Interplay between chromatin conformation and transcription in eukaryotes



Shweta Bhardwaj

University of Oxford

Sir William Dunn School of Pathology

And

Lincoln College

Thesis submitted for the degree of Doctor of Philosophy at the University of Oxford

Trinity Term 2013

ABSTRACT

The three-dimensional organization of the genome is important for various processes such as transcription, replication, and repair. Several studies have shown that the genome is organized into long-range and short-range chromatin loops. Gene loops represent a short-range chromatin loop, synonymous with the juxtaposition of promoter and terminator regions of a gene. In Chapter III, I investigate the mechanism of gene-loop formation in a constitutively expressed gene, mouse serum albumin *(Alb).* The *Alb* locus appears to exist in a clover-leaf structure, with the promoter in close physical proximity with an upstream enhancer and downstream genic sequences. Furthermore, *Alb* forms a promoter-terminator gene loop that is dependent on serine 2 phosphorylation of RNA polymerase C-terminal domain. I also investigate the presence of gene loop conformation at the human Nuclear factor NF-kappa-B (*NFκB1*) gene. In response to cytokine stimulation, *NFκB1* transcription proceeds as a wave, with nascent RNA appearing as RNA polymerase traverses along the gene length. This coincides with formation of transient contacts between *NFκB1* promoter and genic regions.

The cohesin complex is a key mediator of chromatin loops and sister chromatid cohesion. The association of cohesin with chromatin is dependent on the loading complex, Mis4/Ssl3. In Chapter IV, I provide direct evidence for two functionally different populations of cohesin is *Schizosaccharomyces pombe*. Cohesin that co-localizes with Mis4 represents “cohesive” cohesin. In contrast, cohesin alone is unable to maintain stable sister chromatid cohesion, therefore, “less-cohesive” cohesin. Cohesive cohesin ensures stable cohesion because it is acetylated by the Eso1 acetyltransferase, which preferentially interacts with Mis4. In contrast, less-cohesive cohesin may function in recombination/repair. In Chapter V, I have identified a novel interplay between cohesin loading and transcription by RNA polymerase II. Inhibition of transcription initiation results in loss of Mis4 and consequently, cohesin binding on chromosomal arms regions. Furthermore, cohesin and Mis4 physically interact with RNA polymerase II. In Chapter VI, I summarize the above findings and propose a model that describes the stepwise loading of cohesin onto chromosomal arms during the fission yeast cell cycle. To conclude, I discuss the importance of understanding cohesin and its functions in health and diseases.

ACKNOWLEDGEMENT

Whilst this DPhil will forever remain my accomplishment, it would have not been possible without the support, guidance, love and affection of family, friends and colleagues. I take this opportunity to express my sincere gratitude to everyone who helped me realize this goal.

First, I sincerely thank Professor Nicholas J Proudfoot for giving me the opportunity to pursue doctoral research in his lab. The time spent in his lab has immensely contributed towards my growth as an independent researcher. I thank all the lab members for the cheerful environment and continuous encouragement, especially through the tough times.

I am grateful to the Felix Trust for awarding me full scholarship to pursue my DPhil at Oxford, an experience that has taught me to strive for the best.

I am thankful to Akis Papantonis for helpful discussions and selflessly teaching me the little tricks of chromosome conformation capture, my favourite experiment!

I will forever remain grateful to Monika Gullerova, for without her this thesis would not have been possible. She displayed conviction and reinstated my confidence in the one year that we worked together.

I thank Sabya and Navneet for all the wonderful chitchats at our favourite Indian restaurant in Oxford, “4550 miles from Delhi”.

From the bottom of my heart, I thank Vardhman for being an incredibly amazing friend. I adore your sense of humour and am thankful for your calming presence in my life.

At last, I am utmost grateful to my parents and my darling little brother Arjun, for letting me fulfill my dreams. Mom and dad, I will always cherish the values and principles you have bestowed upon me that make me a kind-hearted, hard-working and affable person. Most of all, I owe my success to Arjun, for its his spiced ginger tea that helped me concentrate during my late night studies that eventually brought me to Oxford. You people are my strength and I love you all very much.

TABLE OF CONTENTS

ABSTRACT 1
ACKNOWLEDEGMENT 2
TABLE OF CONTENTS 3
GLOSSARY OF ABBREVIATIONS 7
I. INTRODUCTION 11
I.1 Territorial organization of the nucleus: Chromosome territories 13
I.1.1 Inter-chromosomal interactions: long range associations between chromosome territories 15
I.1.2 Intra-chromosomal interactions: long range associations within chromosome territories 16
I.1.3 Gene loops 17
I.2 Methods to identify long-range chromatin interactions 20
I.2.1 Chromosome Conformation Capture (3C) 20
I.2.2 Chromosome Conformation Capture-on-ChIP (4C) 23
I.2.3 Chromosome Conformation Capture Carbon Copy (5C) 23
I.2.4 Combined Chromosome Conformation Capture-ChIP-Cloning (6C) 24
I.2.5 HiC 24
I.2.6 Chromatin Interaction Analysis by Paired-end Tag Sequencing (CHIA-PET) 25
I.2.7 Is every interaction a signal? 26
I.3 Organization of transcription in 3D 27
I.4 Sister Chromatid Cohesion and Chromosome Segregation 30
I.4.1 The Cohesin Complex 30
I.4.2 The Cohesin Ring Model 32
I.4.3 Cohesin during the cell cycle 33
I.4.3a Cohesin loading occurs during Gap-1 (G1) and depends on a separate loading complex 33
I.4.3b Cohesion establishment occurs during the synthesis (S) phase 35
I.4.3c Cohesion is maintained throughout Gap-2 (G2) phase 37
I.4.3d Resolution of cohesion in the mitotic (M) phase 38
I.4.4 Cohesin and gene regulation 39
I.4.5 Cohesin: An organizer of interphase chromatin 41
I.5 Aims of thesis 42

II. MATERIALS AND METHODS **43**
II.1 Mammalian Techniques 44
II.1.1 Mammalian Cell Culture 44
II.1.2 RNA isolation 44
II.1.3 Reverse Transcription (RT) PCR 44
II.1.4 Standard PCR 44
II.1.5 Quantitative Real-Time (qRT) PCR 45
II.1.6 3'-RACE 45
II.1.7 Chromatin Immunoprecipitation (ChIP) analysis 46
II.1.8 Chromosome Conformation Capture (3C) 48
II.1.9 RNA-Fluorescent *in situ* hybridisation (FISH) analysis 49
II.2 Fission Yeast Techniques 51
II.2.1 *S. pombe* growth conditions 51
II.2.2 Cell synchronisation by Nitrogen starvation 51
II.2.3 DNA isolation 52
II.2.4 RNA isolation 52
II.2.5 Flow Cytometry analysis (FACS) 53
II.2.6 ChIP analysis 53
II.2.7 PCR for FISH probes 55
II.2.8 DNA FISH 55
II.2.9 co-Immunoprecipitation (co-IP) 58
II.2.10 Mass Spectrometry (MS) 59
II.2.10a Co-IP 59
II.2.10b Tryptic Digest 59
II.2.10c C18 Desalting 59
II.2.10d MS data analysis 60
II.2.11 Bioinformatics analysis 60
II.2.11a Gene expression analysis 60
II.2.11b Comparison to the background distribution 61
II.2.11c Permutation Test 61
II.2.11d Boxplots 61
II.3 Western Blot Analysis 62
II.4 Reagents 64
II.5 Enzymes 65
II.6 Antibodies 65

II.7 Primers for *Alb* analysis 66
II.8 Primers for *NFκB1* analysis 68
II.9 Regions selected for DNA-FISH analysis (Chapter-IV) 69
II.10 Primers for *S. pombe* analysis 70
II.11 List of strains 73

III. MECHANISTIC INSIGHTS INTO GENE LOOP FORMATION: CONSTITUTIVE vs INDUCIBLE MODELS OF GENE LOOPING 74

III.1 Constitutive model of gene loop formation: The mouse serum albumin gene 75
III.1.1 Expression of *Alb* mRNA in permissive and non-permissive cell lines 76
III.1.2 Validation of the *Alb* transcription termination region 78
III.1.3 RNA Polymerase II binding profile across the albumin locus 78
III.1.4 *cis* interactions at the albumin locus 80
III.1.5 *cis* interactions at the *Alb* locus are dependent on basal transcription 85
III.2 Inducible model of gene loop formation: The nuclear factor of kappa B light chain 1, *NFκB1* gene 88
III.2.1 A wave of nascent transcription across *NFκB1* 89
III.2.2 A wave of transcribing polymerases along *NFκB1* 93
III.2.3 Gene loop formation at *NFκB1* 93
III.2.4 RNA-FISH analysis of *NFκB1* 96
III.3 Discussion 99

IV. "COHESIVE" VERSUS "LESS-COHESIVE" COHESIN 107

IV.1 Validation of Mis4/Ssl3 and Cohesin binding pattern in *S. pombe* 109
IV.2 "Cohesive" versus "Less-Cohesive" Cohesin 109
IV.3 Loss of sister chromatid cohesion in *rad21ts* 114
IV.4 Loss of sister chromatid cohesion in *mis4ts* 117
IV.5 Stoichiometric differences in cohesin occupancy at "Cohesive" versus "Less-Cohesive" sites 117
IV.6 Eso1 differentiates between cohesive and less-cohesive sites 120
IV.7 "Cohesive" sites preferentially associate with actively transcribed genes 123
IV.8 Discussion 127

V. COHESIN AND TRANSCRIPTION: A REQUIREMENT OF RNAPII FOR CHROMOSOMAL ASSOCIATION OF COHESIN **132**

V.1 RNAPII occupancy in *rad21ts* and *mis4ts* 134

V.2 Loss of cohesin after inhibition of transcription 135

V.3 Loss of MIS4 binding after transcription inhibition 143

V.4 Transcription inhibition impairs cohesin binding in a *Δwapl* background 145

V.5 Loss of cohesin binding after inhibition of transcription by 1,10-phenanthroline 148

V.6 Cohesin interacts with RNAPII 151

V.7 Mass spectrometry confirms the interaction between cohesin complex and RNAPII 154

V.8 Cohesin interacts with transcription elongation factor Spt5 158

V.9 Discussion 158

VI. GENERAL DISCUSSION **162**

VII. REFERENCES **171**

GLOSSARY OF ABBREVIATIONS

A	Adenine
ATP	Adenosine Triphosphate
ABC	ATP-binding cassette
bp	base pairs
C	Cytosine
Cdk	Cyclin-dependent kinase
cDNA	complementary DNA
CFIA	Cleavage Factor IA
CFIB	Cleavage Factor IB
CFIm	mammalian Cleavage Factor I
CFIIm	mammalian Cleavage Factor II
ChIP	Chromatin Immuno-Precipitation
CPF	Cleavage and Polyadenylation Factor
CPSF	Cleavage and Polyadenylation Specificity Factor
CstF	Cleavage stimulatory Factor
CTD	Carboxyl-Terminal Domain
CTCF	CCCTC-binding protein
3D	three- dimensional
3C	Chromosome Conformation Capture
4C	Chromosome Conformation Capture on ChIP
5C	Chromosome Conformation Capture Carbon Copy
6C	Chromosome Conformation Capture-ChIP-Cloning
ChIA-PET	Chromatin Interaction Analysis by Paired-End Sequencing
CdLS	Cornelia de Lange Syndrome
Co-IP	co-immunoprecipitation

DAPI	4’,6-diaminidino-2-phenylindole
DNA	Deoxyribonucleic Acid
DNase	Deoxyribonuclease
dNTP	Deoxyribonucleotide triphosphate
DSB	Double strand break
DTT	Dithiothreitol
DSIF	DRB-sensitivity inducing factor
EDTA	Ethylenediaminetetraacetic Acid (disodium salt)
EMM	Edinburgh Minimal Medium
G	Guanine
GFP	Green Fluorescent protein
HEPES	N-2-Hydroxyethylpiparizine-N’-2-ethanesulphonic acid
HIV	Human Immunodeficiency Virus
IP	Immuno-Precipitation
kbp	kilobase
LTR	Long Terminal Repeat
MEM	Minimal Essential Media
min	minutes
ml	millilitre
mRNA	messenger RNA
MSA	Mouse Serum Albumin
NELF	Negative Elongation Factor
NTP	Ribonuceloside Triphosphate
OD	Optical Density
ORF	Open Reading Frame
PABP	Poly(A) Binding Protein
PAGE	Polyacrylamide Gel Electrophoresis

PAP	Poly (A) Polymerase
PBS	Phosphate Buffered Saline
Pcf11	Polyadenylation cleavage factor 11
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
PET	PBS/EDTA/Trypsin
PIC	Pre-Initiation Complex
PIPES	Piperazine-N, N'-bis 2-ethanesulphonic acid
PMSF	Phenylmethylsulfonyl Fluoride
RNAPII	RNA Polymerase II
Poly(A)	Polyadenylation
PSF	Protein-associated Splicing Factor
P-TEFb	Positive Transcription Elongation Factor b
rpm	revolutions per minute
RNA	Ribonucleic Acid
RNase	Ribonuclease
rNTPs	Ribonucleotide Triphosphates
rRNA	ribosomal RNA
RT	Reverse Transcription
S. cerevisiae	*Saccharomyces cerevisiae*
S. pombe	*Schizosaccharomyces pombe*
SDS	Sodium Dodecyl Sulphate
Ser2P	CTD Ser 2 phosphorylated
Ser5P	CTD Ser 5 phosphorylated
Ser7P	CTD Ser 7 phosphorylated
siRNA	small interfering RNA
snRNA	small nuclear RNA

ssDNA	single stranded DNA
Ssu72	Suppressor of Sua7, gene 2
SMC	Structural Maintenance of chromosomes
T	Thymine
TF	Transcription Factor
TAP	Tandem Affinity Purification
tRNA	transfer RNA
U	Uracil
UV	Ultraviolet light
WT	Wild Type
YES	Yeast Extract with Supplements

CHAPTER I
INTRODUCTION

Genome organization is fundamental to the regulation of gene expression and genome maintenance. A combination of imaging methods and genome-wide molecular approaches is unveiling the underlying organization principle. Broadly, cellular organization is a combination of chromatin compaction into higher-order domains, spatial arrangement of chromosomes and genes within nuclear space and, spatial and temporal organization of nuclear processes such as replication, transcription and repair.

This thesis focuses on the interplay between chromatin conformation, transcription and sister chromatid cohesion, key deterministic players of genome organization. In brief, chapter III focuses on the structure adopted by a gene during active transcription, in mouse and human model systems. Chapter IV employs fission yeast, *Schizosaccharomyces pombe*, as the principle model system, to investigate aspects of sister chromatid cohesion. I find two different populations of the cohesin complex on chromosomal arms, “cohesive” and “less-cohesive” cohesin, and discuss the differences between the two. Chapter V also employs fission yeast as the model system and here; I describe a novel connection between transcription and cohesin loading onto chromosomal arms.

In light of the variety of topics covered in this thesis, my Introduction (Chapter I) provides a general account of genome architecture, emphasizing on chromosome territories and long- and short-range chromatin interactions. This is followed by a comparison of the techniques used to identify such interactions. Subsequently, I provide a brief overview of the three-dimensional organization of transcription. I also discuss the diverse roles played by the cohesin complex, focussing on sister chromatid cohesion, regulation of transcription and organization of interphase chromatin.

I.1. TERRITORIAL ORGANIZATION OF THE NUCLEUS: CHROMOSOME TERRITORIES

The chromosome territory hypothesis germinates from two cardinal observations made by Carl Rabl in 1885. First, he repeatedly counted 24 chromosomes in the mitotic cells from *Salamandra maculata* and postulated "a law of constant number of threads for each cell type". Second, he noticed a polarized pattern of chromosome structure both at the beginning and the end of mitosis. The centromeres would cluster at one side of the nucleus, while the telomeres extended to the opposite side of the nucleus (Rabl 1885). In 1909, Theodore Boveri hypothesized that if chromosomes maintain their individuality and relative arrangements throughout interphase, then each mitotic chromosome should reappear at prophase in exactly the same nuclear region, where it was positioned at the beginning of interphase (Boveri 1909). He compared the arrangement of chromosomes in the blastomeres of horse roundworm, *Ascaris megalocephala*, observing a striking similarity in their orientation between daughter nuclei and thus, coined the term "chromosome territory". However, the unequivocal existence of chromosome territories (CTs) was jeopardized by the lack of methods to visualize organization of the interphase nucleus.

In 1970s, Stephen M. Stack, David B. Brown, and William C. Dewey revisited the concept of CTs (Stack et al. 1977). Fixed cells from onion root tip and Chinese hamster cell line were air dried and treated with NaOH, followed by Giemsa staining. The authors observed clumping of interphase chromatin and concluded "chromosomes remain in distinct domains throughout interphase". Further evidence supporting the existence of CTs was obtained using laser-UV-micro-irradiation experiments (Zorn et al. 1976, Zorn et al. 1979, Cremer et al. 1982a, Cremer et al. 1982b). Chinese hamster nuclei were irradiated to induce localized DNA damage, followed by pulse-labeling with [3H]-thymidine. Radioactive label incorporation at sites of DNA repair was detected either by autoradiography or immunostaining with antibodies against DNA lesions. If chromosomes were randomly distributed, the "damage signal" would appear scattered throughout the nucleus. However, the authors observed a localized pattern of damage signal, favoring a territorial organization of interphase chromatin.

The development of fluorescent *in situ* hybridization came as a major breakthrough in the field of nuclear organization. Three-dimensional-FISH (3D-FISH) or “interphase FISH” provided evidence for a territorial organization of chromosomes in higher organisms (Schardin et al. 1985, Pinkel et al. 1986). For example, in human lymphocytes, it was shown that gene-dense chromosome 19 was often present in the inner nuclear space, while gene-poor chromosome 18 localized to the nuclear periphery (Croft et al. 1999, Cremer et al. 2001, Cremer et al. 2003). Recent evidence strongly suggests non-random gene positioning within a CT and non-random CT arrangement, primarily depending on local gene density, and other factors such as transcriptional activity, replication timing and GC content. Further confirmation for the existence of CTs comes from genome-wide HiC analysis (see section I.2.5) in mouse and human cells (Lieberman-Aiden et al. 2009, Zhang et al. 2012). Gene-rich chromosomes interact at a higher frequency than gene-poor chromosomes and the former preferentially localize to the center of the nucleus, while the latter are present at the nuclear periphery (Croft et al. 1999, Lieberman-Aiden et al. 2009, Zhang et al. 2012).

At higher resolution, a chromosome territory can be subdivided into topologically associating domains (TADs) of active and inactive chromatin (Dixon et al. 2012, Nora et al. 2012, Sexton et al. 2012). Active domains harbor gene promoters, enhancers and transcribing regions, while Polycomb proteins and repressive histone marks are characteristic of inactive domains. Interestingly, genes within such domains exhibit coordinated expression during development and differentiation, with pronounced long-range interactions between loci within the same domain than between adjacent domains. Furthermore, chromatin fibers within active domains reach out to form associations across chromosome territories, suggesting some intermingling between chromosome territories (Branco and Pombo 2006, Dixon et al. 2012).

In the following sections, I discuss a series of inter-territorial and intra-territorial interactions that shape the chromatin environment during cellular differentiation and development.

I.1.1. INTER-CHROMOSOMAL INTERACTIONS: long-range associations between chromosome territories

The *IFN-γ* and cytokine genes and their regulatory T_H2 locus control region (LCR) is a well-studied example of *trans* interactions between regulatory elements. IFNγ is encoded on chromosome 10, while the *IL-4, IL-5, IL-3* genes and the T_H2 LCR are positioned on chromosome 11. In naïve $CD4^+$ T cells, the regulatory T_H2 region on chromosome 11 communicates with *IFN-γ* promoter on chromosome 10, poised for a rapid response to T-cell activation. Upon gene activation and cell differentiation into effector T_H1 or T_H2 types, the inter-chromosomal interactions are lost and intra-chromosomal associations between the LCR and promoters of *IL-4, IL-5* and *IL-3* genes are established (Lee et al. 2004, Spilianakis et al. 2005). The cell-type specific genome organizer protein "SATB1" (special AT-rich sequence binding protein 1), has been implicated in maintaining this intricate network of regulatory interactions concomitant with cytokine gene expression (Cai et al. 2006).

Similarly, inter-chromosomal interactions are present between the imprinting control region of the insulin-like growth factor 2 *(Igf2)/H19* locus on chromosome 7 and *Wsb1/Nf1* locus on chromosome 11 (Ling et al. 2006). These contacts are mediated by CTCF, a zinc finger-binding protein and are essential for expression at the Wsb1/Nf1 locus. Furthermore, human IFN-β expression is regulated by an upstream enhancer element that is recognized by three sets of transcription factors (NFκB, IRFs and c-Jun). Virus infection leads to a coordinated assembly of the above transcription factors, which is either a cause or a consequence of inter-chromosomal contacts between the IFN-β enhancer and regulatory loci *in trans* (Apostolou and Thanos, 2008).

Inter-chromosomal associations involving whole chromosomes are also emerging as key regulators of gene expression states. During early differentiation, the X chromosomes undergo transient pairing in embryonic stem cells, by a less characterized mechanism. This results in monoallelic expression of the *Tsix* antisense transcript and monoallelic upregulation of the *Xist* RNA. *Xist* expression leads to inactivation of the X chromosome in *cis* (Xu et al. 2006, Bacher et al. 2006, Augui et al. 2007, Masui et al. 2011).

I.1.2. INTRA-CHROMOSOMAL INTERACTIONS: long-range associations within chromosome territories:

The importance of genome topology in gene regulation is further strengthened by innumerable examples of intra-chromosomal interactions seen in several organisms. The diverse range of such regulatory chromatin loops includes enhancer-promoter interactions; insulator-mediated interactions; looping between Polycomb response elements; and finally, loops between the promoter and terminator regions (gene loops). While the first three are described in the paragraphs below, gene loops are discussed in the following section I.1.3.

The mammalian *β*-globin locus remains the best-characterized example of an enhancer-promoter interaction. The *β*-globin locus is organized as a multi-gene array, where each gene exhibits a development stage-specific expression in erythroid cells (Tolhuis et al. 2002). The expression is regulated by a LCR located far upstream the gene cluster (40-60kb in mouse; ~50kb upstream *HBB1* in human cells). LCR deletion leads to ~25-100 fold reduction in *β*-globin expression and *β*-thalassemia in humans, emphasizing its significance for *β*-globin expression (Bender et al. 2000, Stamatoyannopoulos 2005). Using chromosome conformation capture (3C) and RNA TRAP (Tagging and Recovery of Associated Proteins), it was shown that the LCR exists in close physical proximity with the *β*-globin cluster, thereby looping out the intervening chromatin (Grosveld 1999, Tolhuis et al. 2002). In adult erythroid cells, the DNaseI hypersensitive sites within the LCR exist in a poised conformation, forming a core “active chromatin hub” (ACH), which associates with the active *β*-globin gene, looping out the intervening inactive gene. Thus, it forms a development stage-specific transcriptional switch (Carter et al. 2002, Tolhuis et al. 2002, Patrinos et al. 2004).

Several proteins function as mediators of such enhancer-promoter interactions. The *β*-globin locus LCR and gene promoters are enriched for transcription factors GATA-1, FOG-1, EKLF and Ldb1 (Drissen et al. 2004, Song et al. 2007, Vakoc et al. 2005). Interestingly, the insulator protein CTCF has also been implicated in maintaining a closed conformation for the ACH in non-erythroid cells (Farrell et al. 2002). Actively transcribed genes in mouse embryonic stem cells also exhibit enhancer-promoter contacts, facilitated by the mediator and cohesin complexes (Kagey et al. 2010).

Contacts between insulator elements also correlate with cellular phenotype. Insulators function either by blocking inappropriate enhancer-promoter contacts, or by acting as barriers to the spreading of heterochromatin (Gaszner and Felsenfeld, 2006). In yeast, spreading of silent heterochromatin from telomeres and mating type loci is prevented by insulator boundaries. In vertebrates, loops between insulator sites have been demonstrated at the *scs/scs'* insulator pair flanking the 87A7 heat shock puff in *Drosophila*, mediated by insulator binding proteins Zeste-white 5 (Zw5) and Boundary Element Associated Factors (BEAF) (Blanton et al. 2003). In mammals, CTCF mediates looping between the 5'-HS4 and 3' insulator sequences of the chicken *β*-globin locus (Bell et al. 1999, Saitoh et al. 2000). Furthermore, CTCF binds to the non-methylated ICR at the *Igf2/H19* locus, resulting in allele-specific locus conformation and imprinted expression (Lewis et al. 2004, Kurukuti et al. 2006).

A third type of long-range interaction includes those between Polycomb response elements, mediated by the Polycomb group (PcG) proteins. The PcG proteins co-localize in the nucleus in structures known was PcG bodies, where they control the developmental repression of the *Hox* gene cluster, both in *Drosophila* and mammals (Bantignies and Cavalli 2011).

I.1.3. GENE LOOPS

Gene loops are dynamic structures formed by juxtaposition of the promoter and terminator regions of RNA polymerase II (RNAPII) transcription units. Evidence supporting the existence of gene loops stems from several studies performed in budding yeast, *Saccharomyces cerevisiae*. The TFIIB interacting protein, Ssu72 was found to be an integral component of the yeast pre-RNA 3'-end processing complex CPF (Sun and Hampsey 1996, He et al. 2003). Ssu72 also interacts with another CPF component, Pta1, which in turn associates with Sub1, a transcriptional co-activator involved in promoter clearance and start site selection (Knaus et al. 1996, Wu et al. 1999, He et al. 2003). Additionally, Sub1 interacts with the polyA signal recognizing protein, Rna15 (Calvo and Manley 2001). Sub1 and Rna15 are homologs of human transcription co-activator PC4 and 3'-end processing protein CstF64, respectively (Calvo and Manley 2001). Furthermore, the mammalian cleavage and polyadenylation factor CPSF, was also shown to interact with transcription initiation factor TFIID (Dantonel et al. 1997).

Further evidence emphasizing a 5'- and 3'- end connection came from ChIP analysis in mammalian cells. 5'-end binding proteins such as RNAPII, capping enzyme and guanidine methyltransferase were enriched at the 3'-end regions. Similarly, pre-RNA 3'-end processing factors, CPSF73 and Cstf64, were detected at the 5' end (Glover-Cutter et al. 2008). This suggested a bi-modal distribution for key transcription proteins, possibly due to looping between the 5'- and 3'- ends.

The ultimate proof for the gene loop phenomenon came from single-gene 3C analysis in *S. cerevisiae*. O'Sullivan et al. demonstrated physical juxtaposition between the promoter and terminator regions of *FMP27* and *SEN1* (O'Sullivan et al. 2004). The authors observed a bimodal distribution for the serine 5 phosphorylated form of RNAPII and Ser 5 kinase Kin28 (mammalian TFIIH homolog). Subsequently, it was shown that gene-loops are dependent on active transcription, mammalian TFIIB homolog Sua7 and components of the pre-RNA 3'-end processing machinery, Ssu72p and Pta1p (Ansari and Hampsey 2005, Singh and Hampsey 2007).

The initial studies were quickly extended to short genes such as *BLM10, SAC3, GAL10* and *HEM3* in *S. cerevisiae* (Singh and Hampsey, 2007). Interestingly, gene loops have also been reported in mammalian systems. For example, high levels of *CD68* expression in human monocytes is linked to intra-genic interactions (O'Reilly and Greaves 2007). In breast cancer cells, *BRCA1* forms a clover-leaf conformation between the promoter, regulatory intronic sequences and terminator regions. However, the promoter-terminator loop is lost upon estrogen stimulation, suggesting a repressive gene-loop at this locus (Tan-Wong et al. 2008). Similarly, a repressive gene-loop is also present at the *SAMD4A* locus in primary endothelial cells (Larkin et al. 2012). In uninduced cells, CTCF and TFIIB colocalize at the ends of *SAMD4A.* Upon stimulation with TNFα, CTCF binding is retained but TFIIB levels drop at the 3' end, correlating with the disappearance of the gene loop.

In plants, a gene-loop has been shown at the floral repressor gene *FLC.* However, PcG-dependent epigenetic silencing of *FLC* during cold temperatures disrupts the loop. Loop disruption parallels upregulation of *FLC* antisense transcripts, followed by heterochromatin formation and transcriptional shut-off. Upon shift to warm conditions, heterochromatin marks spread across the *FLC* locus, resulting in an epigenetically repressed state (Crevillin et al. 2013)

Gene-loops are thought to facilitate efficient recycling of RNAPII from its termination site, back to the promoter. At the physiological level, gene-loops are linked to transcription memory. Two seminal studies in *S. cerevisiae* have shown that gene-loops facilitate rapid activation of galactose inducible genes after a prior cycle of activation and repression (Tan-Wong et al. 2009, Laine et al. 2009). Tan-Wong et al. extend the analysis further, showing that such "memory gene-loops" are dependent on the Mlp1 protein, a component of the nuclear pore complex. Another impressive piece of work implicates gene-loop formation in maintaining promoter directionality (Tan-Wong et al. 2012). Loss of gene-loops in an Ssu72 deletion background results in widespread synthesis of antisense non-coding transcripts, known as SRTs (Ssu72 restricted transcripts) at the promoter regions. Ssu72 inactivation causes increased histone acetylation at otherwise condensed deacetylated regions, thereby providing RNAPII with the opportunity to transcribe promiscuously.

