levels were not increased upon T cell activation (Figure 4.11B). This shows that the T cell activation-associated *Myc* upregulation is a consequence of increased stability of

the *Myc* transcript and not due to an increase in transcription of the *Myc* gene. Overall, this experiment showed that Nipbl is not required for increased *Myc* expression upon $CD4^+$ T cell activation.

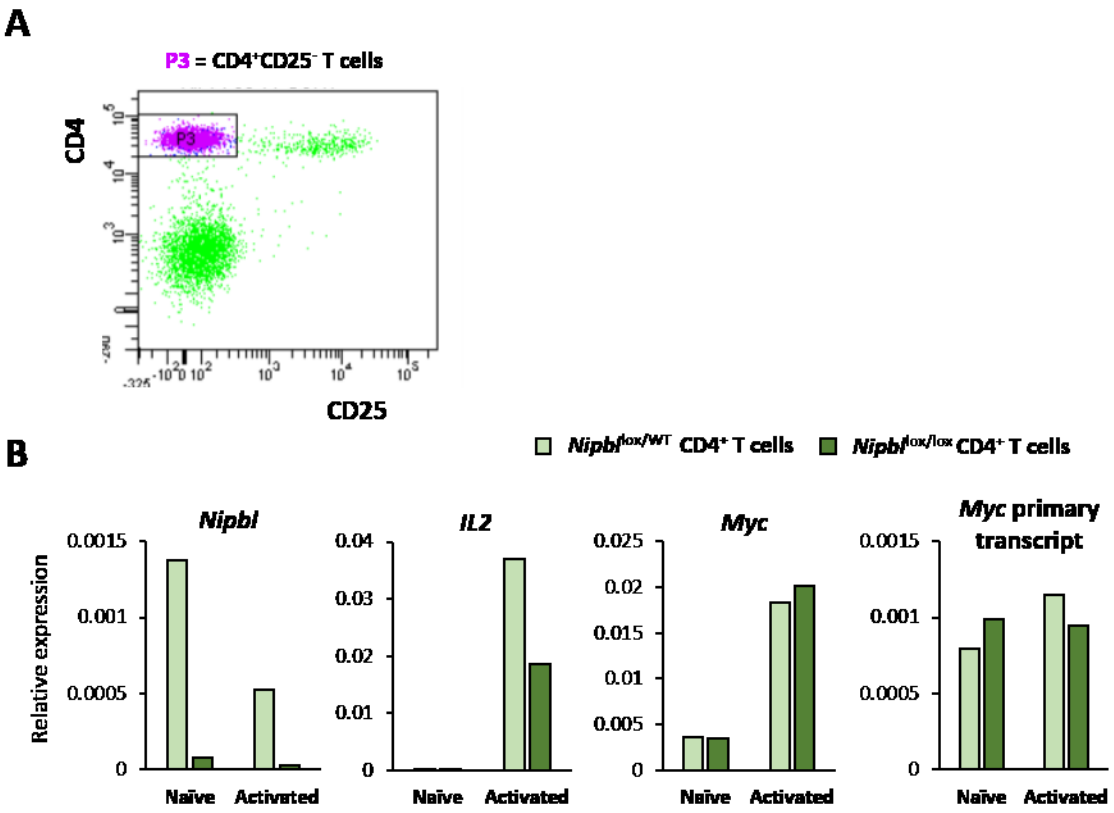


Figure : *Myc* upregulation upon $CD4^+$ T cell activation does not depend on Nipbl. A) FACS profile of the CD4 and CD25 stained cells from lymph node. The sorted $CD4^+CD25^-$ naïve T cells from CD4Cre-*Nipbl*$^{lox/WT}$ and CD4Cre-*Nipbl* $^{lox/lox}$ mice were collected and activated by simultaneous stimulation of the CD3 and CD28 receptors. **B)** qRT-PCR gene expression analysis in naïve and activated $CD4^+$ T cells. *Myc* mRNA levels appropriately upregulated in the absence of *Nipbl* (Normalised to b*Actin*).

Appendix 3: Publications

Parts of the work presented here have been published as follows:

Lavagnolli, T., Gupta, P., Hörmanseder, E., Mira-Bontenbal, H., Dharmalingam, G., Carroll, T., Gurdon, J.B., Fisher, A.G., and Merkenschlager, M. (2015). Initiation and maintenance of pluripotency gene expression in the absence of cohesin. Genes Dev. *29*, 23–38.