I.2. METHODS TO IDENTIFY LONG-RANGE CHROMATIN INTERACTIONS

Three-dimensional organization of chromatin is a manifestation of an underlying network of long-range physical contacts between regulatory DNA elements and target gene loci, both in *cis* and *trans.* Such contacts help establish a spatial and temporal pattern of gene expression. Undoubtedly, identification of such contacts is pivotal for a detailed understanding of genome organization and its functional implications. In this section, I summarize a decade of technological advancement that has allowed researchers worldwide to decipher regulatory interactions, across a variety of organisms. Figure I.1 provides a tabulated version of the techniques discussed below.

1.2.1 CHROMOSOME CONFORMATION CAPTURE (3C)

The first step in a 3C experiment involves crosslinking with formaldehyde to preserve *in vivo* DNA-protein and protein-protein interactions. Next, the fixed chromatin is digested with a suitable restriction enzyme, a four- or six-base cutter, depending on the required resolution. The digested chromatin is ligated under diluted conditions to promote intramolecular ligations. Subsequently, crosslinks are reversed and DNA is purified. This 3C template is a collection of both, random and specific ligation events, which are estimated by polymerase chain reaction (PCR) based methods, using site-specific primer pairs. The amplification efficiencies of different primer pairs are plotted against the genomic distance between them, which represents the corresponding interaction frequencies (Dekker et al. 2002).

Over the past decade, 3C analysis has been instrumental in the understanding of genome organization, both at multi-gene loci and single genes (gene loop). The seminal study by Dekker et al. established a “contorted ring” conformation for yeast chromosome III (Dekker et al. 2002). Subsequently the method was applied to mammalian systems showing, for the first time, that the *β*-globin LCR physically interacts with active globin genes during erythroid development, looping out ~50 kbp of intervening chromatin (Tolhuis et al. 2002, Palstra et al. 2003). Several other examples of regulatory loops that were detected using 3C include the *H19-Igf2* loci (Murrell et al. 2004), interleukin T_H2 (Spilianakis et al. 2005), *α*-globin locus (Vernimmen et al. 2007), apolipoprotein gene cluster (Mishiro et al. 2009) and the *CFTR* gene locus (Gheldof et al. 2010).

Figure I.1 A tabular representation of the different Chromosome Conformation Capture Techniques. Refer to text (section I.2) for the methods in detail. Adapted from Sajan and Hawkins 2012.

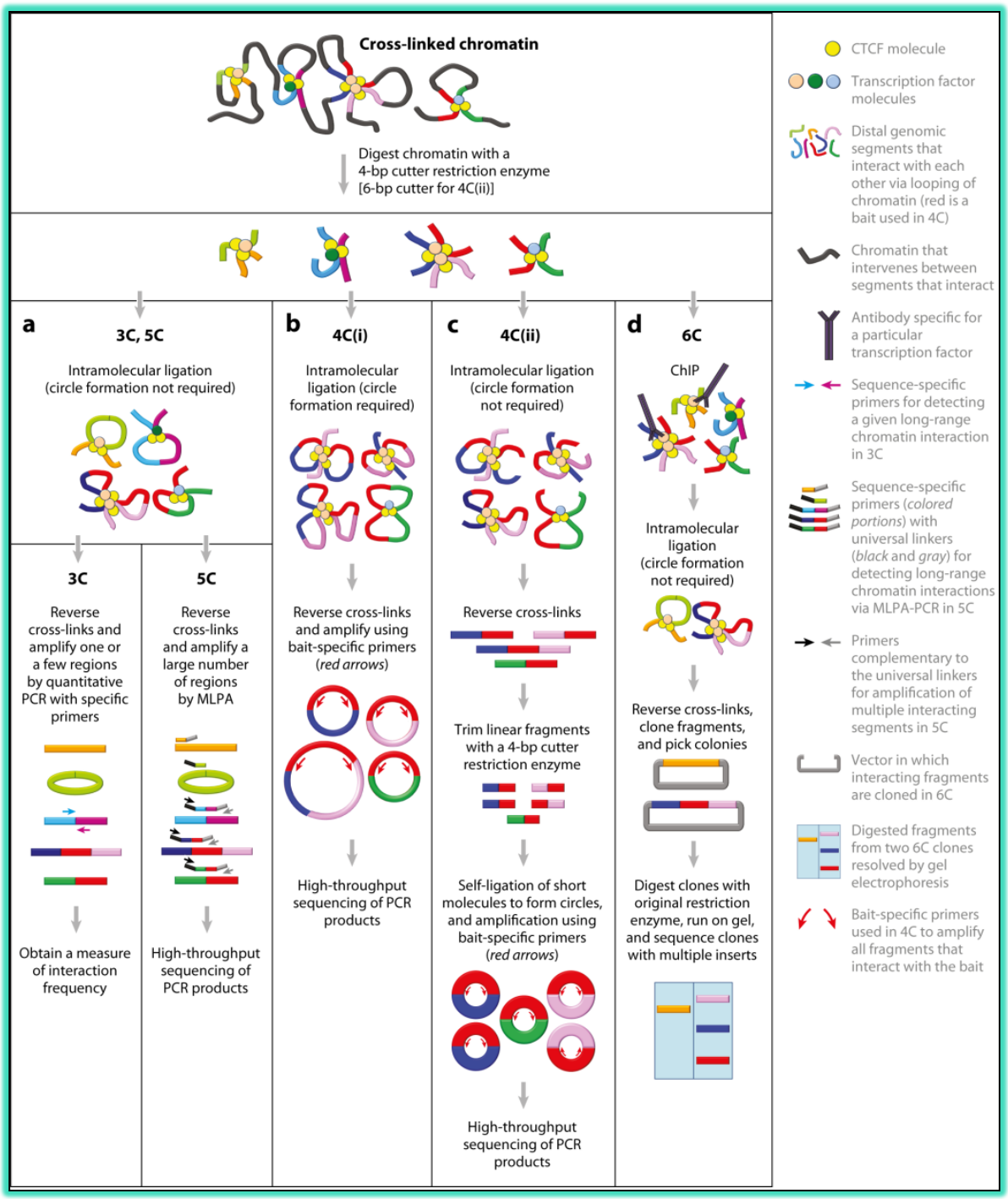
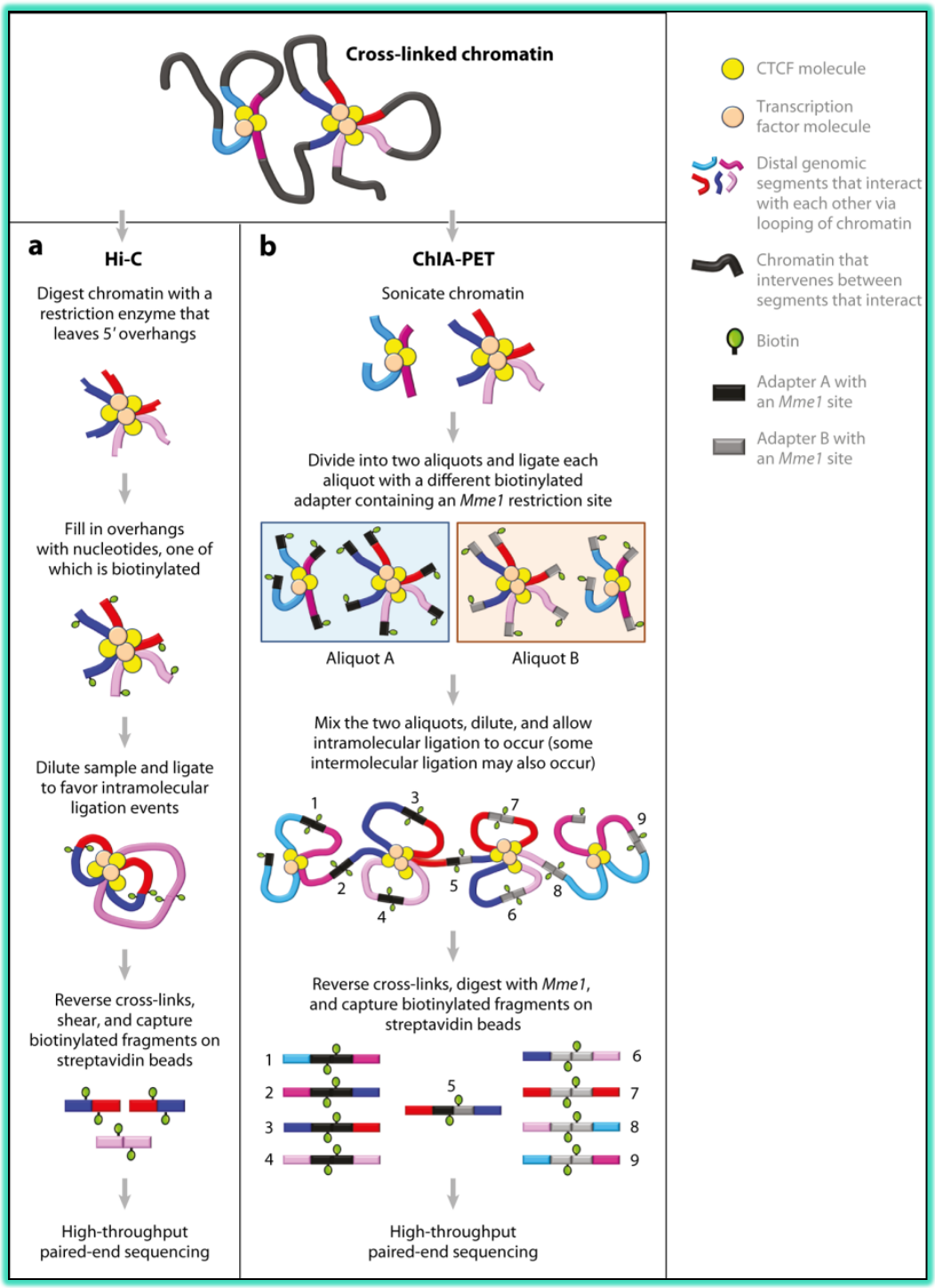


Figure I.1 continued…..

Figure I.1 A tabular representation of the different Chromosome Conformation Capture Techniques. Refer to text (section I.2) for the methods in detail. Adapted from Sajan and Hawkins 2012.



In 2004, 3C was applied to understand the conformation adopted by single genes during the transcription cycle. It was shown that actively transcribed genes display a preferential proximity between the beginning and the end of the gene, a phenomenon referred to as “gene-loop” (O’Sullivan et al. 2004). Likewise gene-loops were identified at *CD68* in monocytic cell lines (O’Reily et al. 2007), HIV-1 proviral integrants (Perkins et al. 2008) and *BRCA1* (Tan-Wong et al. 2008).

1.2.2 CHROMOSOME CONFORMATION CAPTURE-ON-CHIP (4C).

4C is commonly referred as a “one versus all” technique, as the interactions between a single “anchor/bait” are screened across the whole genome. The linear 3C template (as above) is subjected to a second round of digestion with a 4-base cutter, followed by ligation to form circles. Anchor-specific primers are then used to amplify across the ligation junctions into the interacting fragment, which is identified using microarrays or high-throughput sequencing.

4C was first employed to investigate long-range intra- and inter-chromosomal contacts established during tissue-specific and ubiquitous gene expression. In erythroid cells, the β-globin locus contacted actively transcribed regions, whereas in fetal brain, the locus was in proximity with inactive regions (Simonis et al. 2006). In contrast, housekeeping gene *Rad23a* made contacts with active regions, both in *cis* and *trans* (Simonis et al. 2006). 4C analysis in *Drosophila* suggested an extensive network of interactions between genes repressed by the PcG proteins (Tolhuis et al. 2011).

Most 4C studies have a higher propensity to detect intra-chromosomal interactions, in accordance with the idea of chromosome territories. In a clever approach, Splinter et al. employed an allele-specific 4C analysis to understand chromosome topology. The authors show that after *Xist* silencing, the inactive X chromosome adopts a structure similar to the active X chromosome; while in wild type cells, the two exhibit quite different conformations (Splinter et al. 2011).

1.2.3. CHROMOSOME CONFORMATION CAPTURE CARBON COPY (5C)

5C is a high throughput version of 3C, a “many versus many” strategy that allows simultaneous detection of interactions between multiple sequences. 5C

involves a multiplex ligation-dependent probe amplification (MLPA) approach to detect the 3C ligation products. Multiple primers are designed specific to the loci of interest. Each primer includes a T3/T7 adapter, followed by a 24-40 bp long locus-specific sequence and half of the restriction site sequence, however only the reverse primer is phosphorylated at the 5' end. The primer mix is combined with the 3C library to allow annealing of only those primer pairs that span a ligation junction in the 3C library. The annealed primers are ligated to each other and amplified using universal primers. PCR products are detected either by microarrays or high-throughput sequencing.

The 5C technology has been applied to the human *β*-globin locus (Dostie et al. 2006), the human *α*-globin locus (Bau et al. 2011) and the human *HOXA-D* gene clusters (Fraser et al. 2009, Ferraiuolo et al. 2010, Wang et al. 2011). 5C provides a medley of interaction frequencies between many sites and is useful to model the 3D conformation of larger genomic regions.

1.2.4. COMBINED CHROMOSOME CONFORMATION CAPTURE-ChIP-CLONING (6C)

6C is an extension of 3C analysis, where the initial chromatin digestion is followed by ChIP, intramolecular ligation, cloning of the ChIP-isolated DNA into a vector and transformation into bacteria. Several clones are isolated and screened for the presence of multiple fragments, by digestion with the original restriction enzyme. The positive clones are then sequenced to identify interacting partners.

6C was employed to identify interacting partners bound by the Polycomb-group histone methyltransferase EZH2, where the authors identified 5 positive clones (out of 352) (Tiwari et al. 2008). The application of 6C has been limited since it is a labor-intensive and non-genome wide technique.

1.2.5 HiC

HiC is a further modified version of 3C, with a "many versus many" strategy. Chromatin is digested to produce overhangs, which are filled with biotin-labeled nucleotides. Blunt end ligation is performed, followed by shearing of DNA and a streptavidin pull-down to enrich ligation junctions. Sequencing reads are mapped

back to the genome, and an interaction is scored when a pair is found on two different restriction fragments. Unlike the above techniques, HiC is neither dependent on any prior knowledge of interacting regions, nor does it require large-scale primer designing. However, HiC offers a resolution of ~1 Mb and is therefore, only useful to study genome organization between large domains of active and inactive chromatin.

HiC analysis in human and mouse cell lines showed spatial compartmentalization between open (A) and closed (B) chromatin domains throughout the genome (Leiberman-Aiden et al. 2009, Zhang et al. 2012). Sub-chromosomal compartment A was enriched for gene-rich, transcriptionally active and DNaseI hypersensitive loci, whereas compartment B was a reservoir of transcriptionally silent, gene poor and DNaseI insensitive loci. HiC data in yeast validated the existence of chromosomal territories (Duan et al. 2010, Tanizawa et al. 2010), with preferential clustering between functionally related genes in fission yeast (Tanizawa et al. 2010).

1.2.6 CHROMATIN INTERACTION ANALYSIS BY PAIRED-END TAG SEQUENCING (CHIA-PET)

ChIA-PET is a combination of 3C, HiC and ChIP. It allows analysis of genome-wide interactions between regulatory regions bound by a protein. The main steps include: crosslinking and sonication of the chromatin, ChIP to enrich fragments bound to a protein of interest, ligation of biotinylated linkers (contain a MmeI restriction site) to fragment ends, intramolecular ligation under diluted conditions, reversal of crosslinks, second digestion with MmeI, streptavidin pull down to enrich fragments with ligation junctions and finally, sequencing. ChIA-PET is a genome wide version of ChIP-3C or ChIP-loop technique (Horike et al. 2005).

ChIA-PET was first applied to build a chromatin interactome map of estrogen receptor α in estrogen stimulated breast cancer cells (Fullwood et al. 2009). Subsequently, ChIA-PET was applied to investigate regulatory interactions mediated by CTCF in mouse embryonic stem cells (Handoko et al. 2011).

1.2.7 IS EVERY INTERACTION A SIGNAL?

Clearly 3C-based methods have revolutionized our understanding of the three-dimensional organization of the genome. However, the pertinent question that needs to be answered with every such “C” experiment is, “what makes a real signal?” All 3C-derivates suffer from inherent limitations, which are outlined below.

- 3C and its derivatives are limited by the choice of a suitable restriction enzyme. The resolution obtained is a compromise between the genomic distribution of restriction sites and chromatin accessibility.
- Detection of ligation events from the “C” library requires multiple PCR cycles, providing a rather non-quantitative estimation. Furthermore, interpretation of PCR signals depends on critical controls, making these labor-intensive techniques.
- “All versus all” techniques provide a comprehensive statistical model of the 3D genome but the choice of where to draw the threshold lies purely with the computational biologist.
- All “C” experiments identify numerous genome-wide contacts but their contribution to the nuclear architecture can be successfully established only after correlation with features such as CpG methylation, histone modification profiles, transcription factor binding sites, DNaseI hypersensitive sites and repeat elements.
- Most importantly, “C” techniques present an average three-dimensional interactome of the total cell population. They do not consider the dynamic nature of the chromatin fiber and most certainly, do not assess cell-to-cell variations in genome topology. In other words, “Cs” cannot determine whether loci interact simultaneously or sequentially or are mutually exclusive to each other, in 2% or 50% or 100% of the cells analyzed.

Thus, to mine deeper into a structure-function relationship of the genome, we need an integration of the “C” techniques with high-resolution single-cell image analysis.

I.3. ORGANIZATION OF TRANSCRIPTION IN 3D

Transcription is fundamental to gene expression and therefore, requires tightly controlled mechanisms to ensure its smooth operation. The textbook model of transcription portrays promoter-bound factors recruiting RNAPII to gene promoters, which then travels along the transcription unit, simultaneously extruding out the nascent transcript. The strongest support for this conventional model comes from the 'Miller spreads', which illustrate multiple polymerase molecules positioned along DNA, apparently tracking along the genes during transcription (Millller 1984). Although, 'Miller spreads' may well represent experimental artifacts!

Visualization of transcription sites and active genes challenged the conventional tracking model and suggested that transcription occurs in discrete sub-nuclear compartments/ foci, termed as "transcription factories" (Jackson et al. 1993, Iborra et al. 1996a, Pombo et al. 1999, Osborne et al. 2004). FRAP (Fluorescence recovery after Photobleaching) and FLIP (fluorescence loss in Photobleaching) analysis of GFP-RNAPII dynamics revealed two populations of RNAPII (Kimura et al. 2002, Hieda et al. 2005). The majority (~80%) GFP-RNAPII had a short recovery time, implying rapid free diffusion, while ~20% GFP-RNAPII had a $t_{1/2}$ of ~20mins, consistent with relative immobilization. Furthermore, stable RNAPII foci were sensitive to transcription initiation and elongation inhibitors, suggesting the existence of phospho-Ser5 RNAPII and phospho-Ser2 RNAPII foci, respectively.

As expected, RNAPII transcription factories are rich in RNAPII, nascent transcripts and associated ribonucleoproteins, with an average diameter of ~87nm (Eskiw et al. 2008). The number of factories per cell varies from 8000 in Hela cells, 2000 in mouse embryonic fibroblasts, 100-300 in erythroblasts, B-cells, T-cells and fetal brain (Pombo et al. 1999, Osborne et al. 2004, Ragocy et al. 2006, Osborne et al. 2007, Schoenfelder et al. 2010). Such variation in the number of factories between cell types can be attributed to changes in gene expression due to differences in cell and nuclear morphology (Dalby et al. 2007, Chang and Hughes-Fulford et al. 2009).

A combination of DNA and RNA FISH, along with immunostaining and 3C-based molecular approaches have shown that actively transcribed genes colocalize to the same transcription factories (Osborne et al. 2004). This observation is

supported by the fact that mouse erythroid cells transcribe ~6500 genes but have only 100-300 RNAPII foci per nucleus (Schoenfelder et al. 2010). However, genes appear to be transcribed in “bursts”, with individual cells expressing none, one or two alleles at a given time, so the number of active transcription units per factory might be less (Chubb et al. 2006, Raj et al. 2006, Osborne et al. 2004, Osborne et al. 2007). Plasmid-based studies indicate that genes regulated by similar promoters tend to cluster into the same factories, suggesting the presence of “specialized factories” (Xu and Cook 2008). In contrast, another study indicates that formation of transcription domains (TDs) is dependent on the presence of homologous gene sequences (Binnie et al. 2006). Plasmids containing non-homologous gene sequences form separate TDs, independent of homology in the backbone or promoter sequences. Recent evidence favors specialized factories as hotspots for transcribing co-regulated genes. These include interleukins (Cai et al. 2006), *Hox* genes (Noordermeer et al. 2011), steroid receptor-binding genes (Fullwood et al. 2009, Grøntved and Hager 2012), factors involved in *β*-globin expression (Schoenfelder et al. 2010, Noordermeer et al. 2011b) and NFκB-regulated genes (Papantonis et al. 2012).

The mechanistic principle underlying the factory hypothesis is as follows. The factory skeleton anchors the polymerases that reel the gene through the active site and simultaneously extrude the nascent transcript, via the RNA exit channel. A recent study supports this model by observing changing contacts between two inducible genes, as transcription proceeds across the respective gene lengths (Papantonis et al. 2010). Based on 3C analysis, Larkin et al. suggest that transcription initiation and elongation occurs within the same factory (Larkin et al. 2012). However, *in vivo* live cell image analysis indicates differential compartmentalization of initiation and elongation (Ghamari et al. 2013).

A pertinent question concerning the factory hypothesis is, “how do factories originate?” Presumably, they exist as preassembled structures attached to a “nucleoskeleton”, which is stabilized once in contact with specific transcription factors and associated genomic loci (Jackson and Cook, 1985, Kimura et al. 1999, Mitchell and Fraser 2008). Or factories might just be aggregates of self-organized macromolecular complexes, due to “crowding” in the limited volume of the nucleoplasm (Misteli 2007). Perhaps they are both! Answering such a fundamental question has remained a challenge due to lack of information about the formation

and stability of transcription factories in living cells. Using single-molecule light microscopy, Cisse et al. shed some light upon the spatio-temporal organization of RNAPII in live cells, providing evidence against stable transcription factories (Cisse et al. 2013). Using a stable cell line expressing RNAPII tagged with a photoconvertible protein, the authors show that RNAPII forms transient clusters with an average lifetime of ~5 seconds and diameter of ~100 nm.

Currently, the field requires high-resolution live cell image analysis coupled with single gene expression kinetics to gain further insights about the nuclear organization of transcription. Isolation of individual factories and the molecular ties holding them together will also be informative.

I.4. SISTER CHROMATID COHESION AND CHROMOSOME SEGREGATION

Sister chromatid cohesion refers to the process of holding two identical DNA molecules or sister chromatids in close proximity, once they are generated in the synthesis (S) phase of the cell cycle. Cohesion is maintained till the bipolar attachment of sister kinetochores to both spindles in metaphase. Thereafter, cohesion is resolved at the onset of anaphase, allowing separation and segregation of the sister chromatids to opposite cell poles accompanied with the formation of two genetically identical daughter cells. The central player in the cohesion process is a ring-shaped, multi-subunit protein complex known as “cohesin” (Guacci et al. 1997, Michaelis et al. 1997). Cohesin ensures timely separation of sister chromatids by opposing the pulling forces exerted by the microtubules (Tanaka et al. 2000, Furuya et al. 1998; Ciosk et al. 2000). Defective cohesion and chromosome segregation result in loss or gain of chromosomes commonly referred as aneuploidy. Aneuploidy in humans is the leading cause of genetic birth defects, such as Down’s syndrome (trisomy 21), and is also associated with chromosome instabilities during cancer. Without doubt, the cohesin complex lies at the heart of a healthy biological system and in the following sections I will discuss its structure and function during the cell cycle. A comprehensive list of cohesin subunits across species, along with their functions in the cohesin cycle, is presented in Tables I.1 and I.2.

I.4.1. THE COHESIN COMPLEX

The core cohesin complex consists of Smc1, Smc3 and Scc1/Mcd1/Rad21 subunits. Smc1 and Smc3 are members of the “structural maintenance of chromosomes” (SMC) protein family. Other members of the SMC family include condensin (involved in chromosome condensation), Smc5/Smc6 (involved in DNA damage repair) and bacterial Smc proteins (Hirano and Losada 2005). Each SMC subunit is characterized by globular domains at its N- and C- termini (Figure I.2), which are separated by a long α-helical structure, containing another globular domain in the middle called the hinge domain. The two alpha-helical domains fold back onto themselves, forming a rod-like structure known as the “arm” region (Melby et al. 1998, Haering et al. 2002, Anderson et al. 2002). As a result, the N- and C-termini come together forming the globular “head” domain. Smc1 and Smc3 interact via their hinge domains to form 50 nm long V-shaped heterodimers, equivalent to 150 bp of double stranded DNA (Hirano and Losada, 2005). On the opposite end, each SMC

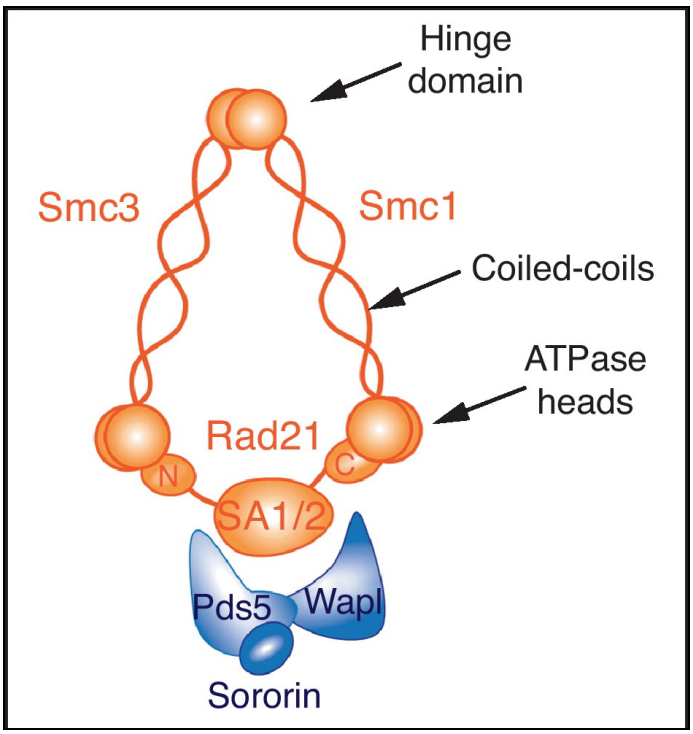


Figure I.2. The ring-shaped cohesin complex.

Cohesin is composed of four subunits, (orange). Smc1 and Smc3 polypeptides fold back on themselves by anti-parallel coiled-coil interactions, to yield molecules with a 'hinge' domain at one end and a globular ATPase 'head' at the other. The hinge domains of Smc1 and Smc3 bind tightly whereas the Smc heads are connected by the α-kleisin Scc1/Rad21. The fourth subunit is called Scc3/SA (either SA1 or SA2 in vertebrate somatic cells) and interacts with the central region of Scc1/Rad21. Pds5, Wapl and Soronon (all blue), associate with the complex once it is loaded on chromatin and modulate its dynamic behavior.
Figure adapted from (Remeiseiro and Losada, 2012).

head forms a nucleotide-binding domain (NBD). Each NBD contains a Walker A and Walker B motif, which together make up a functional ATP-binding cassette (ABC), in addition to an ABC signature motif in the other SMC subunit (Haering et al. 2004). The NBDs of Smc1 and Smc3 are connected by the kleisin ("bridging") subunit Scc1, with the N terminus binding to Smc3 and the C terminus binding to Smc1 (Haering et al. 2002). These interactions between Smc1, Smc3 and Scc1 form a tripartite ring with an inner diameter of ~40 nm, sufficient for entrapping two 10 nm chromatin fibers.

The core cohesin complex is completed by a fourth subunit, known as Scc3 in yeast and stromal antigen (SA/STAG) in mammals. There are three isoforms of SA in mammals; SA1, SA2 (mitotic subunits) and SA3 (meiotic subunit). Emerging evidence suggests that cohesin-SA1 is essential for telomeric cohesion, while cohesin-SA2 is important for centromeric cohesion and both are required for cohesion on the arms (Canudas and Smith 2009). Scc3 is also essential for sister chromatid cohesion (Michaelis et al. 1997, Nishiyama et al. 2010).

I.4.2. THE COHESIN RING MODEL

The ring-like structure of cohesin suggests that cohesin tethers sister chromatids by trapping both chromatin fibers. Electron micrographs provide evidence for the existence of cohesin rings (Anderson et al. 2002). In such a scenario, a break in the ring would result in dissociation of cohesin and separation of chromatids. Indeed, cleavage of cohesin subunits Rad21 or Smc3, triggers cohesin release from dimeric minichromosomes (Gruber et al. 2003, Haering et al. 2008, Eichinger et al. 2013). Such an effect also coincides with enzymatic cleavage of minichromosome DNA (Ivanov and Nasmyth 2005). Thus, this argues against a physical connection and advocates for topological embrace between cohesin and chromatin.

Several versions of the ring model can be envisaged (Nasmyth 2011). The one-ring model suggests that two sisters can be entrapped by a single cohesin monomer. In contrast, the two-ring model advocates dimerization of cohesin subunits, with each monomer holding one sister. One version of the two-ring model (the handcuff model) proposes association of two cohesin rings by virtue of one or two kleisin subunit(s) bound to the NBDs of the Smc3-Smc1 heterodimers. Alternatively, cohesin dimers can also form by concatenation of two rings. A third

model, the bracelet model, suggests that cohesin subunits interact to form oligomeric filaments or bracelets, which intertwine around sister chromatids.

I.4.3. COHESIN DURING THE CELL CYCLE

The association of cohesin and chromosomes begins during G_1 phase of the cell cycle, with newly assembled cohesin ring(s) after *de novo* synthesis of Rad21. DNA replication in S phase coincides with generation of sister chromatids. Cohesion is established in S phase and maintained during G_2. Finally, Rad21 cleavage results in dissolution of the cohesin ring and separation of sister chromatids during anaphase. Each of the steps are described below in further detail. Please refer to Figure I.3 simultaneously.

I.4.3.a. Cohesin loading occurs during Gap-1 (G_1)/ telophase and depends on a separate loading complex

The association of cohesin with chromatin depends on a separate loading complex known as kollerin (Nasmyth 2011). In budding yeast, kollerin consists of two essential subunits, Scc2 and Scc4 (Scc2/4) (Ciosk et al. 2000). Scc2/Scc4 homologs have been discovered in fission yeast (Mis4/Ssl3), *Drosophila* (Nipped-B), *Xenopus* and humans (NIPBL/MAU2) (Furuya et al. 1998, Rollins et al. 1999, Gillespie and Hirano 2004, Tonkin et al. 2004, Tomonaga et al. 2000, Bernard et al. 2006, Seitan et al. 2006, Watrin et al. 2006, Takahashi et al. 2008). Mutations in Scc2/Scc4 result in loss of sister chromatid cohesion (Michaelis et al. 1997, Ciosk et al. 2000). Interestingly, mutations in NIPBL are the leading cause of Cornelia de Lange syndrome (CdLS), a neuro-developmental disorder without defects in sister chromatid cohesion, suggesting a cohesion-independent function for the loading complex (Tonkin et al. 2004, Krantz et al. 2004, Kaur et al. 2005, Dorsett 2009).

Scc2 is a HEAT-repeat protein that interacts with cohesin at the Smc3-Rad21 interface (Arumugam et al. 2003). While Scc4 is composed of tetratricopeptide repeats (Watrin et al. 2006). Scc2/Scc4 binding and ATP hydrolysis at the NBDs is thought to introduce conformational changes in the cohesin ring that result in transient opening of the hinge domain, thus forming a DNA “entry gate” (Gruber et al. 2006, Mishra et al. 2010, Kurze et al. 2011, Hu et al. 2011). Recently, Buheitel and Stemman have demonstrated that blocking the Smc3-Smc1 hinge interface compromises cohesin

TABLE I.1 Subunits of the core cohesin complex.

Mammals	*D. melanogaster*	*S. cerevisiae*	*S. pombe*	Function
SMC1A SMC1B	Smc1	Smc1	Psm1	core cohesin (mitosis) core cohesin (meiosis)
SMC3	Smc3	Smc3	Psm3	core cohesin
RAD21 REC8	Rad21/Vtd C(2)M	Mcd1/Scc1 Rec8	Rad21 Rec8	core cohesin (mitosis) core cohesin (meiosis)
SA1/STAG1 SA2/STAG2 SA3/STAG3	SA (Stromalin) SA2 (Stromalin-2)	Scc3	Psc3	core cohesin (mitosis) core cohesin (meiosis)

TABLE I.2. Regulatory proteins involved in the cohesion cycle

Mammals	*D. melanogaster*	*S. cerevisiae*	*S. pombe*	Function
NIPBL/SCC2	Nipped-B	Scc2	Mis4	Cohesin loading
MAU2/SCC4	Mau2	Scc4	Ssl3	Cohesin loading
ESCO1 ESCO2	Eco/Deco San	Eco1/Ctf7	Eso1	Cohesion establishment
PDS5A PDS5B/APRIN	Pds5	Pds5	Pds5	Cohesion maintenance
WAPL/WAPAL	Wapl	Rad61/Wpl1	Wpl1	Cohesion maintenance
SORONIN/CDCA5	Dmt (Dalmatian)	-	-	Cohesion maintenance
HDAC8	-	Hos1	-	Cohesin dacetylase
Shugosin1	Sse1	Esp1		Protection of centromeric cohesion
Separase	Sse1	Esp1	Separase	Cohesin removal
Polo like Kinase 1 (PLK1)	Polo	Cdc5	Plk1	Cohesin removal

loading even in humans (Buheitel and Stemman, 2013). This emphasizes a conserved mechanism for cohesin loading via the hinge domain.

Centromeres and peri-centromeric regions constitute the strongest cohesin binding sites, although cohesin is also detected on chromosomal arms (Peters 2008). How kollerin chooses a loading site is still a mystery. It is plausible that intermediary proteins assist site selection by kollerin. Examples of intermediary proteins include sub-units of the pre-replicative complex, heterochromatin binding protein Swi6/HP1 or the Ctf19 kinetochore sub-complex (Gillespie and Hirano 2004, Takahashi et al. 2004, Bernard et al. 2001, Nonaka et al. 2002, Fernius and Marston 2009). Surprisingly, cohesin has been suggested to trigger its own loading. Loss of Scc1/Rad21/Mcd1 prevents association of Scc2 with budding yeast centromeres (Fernius et al. 2013). Nonetheless, cohesin loading is a dynamic process that occurs throughout the cell cycle until prophase, when kollerin itself dissociates from the chromosomes.

I.4.3.b. Cohesion establishment occurs during the synthesis (S) phase

Cellular cohesin can be divided into two populations. During G_1 and G_2 stages of the cell cycle, cohesin maintains a dynamic state of association and dissociation on chromosomes, referred to as “soluble cohesin”, with a mean residence time of ~25 min (Gerlich et al. 2006). After S phase, the mean residence time increases to ~6 h, suggesting establishment of stable cohesion and hence, “stable cohesin”. Stable cohesin is particularly essential in mammalian oocytes, where cohesion is established during pre-meiotic S-phase and has to be maintained until completion of meiosis, which only occurs many years later (Hunt and Hassold 2010).

Establishment of stable cohesion is attributed to an enzyme known as Eco1/Ctf7. Eco1 is an acetyltransferase that is necessary for the establishment of cohesion during DNA replication but not for its maintenance during G_2 and M phases (Toth et al. 1999, Skibbens et al. 1999). Eco1 is recruited to the replication fork probably by its physical interactions with the DNA polymerase processivity factor PCNA (Lengronne et al. 2006, Moldovan et al. 2006) and acetylates Smc3 on two lysine residues [K112 and K113 in budding yeast, and K105 and K106 in humans] (Ben-Shahr et al. 2008, Unal et al. 2008, Zhang et al. 2008, Rowland et al. 2009). How acetylation result in stabilization remains unknown. It is suggested that Smc3 acetylation counteracts the “anti-establishment” activity of the *Wapl* gene product

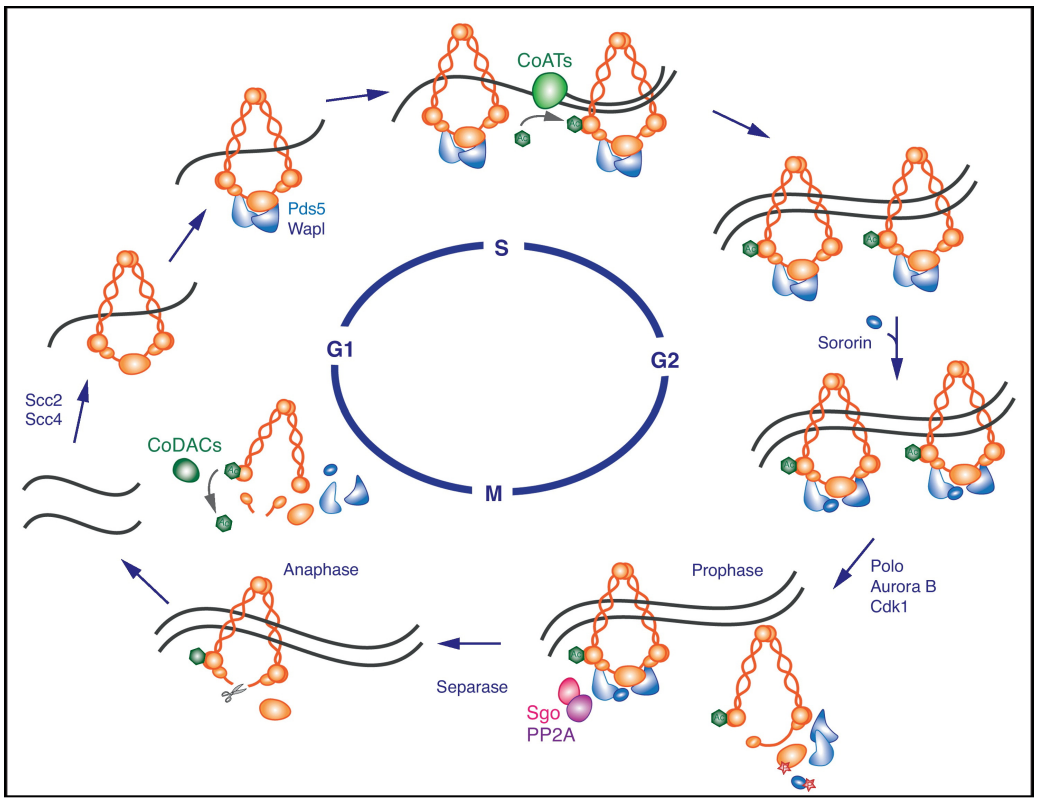


Figure I.3. Regulation of cohesin throughout the cell cycle.

Cohesin is loaded onto chromatin by the Scc2/Scc4 loading complex, in early G1 in vertebrate cells and late G1 in yeast. Upon DNA replication, cohesin mediates cohesion between the newly replicated sister chromatids. Smc3 acetylation by CoATs counteracts the antiestablishment action of Pds5–Wapl. Sororin binding displaces Wapl and stabilizes cohesive complexes. In mitosis, most cohesin is released from chromatin in prophase by phosphorylation of the SA subunit by Polo and Sororin by Cdk1, which facilitates the action of Pds5–Wapl. Aurora B also plays a role in this step. The fraction of cohesin that remains at centromeres is protected by Shugosin-Protein phosphatase (Sgo-PP2A) is cleaved upon activation of Separase at the onset of anaphase, allowing the separation of the chromatids to opposite spindle poles. Once cohesin is released from chromatin, Smc3 is deacetylated by soluble CoDACs so that it can be reused in the next cycle.
Figure adapted from (Remeiseiro and Losada, 2012).

(Rowland et al. 2009, Sutani et al. 2009, Terret et al. 2009). In mammals, acetylation promotes recruitment of soronin, which neutralizes the destabilizing activity of Wapl and Pds5 (Lafont et al. 2010, Nishiyama et al. 2010).

Accumulating evidence suggests a tight coupling between replication and establishment of stable cohesion. Mutations in replication proteins such as Ctf4, Ctf18, Tof1, Csm3, Chl1 and Mrc1, lead to defects in sister chromatid cohesion (Hanna et al. 2001, Mayer et al. 2001, Williams et al. 2002, Mayer et al. 2004, Petronczki et al. 2004, Xu et al. 2004, Errico et al. 2009). Recently, it has been shown that Ctf18, Tof1, Csm3 and Mrc1 promote cohesin acetylation during S phase (Borges et al. 2013). Either these factors facilitate Eco1 recruitment at the replication fork or they might protect acetylated Smc3 from deacetylases. On the other hand, Ctf4 and Chl1 contribute to cohesion establishment in an acetylation-independent manner, probably by regulating the passage of the replication fork through the cohesin ring (Borges et al. 2013).

Despite being interconnected, the molecular mechanism coupling cohesion and replication still remains elusive. Nonetheless, several models have been proposed (Uhlmann 2007). First, the replisome passes through the cohesin rings, aided by the replication proteins discussed above. Second, cohesin temporarily disassembles from the chromatin to allow replication fork progression and the replication proteins might help re-assemble the cohesin rings. Third, analogous to the mitotic condensin complex, cohesin and replication proteins might initiate a series of DNA-DNA and DNA-protein interactions, leading to chromatin compaction and cohesion between sister chromatids (D'Ambrosio et al. 2008, Lopez-Serra et al. 2013, Borges et al. 2013).

I.4.3.c. Cohesion is maintained throughout G_2 phase of the cell cycle

Once S phase is completed, Eco1 is phosphorylated by Cdk1 and degraded by the SCF ubiquitin ligase (Lyons et al. 2011). Consequently, further establishment of cohesion ceases.

Maintenance of sister chromatid cohesion is independent of Eco1 but depends on Wapl and Pds5. Null alleles of Wapl and point mutations within Pds5, Smc3 and Scc3, enable cell growth even in the absence of Eco1, suggesting an

antagonistic interplay between Eco1 and Wapl-Pds5 (Tanaka et al. 2001, Ben-Shahar et al. 2008, Unal et al. 2008, Rowland et al. 2009, Sutani et al. 2009). Both proteins are conserved from yeast to humans (Sumara et al. 2000, Losada et al. 2000, Losada et al. 2005). Wapl and Pds5 interact *in vitro* and Wapl is recruited to the cohesin ring via Pds5 (Chan et al. 2012). Both proteins interact with the cohesin ring at the Smc3 NBD and α-kleisin subunit (Scc1) interface. Wapl exerts a destabilizing effect on cohesin association because Wapl deletion results in cohesin stabilization (Sutani et al. 2009). Pds5, on the other hand, is a multifaceted protein. Pds5 is crucial for cohesin release via the recruitment of Wapl (Shintomi and Hirano 2009, Chan et al. 2012). In contrast, Pds5 is essential for *de novo* acetylation of Smc3, possibly via a direct interaction with Eco1 (Tanaka et al. 2001, Vaur et al. 2012, Chan et al. 2013). Additionally, Pds5 protects cohesin from Hos1-mediated deacetylation during G_2/M (Chan et al. 2013). Furthermore, Pds5 might have a yet uncharacterized role at the Smc3-Smc1 hinge region (Mc Intyre et al. 2007).

I.4.3.d. Resolution of cohesion in the mitotic (M) phase

The dissociation of cohesin from chromosomes occurs during the mitotic phase and is pivotal for separation of sister chromatids into the daughter cells. In most eukaryotic cells, cohesin complexes are removed in a two-step manner. The first step, known as the “prophase pathway”, involves cohesin removal from the chromosomal arms during prophase and prometaphase; while the second involves cohesin removal from the centromeres by cleavage of the α-kleisin subunit (Scc1) at the onset of anaphase (Uhlmann et al. 1999, Waizenegger et al. 2000).

The prophase pathway is facilitated by phosphorylation of the SA subunit and soronin by Aurora B kinase and polo-like kinases (Plk1) (Losada et al. 2002, Sumara et al. 2002). Plk1 also phosphorylates Wapl and Pds5 (Hegemann et al. 2011). Phosphorylation might empower Wapl to exert a destabilizing effect, possibly by opening the “exit gate” located at the Smc3 NBD and α-kleisin interface (Nasmyth 2011, Chan et al. 2012). Additionally, the kollerin complex also dissociates from chromosomes during prophase, impeding any further cohesin loading (Hauf et al. 2005). Centromeric cohesin is maintained during prophase by the action of shugoshin, which prevents soronin phosphorylation by recruiting protein phosphatase 2A (PP2A) (Salic et al. 2004, McGuinness et al. 2005, Riedel et al. 2006, Xu et al. 2009).

The metaphase to anaphase transition is accompanied by destruction of centromeric cohesin and resolution of sister chromatid cohesion. Cohesin at the centromeres is removed by separase-mediated cleavage of Scc1. Once chromosomes are bioriented at the metaphase plate, the anaphase-promoting complex (APC) leads to ubiquitin-mediated degradation of securin and cyclin B, thereby releasing Separase (Uhlmann et al. 1999, Hauf et al. 2001). Separase now cleaves Scc1, inducing cohesin dissociation and sister chromatid dissolution.

I.4.4. COHESIN AND GENE REGULATION

The first evidence that cohesion factors regulate gene expression and development came from studies on the *Drosophila cut* and *Ultrabithorax* homeobox genes (Rollins et al. 1999, Rollins et al. 2004, Dorsett et al. 2005). Heterozygous *Nipped-B* mutants showed reduced *cut* expression, while loss of Rad21, Smc1 or SA resulted in increased *cut* expression. Mutations in *NIPBL* are the leading cause of Cornelia de Lange syndrome (CdLS), a neuro-developmental disorder with upper extremity malformations, hirsutism and other dysfunctions affecting cardiac, renal and gastro-esophageal systems (Krantz. et al. 2004, Muto et al. 2011). Hundreds of genes are indeed dysregulated in cell lines derived from CdLS patients but significant cohesion defects are absent (Liu et al. 2009). Cohesin also binds and facilitates expression of the *c-myc* gene, a function conserved across *Drosophila*, zebrafish, mouse and humans (Rhodes et al. 2010). An obvious explanation for gene expression changes upon cohesin mutation is impaired cohesion and chromosome segregation. However, homozygous mutations in the *Smc1* and *SA* genes block ecdysone steroid hormone signaling, axon pruning and dendrite formation in *γ* neuron of the mushroom body in *Drosophila* (Schuldiner et al. 2008, Pauli et al. 2010). Furthermore, genetic deletion of cohesin in non-dividing mouse thymocytes results in reduced transcription and rearrangement at the T cell receptor *α* locus *(Tcra*), thereby affecting thymocyte differentiation (Seitan et al. 2011). Clearly, the phenotype in non-dividing cells is independent of cohesin's role in cell division.

Genome wide analysis of cohesin binding sites have been performed in *Drosophila* cell lines (Misulovin et al. 2008), *Drosophila* salivary glands (Pauli et al. 2010), human lymphocytes (Liu et al. 2009), HeLa cells (Wendt et al. 2008), MCF-7 and HepG2 tumor cell lines (Schmidt et al. 2010), mouse pre-B cells, mouse embryonic stem cells (ESCs) (Kagey et al. 2010) and mouse thymocytes (Parelho et

al. 2008). In *Drosophila* and mammals, Nipped-B/NIPBL and cohesin colocalize to actively transcribed genes, with peaks near the transcription start site (Misulovin et al. 2008). Additionally, in mammals there is an 89% overlap between cohesin and CTCF binding sites but no overlap between CTCF and NIPBL occupancy (Wendt et al. 2008). In a CTCF-independent manner, RAD21 also colocalizes with tissue-specific transcription factors such as estrogen receptor *α* in breast cancer cell lines, hepatic nuclear factors in liver cells, and pluripotency factors in stem cells (Schmidt et al. 2010, Nitzsche et al. 2011). Additionally, cohesin co-purifies with the mediator complex, which itself is important for transcription (Kagey et al. 2010). It is plausible that cohesin is loaded at NIPBL sites and slides away, as observed in yeast (Lengronne et al. 2004, Schmidt et al. 2009), accumulating at CTCF or transcription factor sites. Alternatively, cohesin might recruit TFs to gene promoters or the TFs could recruit NIPBL, which subsequently loads cohesin.

Cohesin depletion reduces *myc* transcription and MYC has emerged as the universal regulator of transcriptional programs in mammalian cells (Rhodes et al. 2010, Nie et al. 2012). Interestingly, genes with the highest cohesin levels produce nearly 2-fold more steady-state RNA than genes lacking cohesin (Schaaf et al. 2013). Yeast two-hybrid and mass spectrometry experiments in *Caenorhabditis elegans* identified novel interactions between RAD21 and RNA processing factors. However, their functional implication and species conservation has not been explored (Panigrahi et al. 2012). In fission yeast, cohesin promotes efficient transcription termination between convergent genes during the cell cycle (Gullerova and Proudfoot 2008). During G_1 phase, overlapping transcripts at convergent gene loci produce double strand RNA, followed by RNA interference-dependent transient heterochromatin formation. Subsequently, Swi6 recruits cohesin either via a direct interaction or by first recruiting Mis4 (fission yeast cohesin loader), which then loads cohesin. In G_2 phase, cohesin is concentrated at the intergenic regions of convergent genes where it promotes gene-proximal transcription termination.

Indisputably, cohesin exerts a direct influence on transcription. A recent PRO-Seq and ChIP-chip analysis in *Drosophila* shows a positive correlation between Rpb3 (Pol II subunit) and cohesin binding at actively transcribing genes (Schaaf et al. 2013). Cohesin preferentially binds to genes with high promoter-proximal paused RNAPII and cohesin depletion increases RNAPII pausing at cohesin-bound genes, suggesting that cohesin might influence the transition from paused polymerase to

elongation. However, cohesin-bound genes lack H3K36me3 modification (a mark of elongation) and pause release is independent of pausing factors DSIF and NELF (Fay et al. 2011). Cohesin depletion coincides with either an increase or a decrease at paused RNAPII loci, implying that cohesin plays a context-dependent role.

I.4.5. COHESIN: AN ORGANIZER OF INTERPHASE CHROMATIN

Cohesin and CTCF play a pivotal role in the structural and functional organization of the nucleus. Direct evidence from knockdown experiments implicates CTCF and cohesin in maintaining long-range interactions at the imprinted *Igf2/H19* locus (Nativio et al. 2009), interferon gamma locus (Hadjur et al. 2009), *β*-globin locus (Chien et al. 2011), *APO* gene cluster (Mishiro et al. 2009) and *MHC* class II cluster (Majumder and Boss 2011). Depletion of cohesin and mediator subunits results in reduced enhancer-promoter contacts of key pluripotency genes and changes in the gene expression program (Kagey et al. 2010). However, as ES cells are actively dividing, these effects may arise through cell cycle defects. Subsets of CTCF/cohesin sites are also bound by transcription factor TAF3, where long-range looping is imperative for endoderm lineage differentiation and prevention of premature differentiation of neuroectoderm and mesoderm in mouse ESCs (Liu et al. 2011). Upon induction of differentiation, cohesin and NIPBL bind to the *β*-globin locus control region (LCR) and regulatory sites, facilitating long-range interactions that are lost upon partial reduction of NIPBL (Chien et al. 2011).

The mechanism behind the function of cohesin in long-range interactions remains a mystery. However, the simplest explanation is that cohesin entraps regulatory elements in *cis*, similar to how it holds sister chromatids together in *trans.*

AIMS OF THESIS

I. To understand the mechanism-underlying gene loop formation in two different mammalian transcription units. The first model system is a constitutively expressed gene, mouse serum albumin *(Alb).* The second model is an inducible gene, human nuclear factor κB factor 1*, (NFκB1).*

II. To provide single-cell evidence supporting gene loop formation in mammals, using high-resolution image analysis.

III. To characterize functional differences between the cohesin populations in fission yeast, *Schizosaccharomyces pombe.* This analysis focuses on cohesin sub-sets colocalizing with or without the loading complex, Mis4/Ssl3.

IV. To understand the mechanism of cohesin loading onto chromosomes during the fission yeast cell cycle. This analysis focuses on the requirement of RNA Polymerase II transcription for chromosomal association of cohesin.

CHAPTER II

MATERIALS AND METHODS

II.1 MAMMALIAN TECHNIQUES

II.1.1 Mammalian cell culture

Hepa-1-6 and C2C12 cells were grown in DMEM media supplemented with 10% FCS, 100 units/ml penicillin and streptomycin at 37^{o}C with 5% CO_2. Hepa-1-6 cells were treated with Flavopiridol (200 µM) or KM05283C (100 µM) for 1 h and with α-amanitin (2.5 µg/ml) for 12 h and processed for RNA isolation, ChIP and 3C. HUVECs were grown in EBM2 supplemented with 2% FCS, growth factors and antibiotics. For *NFκB1* induction, HUVECs were serum starved in EBM2 + 0.5% FCS (no supplements) for 14-16 h and then induced with TNFα at a final concentration of 10 ng/ml. Subsequently, cells were harvested for RT-PCR and ChIP analysis at different time points.

II.1.2 RNA Isolation

Cells were washed with PBS and harvested by adding 1 ml Trizol reagent (Invitrogen). RNA was isolated according to the manufacturer's instructions.

II.1.3 Reverse Transcription (RT)-PCR

Total RNA was isolated using Trizol reagent, DNaseI treated and reverse transcribed with SuperScript III Reverse Transcriptase using 1 µM gene-specific reverse primer following the manufacturers' instructions. Non-reverse transcribed controls were performed simultaneously.

II.1.4 Standard PCR

50-100 ng DNA template was amplified in the presence of 0.5 µM primer pairs, 0.2 mM dNTPs, 1.75 mM $MgCl_2$, 1X GoTaq buffer and GoTaq polymerase. Primer annealing temperatures were chosen to be approximately 3^{o}C lower than the primer melting temperature. Primers were designed to anneal at 62°C. Cycling conditions were determined according to manufacturers' guidelines.

II.1.5 Quantitative Real-Time PCR (q-PCR)

Real-Time PCR is used to simultaneously amplify and quantify a DNA molecule. In contrast to standard PCR, where the PCR product is detected at its amplification end point, qPCR detects amplified DNA as the reaction progresses in real time. qPCR measures the increase in fluorescence generated by the fluorescent dye SYBR Green that intercalates into the double-strand DNA product as it is generated. The general principle of DNA quantification by qPCR relies on plotting fluorescence against the number of cycles on a logarithmic scale. qPCR was performed using the Sensimix (Rel Time PCR reaction mix) in a Corbett Rotogene 3000 machine. Reactions were performed in triplicates to minimise experimental error and quantified using the Rotogene 3000 software. A SYBR Green melting curve was performed simultaneously to confirm the integrity of the PCR product. *Gapdh* and 28S rRNA were used as normalization controls for qPCR analysis of *Alb* and *NFκB1,* respectively.

In experiments comparing the efficiency of different transcription inhibitors, comparative quantification was performed between the untreated and treated samples. The values thus obtained were plotted relative to the untreated control that was set to 1.

In *NFκB1* transcription analysis, the raw values for each time point were first normalized to corresponding 28S rRNA values and then plotted as relative fold change to the uninduced (t =0 min) sample.

II.1.6 3'RACE

Hepa-1-6 RNA was reverse transcribed with a phased oligo-dT primer. 1 µl of the first strand cDNA was PCR amplified with *Alb*-Ex12F (forward primer) and an adapter specific reverse primer. The PCR product was gel purified, cloned in a linearized pCR 2.1 TOPO vector (Invitrogen), transformed into NM554 competent cells and grown under ampicillin (50 µg/ml) selection. Multiple colonies were screened with colony PCR, plasmids isolated from positive clones and sequenced with M13F-M13R primers. All sequences were aligned/analyzed using the eBioX software program (V1.5.1)

II.1.7 Chromatin immunoprecipitation (ChIP) analysis

ChIP analysis was carried out as previously described (Kadener et al. 2002; West et al. 2004) with some modifications. Nearly 10^7 cells were fixed for 10 min using 36.5% methanol-stabilised formaldehyde solution to a final concentration of 1%. Formaldehyde was quenched with 0.125 M glycine, cells were washed and harvested in PBS at 4°C. Cells were lysed in 500 µl of cell lysis buffer and incubated for 10 min on ice. The lysed cells were centrifuged at 500 g for 5 min to pellet nuclei. The nuclear pellet was resuspended in 400 µl of nuclear lysis buffer. Sonication was performed in 1.5 ml tubes in a diagenode Bioruptor for 15 min at high power with pulses of 30 s. Lysates were pre-cleared with 50 µl of Protein A/G agarose beads for 1 h at 4°C and then incubated with 3-5 µg of antibody overnight. 10% of pre-cleared material was kept as total input. Washes were performed with buffers A, B, C and 2x D and elution twice with elution buffer. Eluted samples were reverse crosslinked for 4-5 h at 65°C with 0.3 M NaCl and 3 µg/ml RNase A. Proteinase K treatment was performed for 2 h at 45°C with 10 mM EDTA, 40 mM Tris-HCl pH 6.5 and 20 µg proteinase K. The chromatin was purified through Qiagen PCR-clean up columns. The immuno-precipitated, no antibody, and input DNAs were used as templates for qPCR. The PCR mixture contained Sensimix, 2 ul of template DNA and primers from the Table. Cycling parameters were 95°C for 15 min, followed by 45 cycles of 95°C for 15 s, 62°C for 20 s, and 72°C for 20 s. ChIP quantification was performed by the $\Delta\Delta C_t$ method (Livak et 2001). First, the qPCR values for the immunoprecipitated (IP) samples were subtracted from the input value, giving us the ΔCt value for each primer pair. Subsequently, the ΔCt value of the negative control was subtracted from the ΔCt value of the positive control and regions of interest, giving us corresponding ΔΔCt values. The ΔΔCt values were plotted to the power of 2 to calculate the relative fold change over the negative control, i.e., Relative Fold Change= $(2)^{\Delta\Delta Ct}$.

Cell lysis buffer

NP40	0.5%
Potassium Chloride	85 mM
PIPES pH 8	5 mM
PMSF	1 mM
Leupeptin	1 µg/ml
Pepstatin	1 µg/ml

Nuclear lysis buffer

SDS	1%
EDTA	10 mM
Tris HCl pH 8	50 mM
PMSF	1 mM
Leupeptin	1 µg/ml
Pepstatin	1 µg/ml

IP dilution buffer

SDS	0.01%
EDTA	1.2 mM
Tris HCl pH 8	16.7 mM
Triton X-100	1.1%
Sodium Chloride	167 mM
PMSF	1 mM
Leupeptin	1 µg/ml
Pepstatin	1 µg/ml

Wash buffer A

SDS	0.1%
Triton X-100	1%
EDTA	2 mM
Tris HCl pH 8	20 mM
Sodium Chloride	150 mM

Wash buffer B

SDS	0.1%
Triton X-100	1%
EDTA	2 mM
Tris HCl pH 8	20 mM
Sodium Chloride	500 mM

Wash buffer C

Lithium Chloride	0.25 M
NP40	1%

Sodium deoxycholate	1%
Tris HCl pH 8	10 mM
EDTA	1 mM

Wash buffer D

10:1 Tris HCl pH 8: EDTA

Elution buffer

SDS	1%
Sodium hydrogen carbonate	0.1 M

II.1.8 Chromosome Conformation Capture (3C)

Hind III 3C analysis was performed with Hepa-1-6 and HUVECs according to a previously published protocol (Simonis et al. 2006) with slight modifications. Nearly 10^7 cells were cross-linked with 1% formaldehyde, quenched with 0.125 M glycine, washed with PBS and centrifuged at 3000 rpm for 5 min to obtain cell pellets. Cells were lysed in cold lysis buffer and incubated on ice for 10 min. Nuclei were obtained after centrifugation at 2000 rpm, 2 min and nuclear pellets were re-suspended in PBS, either stored at -80°C or processed immediately. Nuclei were then resuspended in 1.2X restriction enzyme buffer (Roche) supplemented with 0.3% SDS and incubated at 37°C for 1 h with gentle shaking. Subsequently, SDS was quenched with 1% Triton-X-100 and further incubated at 37°C for 1 h. A total 800 U of HindIII was added in two intervals separated by ~6-8 h and restriction digestion was performed overnight at 37°C. Digestion was stopped with 2% SDS at 65°C for 15 min. The digested chromatin was diluted with the ligation buffer to a total volume of 8 ml and SDS was quenched by 1% Triton-X-100 at 37°C for 1 h. For each 3C ligation 6000 cohesive end units of T4 DNA ligase were used. The ligation mixture was incubated at 16°C for at least 4 h and maximum overnight. Subsequently, the samples were treated with 300 µg of Proteinase K and 300 µg of RNase A at 65°C for at least 6 h to overnight. The 3C DNA template was extracted either by phenol chloroform or by Qiagen PCR purification columns and resuspended in 50-100 µl MilliQ water. As a ligation control, two adjacent fragments from the *Gapdh* (Albumin 3C) and *GAPDH* (*NFκB1* 3C) gene loci were used.

Positive controls for 3C analysis were generated by digesting equimolar quantities of ~500 bp PCR products encompassing HindIII restriction sites across the

Alb and *Gapdh* locus for Albumin 3C control or *NFκB1* and *GAPDH* for *NFκB1* 3C control. The digested products were ligated in 100 µl reaction using the QuickLigase Kit (NEB) at 16°C for 4-6 h. Serial dilutions of the ligated products were used as standards for optimizing primer efficiency and quantification in each 3C experiment. 3C PCRs for the test and control templates were performed using GoTAQ DNA polymerase as follows: 95°C for 2 min, followed by 35 cycles of 95°C for 30 s, 62°C for 30 s, 72°C for 30 s and 72°C for 5 min. PCR products were separated on 2.5% agarose gels and quantified using the GelQuant Software. The raw values for each primer pair from the experimental samples were first normalized to a loading control and the relative interaction frequency between two primer pairs was calculated as below. Final graphs were plotted relative to the immediate adjacent fragment to the anchor.

Relative Interaction Frequency (X) = $[\text{Primer A/ Gapdh}]_{\text{3C template}} \div [\text{Primer A/ Gapdh}]_{\text{3Ccontrol}}$

Cell Lysis Buffer

Tris pH8.0	10 mM
NaCl	10 mM
NP-40	0.2%
PMSF	1 mM
Leupeptin	1 µg/ml
Pepstatin	1 µg/ml

Electrophoresis buffers:

50x E buffer (for agarose gels):

Sodium Acetate	250 mM
EDTA	50 mM
Tris	2 M
Glacial acetic acid	91 ml/L

II.1.9 RNA-Fluorescent *in situ* hybridization (FISH) analysis

The protocol is obtained from the http://www.singerlab.org/protocols. The RNA FISH analysis described below was performed by Dr.Timothee Lionnet (Post

doctoral researcher in Prof. Robert Singer's Group, Albert Einstein College of Medicine, Bronx, NY, USA).

HUVEC cells were grown on coverslips, washed once in 1X PBS and fixed for 10 min at room temperature in 4% formaldehyde (40% liquid stock from Electron Microscopy Science), 10% acetic acid and 1X PBS. After two washes in PBS, cells were permeabilized by overnight treatment with 70% ethanol at 4°C. Cells were rehydrated at room temperature in 2X SSC and 50% formamide. Cells were hybridized overnight at 37°C in 40 µl of a mixture containing 10% dextran sulfate, 2 mM vanadyl-ribonucleoside complex, 0.02% RNase-free BSA, 40 µg *E.coli* tRNA, 2X SSC, 50% formamide and 30 ng of probe. The following day, cells were washed twice for 30 min at the appropriate stringency: 2X SSC + 50% formamide at 37°C. Slides were washed in 1X PBS and mounted in 90% glycerol, 1 mg/ml p-phenylendiamine, 0.1 µg/ml DAPI.

Oligomer probes were designed with OLIGO software from Molecular Biology Insights and synthesized on an Applied Biosystems automated DNA/RNA synthesizer. The probes covering ~5 kbp sequences in *NFκB1* intron 1, intron 5 and intron 23 were modified by incorporating amino-modified nucleotides (amino-allylT), the modified DNA was resuspended in 0.1 M $NaHCO_3$ buffer and labeling was performed by adding DMSO containing the activated fluorophore (Amersham labeling kit). Labeled probe was purified using G50 columns and resuspended in MilliQ water. Probes for intron 1, 5 and 23 were coupled to Cy3, Cy3.5 and Cy5 fluorophores respectively.

Image acquisition was performed with Upright fluorescence microscopes (Olympus). Illumination was provided by a 100 W Mercury arc lamp. Microscopes were fitted with Olympus PlanApo 60x, 1.4 NA objectives and Chroma HiQ band pass filters to separate fluorescence signals. Different transcripts were identified using appropriate filters for the following channels: DAPI, Cy3, Cy3.5, and Cy5. Acquisition and data analysis were performed using IPLab software from Scanalytics. Custom coded algorithms were used for data analysis using the JAVA Programming Language, the Java Development Kit and Advanced Imaging Library from Sun Microsystems.

II.2 FISSION YEAST TECHNIQUES

II.2.1 *S. pombe* growth conditions

Haploid strains were used for all applications. Commercially available rich medium (YES) was obtained from Formedium and supplemented with 225 mg/ml of leucine, adenine, histidine, uracil and lysine hydrochloride. Rich medium plates were prepared using commercially available YES medium, supplemented with amino acids as above. Most experiments were performed in cycling G_2 cells. When cells were synchronized in G_1, they were grown in EMM medium (14.7 mM potassium hydrogen phthalate, 15.5 mM Na_2HPO_4, 93.5 mM NH_4Cl, 2% glucose, 20 ml/l of 50X salts, 1 ml/l of 1000X vitamins, 0.1 ml/l of 10,000X.

50X salts solution= 0.26 M magnesium chloride, 4.99 mM calcium chloride, 0.67 M potassium chloride and 14.1 mM sodium sulphate.

1,000X vitamins solution= 4.3 mM pantothenic acid, 81.2 mM nicotinic acid, 55.5 mM inositol and 40.8 µM biotin.

10,000X minerals solution= 80.9 mM boric acid, 23.7 mM manganese sulphate, 13.9 mM zinc sulphate, 7.4 mM iron chloride, 2.47 mM molybdic acid, 6.02 mM potassium iodide, 1.6 mM copper sulphate and 47.6 mM citric acid.

Strains were kept as frozen stocks (-80°C) and thawed on agar plates when required. Liquid cultures were inoculated from a patch into the desired medium and grown overnight at 30°C for wild type and 25°C for temperature-sensitive strains.

II.2.2 Cell synchronization by nitrogen starvation

In response to nitrogen starvation, fission yeast cells arrest in G_1 phase of the cell cycle. For DNA FISH analyses in G_1, wild type *(972), rad21-K1ts* and *mis4-367ts* cells were synchronized in G_1 by this method. Approximately, 30 ml of mid-log phase cultures (OD_{600}= 0.5) were exponentially grown in EMM. Cells pellets were washed twice with sterile water and twice with EMM medium lacking ammonium chloride (NH_4Cl). Subsequently, cells were cultured in EMM-NH_4Cl for at least 16 h. Thereafter, cells were resuspended in pre-warmed rich (YES) medium at an

OD_{600}=0.2 and shifted to 37°C, for a total of 8 h. Aliquots were taken at 1 h intervals for FACS analysis. Most cells were arrested in G_1 after nitrogen starvation. After release into rich medium, cells completed replication by 4 h. The wild type continued the cell cycle normally but the mutants progressed through G_2 and M phases with segregation defects.

II.2.3 DNA Isolation

This protocol is adapted from Hoffman and Winston, 1987. Cell pellets from ~10^8 cells were washed with sterile water and resuspended in DNA isolation Buffer A, 200 µl acid washed glass beads and 200 µl phenol: chloroform: isoamyl alcohol (25:24:1). Cells were lysed thrice in a MagnaLyzer at 7000 rpm for 1 min and 200 µl of TE was then added. The beads were collected by a quick spin and supernatant transferred to a new Eppendorf tube. DNA was precipitated with 1 ml 100% ethanol, 10 µl of 4 M ammonium acetate and 2 µl glycogen (20 mg/ml stock). The pellet was washed once in 70% ethanol, air-dried and re-suspended in 50 µl TE. DNA yield was measured with a Nanodrop photometer.

DNA Isolation Buffer A

Triton-X-100	2%
SDS	1%
NaCl	100 mM
Tris-HCl pH 8.0	10 mM
EDTA	1 mM

II.2.4 RNA Isolation

Cell pellets from ~10^9 cells were washed in sterile water, resuspended in AE buffer and transferred to sterile Eppendorf tubes. 200µl 25% SDS and 3 ml acidic phenol: chloroform: isoamyl alcohol were then added, vortexed and incubated at 65°C for 10 min, with intermittent mixing every 2 min. The samples were centrifuged for phase separation. The aqueous phase was transferred to a fresh tube and RNA was precipitated using one-tenth volume of 3 M NaAcetate, 100% isopropanol and 2 µl glycogen (20 mg/ml stock). The RNA pellet was washed in cold 70% ethanol and resuspended in DEPC-water. qRT-PCR was performed as described in section II.1.3 and II.1.5.

In experiments with chemical treatment, cells were treated either with thiolutin (20 µg/ml in DMSO) or 1,10-phenananthroline (100 µg/ml in ethanol) and RNA was isolated as above.

Buffer AE

NaAcetate pH 5.2	50 mM
EDTA	10 mM

II.2.5 Flow cytometry analysis (FACS)

To determine cell cycle progression, the DNA content was measured by propidium iodide staining of ethanol-fixed cells (Moreno et al. 1991). 1 ml aliquots were pelleted and fixed in cold 70% ethanol. Cells were resuspended in 500 µl of 50 mM sodium citrate and treated with 1 mg/ml RNaseA for at least 2 h at 37°C. Cells were washed twice with 50 mM sodium citrate and resuspended in 50 mM sodium citrate containing propidium iodide to a final concentration of 8 µg/ml. Cells were stained at 4°C for 2 h (to overnight), washed with 50 mM sodium citrate and DNA content analyzed using the CellQuest pro software for the FACS Calibur machine.

II.2.6 ChIP analysis

Nearly 100 ml of exponential phase cells (OD_{600}= 0.350) were crosslinked with 1% formaldehyde at 25°C for 10 min. The formaldehyde was quenched with 0.125 M glycine and cells were centrifuged at 3000 rpm for 5 min. Cell pellets were washed twice with sterile water and once with Buffer I. Subsequently, cells were resuspended in 500 µl cold Buffer I supplemented with 1X protease inhibitor cocktail and 1 mM PMSF. An equal volume of acid washed glass beads were added and cells were spun at 7000 rpm for 1 min at 4°C in the MagnaLyzer. This step was repeated thrice to ensure maximum cell wall lysis. The whole cell extract was filtered through a 25G needle and collected in a 1.5 ml Eppendorf tube. The samples were sonicated in a diagenode Bioruptor at high intensity for 15 min, with 30 s ON/OFF intervals. The sonicated samples were centrifuged to remove cellular debris. The supernatant was diluted with Buffer I and pre-cleared with 50 µl Protein A agarose beads. 5% of the pre-cleared material was kept as input and the rest was equally divided into immunoprecipitation (IP) and a no antibody control. IP was performed overnight with of 3-5 µg of antibody. The following day, crosslinked immuno-

complexes were captured with 50 µl of Protein A agarose beads. Beads were washed once each with Buffer I, II, III and TE. Immunoprecipitated DNA was eluted from the beads in 1 mM TE + 1% SDS. Samples were reverse crosslinked with 3 µg/ml of RNaseA at 65°C overnight. Thereafter, samples were treated with 20 µg Proteinase K at 45°C for 2 h and purified using Qiagen PCR purification columns. DNA was eluted twice with 35 µl MilliQ water. 2 µl of each sample was used for qPCR. ChIP quantification was performed by the ΔΔCt method, as described in section II.1.7.

In experiments with chemical treatment, cells were treated either with thiolutin (20 µg/ml in DMSO) or 1,10-phenananthroline (100 µg/ml in ethanol) and ChIP analysis was performed as above.

ChIP Buffer I

HEPES pH 7.5	50 mM
NaCl	140 mM
EDTA (pH 7.5)	1 mM
Triton-X-100	1%
Sodium Deoxycholate	0.1%
PMSF	1 mM
Leupeptin	1 µg/ml
Pepstatin	1 µg/ml

ChIP Buffer II

HEPES pH 7.5	50 mM
NaCl	500 mM
EDTA (pH 7.5)	1 mM
Triton-X-100	1%
Sodium Deoxycholate	0.1%

ChIP Buffer III

Tris-Cl pH 8.0	10 mM
LiCl	250 mM
EDTA (pH7.5)	1 mM
NP-40	0.5%
Sodium Deoxycholate	0.5%

TE Buffer

Tris-Cl (pH 7.6)	10 mM
EDTA (pH 7.5)	1 mM

II.2.7 PCR for FISH probes

50-100 ng DNA template was amplified in presence of 0.5 µM primer pairs, 0.2 mM dNTPs, 1X Phusion HF or GC buffer, 3% DMSO and Phusion DNA polymerase. Primer annealing temperatures were chosen to be approximately 3^{o}C lower than the primer melting temperature. Cycling conditions were determined according to manufacturers' guidelines.

II.2.8 DNA FISH

Wild type (or *ts* mutants) cells were grown overnight to mid-log phase (OD_{600}= 0.5). The following day, cells were re-inoculated into fresh medium and incubated for an hour till the culture had an OD_{600}=0.30-0.35. Nearly 3 x 10^7 cells were resuspended in 1.2 M sorbitol and gently swirled for 5-10 min. Cells were fixed in freshly prepared para-formaldehye (4% final concentration) at 30°C for 30 min and quenched with 0.104 mM glycine. Subsequently, cells were harvested at 3000 rpm for 5 min and washed twice with PEM buffer. Cells were transferred to 1.5 ml sterile screw cap tubes and then washed once with PEMS. Cells were spheroplasted with 100 U of Zymolyase 100T in PEMS at 37°C for 1 h and pelleted twice at 2000 rpm for 1 min, by turning the caps for each spin such that all cells can be collected. To permeabilize the cytoplasm, spheroplasts were resuspended in PEMS +1% Triton-X-100 for exactly 6 min. Cells pellets were washed thrice with PEM, resuspended in PEMBAL + 1 mg/ml RNase A and incubated at 37°C for 2-4 h.

Hybridization was performed in solution. After RNase treatment, cells were resuspended in 200 µl 2X SSC for 15 min and allowed to sit down at room temperature. Cell pellets were obtained by two spins at 2000 rpm for 1 min each, turning the caps in between. Next, cells were resuspended in 200 µl 2X SSC + 10% formamide for 15 min at room temperature. This was followed by once each in 2X SSC + 20% formamide and 2X SSC + 40% formamide. Finally, cells were resuspended in 100 µl hybridization buffer and labeled FISH probe was added at a final concentration of 1 ng/µl. The hybridization mix (cells + probe) was denatured at

75°C for 5 min, followed by 2 min on ice and finally, incubated at 40°C overnight in a thermomixer at 700 rpm. The next day, cells were washed thrice with 2X SSC for 15 min at 700 rpm, room temperature and DNA stained with 0.2 µg/ml of DAPI in PBS. Excess DAPI was removed by a wash in PBS only and cells were resuspended in 30-50 µl PBS.

For microscopy, ~7 µl cells were dropped on poly-L-lysine coated slides and covered with 22 x 22 mm coverslips. Excess liquid was removed by blotting and slides were sealed with nail varnish. When imaging was not performed immediately, slides were stored at 4°C in the dark.

Images were acquired using the SoftWorx program (Applied Precision) on the DeltaVision microscope (Applied Precision) at the Micron facility, Department of Biochemistry, University of Oxford, UK. Cells were visualized with a 100X 1.3NA objective lens. For each experiment, z-sections were taken at 200 nm intervals. Background light was corrected by the built-in deconvolution algorithm of the SoftWorx program. Final images represent maximum intensity projections of z-stacks obtained using the ImageJ/Fiji software (NIH).

DNA FISH probes were prepared using the FISH-Tag Kit (Molecular Probes, Invitrogen) with either Alexa-488 or Alexa-555 fluorophores. All FISH probes (except the centromeres) were ~10 kbp in length. The centromeric probes targeted the *dg* region in the outer most centromeric repeats and were ~2.2 kbp in length. Overlapping PCR fragments (~2.5 kbp) were amplified using Phusion DNA polymerase and pooled in equimolar quantities. The labeling reaction was performed according to the manufacturers' instructions. The strategy involved modifying the template DNA by incorporating amino-allylT nucleotides, which then acted as a bait to couple the reactive dye. The labeled probes were always stored at -80°C.

PEM

PIPES	100 mM
EGTA	1 mM
$MgSO_4$ (pH 7.6)	1 mM

PEMS

PIPES	100 mM

EGTA	1 mM
$MgSO_4$ (pH 7.6)	1 mM
Sorbitol	1.2 M

PEMBAL

PIPES	100 mM
EGTA	1 mM
$MgSO_4$ (pH 7.6)	1 mM
BSA	1%
NaN_3	0.1%
Lysine monohydrochloride	100 mM

Para-formaldehyde (PFA) solution in PEM

PFA	30%
NaOH	10 mM

Dissolve by heating at 65°C and filter sterilize.

20X SSC

Sodium chloride	3 M
Sodium citrate	300 mM

Adjust pH to pH7 by addition of hydrochloric acid

Zymolyase solution

Zymolyase 100T, 10 mg/ml in PEMS

Hybridization Buffer

Dextran sulphate	10%
Denhardt's	5X
Formamide	50%
Salmon sperm DNA	0.5 mg/ml
SSC	2X

100X Denhardts solution:

Ficoll	2%
Polyvinylpyrrolidone	2%
Bovine serum albumin	2%

II.2.9 Co-Immunoprecipitation (co-IP)

Co-IP was preformed as previously described (Asp and Sunnerhagen 2003). Cell pellets from ~100 ml cultures (OD_{600}= 0.8) were washed twice with sterile water, once with Co-IP buffer A + 1% NP-40 and once with Co-IP buffer A + 0.25% NP-40. Finally, cells were resuspended in Co-IP buffer A + 0.25% NP-40 supplemented with 1X protease inhibitor cocktail and 1 mM PMSF. Cells were lysed using acid washed glass beads in a MagnaLyzer at 7000 rpm for 1 min, repeated three times. The whole cell extract (WCE) was filtered through a 25G needle and collected in a 1.5 ml Eppendorf tube. The filtered WCE was diluted with Co-IP buffer A + 0.25% NP-40 and split as follows. 10% of the WCE was kept as input (in) and the remaining was used for immunoprecipitation (IP). IP was performed overnight at 4°C.

For anti-TAP pull-down, IgG Sepharose beads were prepared as follows. 50 µl beads were first washed in MillQ at 4°C for 10 min. This was followed by three washes in Co-IP buffer A + 0.25% NP-40 at 4°C for 10 min. Beads were then added to TAP pull-down samples and incubated overnight at 4°C. The protein-IgG-Sepharose complexes were washed twice with Buffer A + 0.25% NP40 and resuspended in the same buffer supplemented with 1X SDS-PAGE loading dye. Along with the input (mixed with loading dye), the TAP pull-down samples were boiled at 95°C for 10 min. For each sample, the corresponding input and pull-down were loaded on the gel and western blot was performed as described below (section II.3).

For anti-RNAPII Co-IP, WCE was pre-cleared with 50 µl Protein A agarose beads and then split into 10% input and 90% IP material. After overnight IP (8WG16 or anti-Phospho Ser5 RNAPII antibodies, Abcam), the protein-antibody complexes were captured on 50 µl Protein A agarose beads, gently washed with co-IP Buffer A + 0.25% NP-40, mixed with loading dye, denatured and run on the gels as in section II.3.

Co-IP Buffer A

Tris-Cl (pH8.0)	50 mM
NaCl	50 mM
Triton-X-100	0.2%
NP-40	0.25% or 1%

PMSF	1 mM
Leupeptin	1 µg/ml
Pepstatin	1 µg/ml

II.2.10 Mass Spectrometry (MS)

II.2.10a Co-IP

Co-IP was performed as in section II.2.9. The samples were run in a 4-15% bis-tris gel at 180 V, till they descended ~5 mm into the wells. Gel bands were cut and sent to the Mass Spectrometry Facility at Sir William Dunn School of Pathology, University of Oxford. The MS specialist at the above facility performed the following steps.

II.2.10b Tryptic Digest

The gel bands were washed twice in 100 µl of Solution B for 30 min. The bands were then washed in 100 µl of 100% acetonitrile for 10 min. The dehydrated gel pieces were dried in a SpeedVac, 100 µl 10 mM DTT was added and incubated at 37°C for 30 min. The gel pieces were washed twice with 25 mM ammonium bicarbonate and once in acetonitrile. 100 µl 55 mM iodoacetamide was added for 1 h in the dark. Subsequently, bands were washed twice in 100µl solution B, 10 mins, 1x 100% acetonitrile and vacuum dried. Tryptic digestion (20 µl of trypsin) was performed overnight and stopped with 1 µl formic acid. The gel was incubated in extraction buffer and supernatants containing peptides were all pooled together and vacuum dried.

II.2.10c C18 Desalting

The peptide containing samples were acidified by formic acid. The C18 tip was prepared by loading 30 µl of wetting buffer in to the top of the C18 tip. The column was equilibrated with 30 µl of equilibration buffer. Subsequently, the sample was loaded onto the C18 tip. The flow through was collected in Eppendorf tubes and peptides bound to C18 were desalted by 30 µl wash buffer. The desalted peptides were eluted by 30 µl elution buffer into an Eppendorf tube and vacuum dried.

11.2.10d MS data analysis

Samples were analysed on an Ultimate 3000 RSLCnano HPLC (Dionex, Camberley, UK) system run in direct injection mode coupled to a QExactive Orbitrap mass spectrometer (Thermo Electron, Hemel Hempstead, UK). Samples were resolved on a 25 cm by 75 micron inner diameter picotip analytical column (New Objective, Woburn, MA, USA) which was packed in-house with ProntoSIL 120-3 C18 Ace-EPS phase, 3 µ bead (Bischoff Chromatography, Germany). The system was operated at a flow-rate of 300 nL/min. A 120 min gradient was used to separate the peptides. The mass spectrometer was operated in a "Top 10" data-dependent acquisition mode. Precursor scans were performed in the Orbitrap at a resolving power of 70,000, from which the ten most intense precursor ions were selected by the Quadrupole and fragmented by HCD at a normalised collision energy of 25%. The Quadrupole isolation window was set at *3 m/z.* Charge state +1 ions and undetermined charge state ions were rejected from selection for fragmentation. Dynamic exclusion was enabled for 40 s. Data were converted from .RAW to .MGF using ProteoWizard (Chamber et al. 2012). CPFP analysis was done according to Trudgian et al. 2010 and SINQ normalization was performed according to Trudgian et al. 2011.

II.2.11 Bioinformatics analysis

The bioinformatics analysis described below was performed by Margarita Schlackow, a fellow DPhil student at the Mathematics department, University of Oxford.

II.2.11a Gene expression analysis

Gene expression data was extracted from RNA-Seq data (Marguerat et al. Cell 2012), while Rad21 (224 in total) and Mis4 peaks (67 in total) were taken to be the ones detected by Schmidt et al. (Supplementary tables, Genome Biology, 2009). Peaks in centromeric regions and in the region labeled as "pi-transcript" were not considered (Schmidt et al. 2009). To adjust the genomic annotation to the current release, all coordinates < 1,280,000 (as in Schmidt et al. 2009) were shifted to the right by 80 kbp and > 1,29,750 by 41.25 kbp (the region in between was

excluded). The gene expression value for a Rad21 peak was taken to be the highest RPKM of the genes overlapping it. We compared the gene expression of Rad21 peaks which overlap a Mis4 peak (from now on M-peaks) and which do not (from now on C-peaks). We detected 44 M-peaks out of 224 Rad21 peaks in total.

II.2.11b Comparison to the background distribution

Let A be the average gene expression of the M-peaks and B of the C-Peaks. We then randomly sampled 44 Rad21 peaks 10^6 times. A Histogram with 100 bars presents the distribution. The right hand p-value was evaluated by summing the bin counts of the bin where A falls within and all the bins above divided by 10^6. Equivalent analysis was performed for the C-peaks.

II.2.11c Permutation Test

We computed the significance in the difference of gene expression between the M-peaks and C-peaks. The statistic was A-B. A permutation test was performed by randomly splitting the 224 peaks into groups of 44 and 180, 10^6 times, each time the difference in averages of gene expression was computed. We hence obtained a distribution for the statistic and the significance of the value A-B. The Histogram presentation and computation of the p-value was performed as for the background distribution.

II.2.11d Boxplots

Boxplots were created using MATLAB. The median (red) is presented as a line, upper and lower quartiles (blue, q3 and q1 respectively) are presented as a box. The whiskers are given by q3 + 1.5(q3 – q1) (upper) or q1 – 1.5(q3 – q1) (lower). The default of 1.5 corresponds to approximately +/–2.7σ (where σ is the standard deviation) and 99.3% coverage if the data are normally distributed. Outliers are plotted as red crosses if their value is higher than the upper whisker or smaller than the lower. The plotted whiskers extend to the adjacent value, which is the most extreme data value that is not an outlier.

Please note that average and median are different statistical terms, hence the value of the average as computed for the histograms does not coincide with the median in the boxplot.

II.3 WESTERN BLOT ANALYSIS

To prepare yeast protein whole cell extracts cells pellets were washed twice with sterile water, once with Co-IP buffer A + 1% NP-40 and finally, resuspended in Co-IP buffer A + 0.25% NP-40, supplemented with 1X protease inhibitor cocktail (Roche) and 1 mM PMSF. Cells were lysed using acid-washed glass beads in a MagnaLyzer at 7000 rpm for 1 min, repeated three times. The lysed whole cell extract (WCE) was filtered through a 25G needle and collected in a 1.5 ml Eppendorf tube. The filtered lysates were briefly sonicated in a bioruptor Diagenode sonicator at high intensity for 5 min, with 30 s ON/OFF intervals. Cellular debris were removed by centrifugation at 13,000 rpm for 10 min.

To prepare mammalian WCE, cells were washed with PBS and scraped off the culture dish in 200 µl RIPA buffer supplemented with 1X protease inhibitor cocktail (Roche) and 1 mM PMSF. Cells were resuspended in RIPA buffer and incubated on ice for ~20 min. A brief sonication was performed with bioruptor Diagenode sonicator at medium intensity for 5 min, with 30 s ON/OFF intervals. Cellular debris were removed by centrifugation (13,000 rpm, 10 min).

Protein estimation was performed using Bradford reagent, with serially diluted BSA standards ranging from 0.2-3 mg/ml. WCE was mixed with 1X SDS-PAGE gel loading dye and boiled at 95°C for 10 min. Alongwith a pre-stained marker, denatured samples were run on 4-15% bis-tris precast gels in 1X Tris-Glycine-SDS running buffer at 180 V.

To set up the transfer a PVDF membrane was used. The membrane was activated in 100% methanol (15 s) and gently shaken in the transfer buffer for ~10 min along with the gel. Transfer was performed in the Semi-Dry Transfer apparatus (BioRAD) at 15 V for 2 h.

The membrane was either blocked overnight in PBST + 5% milk at 4°C or for 1 h at room temperature. Primary antibody was added in PBST + 5% milk, either

overnight (4°C) or room temperature (1.5 h). The membrane was washed three times with PBST for 15 min at room temperature. Secondary antibody was added at a dilution of 1:5000 (anti-rabbit) or 1:10,000 (anti-mouse) for 1 h at room temperature. Washes were repeated three times with PBST for 15 min at room temperature. After a final wash with PBS, the proteins were detected using Millipore detection reagent.

RIPA lysis buffer (mammalian WCE)

Tris HCl pH 7.5	50 mM
Sodium Chloride	150 mM
NP40	150 mM
Deoxycholate	0.5%
SDS	0.1%
Glycerol	10%
PMSF	1 mM

Co-IP Buffer A (yeast WCE)

Tris-Cl (pH8.0)	50 mM
NaCl	50 mM
Triton-X-100	0.2%
NP-40	0.25% or 1%
PMSF	1 mM
Leupeptin	1 µg/ml
Pepstatin	1 µg/ml

SDS Gel Loading Dye

Tris HCl, pH6.8	62 mM
Glycerol	25%
SDS	2%
Bromophenol blue	0.25%

Tris Glycine Running Buffer

Tris-Cl	250 mM
Glycine	1.92 M
SDS	1%

Transfer Buffer

Tris-Cl	250 mM
Glycine	1.92 M
SDS	1%
Methanol	20%

PBST

0.1% Tween in PBS

Blocking solution

PBST + 5% skimmed milk

II.4 REAGENTS

Most materials were obtained from either Sigma-Aldrich or VWR for BDH products and were 'AnalaR' grade. Materials described in this chapter are listed below.

α-amanitin	Sigma
Flavopiridol	Sigma
KM05283C	Fermentas
Agarose	Melford
Bradford reagent	Biorad
Deoxyribonucleotide triphosphates (dNTPs)	Invitrogen
Dithiothreitol (DTT)	Invitrogen
Dulbecco's Modified Eagle Medium	PAA
Dulbecco's PBS	PAA
ECL Western Blot Detection Reagents	Millipore
Foetal Calf Serum	Gibco BRL
Glycogen	Roche
HEPES	Melford
Hybond-PVDF protein transfer membrane	GE healthcare
Maxi-prep kit	Qiagen
Penicillin/Streptomycin	PAA
Trizol	Invitrogen
RNase-free tRNA	Roche

TOPO TA Cloning Kit	Invitrogen
Trypsin	Gibco BRL
Thiolutin	Tocris Bioscience
Phenanthroline	Sigma
Zymolyase 100T	Amsbio
FISH Tag Green KIT-Alexa 488	Life Technologies
FISH Tag Orange KIT-Alexa 555	Life Technologies
DAPI	VWR
YES medium	Formedium
YES Agar	Formedium

II.5 ENZYMES

The enzymes described in this chapter were obtained from the following suppliers.

RNase out	Invitrogen
Proteinase K	Roche
RNase A	Roche
T4 DNA Ligase	Roche
QuickLigase	NEB
SuperScript III reverse transcriptase	Invitrogen
DNase I-RNase free	Roche
GoTaq DNA Polymerase	Promega
Phusion Taq DNA Polymerase	NEB
Sensimix (Real Time PCR reaction mix)	Bioline
Hind III	Roche

II.6 ANTIBODIES

Anti-Actin (A2066)	Sigma
Anti-Mouse (M4155)	Sigma
Anti-Rabbit (R5506)	Sigma
Anti-RNAPII (N20X)	Santa Cruz Biotech
Anti-Ser2 H5 (MMS-129R)	Covance
Anti-RNAPII 8WG16 (MMS-126R)	Covance

Anti-RNAPII Ser5 (ab5131)	Abcam
Anti-RNAPII Ser2 (ab5095)	Abcam
Anti-RNAPII Ser5 (ab24759)	Abcam
Anti-H3 (ab1791)	Abcam
Anti-GFP (ab290)	Abcam
Anti-TAP (CAB1001)	Pierce Antibodies

II.7 PRIMERS FOR *Alb* ANALYSIS

	***Alb* Pol II ChIP**
A_F	GGCAGTCTTACACATTTTTGGCAGC
A_R	TGGTCCCCGTGTACTCATTCCCA
B_F	TTGAGACACTGTGTGCTTTATGGAA
B_R	TCTCTTTGTCTCCCATTTGAGCAT
C_F	GGGGTAGGAACCAATGAAATGCGAGGT
C-R	TGCCAGAGGCTAGTGGGGTTGA
D_F	GTTGAGGACCAGGAAGTGTGCA
D_R	GTCCTGCTAACACCAAAGGAAAGC
E_F	CAGAGACTGCCTTGTGTGGAAGACT
E_R	GGGCCATGACCCAGATACCA
F_F	TCAGAAATGGTGAGACCCCTG
F_R	ACCCTCTCACCCCAACTCGT
G_F	AAGACATGAAGACTCAGGACTCATCTT
G_R	CGCGCCTTTGGAAGATAGAGTT
H_F	TAATAACGAGCCAGGTGTGGAA
G_R	AATTATTTCCCCATGGATGCC
H_F	GGAAATAATTCCTATCTGTCTGCCAT
H_R	CAAGTCACGCTGACAACATAACCT
I_F	GGCGGTGTTTGCCAAGTTTCCT
I_R	CCCCTTCCAGAAGCTTAGTACCTTTCA
J_F	CGGATTCTGGAGTTAAGGACAGGGT
J_R	TGGGAGTGCAGGGTGAATCA
K_F	GGGAGCTGAGTTGAGGAGCAGT
K_R	TGGGACCTTCATAAATCACAAGCC

	***Alb* 3C Analysis control template primers**
control-1F	TCCCCAAACTTAGCTTAGAAGCC
control-1R	TGTCACCTCATGCTCAAGGGA
control-2F	CGCCACACGGGAAGCACACAAT
control-2R	TCTTGACACTTGGCCAGCCTGG
control-3F	CCCACAAGCACGAGAAAACA
control-3R	CAGCCAAAGCCTGGAGAGAA
control-4F	TCGCCGAGAAGCACGTAAGA
control-4R	CCCGTTAAGTGTTTATATCTGTGCA
control-5F	TCACAGTTCGCATCTTTCTGTTGA
control-5R	GCACCCAGACTCAGCTACAGCTAT
control-6F	GAATCAGGACTCCTCTCAGCTCATA
control-6R	CCATCCCTCTGTCCAAAGACA
control-7F	AAAAGGAGACCAAACCAGGGA
control-7R	GCCTTATGTCTATAGGTTCTTTGCTTGA
control-8F	CTTTGGCTCAGAGGAATCCCT
control-8R	GTATATGAACAACTGGCTCCTGGA
control-9F	AACAGGTTATGTTGTCAGCGTGA
control-9R	TCACAATCAAAAGGACACTTATAGATCA
control_Gapdh Site1F	TTTCTCAGTCTTTCCCCTTAGTTCGA
control_Gapdh Site1R	GACCAGGCGCCCAATACG
control_Gapdh Site2F	AGTTCTGATCTCAGCTCCCCTGT
control_Gapdh Site2R	GGCGGAGATGATGACCCTT
	***Alb* 3C primers**
3C_1F	AGGTGGAGCTTACTTCTTTGATTTGA
3C_2F	GCATGGAGACAATAACAGATCTGC
3C_3F	TTTTGCATTTATGCTCCTGCCA
3C_4F	ATTTGCATCTGAGAACCCTTAGGTGG
3C_5F	GACAATAGAGGTGAAGTCAGAATAGCAG
3C_6F	GATGGCCACTATCTAGTGAACTTCTGC
3C_7F	AAACCAGAGCAAAGGAGTTAGCCC
3C_8F	CCATGTGGTTTGACCCAATCATTC
3C_9F	GTAGTTCCACATGGGCAACACAAT
Gapdh 1R	CCTGGCAGGGCTTTTATAGCA
Gapdh 2R	GAAGACACCAGTAGACTCCACGAC

Loading_Gapdh Ex6F	TGACATCAAGAAGGTGGTGAAGC
Loading_Gapdh Ex7R	CCGAGTTGGGATAGGGCCTCT

II.8 PRIMERS FOR *NFκB1* ANALYSIS

	***NFκB1* 3C control primers**
1_F	CATCTTTCACCATTAATCCAGAAACCT
1_R	CTGAAGAGACCGCTACTGGCA
2_F	CTTGTCCTTTCATAAGAAAAATTTCCCAT
2_R	AGCTGGTTCAGCAGGGACTACA
3_F	GGAGGATTTTGAGTGAAGGAGTGATA
3_R	GAGTCTGAATAAACTGTGCTTAACATAAATGT
4_F	AGGCTGCTTTGGAAGAGAAGAAAGTA
4_R	ATCCAGCTGTCCTGTCCATTCTTA
5_F	CCCTCAGTTAAAATGTCAGTATTGCCA
5_R	GTAATTCCTTGAATATGACTGTTAAGCAGG
6_F	GATCATTCCGTATAGCCTGGTTATCAC
6_R	CCCTTAATTTGAATGCAAAATGATGCT
7_F	GTGTCTTCCTAGGCCACATCCA
7_R	TTCTAGATCCCTGAGGAATCGCC
8_F	CATTGTGTTCTGAGGGAGCCA
8_R	CTCGGGGAGACTGACTCTGAGTT
	***NFκB1* 3C primers**
3C_1F	TGCCTGGTGAGGACCTGATTAC
3C_2F	GTAATACAAATTTCATCTGCTTTAATTTATTGCC
3C_3F	AGCCCACAGTAGTTTATATAGTGAAGGC
3C_4F	CCTATTGATTCCTCTTTGCTGTCTAGA
3C_5F	CTGCAAATGAAGAATTAAAAGGCC
3C_6F	ATGGTCAGATAGCCTAGGGCCTT
3C_7F	TGTGTTAGAGGGAGAAACAAATACCTGT
3C_8F	GCCTGTAGCCATCTCTCTCTTTCC
control-GAPDH_1F	AATACAGCTTCCCCTCTTCCCA
control-GAPDH_1R	ATATGTGAGCAGCCCTAGGCCA
control-GAPDH_2F	AGATTTGGTCGTATTGGGCGC
control-GAPDH_2R	AGCATCGCCCCACTTGATTTT
GAPDH_3C_1F	TCTGACCTTTACTCCTGCCCTTTGA
GAPDH_3C_2F	AGTGCTACATGGTGAGCCCCA

	***NFκB1* Pol II/RT-qPCR ChIP primers**
A_F	AAGGTCGACTGTTGGGGCCT
A_R	TTTCGTGCCCACTCTCCATC
B_F	ACCAAAGAGGAACTACATCATGTCCT
B_R	TTCTGCCATTCTGAAGCTGTGTATTAA
C_F	CACCCTGTCATTTGTTCATCACA
C_R	CCACATAGTTGCAGATCTATATAAGAAGCA

D_F	CTAATAACATCCTCACAGTATGTCCCAA
D_R	CTCGTCTGAGGGGAGAAAGCA
E_F	AGTCAGTCTCCCCGAGAACCTT
E_R	GGAGCTAGACAGAGCTGGGTGT
F_F	ACAACACGTGGGAATTGTGGA
F_R	TCAGATGGATTGTTTTTAAGTGATTAGCC
28S LF	GCAGGAGGTGTCAGAAAAGTTACCACAG
28S LR	CTAACCTGTCTCACGACGGTCTAAACCC

II.9 REGIONS SELECTED FOR DNA-FISH ANALYSIS (CHAPTER-IV)

Table II.9.a **Cohesin peaks used for FISH analysis, overlapping with Mis4**

FISH REGION	Rad21 PEAK (Schmidt et al. 2009)	LOCUS	PEAK HEIGHT	PEAK LENGTH (kbp)
dis2	158	3132251	0.77	4000
tif11	163	3231251	0.89	15500
ubp9	147	2896751	1.20	6500
srp1	169	3323001	0.87	8000
rpl3801	34	676751	0.97	7000
cendg	103	2078501	1.13	71500

Table II.9.b **Mis4 peaks used for FISH analysis, overlapping with cohesin**

FISH REGION	Mis4 PEAK (Schmidt et al. 2009)	LOCUS	PEAK HEIGHT	PEAK LENGTH (kbp)
dis2	52	3132376	0.73	2750
tif11	54	3234751	0.45	3500
ubp9	49	2899126	0.63	2750
srp1	55	3322126	0.73	6750
rpl3801	13	675126	0.63	2750

Table II.9.c **Rad21 peaks used for FISH analysis, non- overlapping with Mis4**

FISH REGION	PEAK (Schmidt et al. 2009)	LOCUS	PEAK HEIGHT	PEAK LENGTH (kbp)
thi5	187	3652001	0.84	9500
mug37	191	3720876	0.80	12750
cut3	44	921751	1.00	7500
swi3	154	3048376	0.78	14750
pla1	40	850001	0.93	7500
psh3	54	1089126	0.44	8750
med8	162	3178376	0.63	8750

II.10 PRIMERS FOR *S. pombe* ANALYSIS

	***S.pombe* DNA FISH primers**
FISHDis2-F1	TGGATCAACTTTTCAGCCTTG
FISHDis2-R1	TATACTGGTCTCACGTTGGA
FISHDis2-F2	AAGCACCTTACTCTCTCTTAC
FISHDis2-R2	ATCATAAGCAGGTGCAGCTT
FISHDis2-F3	GAGATAGCAACCTACTTCAAGA
FISHDis2-R3	GCTGGTATTAATTAGTGGACATGC
FISHDis2-F4	ATGATGAAGGCTATACTCCCTTGT
FISHDis2-R4	TCATTAGGCGTCTTGTGACCTT
FISHTif11-F1	ATTGTCGCGGTTTCTCAGTT
FISHTif11-R1	TTGACGGTGTCAAGCGATTA
FISHTif11-F2	GTGCCTCATCAGCAGTATAT
FISHTif11-R2	CCATCATGTCTCTTACCATG
FISHTif11-F3	CGTATTCATGACTGACTAAAGCTGT
FISHTif11-R3	TAGAATCCCCGGTCTTCACA
FISHRif11-F4	CAGGCTCTAATAATTGTAAGCGCA
FISHTif11-R4	TACTTGTTCGTGGTTTAACCCA
FISHTif11-F5	GGCACTCACCTTTCTAATTCGTA
FISHTif11-R5	TATGCTGGTAAATGGCCTTTCT
FISHubp9-F1	GCCTAAGATTATGATCACCAACGTA
FISHubp9-R1	TTCGATGTATGAAGGTTGCC
FISHubp9-F2	CTCGTTGTTCCCAATTTGAAC
FISHubp9-R2	GCAGTTTCTCAATAGGAGATAGCC
FISHubp9-F3	CCTTCGAGTTTGCGTTATAAAC
FISHubp9-R3	GAAGAGAAGAAGAGTATGCCGAA
FISHubp9-F4	CTGCTTTATTGCCTTCTTCTGTTC
FISHubp9-R4	TGGAAGGAGAGATTTCTGCTCT
FISHsrp1-F1	GTGTACATCAATAGTGTTCGTTCCC
FISHsrp1-R1	ATTGGATGATGAGACTCTCCGT
FISHsrp1-F2	CAGCAACATGACCACCATAGATAT
FISHsrp1-R2	AAGAGAAATTAGGGACTCGAGTGTA
FISHsrp1-F3	AGGAAGGTATGCGATCGGATA
FISHsrp1-R3	TGCTAGTTTAATACACAGGCGCT
FISHsrp1-F4	TTATCAGTATTAGATTCCCAGCGC
FISHsrp1-R4	TGGGCTAATGCAACATTCGT
FISHrpl38-F1	AACTTCAATCTAACGCCCATTC
FISHrpl38-R1	GGAACTTTCACGTACTCCTTTTCT
FISHrpl38-F2	GCTTCAGGCGTTACAGTACAA
FISHrpl38-R2	TTGACGTTGGTGTTTATTTGCC
FISHrpl38-F3	CCTACGATTACACCATCAGGATTA
FISHrpl38-R3	TCGAATAAGCTTTCTGCGCTT
FISHrpl38-F4	AGTTGACTGTCTCTCAAAGATGGAT
FISHrpl38-R4	CCGGAGATTTCACATTTAGACCT

FISHmed8-F1	TGCGTTTGAAGAAGTAAGGACT
FISHmed8-R1	CTCCATCAAAGTTGACAACAC
FISHmed8-F2	ACTTGATGTAGTTCTGTGAGGA
FISHmed8-R2	AGTCCAGGTGCTTTCATCTAA
FISHmed8-F3	GAAATCCCTCCAGGTAGATAT
FISHmed8-R3	ATTCAAGCACTGTGTGAACGA
FISHmed8-F4	GGCAACTATAGTTGTATTCGGA
FISHmed8-R4	CATGTTCAGCAACAGAAGAGTT
FISHpsh3-F1	CTGCCAACCTTTTGAGTCAACT
FISHpsh3-R1	CGTATAGCTTACTTGACTGCATAGCAT
FISHpsh3-F2	CTAGTCCAACTTACCAAATTCGCT
FISHpsh3-R2	GCGACCATTGTGAAAACTGA
FISHpsh3-F3	GTTTCGTCAGTTTTCACAATGGT
FISHpsh3-R3	AGCTTTGAATGCAGCAGACT
FISHpsh3-F4	TCACTGCTAGGAACAAGGGATACT
FISHpsh3-R4	CATGCAAATCGCTTACCCAT
FISHpla1-F1	GTTCAATACAGTCGAACAGTTGCA
FISHpla1-R1	CCCCAACATCGTAATTACTATCC
FISHpla1-F2	ATGGGAGTCAAGAGCGAAAAC
FISHpla1-R2	TTGGCAAATTCGAGCTACCAT
FISHpla1-F3	CGCTAATGTCGTAGGATTTCCT
FISHpla1-R3	TGGTATGGCTTATTATGCCTCTCT
FISHpla1-F4	CAATCCGGCAACATAATGACA
FISHpla1-R4	TCGCTATCTGAAAGAGTTTCCCA
FISH-thi5-F1	ACTGGCAAATTAAGTCAGCAGCT
FISH-thi5-R1	CTATTAATCGAACCGAACGGTAGT
FISH-thi5-F2	GTCATTGAGATCTTTACGCTGA
FISH-thi5-R2	GCTGATCGGATAATAACGTTGAACT
FISH-thi5-F3	ACGCTTGGTTATTGGGATGTAA
FISH-thi5-R3	AACCCATACTTCCTGTACGTTCAT
FISH-thi5-F4	GGTTCTCCTAGAGCATCTAAGAATGA
FISH-thi5-R4	AACATGACTGAAAAGGATACTGAGC
FISH-mug37-F1	AAGCGTTTCTATCCAACCAGACT
FISH-mug37-R1	CACGTAAAAGTCGTGTTTTAACGTC
FISH-mug37-F2	GAATTAGAAGCTGAAGTTAGGAAGC
FISH-mug37-R2	CATGACGCTTCATTGTTTTATGC
FISH-mug37-F3	GCAAGCACTGTTCGTTTTTTCT
FISH-mug37-R3	GCCCTGTTATTAACTGAGCGTAGT
FISH-mug37-F4	TCAGTTAATAACAGGGCGCAAT
FISH-mug37-R4	GGTTAAGAAGGTAATCTGATAGCTTCTC
FISH-cut3-F1	CAAGCATTTGCTCAGCAACA
FISH-cut3-R1	TTCGAGATAGCTATTTGGCGGT
FISH-cut3-F2	GGTTCAAGAGCTTTGGTTTAACCT
FISH-cut3-R2	CCAGACAACAATGGACTCAACTT
FISH-cut3-F3	CAGTTCATCTTCAGAATATTCGGGA

FISH-cut3-R3	TCAAGGAGAAGTCGAATCAATAGC
FISH-cut3-F4	AGACCATCATCACCTTCTGATATAGC
FISH-cut3-R4	GGAAGCGATTTAATTTCAGGCA
FISH-swi3-F1	CGATGTTTATGTGCGTGTTGA
FISH-swi3-R1	TTCTGCAGTAGCCAATCCAATA
FISH-swi3-F2	TTGACTGGATCCTCTAAGGGTATT
FISH-swi3-R2	ACCTCCTAATATTGCCCTGACAT
FISH-swi3-F3	CACCGAGTTTTAACGATGTCA
FISH-swi3-R3	CGACGTTGTTTTCTTTGCTCA
FISH-swi3-F4	GGGTAAACAAAGATGGCACCA
FISH-swi3-R4	CACAATACCTACAATTTGCTTCC
cen-dg-F1	CAAGTCTCTTAATAAGTAAGGAAGTCGC
cen-dg-R1	TTTGTACTTGCTAAGCATGTCCTGT
cen-dg-F2	TCAGGAGTGGAGACCATAATCATT
cen-dg-R2	CGCAGCATATTGGTCGTTTTTA

	***S.pombe* ChIP primers**
dis2-ChIP-F	GCTAGACTGATGATTCGCCCTTATA
dis2-ChIP-R	TTTACCGGCGAGTTAGCAGAGTA
776-ChIP-F	AACGCTTCGAAATGTCTGATCTT
776-ChIP-R	TCATGTATACATCAACTCCCAGCA
med8-ChIP-F	AGGCAGGAATCAGAAGGAGAGTT
med8-ChIP-R	CAATCTCACCCACTTCTCCACTT
mrs2-ChIP-F	GTGAACGAAGAGCCTATTTACCGT
mrs2-ChIP-R	ATTGCGTATCTTACGTCGAGTAGCT
hrf1-ChIP-F	CTACTTCTGTTGCCTAATAGTAAAAACCGT
hrf1-ChIP-R	ACTGTCCGTGAATCAATCGTG
srp1-ChIP-F	ATTCGTGGCAGATTTTGCTCT
srp1-ChIP-R	AAACAAGCGTTTGGCAGTAGC
ubp9-ChIP-F	CAGGACAATTGGATGCTAAGTGTATT
ubp9-ChIP-R	GGAACCTCTTCATTGCTTAAGCAT
negative control_F	AAGGTATAAACAGCAAATCGGCA
negative control_R	GCACCCATCATGTCGTTGTA
pla1-ChIP-F	GTTGCTTGCTTAACACTCACGAGTA
pla1-ChIP-R	CTGGTTCAATATCGATAATATACAAGCAGT
thi5-ChIP-F	CTTCCTTCAAGAATCGTATAGATATCGGT
thi5-ChIP-R	CGCTAGCTATATTGACTGTTCCGTAT
cut3-ChIP-F	GGATACGTTCTATATGGTATGCATACGA
cut3-ChIP-R	ATCTTCACTCGCTTTTAAAGCCTGT
rpl3801-ChIP-F	TTACTGAGGTTGGAAAGAAGGCAT
rpl3801-ChIP-R	GGGTCTGCTTATAAAATTCGTGAAC
rpl29Prom-F	GGTATTCGCTTTAGTAAACTACGCT
rpl29Prom-R	TGAGCCTTCTTGTTTTGGTTGT
Dis2Prom-F	GTCTTACTTTTCCTCGGAACACTTTACT
Dis2Prom-R	TCAAGTCTGCTTTTCCCCACT
egd2Prom-F	TTTGCTATGCCCTACTACCGCT
egd2Prom-R	AGCAGGAAGCTCAGAAATTTTCTCT

rpl1701Prom-F	TACCCTAAGTGTCCAAACCAGTCA
rpl1701Prom-R	CTTGAAATGGGTTCTCAAGTAAGCT
nuc1Prom-F	CACCGGACTTTAGCAAATTACAGTC
nuc1Prom-R	ATCGTCAACATCGTAAATCCCA
cendg_F	AAGACTGTTGTTGAGTGCTGTGGA
cendg_R	CCATGCTTTTAGTGCGGTCA
telo_F	GCCGACAGTGATATGGCTTA
telo_R	CTGCAATAGTCCTTCGGTATTG

	***S. pombe* qRT-PCR primers**
dis2-RT_F	AGTTTGATAATGCAGGTGCGAT
dis2-RT_R	GGAGTCATGTGCCAATTTTGA
act1-RT_F	CGCTGTGATTATCGATTAAAGAACTGT
act1-RT_R	ACACATTCAACCATTTGGCTTC

II.11 LIST OF STRAINS

	***S.pombe* strains used**
3678	*h-ura4 psm3-GFP-natR*
3505	*h-ura4 psm3-GFP-ura4+ wpl1delta::kanR*
psm3-TAP	*h+ura3 psm3-TAP-kanMX6 leu2 trp1 his3 his4 ura4-D18 leu1-32 ade6-M216*
mis4-TAP	*h+ura3 mis4-TAP-kanMX6 leu2 trp1 his3 his4 ura4-D18 leu1-32 ade6-M216*
Sssl3-TAP	*h+ura3 ssl3-TAP-kanMX6 leu2 trp1 his3 his4 ura4-D18 leu1-32 ade6-M216*
pds5-TAP	*h+ura3 pds5-TAP-kanMX6 leu2 trp1 his3 his4 ura4-D18 leu1-32 ade6-M216*
wapl-TAP	*h+ura3 wpl-TAP-kanMX6 leu2 trp1 his3 his4 ura4-D18 leu1-32 ade6-M216*
rad21-TAP	*h+ura3 rad21-TAP-kanMX6 leu2 trp1 his3 his4 ura4-D18 leu1-32 ade6-M216*
psc3-TAP	*h+ura3 psc3-TAP-kanMX6 leu2 trp1 his3 his4 ura4-D18 leu1-32 ade6-M216*
mis4-GFP	*h-leu1 mis4-GFP-LEU2+*
rad21-GFP	*h- rad21-GFP:LEU2 leu1-32*
rad21-K1ts	*h- rad21-K1:ura4+ ade6-M216 leu1-32 ura4-D18 his7-366*
mis4-367	*h- mis4-367*
eso1-GFP	*h- eso1-GFP-kanR+*
972	*h+*

CHAPTER III.

MECHANISTIC INSIGHTS INTO GENE LOOP FORMATION:

CONSTITUTIVE vs INDUCIBLE MODELS OF GENE LOOPING.

Gene looping has emerged as an evolutionary conserved phenomenon, with evidence across yeast, mammals and plants (section I.1.3). Key proteins implicated in gene loop formation include Ssu72, TFIIB, Kin28 and Pta1p (Ansari and Hampsey 2005; Singh and Hampsey 2007). Interestingly, the above factors are ubiquitously expressed, raising the question as to why gene looping is detectible only at specific loci? Moreover, gene loops are present at both actively transcribed and repressed genes. Further complicating the scenario is the methodology used to identify gene loops. As discussed in section I.2, the "C" techniques represent the average interactions of a heterogeneous population, questioning the biological significance of such interactions.

The results presented in Chapter III deal with the above concerns. To address the generality of gene loop formation and identify additional, perhaps gene-specific molecular factors, I studied the constitutively expressed mouse serum albumin gene as a model system. To study the dynamics of gene loop formation using single-cell image analysis, I studied the inducible *NFκB1* gene. While 3C analysis was used to understand gene-loop mechanism for the albumin gene, RNA-FISH combined with super-resolution microscopy was used to obtain single-cell evidence for gene-loop formation.

III.1. CONSTITUTIVE MODEL OF GENE LOOP FORMATION: The mouse serum albumin gene

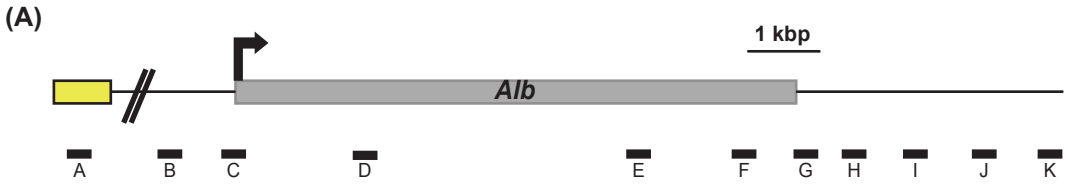
The albumin gene family comprises of albumin (*Alb*), α-fetoprotein (*Afp*), α-albumin (*Afm*) and vitamin-D binding protein (*Dbp*) genes, thought to have evolved from duplication and subsequent diversification of a single ancestral gene, millions of years ago (Tilghman 1985; Yang et al. 1985; Cooke and David 1985; Lichenstein et al. 1994; Belanger et al. 1994). All genes exhibit developmentally regulated liver-specific expression (Belanger et al. 1994). Alpha-fetoprotein (AFP) is expressed at high levels in the visceral endoderm of the yolk sac and fetal liver and low levels in the fetal gut (Tilghman 1985; Tyner et al. 1990). Albumin (ALB) is co-expressed with AFP in fetal liver but is absent from extra-embryonic tissue (Tilghman and Belayew 1982). After birth, AFP synthesis is downregulated and ALB becomes the major serum protein of the adult liver (Tilghman and Belayew 1982). *Dbp* and *Afm* transcription is also initiated during fetal stages and continues during adulthood (Song

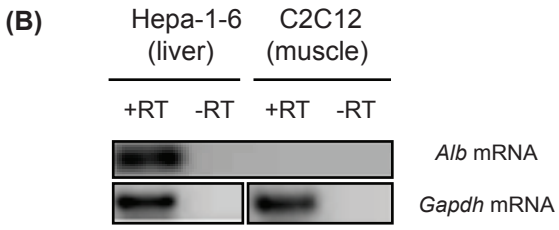
et al. 1998a; Belanger et al. 1994). Interestingly, AFP expression reinitiates during liver regeneration and hepatocellular carcinoma and therefore, AFP is commonly used as a cancer biomarker (Belanger 1994; Tilghman and Belayew 1982; Abelev 1971; Farinati et al. 2006).

The albumin gene family is located on chromosome 5, 4 and 14 in mouse, human and rat, respectively; with *Alb, Afp* and *Afm* present in a tandem orientation and *Dbp* present upstream *Alb* in the opposite direction. *Alb* and *Afp* are separated by a 14 kbp stretch of intervening sequence that harbors three cis-regulatory enhancer elements responsible for *Alb* and *Afp* expression during fetal stages (Godbout et al. 1986; Godbout et al. 1988; Jin et al. 2009). Transgenic mice studies indicate that all three enhancers are active in the fetal and adult liver even though AFP synthesis declines (Ramesh et al. 1995; Camper and Tilghman 1989). Post-natal silencing of *Afp* is attributed to two transcription factors, Zhx2 and Zbtb20 (Xie et al. 2008; Perincheri et al. 2005). Additionally, alpha-fetoprotein regulator 2 (*Afr2*) also prevents *Afp* reactivation by an unknown mechanism (Belayew and Tilghman 1982). Tumor suppressor p53 has been implicated in post-natal *Afp* repression by recruiting chromatin modifiers (Thi T Nguyen et al. 2005). On the other hand, an enhancer located ~10 kbp upstream of the *Alb* promoter maintains high levels of *Alb* transcription in adult liver (Pinkert et al. 1987). Such interplay of cis-regulatory elements and high levels expression made the albumin locus an appropriate system to investigate aspects of chromosome conformation, specifically gene-loop formation.

III.1.1. EXPRESSION OF *Alb* mRNA IN PERMISSIVE AND NON-PERMISSIVE CELL LINES

To select an appropriate cell line, I determined the expression levels of *Alb* mRNA in hepatocellular carcinoma cell line (Hepa1-6) and muscle cell line (C2C12). Total RNA was isolated and reverse transcribed using oligodT reverse primer, followed by PCR amplification with a primer pair across the open reading frame. As expected (Figure III.1B) Hepa-1-6 cells demonstrated high levels of *Alb* expression, which was absent in C2C12. DNA contamination was ruled out since reactions without reverse transcriptase (-RT) did not yield any PCR products. *Gapdh* was used as a positive control to compare between the two cell lines.





(C)

	545	555	565	575	585	595	
	CTCTGTGCTG	CAATTAATAA	AAATGGAAA	GAATCTACTC	TGTGGTTCAA	GAACTCTAT	Genomic DNA
Clone 1	CTCTGTGCTG	CAATTAATAA	AAAATGGAAA	GAACCT~~AA	AAGAAAAAAA	AAAAAAAAA~	
Clone 2	CTCTGTGCTG	CAATTAATAA	AAAATGGAAA	GAACCT~~AA	AAGAAAAAAA	AAAAAAAAA~	
Clone 3	CTCTGTGCTG	CAATTAATAA	AAAATGGAAA	----------AA	AAAAAAAAAA	AA~~~~~~~~	
Clone 4	CTCTGTGCTG	CAATTAATAA	AAAATGGAAA	----------AA	AAAAAAAAAA	AA~~~~~~~~	
Clone 5	CTCTGCGCTG	CAATTAATAA	AAAATGGAAA	GAATCC----AA	AAAAAAAAAA	AAAAA~~~~~	
Clone 6	CTCTGCGCTG	CAATTAATAA	AAAATGGAAA	GAATCC----AA	AAAAAAAAAA	AAAAA~~~~~	
Clone 7	CTCTGTGCTG	CAATTAATAA	AAAATGGAAA	GAATCT-----AA	AAAAGAAAAA	AAAAAAAAAA	
Clone 8	CTCTGTGCTG	CAATTAATAA	AAAATGGAAA	GAATCT-----AA	AAAAGAAAAA	AAAAAAAAAA	
Clone 9	CTCTGTGCTG	CAATTAATAA	AAAATGGAAA	GAATCTG---AA	AAAAAAAAAA	AAAAA~~~~~	
Clone 10	CTCTGTGCTG	CAATTAATAA	AAAATGGAAA	GAATCTG---AA	AAAAAAAAAA	AAAAA~~~~~	
Clone 11	CTCTGTGCTG	CAATTAATAA	AAAATGGAAA	GAATCCA---AA	AAAAAAAAAA	AGAA~~~~~~	
Clone 12	CTCTGTGCTG	CAATTAATAA	AAAATGGAAA	GAATCCA---AA	AAAAAAAAAA	AGAA~~~~~~	

polyA signal **Cleavage site**

Figure III.1 Mouse serum albumin *(Alb)* expression analysis.

(A) Diagrammatic representation of the mouse serum albumin gene. The black arrow indicates the transcription start site (TSS) and the grey bar represents the open reading frame. The hatched line represents ~10 kbp separation between the enhancer (yellow rectangle) and promoter regions. Primer pair A probes the enhancer region, primers C to G are positioned across the albumin transcription unit and primer H to K probe the transcription termination region. Primer B is positioned between the enhancer and promoter regions.

(B) *Alb* expression in permissive and non-permissive cell lines. qRT-PCR analysis was performed to confirm *Alb* expression in liver hepatoma cell line Hepa-1-6, which was clearly absent from the non-expressing muscle cell line C2C12. Housekeeping gene *Gapdh* was used as a positive control to compare between the two cell lines.

(C) 3'-RACE to confirm a single polyA signal. The figure represents an alignment of the genomic DNA (top line) with the cDNA sequences obtained after oligodT RT-PCR of total RNA. The yellow column denotes the transcribed polyA signal in the cDNA species, suggesting a single polyA signal. The blue column denotes the GT tracts where the nascent transcript is potentially cleaved. (A)n represents the polyA tail.

Figure III.1 continued....

III.1.2. VALIDATION OF THE *Alb* TRANSCRIPTION TERMINATION REGION

Based on nuclear run and RT-PCR analysis, it has been shown that *Alb* transcription terminates nearly 0.8-1.4 kbp downstream of the annotated polyA signal (West et al. 2006). To identify a gene loop at *Alb*, it was important to confirm the annotated polyA signal and perhaps, identify potential alternative polyA signals. A phased oligodT primer, where the $(T)_n$ repeats are followed by an adapter sequence at the 5' end, was used for cDNA synthesis from total RNA. Subsequently, PCR was performed with a forward primer ~400bp upstream the reported polyA signal and an adapter-specific primer. The PCR product was ligated to the pCR2.1-TOPO cloning vector and transformed into *E.coli* competent cells. Transformants were screened using colony PCR with M13 forward and reverse primers, followed by analysis of PCR products on an agarose gel. Twelve positive clones were sequenced and the sequences were aligned with the *Alb* DNA sequence assembly July 2007 (NCBI37/mm9), available at the UCSC genome browser (www.genome.ucsc.edu/). In accordance with previously published reports, sequence alignment confirmed that *Alb* has a single polyA signal (Figure III.1C). This observation was further strengthened by the PolyA-sequencing (PolyA-Seq) data available at UCSC, which also shows a single polyA signal in the forward direction.

III.1.3. RNA POLYMERASE II BINDING PROFILE ACROSS THE ALBUMIN LOCUS

RNAPII occupancy across *Alb* was analyzed by chromatin immunoprecipitation in liver and muscle cell lines, using anti-RNAPII 8WG16 antibody. Enrichment for RNAPII was tested by quantitative real time PCR (qPCR) with primer pairs across the gene locus. The raw data was first normalized to the input and then plotted as fold enrichment over the negative control region (Figure III.1D).

Alb expression in the adult liver and hepatocarcinoma cell lines is regulated by an enhancer located ~10 kbp upstream the promoter. As expected, RNAPII binding was certainly detected at the enhancer region (primer A). High levels of RNAPII binding were also observed at the transcription start site (primer C), indicating promoter-proximal pausing. RNAPII occupancy reduced in the open reading frame (primers D and E) but then increased in levels at the 3' end (primers F, G, H and I, respectively). In mammals, RNAPII can transcribe over 1 kbp downstream the polyA

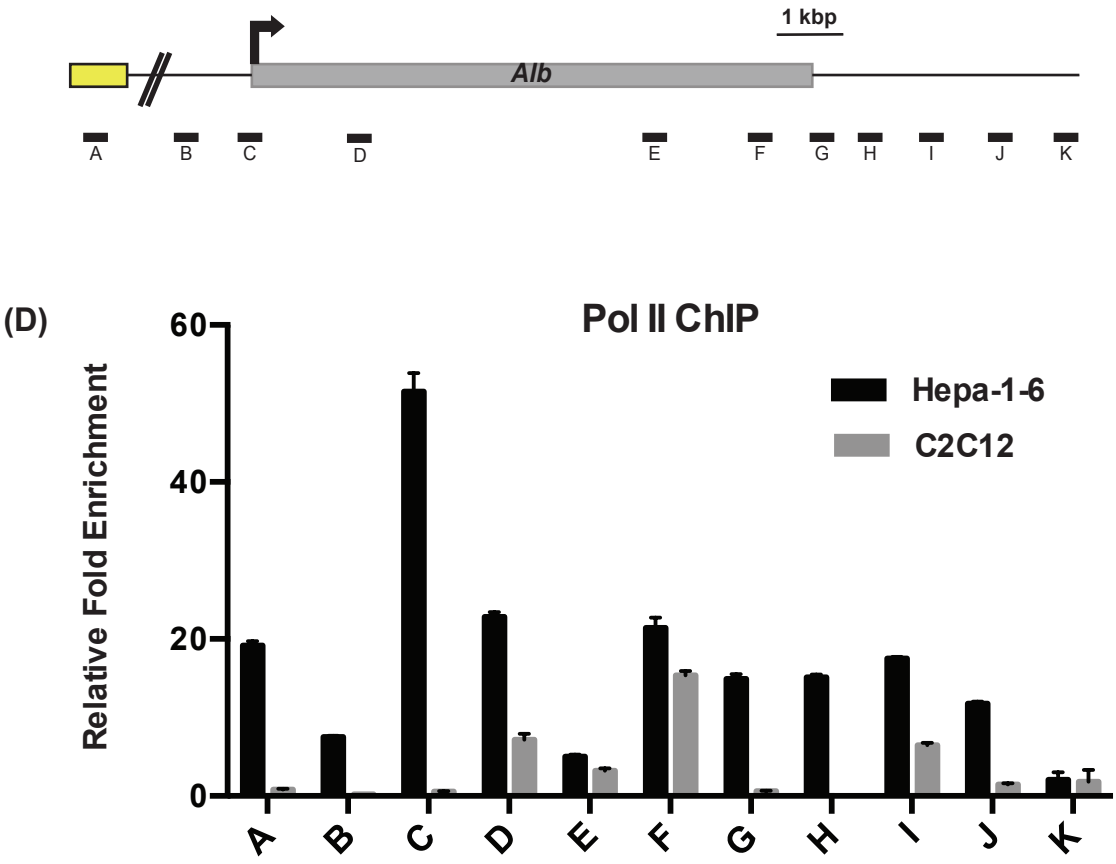


Figure III.1 Mouse serum albumin *(Alb)* expression analysis.

(D) RNAPII ChIP analysis on mouse serum albumin gene using anti-RNAPII 8WG16 antibody, in *Alb* expressing (Hepa-1-6) and non-expressing (C2C12) cell lines. High levels of RNAPII were detected at the TSS, which declined in the gene body (D-E) and peaked up again in the 3' end region (F-I). RNAPII signal dropped in region J and further in region K, suggesting transcription termination and RNAPII release. As expected RNAPII binding was also detected at the enhancer region (A). Error bars represent SEM from three independent experiments.

site prior to termination (Moore and Proudfoot 2009). Transcription through the polyA signal may reduce the rate of RNAPII elongation and 3' flanking sequences downstream the polyA signal can promote RNAPII pausing. RNAPII pausing is functionally coupled to 3' end processing of the nascent pre-mRNA. The transcribed polyA signal is recognized by CPSF, following which CstF binds to the downstream GU-rich sequence. This leads to RNAPII pausing, followed by cleavage and polyadenylation of the nascent transcript. The 5-3' exonuclease Xrn2 then degrades the exposed 5' end of the RNA, leading to termination and RNAPII release from the DNA template. RNAPII termination of *Alb* has been reported to occur 0.8-1.4 kbp downstream the polyA site and indeed primer K detected only background levels of RNAPII, suggestive of transcription termination and RNAPII release from the DNA template. RNAPII enrichment at *Alb* was absent in C2C12 cells where albumin is not expressed.

III.1.4. *cis* INTERACTIONS AT THE ALBUMIN LOCUS.

To ascertain the incidence of a gene loop, I designed and conducted 3C analysis at the *Alb* gene.

Briefly, formaldehyde crosslinking is used to preserve DNA and protein interactions. Subsequently, solubilised chromatin is fragmented by an appropriate restriction enzyme, diluted (at least 10 fold) to prevent random inter-molecular interactions and ligated to covalently join DNA fragments that are cross-linked to the same complex (intramolecular ligation). This "3C template" represents a pool of ligation products between genomic loci that exist in close physical proximity. Interactions between regions of interests can be verified using qPCR with specific primer pairs. The abundance of each ligation product is a measurement of the interaction frequency between two loci. 3C based *cis* and *trans* interactions can also be identified using normal PCR, followed by separation on an agarose gel and quantification of gel bands with software programs such as ImageJ or GelQuant.

The choice of an appropriate restriction enzyme is of utmost importance for a successful 3C experiment. An unambiguous recognition sequence, insensitivity to CpG methylation, robust digestion activity and the presence of well-spaced restriction sites made Hind III a suitable choice. Furthermore, Hind III is inactivated within 20 min

at 65°C, unlike enzymes (such as EcoRV, DpnI) that are inactivated at much higher temperatures and pose a risk of affecting formaldehyde crosslinks.

A single 3C experiment can generate vast amounts of data, but controls are essential for correct interpretation of the same. First, it is essential to normalize for differences in PCR efficiency of different primer pairs, such that a direct comparison of the interaction frequency between two loci can be made. I amplified ~500 bp regions encompassing all HindIII sites at the *Alb and Gapdh* (as a control, discussed below) loci and pooled them in equimolar concentrations. The pooled products were digested for 2 h and ligated overnight to obtain a control template (referred to as "control"). A series of dilutions of the control were prepared and tested for PCR amplification in the exponential phase. This control template represents a pool of all possible ligation products between the Hind III digested fragments. Second, to compare the interaction frequencies between two conditions (different cell lines or chemical treatment), it is important to have an internal control such as a housekeeping gene, where a change in interactions is unanticipated. I chose two adjacent Hind III fragments at the *Gapdh* gene as an internal control for all my 3C experiments.

With the above set of controls, I performed 3C analysis at the *Alb* gene, with restriction sites across the gene body and upstream enhancer region. Nuclei were isolated from crosslinked cells, digested overnight, ligated for 4 h, reverse crosslinked and column purified, to obtain the 3C template. 2 µl of the 3C template was used for each PCR, along with a primer efficiency control. All primers were designed in the forward direction, such that a 3C product was possible only after ligation of regions in physical proximity. 3C products were run on a 2.5% agarose gel and band intensity was quantified using GelQuant. It is important to consider that mere gel bands do not represent significant interactions until they are normalized for the PCR efficiency of primers. All values were normalized for the total input DNA (loading control) and the 3C products were verified by sequencing. The relative crosslinking frequency for two loci was calculated as follows:

$$\textbf{Relative Interaction Frequency (X)} = [\text{Primer A/ Gapdh}]_{\text{3C template}} \div [\text{Primer A/ Gapdh}]_{\text{3Ccontrol}}$$

Using the promoter as the anchor (primer 4), I observed an interaction between the promoter and enhancer regions (4+2) (Figure III.2B) but none between the intervening or upstream regions (4+3 and 4+1). Looping between the *Alb* enhancer and promoter was previously suggested, when promoter-specific transcription factors HNF1α, NF-Y, FTF and Nkx2.8 were detected at the enhancer region and enhancer specific transcription factor Foxa3 was detected at the promoter (Kajiyama et al. 2006). However, this study presents the first experimental evidence for such an interaction. A 3C product was also obtained between the promoter and polyA region (primer 4+8) (Figure III. 3B), suggestive of a novel gene loop between the promoter and terminator regions. The specificity of this interaction was further established in a reciprocal PCR with the polyA site fragment as an anchor. No interaction was detected between the polyA and enhancer regions (8+2) or the open reading frame (8+5, 8+6 and 8+7). A 3C product between primers 4 and 3, and primers 8 and 9 represents ligation of adjacent fragments and serves as technical control along with *Gapdh* 1R-2R. Primer 9 is located ~1.4 kbp downstream the polyA signal and the reduced interaction frequency between 4+9 coincides with transcription termination and RNAPII release. Interestingly, the *Alb* promoter also associates with sequences present in introns 6 (4+5), 9 (4+6) and 14 (4+7), a scenario also observed for *BRCA1*. However, unlike *BRCA1* where the gene-loop is associated with transcriptional repression, the *Alb* gene-loop correlates with high levels of expression (Tan-Wong et al. 2008).

Further support for a gene-loop at *Alb* comes from RNAPII ChIP analysis (Figure III. 1D). 8WG16 recognizes hypo-phosphorylated RNAPII present at gene promoters prior to transcription initiation, however, RNAPII recognition by 8WG16 is affected upon CTD phosphorylation, such that the antibody is unable to recognize phospho-Ser 2 or phoshpo-Ser 7 epitopes on RNAPII CTD but is reactive to phospho-Ser 5 RNAPII (Doyle et al. 2002; Patturajan et al. 1998; Stock et al. 2007; Xie *et al*. 2006). Promoter escape and abortive initiation at the 5'-end of a gene coincide with phosphorylation of serine 5 residues of the RNAPII CTD. In contrast, productive elongation is associated with phosphorylation of serine 7 and serine 2 residues, peaking at the 3'-end of most genes (O'Sullivan et al. 2012). Accordingly, 8WG16 signal at the 3' region of *Alb* indicate RNAPII phosphorylated at Ser 5, which itself is a well-documented marker of gene-loop formation (O'Sullivan et al. 2004).

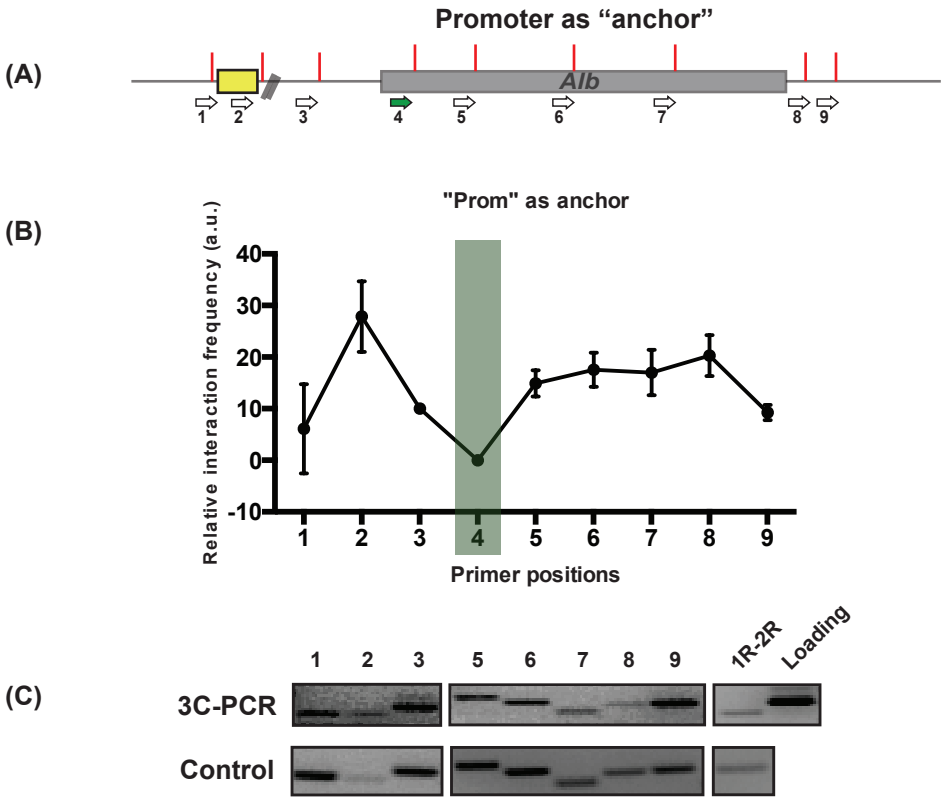


Figure III.2 HindIII-3C Analysis using promoter (”prom”) as an anchor

(A) Schematic representation the *Alb* gene locus. The grey bar depicts the *Alb* open reading frame, yellow box is the enhancer, red lines denote HindIII restriction sites and forward arrows (with numbers below) represent the 3C primers. Primer 4 (green arrow) is used as the anchor and its interactions are probed with different regions across the locus.

(B) Hind III 3C analysis with the promoter “Prom” as anchor. The promoter and enhancer regions are present in close proximity with each other (primer 4+2). Promoter-terminator juxtaposition is exemplified by primers 4+8. The promoter also exists in proximity with potential regulatory intronic regions, 4+5, 4+6 and 4+7.

(C) 3C PCR products. The 3C products were run on a 2.5% agarose gel and quantified using Gel Quant software. 1R-2R represents two adjacent fragments at the *Gapdh* locus used as a technical control. “Loading” represents a PCR within the *Alb* open reading frame, as a loading control. Control represents PCRs performed on a control template using the same primer combinations and is used to control for primer efficiency.

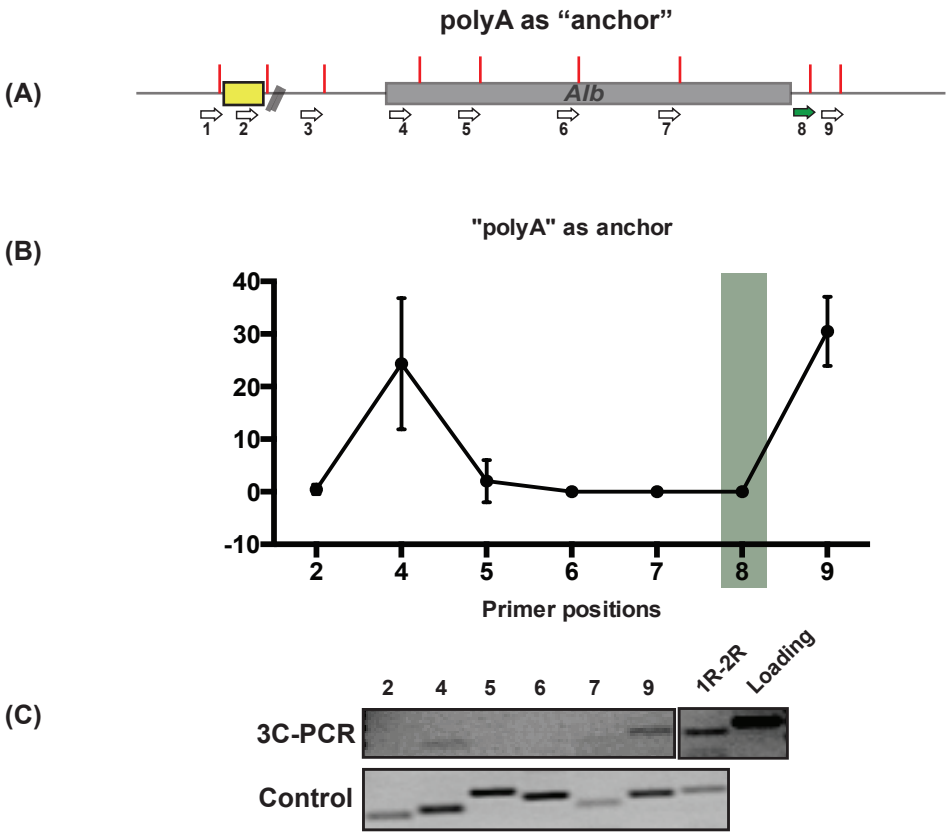


Figure III.3 HindIII-3C Analysis using polyA (”polyA”) as an anchor.

(A) Schematic representation the *Alb* gene locus. The grey bar depicts the *Alb* open reading frame, yellow box is the enhancer, red lines denote HindIII restriction sites and forward arrows (with numbers below) represent the 3C primers. Primer 8 (green arrow) is used as the anchor and its interactions are probed with different regions across the locus.

(B) Hind III 3C analysis with the “polyA” as anchor. The promoter and polyA regions are present in close proximity with each other (primer 4+8), suggesting gene loop formation at *Alb*. Unlike the promoter region, the polyA region does not contact the enhancer (8+2) or any intronic sequences upstream (8+5, 8+6, 8+7). A product between primers 8+9 represents ligation of adjacent fragments and serves as a technical control.

(C) 3C PCR products. The 3C products were run on a 2.5% agarose gel and quantified using Gel Quant software. 1R-2R represents two adjacent fragments at the *Gapdh* locus, used as a technical control. “Loading” represents a PCR within the *Alb* open reading frame, as a loading control. Control represents PCRs performed on a control template using the same primer combinations and is used to control for primer efficiency.

III.1.5. *cis* INTERACTIONS AT THE *Alb* LOCUS ARE DEPENDENT ON BASAL TRANSCRIPTION

To extend the analysis further, I performed 3C analysis after transiently inhibiting transcription using chemical inhibitors. First, I tested a range of potent transcription inhibitors to identify an appropriate combination of dosage and time. Hepa-1-6 cells were treated with Flavopiridol (200 μM) or KM05283C (100 μM) for 1 h, or α-Amanitin (2.5 μg/ml) for 12 h. The efficiencies were compared using qRT-PCR (Figure III.4B) and western blot (Figure III.4C).

Both Flavopiridol and KM inhibit CDK9-mediated Ser 2 phosphorylation of RNAPII CTD, impeding productive elongation (O'Sullivan et al. 2012). α-amanitin binds and degrades the CTD in a dose-dependent manner. Loss of phospho-Ser 2 RNAPII was measured by western blot using anti-Ser 2P H5 antibody (Covance); while total RNAPII was measured using N20X (Santa Cruz), which detects both phosphorylated (RNAPII H_O) and non-phosphorylated (RNAPII H_A) isoforms of RNAPII. Treatment with KM resulted in a huge drop in Ser 2P levels (RNAPII H_O), while RNAPII H_A was unaffected (Figure III.4C). α-amanitin treatment had no affect on Ser2P (H5 blot) but the RNAPII H_O band was absent in the N20X blot. This possibly implies that α-amanitin downregulates Ser 5P without affecting Ser2P. However, if α-amanitin degrades the CTD, both H5 and N20X signals should have reduced considerably. Furthermore, qRT-PCR analysis of nascent pre-mRNA from α-amanitin treated samples showed a slight decrease for intron 8 probe but not for intron 2. Being a bulky compound, α-amanitin has a slow rate of diffusion across the cell membrane. Therefore, it is possible that a 12 h treatment at a concentration of 2.5 μg/ml was insufficient for significant inhibition of transcription. Flavopiridol did not show any effect at the protein or RNA levels. The effectivity of KM was further substantiated by ~70% reduction in nascent transcripts (Figure III.4B) and therefore, KM was selected as the inhibitor of choice for subsequent experiments.

Chromatin from control (100 μM DMSO) and KM treated cells was used to analyze RNAPII occupancy (Figure III.4D) and perform 3C analysis (Figure III. 4E) across *Alb*. To avoid ambiguous results because of chemical treatment and antibody efficiency, RNAPII levels were detected with N20X antibody (and not 8WG16), which reacts with an epitope different from the CTD. Any change in RNAPII density would likely reflect loss of Ser 2P due to KM treatment. As expected, RNAPII levels dropped

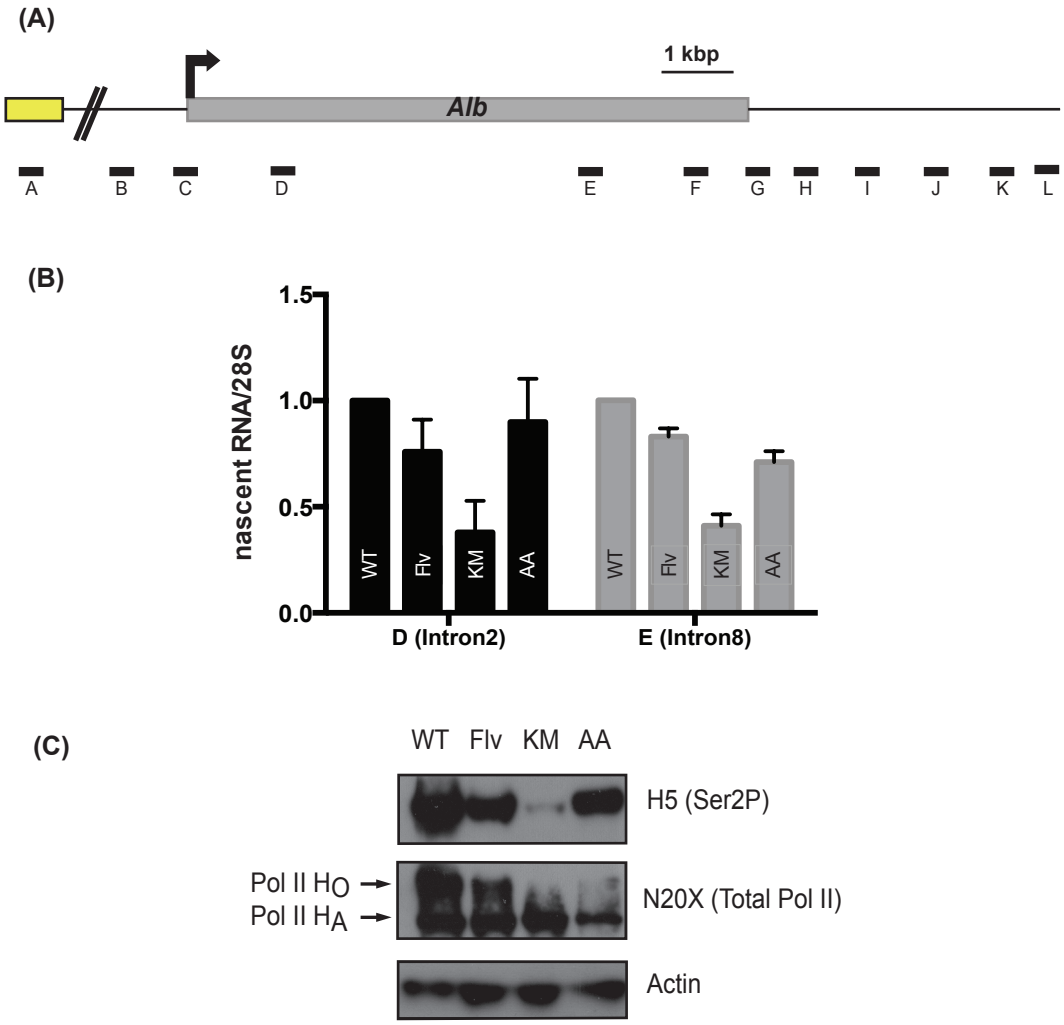


Figure III.4 *Alb* gene loop formation is dependent on transcription.

(A) Schematic representation of the mouse serum albumin gene. The black arrow indicates the transcription start site (TSS) and the grey bar represents the open reading frame. The hatched line represents ~10 kbp separation between the enhancer (yellow rectangle) and promoter region. Primer pair A probes the enhancer region, primers C to G are positioned across the albumin transcription unit and primer H to K probe the transcription termination region. Primer B is positioned between the enhancer and promoter region. Primers D and E represent the intron-exon junction primers used for qRT-PCR analysis in (B).

(B) qRT-PCR analysis to compare the efficiencies of transcription inhibitors Flavopiridol (Flv), KM05832c (KM) and alpha-amanitin (AA). Hepa-1-6 cells were treated with KM (200 μM), Flv (200 μM) for 1 h and AA (2.5 μg/ml) for 12 h prior to total RNA isolation. Primers positioned across intron 2-exon 3 junction (D) and intron 8-exon 9 (E) were used to detect nascent transcript levels by reverse transcription and subsequent qPCR. The raw values were normalized to 28S and then plotted relative to the control (WT) set as 1. Nascent transcript levels dropped maximally after KM treatment, while Flv and AA treatment did not produce any significant change.

(C) Western blot to compare the efficiencies of Flv, KM and AA. Whole cell extract was prepared from Hepa-1-6 cells after inhibitor treatment as in (B). Total RNAPII was probed using N20X (Santa Cruz) antibody, while phosphoSer2-RNAPII was probed with anti-Ser2 antibody H5 (Covance). Actin (Sigma) was used as a loading control. PhosphoSer2-RNAPII level was unchanged after Flv treatment but dropped dramatically after KM. In contrast, AA results were not comparable between H5 and N20X antibodies. While H5 signal was intact, the hyper-phosphorylated (Pol II Ho) signal was missing in the N20X blot. Therefore, Flv and AA were discarded and KM was selected as the inhibitor of choice.

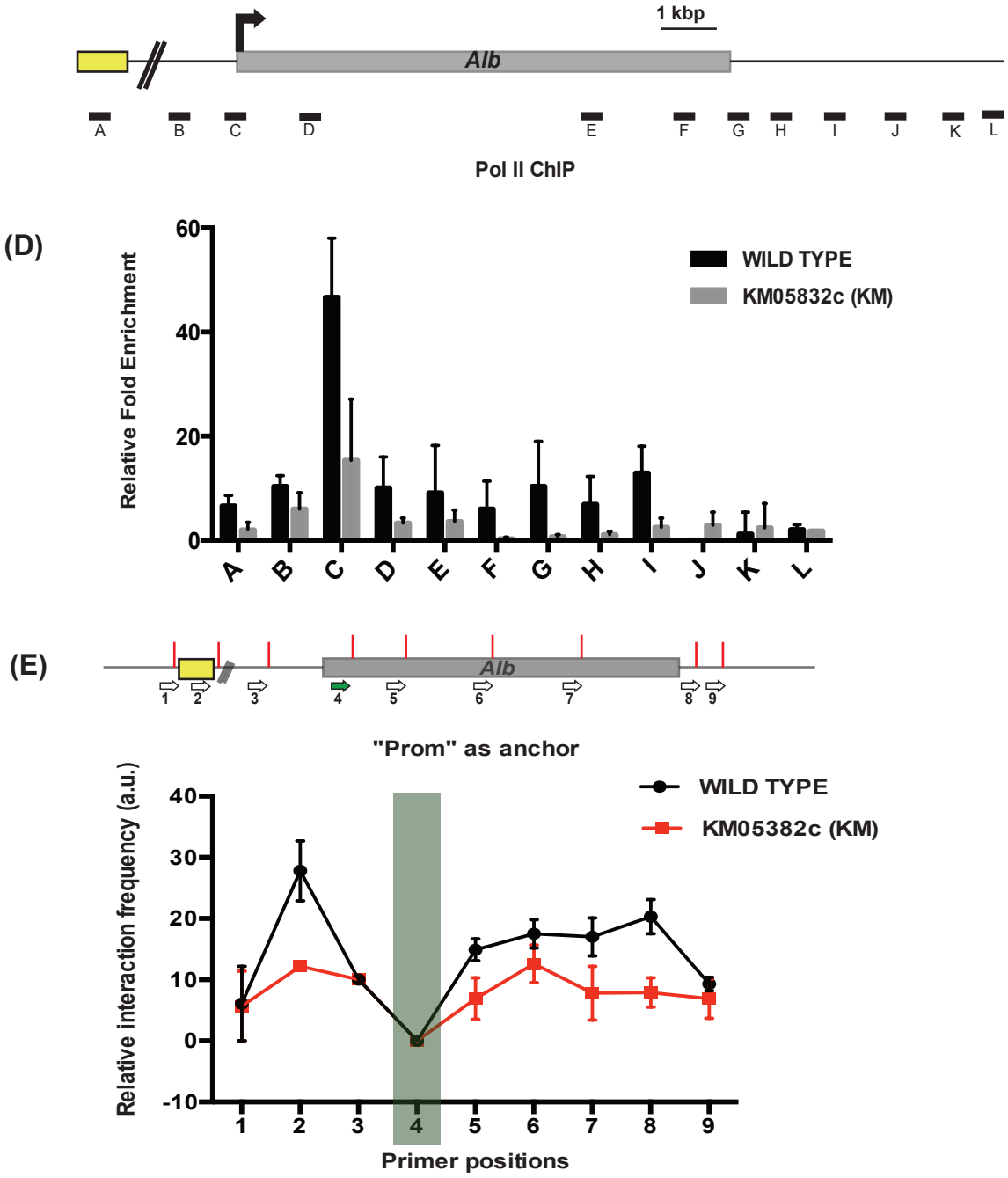


Figure III.4 *Alb* gene loop formation is dependent on transcription.

(D) RNAPII ChIP analysis across the *Alb* gene locus. RNAPII occupancy across the *Alb* gene locus was examined in untreated (Wild type) and cells treated with transcription inhibitor KM05832c (KM). KM inhibits PTEFb mediates Ser2 phosphorylation of RNAPII CTD. As expected, RNAPII occupancy was significantly reduced after KM treatment. Error bars represent SEM from three independent experiments.

(E) 3C analysis between untreated and KM treated cells. The grey bar depicts the *Alb* open reading frame, yellow box is the enhancer, red lines denote HindIII restriction sites and forward arrows (with numbers below) represent the 3C primers. Primer 4 (green arrow) is used as the anchor and its interactions are probed with different regions across the locus. Inhibition of transcription reduces the interaction frequency between the promoter and upstream enhancer and downstream regions. This suggests that looping at the *Alb* locus is dependent on active transcription.

at the 3' end of *Alb* (primers F, G, H and I), indicating lack of productive elongation. RNAPII binding was also reduced at the transcription start site (primer C) and enhancer (primer A). The RNAPII signal at the enhancer region (primer A) was relatively low in wild type cells, in comparison to RNAPII occupancy (Figure III.1D), probably due to differences in the efficiency of N20X and 8WG16 antibodies.

Next, I tested for 3C interaction between the promoter and different regions across the *Alb* locus, in control and KM treated samples. The promoter-terminator (4+8) juxtaposition was reduced upon transcription inhibition (Figure III. 4E), implying that *Alb* gene-loop is transcription dependent. In other words, gene-loop formation is dependent on active transcription, as loss of Ser 2 phosphorylation, a mark of productive elongation, resulted in reduced frequency of gene loop formation. Interactions between the promoter and intron 6 (4+5), intron 9 (4+6), and intron 14 (4+7) were also diminished.

The above results indicate that gene-loop formation is a feature of actively transcribed genes and is dependent on transcription elongation. Therefore, gene-loop may well be a consequence of the pioneer round of transcription and facilitate the recycling of RNAPII from terminator to promoter regions, to maintain high levels of transcription. To obtain single-cell evidence for gene-loop formation, I switched from *Alb* (~16 kbp in length) to a >100 kbp long human gene, *NFκB1*.

III.2. INDUCIBLE MODEL OF GENE LOOP FORMATION: The Nuclear factor of kappa B light chain 1, *NFκB1* gene

3C experiments have provided critical insights about gene conformation associated with active (O'Sullivan et al. 2004; O'Reilly et al. 2007; Perkins et al. 2008) and repressive (Tan-Wong et al. 2008) states of gene expression. However, 3C analysis represents the average transcription and chromatin interactions of a heterogeneous cell population, questioning the frequency and significance of such three-dimensional interactions. A typical 3C experiment is unable to detect cell-to-cell differences in genome topology and determine whether an interaction between two loci is simultaneous, sequential or mutually exclusive. Therefore, the ultimate proof for physiologically significant 3D interactions lies in single cell experiments performed with a combination of microscopic and "C" techniques. The "C" techniques screen large cell numbers and therefore, are useful in identifying potential gene-loop

candidate genes. Further evidence can be obtained by microscopic visualization of gene-loops using fluorescent *in situ* hybridization. I collaborated with the group of Professor Robert H Singer (Albert Einstein School of Medicine, New York, USA), where I employed 3C to identify a candidate for gene looping and Timothee Lionnet (Postdoctoral fellow, Singer Lab) validated my 3C data using RNA-FISH.

It has been shown that stimulation with tumor necrosis factor α (TNFα), a potent cytokine, orchestrates an inflammatory response associated with rapid and synchronous activation of a subset of genes in human umbilical endothelial cells (HUVECs) (Wada et al. 2009). Following abortive initiation, RNAPII enters productive elongation, with nascent transcripts emerging as RNAPII traverses across the gene length. I chose *NFκB1* (Nuclear factor of kappa B light chain 1) for 3C and FISH analysis as it is >100 kbp long and so should provide ample spatial and temporal resolution for 3C and FISH analysis between different regions across the gene length. NFκB1 is a part of a highly conserved signaling pathway, with key functions in the regulation of the immune and inflammatory responses in higher eukaryotes. To study the *NFκB1* gene in my experiments it was first necessary to evaluate its transcription kinetics.

III.2.1. A WAVE OF NASCENT TRANSCRIPTION ACROSS *NFκB1*

The *NFκB1* induction profile after stimulation of HUVECs with TNFα has been described (Wada et al. 2009). A wave of intronic signal appears to “sweep down” the gene from ‘‘Start’’ (at 15 min) to ‘‘End’’ (at 60 min). Exonic signal increases significantly above the basal level only after 60 min, when the polymerase reaches the end of *NFκB1*. To confirm the above dataset, I performed qRT-PCR analysis on total RNA extracted from HUVECs. Cells were synchronized in G_0 state of the cell cycle by serum starvation and stimulated with 10 ng/ml TNFα. Total RNA was extracted for an hour at 15 min intervals, subsequently at 2 h and 4 h, and reverse transcribed with primers positioned across intron-exon junctions to identify primary, unspliced transcripts. RNA levels were measured using qPCR and raw values were first normalized to 28S rRNA levels and then plotted relative to uninduced time point zero.

As shown in Figure III.VB, the pioneering round of transcription for *NFκB1* completed in an hour, as transcripts emerging from the last intron (primer D) were

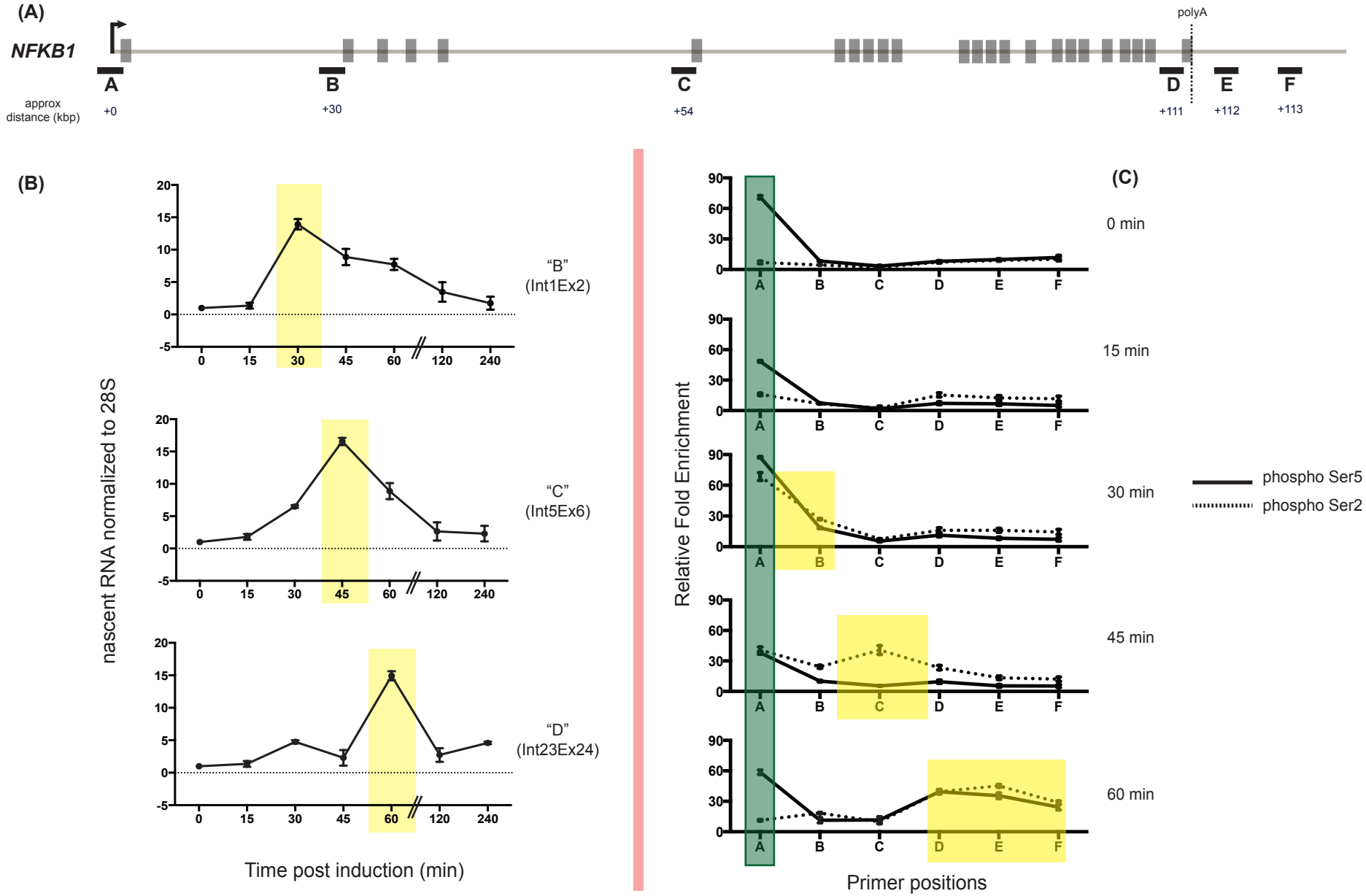

(A)
NFKB1
polyA
A
B
C
D
E
F
approx distance (kbp)
+0
+30
+54
+111
+112
+113
(B)
nascent RNA normalized to 28S
"B" (Int1Ex2)
"C" (Int5Ex6)
"D" (Int23Ex24)
Time post induction (min)
(C)
Relative Fold Enrichment
0 min
15 min
30 min
45 min
60 min
phospho Ser5
phospho Ser2
Primer positions

Figure III.5 *NFκB1* induction analysis in HUVECs.

(A) Schematic representation of the *NFκB1* locus. Thick black arrow represents the transcription start site. Grey boxes represent exons and grey line denotes the intervening introns. Black rectangles depict the primers used for RT-PCR and ChIP analysis.

(B) RT-PCR analysis of *NFκB1* transcription. *NFκB1* transcription was analyzed in serum starved, synchronous HUVECs after stimulation with 10 ng/ml of TNFα. Primers positioned across intron-exon junctions were used for reverse transcription and subsequent qPCR to detect nascent RNA. The raw values were first normalized to 28S rRNA and then plotted relative to t = 0 min. Yellow columns depict peaks of nascent transcripts detected at different positions across the gene length. No transcripts were detected between t = 0-15 min. A signal from intron 1-exon 2 junction (primer B) suggested *NFκB1* enters into productive elongation between 15-30 min after induction. The pioneering polymerase reaches intron 5-exon 6 junction at t = 45 min and completes the first round at t = 60 min, indicated by intron 23-exon 24 (primer D). No transcripts were detected 1 h post induction, suggesting activation of the *IκB* mediated feedback regulation. As transcripts emerge from intron 5, the signal at intron 1 decreases. Similarly, as transcripts emerge from intron 23, the signal at intron 5 decreases. This suggests cotranscriptional splicing of the *NFκB1* nascent RNA.

(C) RNAPII occupancy analysis across *NFκB1*. RNAPII binding was assessed using antibodies against two forms of RNAPII. Phospho-Ser5 RNAPII, a mark of poised and initiating polymerases is prevalent at the 5' end. Whereas, phospho-Ser2 RNAPII coincides with productive elongation and termination and is prevalent at the 3' end. The green column represents poised polymerase present at the promoter region at all time points. The yellow column depicts RNAPII traveling across the gene length in an induction dependent manner. RNAPII levels across the gene body were absolutely negligible between 0-15 min. A high peak for phospho-Ser2 was detected at the promoter and intron 1-exon 2 junction at t = 30 min, indicating productive elongation into the gene body. At t = 45 min, the phospho-Ser2 signal moved further into the gene body, coinciding with the emergence of nascent transcripts from intron 5-exon 6 junction in Figure III.5B. By t = 60 min, RNAPII reaches the 3' end of *NFκB1,* suggesting completion of the pioneering round of transcription.

detectable at 60 min post induction. If RNAPII is thought to transcribe at a rate of 3 kbp/min (Wada et al. 2009), one round of *NFκB1* transcription would take ~40 min, considering the gene spans 115.974 kbp of DNA sequence. However, the first round of *NFκB1* transcription finishes in ~60 min, suggesting a lag in transcription that can be attributed to abortive initiation. Abortive initiation refers to the repetitive synthesis and release of short nascent transcripts by RNAPII before it enters into productive elongation. The first evidence for abortive initiation came from *in vitro* transcription experiments using the lactose operon (lac) UV5 promoter, where RNAPII was shown to synthesize 2-8 nucleotide RNAs without disengaging from the template (Carpousis and Gralla, 1980). Since then abortive initiation is accepted as an inherent characteristic of RNAPII transcription (Guenther et al. 2007; Core and Lis 2008; Core et al. 2008; Seila et al. 2008). A primer positioned within exon1 or at exon1-intron1 junction should be able to detect abortive initiation for *NFκB1*. However, the sequence of this region is extremely GC rich and hence it was difficult to design an efficient primer pair.

Promoter escape occurs between 15-30 min and productive elongation is accompanied by the emergence of transcripts from primer B (Figure III.VB, yellow shaded region). At t = 45 min, nascent RNA can be detected at primer C, suggesting a transcription rate of ~3 kbp/min because primers B and C are separated by ~42 kbp. RNAPII further advances along the gene and reaches primer D at t = 60 min, maintaining the kinetics of ~3 kbp/min. As RNAPII transcribes intron 5 (primer B, t = 45 min), the signal from intron 1 decreases. Likewise, the signal at intron 23 coincides with a decline in signals from intron 1 and intron 5. This observation is consistent with co-transcriptional splicing and degradation of RNA from proximal introns, while the polymerase is still transcribing the distal introns (Richard and Manley 2009).

The NFκB transcription factors are key regulators of cell survival, differentiation and proliferation. There are five NFκB family members in mammals: RelA/p65, RelB, c-Rel, p50 (NFkB1), and p52 (NFkB2); with only p65, RelB and c-Rel possessing transactivation domains (Oeckinghaus and Ghosh 2009). In an uninduced state, NFκB dimers are sequestered in the cytoplasm by the inhibitor kappa protein IκB. Stimulation with lipopolysaccharide, TNFα or interferons, results in proteasomal degradation of IκB and translocation of free NFκB complexes to the nucleus, where they influence target gene expression by binding to κB sites (Ghosh and Baltimore 1993). Interestingly, nuclear NFκB also drives *IκB* expression, thereby generating a

negative feedback loop. The *IκB* gene is <10 kbp in length, so that the cellular pool of IκB is rapidly replenished, whilst *NFκB1* continues with the pioneer round of transcription. The sharp decline observed in primary *NFκB1* transcription at t = 60 mins (Figure III.5B), suggested activation of the IκB feedback pathway.

III.2.2. A WAVE OF TRANSCRIBING POLYMERASES ALONG *NFKB1*

RNAPII occupancy across *NFκB1* was measured at relevant time points by ChIP-qPCR using antibodies against the phosphorylated forms of RNAPII CTD. As mentioned above (section III.1.4.), promoter escape and abortive initiation at the 5' genic regions coincide with Ser 5 phosphorylation of the CTD, while productive elongation is associated with Ser 7 and Ser 2 phosphorylation at the 3' ends (Egloff et al. 2012). In Figure III.5C, the green bar depicts constant RNAPII occupancy at the promoter, while the yellow bars correspond to active phopsho-RNAPII as it proceeds along the gene body. The promoter (primer A) showed enrichment of phospho-Ser 5 RNAPII at all times, coinciding with abortive transcription and promoter-proximal paused polymerase (Core and Lis 2008; Core et al. 2008; Li and Gilmour 2012; Adelman and Lis 2012). Only after t = 30 min, a peak for phospho-Ser 2 RNAPII was detected at intron 1-exon 2 junction (primer B), in concert with the emergence of nascent RNA (Figure III.5B). Similarly, phoshpo-Ser2 RNAPII reached intron 5 (primer C) and intron 23 (primer D) at t = 45 min and t = 60 min, respectively. Primers E and F also detected phospho-Ser2 RNAPII, suggesting that RNAPII was yet to terminate. A primer pair positioned further downstream would presumably have shown a significant drop in RNAPII signal. In addition to phospho-Ser 2, phospho-Ser 5 RNAPII was also detected at the 3' end of *NFκB1* (primers D, E and F), which may reflect a gene loop conformation (similar to *Alb*, section III.1.4). However, the gene loop here would be concomitant with repression as *NFκB1* transcription is downregulated 1 h post induction.

III.2.3. GENE LOOP FORMATION AT *NFκB1*

I next performed 3C analysis to investigate promoter-terminator gene-loop at *NFκB1*. HUVECs were synchronized and induced with TNFα and chromatin was prepared for the given time points. As a technical control (discussed in detail in section III.1.4.), two primers positioned on different restriction fragments at the *GAPDH* gene, 1F-2F (similar to *Gapdh* 1R-2R in *Alb* 3C analysis, section III.1.4) were

used for each time point. Additionally, the amplification efficiency for each primer pair was corrected using a control template, generated as described in section III.1.4.

No 3C signals were observed between the promoter (anchor) and any sites across *NFκB1* between t = 0-15 min (Figure III.6B). At t = 30 min, juxtaposition between the promoter and polyA region (1+6) was observed. At t = 45 min, the promoter-polyA (1+6) interaction intensity reduced and a strong interaction between the promoter and 3'-flanking region (1+8) was obtained. Additionally, promoter contacts along the gene, primer combinations 1+4, 1+5, 1+6 and 1+7 were also detected at t = 45 min. Furthermore, a loop between primers 1+2 was also detected at t = 45 min. No contacts were detected at 60, 80 and 100 min. These were rather puzzling observations, as they negate the qRT-PCR and ChIP-qPCR profiles, which show that at t = 30 min only half the gene is transcribed and bound by RNAPII. Even an assumption that *NFκB1* engages more than one polymerase, such that the pioneering polymerase finishes the first round within 30 min (1+6) and the subsequent contacts 1+4, 1+5, 1+6, 1+7 and 1+2, are due to the second and third polymerases, respectively, does not coincide with the wave of nascent transcription and polymerase occupancy observed previously. Evidence supporting for the qRT-PCR and ChIP-qPCR datasets (Figure III.5B and III.5C) comes from a similar study performed at murine *Nfκb1* in LPS-stimulated macrophages (Bhatt et al. 2012). The authors also document a minimum duration of 60 min to complete the first round of transcription.

The 3C dataset is opposed by the transcription kinetic profile of *NFκB1*. However, the possibility of dynamic reconfiguration at the *NFκB1* gene locus, in response to TNFα and IκB cannot be ignored. The above argument is supported by the fact that 3C analysis, specifically demonstrated promoter-terminator and promoter-gene body interactions at t = 30 min and t = 45 min, respectively (Figure III.6B). Auto-regulation of *NFκB1* is a well-documented pathway and perhaps, *NFκB1* has adopted dynamic re-arrangement as a strategy to mount a rapid and possibly, a synchronous transcription response upon cytokine stimulation. Recently, it has been shown that long genes (>100 kbp) demonstrate dynamic reconfiguration in response to environmental stimuli (Larkin et al. 2012).

Unfortunately, reproducibility of the above dataset in HUVECs was challenged by inherent limitations of the 3C technique. *NFκB1* is a ubiquitously expressed transcription factor present in all differentiated cells and is particularly active in tumor

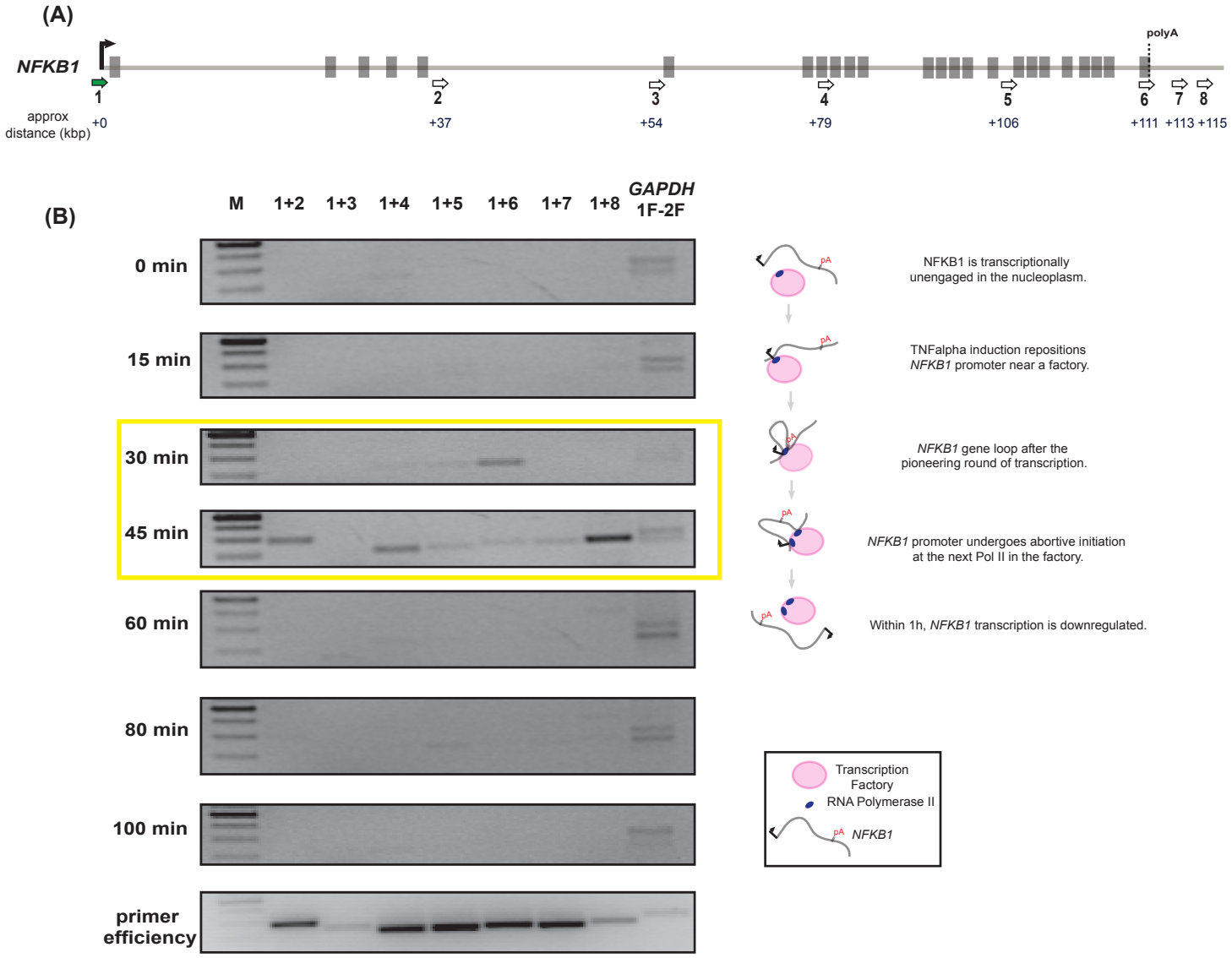


Figure III.6 HindIII 3C analysis across the *NFκB1* locus.

(A) Schematic representation of the *NFκB1* locus. Thick black arrow represents the transcription start site. Grey boxes represent exons and grey line denotes the intervening introns. The 3C primer positions in accordance with their respective Hind III sites are shown as forward, numbered arrows. The green arrow (1) represents the promoter fragment used as an anchor for 3C analysis.

(B) Changing contacts across *NFκB1* using 3C PCRs. Interactions between the *NFκB1* promoter and different regions across the gene length were analyzed before and after cytokine stimulation. No contacts were observed between the promoter and any region across *NFκB1* between t = 0-15 min. At t = 30 min, juxtaposition between the promoter and polyA region (1+6) was observed. At t = 45 min, the promoter-polyA (1+6) interaction intensity reduced and a strong interaction between the promoter and 3'-flanking region (1+8) was obtained. Promoter contacts along the gene, primer combinations 1+4, 1+5, 1+6 and 1+7 were also detected at t = 45min. Furthermore, a loop between primers 1+2 was detected at t = 45 min. No contacts were detected at 60, 80 and 100 min. Primer efficiency was normalized by performing simultaneous PCRs on a control template. *GAPDH* 1F-2F serves as a ligation control. The adjacent diagram represents a possible scenario to interpret the 3C contacts.

cells, where it upregulates anti-apoptotic transcripts (Ben-Neriah and Karin 2011). Therefore, I tried 3C analysis in synchronized HeLa (cervical) and HepG2 (hepatic) cancer cell lines but no interactions were observed. To further investigate the possibility of a stimulus-dependent gene loop at *NFκB1*, RNA FISH analysis was performed by my collaborator, Timothee Lionnet and is discussed in the following sections.

III.2.4. RNA-FISH ANALYSIS OF *NFκB1*

The experiments described in this section were designed, conducted and analyzed by Dr. Timothee Lionnet, a post-doctoral research fellow at Prof. Robert H. Singer's Group, Albert Einstein College of Medicine, Bronx, NY, USA.

My 3C experiments on the *NFκB1* gene gave inconsistent results. Therefore, to clarify the role of a gene loop in this gene, an independent method was employed. RNA FISH was performed in HUVECs with probes spanning ~5 kbp in intron 1, 5 and 23, labeled with Cy3, Cy3.5 and Cy5, respectively (Figure III.7A). 1 h was selected as the appropriate time point (Figure III.7B) after a preliminary time-course experiment in which no signal was detected at 0, 2, 3 and 4 h of induction. Most cells showed two spots suggesting activation of both alleles.

Our hypothesis was based on two extreme conformations (Figure III.8A) that can be adopted during the transcription cycle, "extended" or "looped". If *NFκB1* maintains an extended conformation, then the distance between probes targeting intron 1 and 23 (5'-3') would be greater than distance between intron 1 and 5 (5'-center) or intron 5 and 23 (center-3'). Contrarily, the 5'-3' distance would be smaller if *NFκB1* embraced a looped conformation. The mean interlocus distance between 5' and 3' ends of *NFκB1* was ~358 nm (Figure III.8C), while the distances between 5'-center and center-3' were 267 nm and 302 nm, respectively. This suggested the absence of an induction dependent gene loop for *NFκB1*.

Subpixel localization and intensity quantification of the spots was performed using MatLab (Mathworks) programs. To identify a spot (Figure III.8B), a manually set threshold was applied and the local maxima within a 2-pixel radius were defined as a "FISH spot". A custom-written Matlab program was used for colocalization analysis. The center positions of fluorescence spots in each channel were identified in 3-

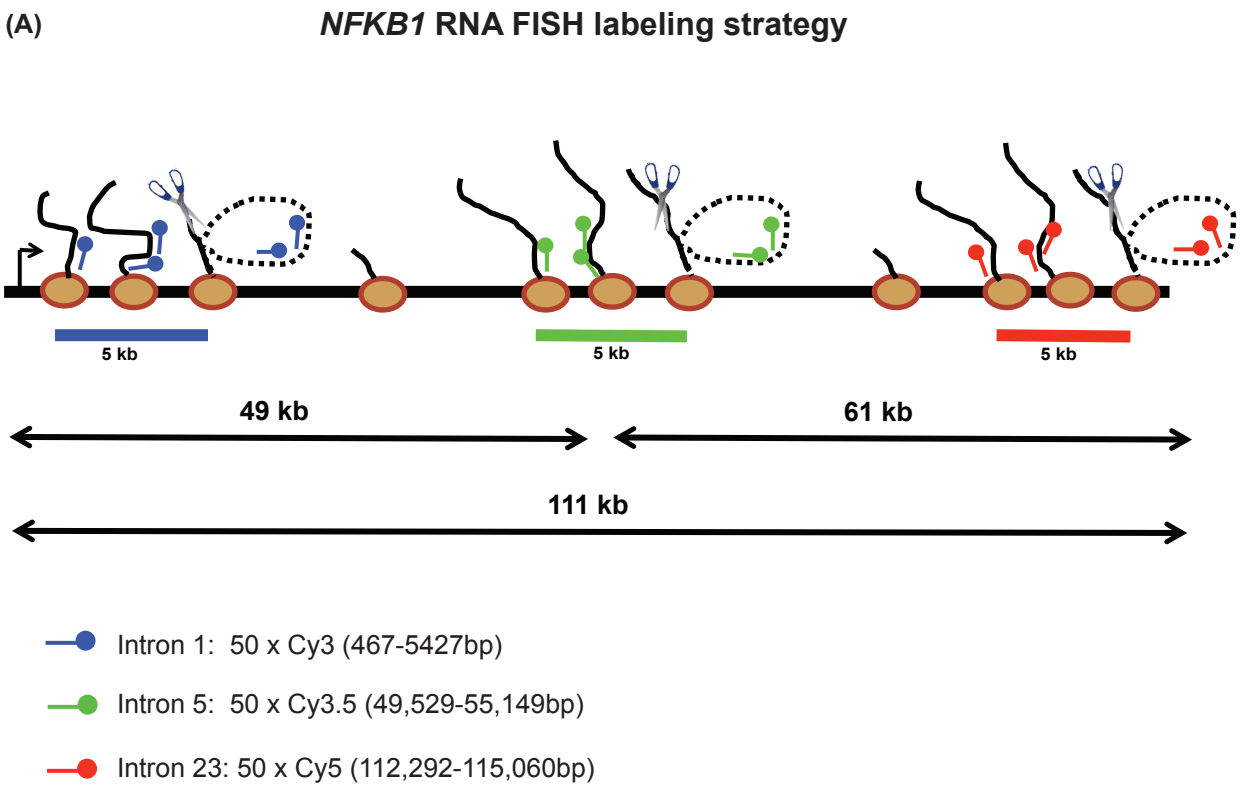


Figure III.7 *NFκB1* induction time course using RNA-FISH.

(A) The RNA FISH labelling strategy. Nearly 5 kbp long oligonucleotide probes were designed specific for intron 1, intron 5 and intron 23; and labelled with Cy3, Cy3.5 and Cy5, respectively. The orange circles depict transcribing RNAPII, while the black lines depict emerging nascent transcript. The dotted lines denote cotranscriptional splicing of intronic sequences. The linear distances between the different probes are shown below.

Figure III.7 continued.....

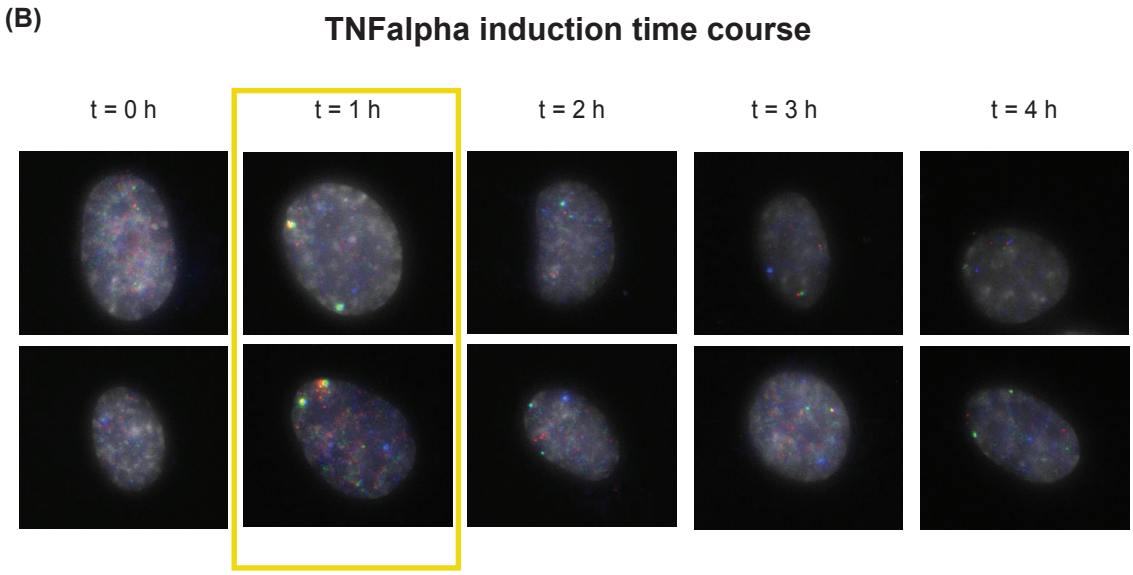


Figure III.7 *NFκB1* induction time course using RNA-FISH.

(B) TNFα induction time course. HUVECs were induced with 10 ng/ml of TNFα and RNA FISH analysis was performed at the indicated time points. No signal was detected at t = 0 min, consistent with the absence of transcription. Strong signals were detected for all three probes at t =1 h and declined thereafter. 1 h was selected as the suitable time point to pursue further RNA FISH analysis.

dimensional space using the spot detection program and distances between a spot of one color with the nearest neighboring spot in the other color were calculated. The distances obtained were plotted relative to the number of transcription sites and the data was fitted to a normal distribution curve to identify the mean separation between two spots. Chromatic aberration between two channels within a z-stack was corrected using sub-diffraction fluorescent beads (Figure III.8B, bright green spots) as fiduciary markers. Distance between spots was measured with ~25 nm precision, when 10 or more beads were used for correcting the 3D offset.

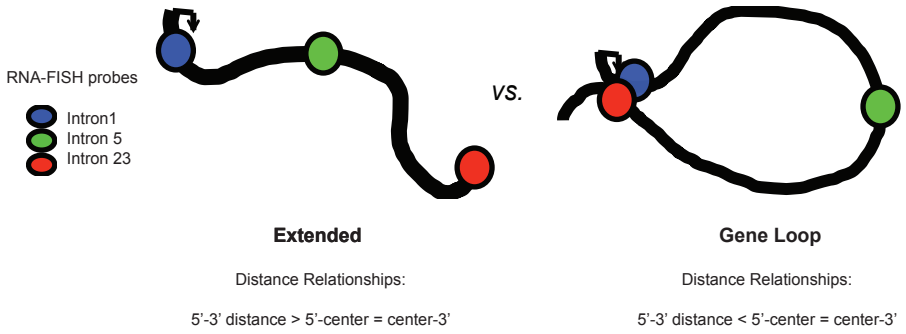
The dataset was also compared to FISH analysis performed in *S. cerevisiae* (Bystricky et al. 2004). The authors assessed colocalization between genomic intervals of 14-100 kbp along chromosomal arms, using high-resolution fluorescence *in situ* hybridization. The interphase chromatin was suggested to exist in higher-order chromatin conformation, with a persistence length of 197 ± 62 nm and a mass density of 144 ± 17 bp/nm. We used the above values to calculate the end-to-end distance between *NFκB1* probes using the Porod-Kratky model (Kratky and Porod 1949). The distance for shorter fragments, 5'-center (49 kbp) and center-3' (61 kbp) was in agreement with the budding yeast data. However, 5'-3' distance (111 kbp) showed a drift from the expected value of ~470-580 nm, to an observed value of 358 nm. This suggests a rather constricted geometry for *NFκB1* but still no gene-loop. Additionally, there was no correlation between the lengths of the 5'-center and center-3' fragments, suggesting a random polymer walk model for *NFκB1* conformation. The analysis was also confirmed with a Gaussian model, which suggested that *NFκB1* behaves as a linear polymer, with each node (chemical bonds connecting nucleotides) free to orient itself in nuclear volume, whilst still being joined to its neighboring subunits.

III.3. DISCUSSION

The work presented in this chapter investigates the phenomenon of intra-genic looping in constitutive and inducible mammalian transcriptional systems. Diagrammatic representations often present an oversimplified, linear version of transcription. Briefly, the general transcription factors recognize the promoter elements and facilitate RNAPII binding at promoter. Transcription activators bind enhancer elements and enhancer-promoter contacts are established by coactivators such as the mediator complex and cohesin. RNAPII undergoes several rounds of

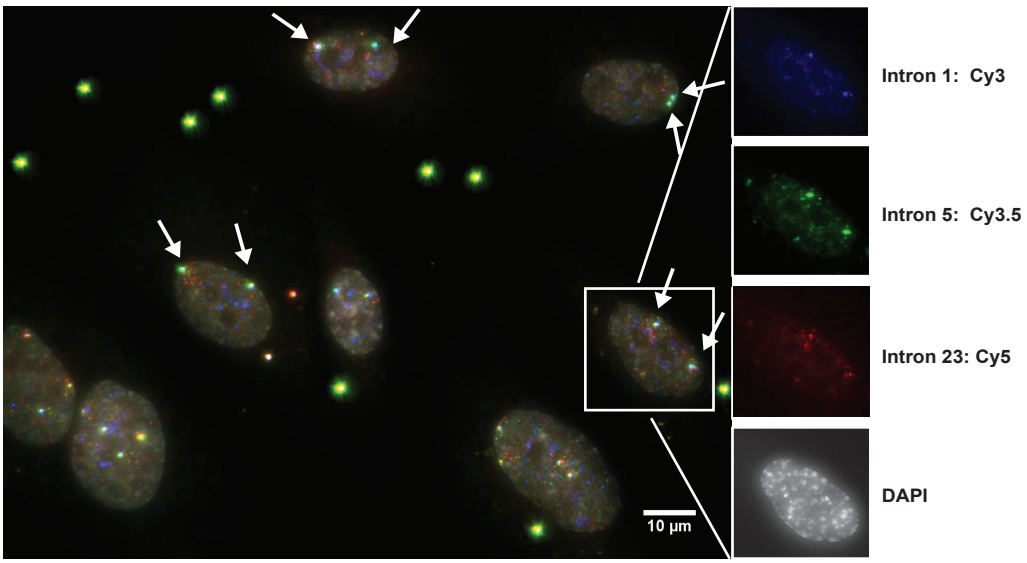
(A)

Two extreme conformations possible for *NFkB1* upon induction



(B)

RNA-FISH spots at *NFkB1* after 1 h induction with TNFalpha (10 ng/ml)



(C)

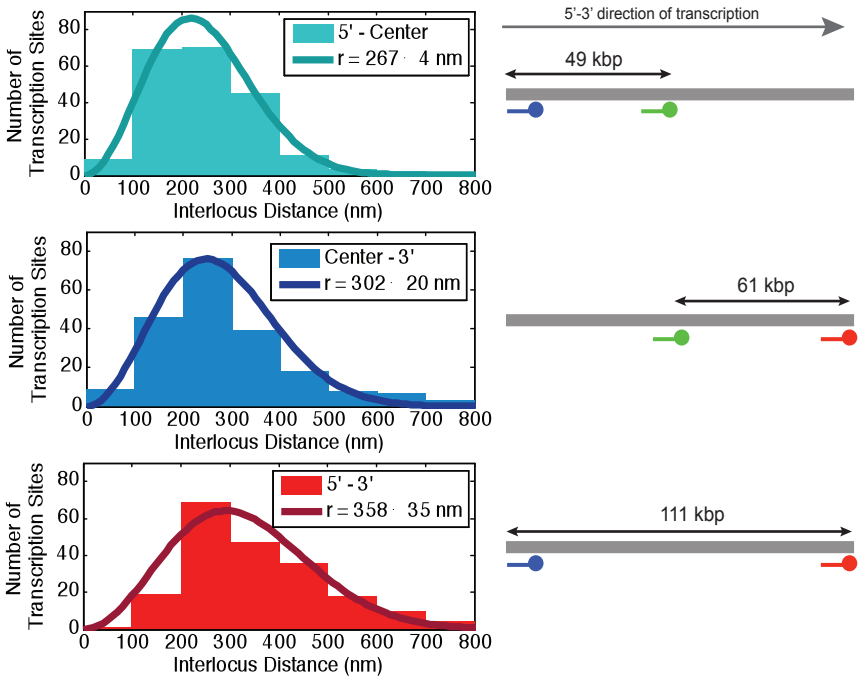


Figure III.8 Image analysis of *NFκB1* conformation.

(A) Two hypothesised conformations for *NFκB1*. If *NFκB1* forms a gene-loop, the distances between the 5' and 3' probes would be smaller than the distances between the 5'-center and center-3'. However, if *NFκB1* was randomly oriented, the 5'-3' distance would be greater than the 5'-center and center-3'.

(B) RNA FISH spots for *NFκB1* transcripts. A microscopic image showing FISH signals for intron 1 (blue), intron 5 (green) and intron 23 (red) probes. The nuclear outline is shown by DAPI staining. Green bright spots represent fluorescent beads used as fiduciary markers to correct for chromatic aberrations between the three channels. Most cells have two FISH spots for each probe, indicating transcriptional induction of both alleles.

(C) Colocalization analysis. Distance between the center of the spots was calculated between two probes at a time and is plotted against the number of transcription sites analyzed. The linear distance of 49 kbp between the 5' and center translates into 267 nm in the nuclear volume. Similarly, 61 kbp of linear distance between the center-3' equals 302 nm and 111 kbp between the 5' and 3' ends is equivalent to 358 nm. Clearly, *NFκB1* exhibits an extended conformation without forming any gene-loop.

abortive initiation, followed by productive elongation that coincides with promoter clearance and RNAPII escape into the gene body. As RNAPII proceeds, the nascent transcript is cotranscriptionally spliced. Once RNAPII encounters the termination region, it falls off the DNA template and transcription is terminated (Richard and Manley 2009).

In vivo measurements of transcription kinetics demonstrate a residence time of between few milliseconds - 100 seconds for transcription factors and an elongation rate between 0.4 kb/min to 4.3 kb/min for RNAPII (Darzacq et al. 2007). Most transcription factors randomly diffuse inside the nucleus, making non-specific contacts with the chromatin fiber, until they encounter specific binding sites. Additionally, the chromatin fiber exhibits constrained diffusion over a range of 0.1-1 μm (Misteli 2007). In light of such a scenario, one round of successful transcription would require an orchestration of chromatin remodeling, along with recruitment and stabilization of the transcription machinery at the target genes. Furthermore, quantitative modeling of the RNAPII FRAP data suggests that only 1 in 90 polymerases proceed to elongation (Darzacq et al. 2007).

Considering such a stochastic nature of transcription and the challenge to maintain high levels of housekeeping and cell-specific gene expression programs, intra-genic looping has been suggested as a mechanism to ensure rapid re-initiation after the termination of transcription. The initial results in this chapter indicate an interaction between the *Alb* promoter and terminator region, suggesting gene-loop formation. A possible mechanism for the preferential promoter-polyA juxtaposition is polymerase pausing associated with the polyA signal (Ashfield et al. 1991; Enriquez-Harris et al. 1991; Glover-Cutter et al. 2008; Gromak et al. 2006). It has been suggested that CPSF bound to RNAPII scans the nascent RNA for AT-hexamers and once it encounters a functional polyA signal, CPSF results in polymerase pausing. Pausing also provides time to recruit CstF to CPSF (Nag et al. 2007).

The *Alb* enhancer region is known to influence albumin transcription in adult liver. Consistent with this, I observed an interaction between the enhancer and promoter regions. The enhancer-promoter and promoter-terminator interactions were transcription-dependent because inhibition of RNAPII elongation resulted in a drop in 3C signals. Transcription-dependent gene-loop formation shown here is consistent with previous observations of gene-loops across yeast, mammals and plants,

discussed in section I.1.3. Surprisingly, the enhancer-promoter interaction was substantially reduced upon transcription inhibition. This is contrary to the current models of enhancer function that argue for a transcription-independent role of enhancers in target gene activation (Schoenfelder et al. 2010). Enhancers communicate with target gene promoters either through chromatin looping or by tracking of enhancer-bound transcription factors and the mediator complex, through intervening chromatin to a target gene promoter (Bulger and Groudine 2011; Ong and Corces 2011; Kolovos et al. 2012). An exciting possibility is that enhancer RNA (eRNA) might act as a binding platform for transcription factors and assist in enhancer-promoter looping (Li et al. 2013). It is important to note that *Alb* is expressed at high levels during fetal stages, long before any contact with its own enhancer is established and therefore, an eRNA might assist such communication.

Enhancer RNAs represent a novel class of small, uni- or bi-directional, non-coding RNAs that are expressed during signal-dependent gene activation programs (Kim et al. 2010; Wang et al. 2011; Wenbo et al. 2013; Li et al. 2013; Hah et al. 2013; Melo et al. 2013). In two separate studies, Li et al. (2013) and Hah et al. (2013) have shown genome-wide induction of eRNA transcription upon estrogen stimulation in breast cancer cells. Using 3C analysis, both groups establish that eRNAs were associated with enhancer-looping between target genes. In particular, Li et al. further demonstrated a requirement for cohesin and mediator subunit MED12 in maintaining such loops. Direct depletion of eRNAs with locked nucleic acid antisense oligonucleotide and siRNA reduced target gene expression, diminished cohesin recruitment to activated enhancers and perturbed enhancer-promoter looping (Li et al. 2013). In contrast, Hah et al. do not observe any loss in enhancer-promoter loops after inhibiting eRNA transcription using Flavopiridol. However, targeted depletion of a transcript is a more direct approach than chemical inhibition. A similar role for eRNA in long-range interactions has also been described for p53-bound enhancer regions (Melo et al. 2013). However, the existence and functional significance (if any) of an eRNA at the *Alb* locus was not investigated.

As the linear distance between two regions increases, their interaction frequency decreases (Dekker et al. 2002). However, 3C analysis showed that *Alb* exists in a ‘cloverleaf’ conformation during transcription, with the promoter in contact with intronic sequences. Such conformation has been proposed for the *BRCA1* gene, albeit under transcription repression (Tan-Wong et al. 2008). Similar conformations

have also been proposed for multi-gene loci, often anchored by a cell-type specific regulatory protein, such as SATB1 (Cai et al. 2006) or cohesin (Hadjur et al. 2009). Therefore, I was interested in identifying a potential 'anchor' or 'mediator', required for anchoring such contacts across the *Alb* gene locus. Considering the diverse array of proteins involved in intra- and inter-chromosomal interactions, a number of proteins seemed as possible candidates. These included, liver-specific transcription factors such as (Foxa3, CEβP), the cohesin complex, the mediator complex, transcription factor TFIIB or maybe RNAPII itself. Accumulating evidence suggests that actively transcribed genes accumulate at specialized transcription factories. However, single gene dynamics, presumably within a factory, had not been investigated at the time of this thesis. Therefore, I chose RNAPII as the candidate protein to perform ChIP-3C analysis (data not shown). If the promoter were tethered to the polymerase(s), then the 3C and ChIP-3C profiles would have been identical or else, a different protein would anchor the *Alb* loop. However, the ChIP-3C dataset was limited by the technical difficulties in combining the ChIP and 3C techniques.

Microscopic evidence suggests that every cell is unique with respect to the overall spatial arrangement of chromosomes and long-range interactions (Parada et al. 2003; Walter et al. 2003). As mentioned earlier, 3C analysis represents the average interactions of a heterogeneous population. Therefore, to understand the kinetics of gene-loop formation, I switched to image analysis. *NFκB1* was selected as the candidate gene because it is sufficiently long (115 kbp) and rapidly responds to exogenous stimuli, implying that its expression can be tightly controlled. The pioneering polymerases initiate within 15-30 min post induction and transcribe steadily to terminate >1 h. A combination of two techniques, 3C and RNA-FISH, was employed to independently investigate *NFκB1* conformation during transcription. 3C protocols screen nearly one-ten million cells and therefore, have a higher propensity of identifying interaction(s) between different loci. The first 3C dataset suggested promoter-terminator juxtaposition at *NFκB1* but surprisingly, the kinetics of loop formation deviated from the transcription kinetics observed with qRT-PCR and ChIP analysis. Even after several repeats using chromatin from HUVECs and other cell lines (HeLa and HepG2), conclusive evidence for a gene-loop at *NFκB1* was not obtained. This could be due to technical difficulties in optimizing the 3C protocol or due to physiological controls underlying *NFκB1* transcription (see below).

High-resolution RNA-FISH analysis was performed using probes positioned in intron1, 5 and 23 respectively. If *NFκB1* were to exist in a looped conformation, intron1 and intron 23 probes would colocalize after the pioneer round of transcription. However, the distance between the 5' and 3' ends was greater than the distances between the 5'-center and center-3' probes. This suggested the absence of a promoter-terminator gene-loop. Interestingly, maximal level of *NFκB1* induction was obtained 1 h after induction and transcription declined thereafter. This coincides with the IκB mediated feedback regulation (Oeckinghaus and Ghosh 2009). Briefly, TNFα stimulation induces IκB degradation, releasing active NFκB dimers that translocate to the nucleus and activate target gene expression. Surprisingly, *IκB* expression itself is regulated by NFκB. *De novo* transcription and translation rapidly replenishes cellular IκB that sequesters NFκB back to the cytoplasm and shuts down the inflammatory response.

RNA-FISH was reasoned as the method of choice to perform image analysis, as nascent RNA is present in multiple copies (unlike two copies of the DNA sequence) and therefore, would provide greater signal intensity in comparison to DNA FISH. However, it is difficult to discern the dwelling time of nascent RNA species at the transcription sites due to co-transcriptional splicing of intronic sequences (Richard and Manley 2009). Perhaps the best approach is DNA-FISH. However, gene loop signatures such as promoter-terminator phosho-Ser5 RNAPII peaks are detectable only in a sub-set (1-4%) of actively transcribed genes (Grosso et al. 2012). Highly transcribed genes tend to be shorter in length and thus, are limited by spatial and temporal resolution.

Even though the 3C, ChIP-3C and RNA-FISH analyses were unable to provide any conclusive information, I would like to emphasize the fundamental nature of the questions I addressed. ChIP-3C analysis with RNAPII as the bait had not been done when I initiated my studies and therefore, at the local and probably genome-wide level would have provided novel insights about transcriptionally engaged chromatin architecture. Indeed in 2012, Li et al. reported a promoter-centric RNAPII interactome, identifying a variety of promoter-promoter, promoter-enhancer and promoter-terminator interactions across five different cell lines (Li et al. 2012). Similarly, 3C-based gene-loop studies lacked any microscopic evidence as most genes did not exceed the microscopic limit of resolution. Had it been successful, the image analysis of *NFκB1* loop formation would have been the first *in vivo* evidence for

intragenic looping. Yet again, Larkin et al. performed similar investigations on three TNFα-inducible genes (Larkin et al. 2012). Using 3C analysis, the authors show a gene-loop for *SAMD4A* (>200 kbp) in uninduced conditions, dependent on CTCF, TFIIB and, surprisingly, transcription itself. It's possible that the loop maintains the gene in a "poised" conformation, so that it can respond rapidly to exogenous stimuli. Interestingly, *SAMD4A* also forms sub-gene loops between the promoter and different regions across the gene body, coinciding with kinetics of RNAPII elongation.

CHAPTER IV

“COHESIVE” versus “LESS-COHESIVE” COHESIN

The cohesin complex was discovered more than a decade ago, and has emerged as a key player in sister chromatid cohesion. Interestingly, cohesin also plays profound roles in gene expression, recombination and DNA damage repair. Genome-wide analyses of cohesin binding sites have been performed in numerous eukaryotes. These include *Saccharomyces cerevisiae* (Glynn et al. 2004; Lengronne et al. 2006), *Schizosaccharomyces pombe* (Schmidt et al. 2009), *Drosophila* cell lines (Misulovin et al. 2008), *Drosophila* salivary glands (Pauli et al. 2010), human lymphocytes (Liu et al. 2009), HeLa cells (Wendt et al. 2008), MCF-7 and HepG2 tumor cell lines (Schmidt et al. 2010), mouse pre-B cells, mouse embryonic stem cells (ESCs) (Kagey et al. 2010) and mouse thymocytes (Parelho et al. 2008). In higher eukaryotes, cohesin preferentially colocalizes with the kollerin complex, with strong signals near gene transcription start sites (Dorsett et al. 2013) Additionally, in mammals there is an 89% overlap between cohesin and CTCF binding sites (Wendt et al. 2008). Cohesin also colocalizes with tissue-specific transcription factors in a CTCF-independent manner (Schmidt et al. 2010, Nitzsche et al. 2011) and interestingly, it co-purifies with the mediator complex (Kagey et al. 2010).

In contrast to higher eukaryotes, centromeric, pericentric and telomeric regions constitute the major cohesin binding sites in yeast (Glynn et al. 2004; Lengronne et al. 2004; Schmidt et al. 2009). In *S. cerevisiae*, cohesin on the chromosome arms localizes exclusively between convergent genes (Glynn et al. 2004; Lengronne et al. 2004). In *S. pombe*, cohesin is detected at chromosomal loading sites, (i) in association with the kollerin complex at regions with strong transcriptional activity and (ii) at the transcription termination regions between convergent genes (Schmidt et al. 2009; Gullerova and Proudfoot 2008). It is widely believed that cohesin repositions from its loading site to intergenic regions, pushed by the transcription machinery. While the majority of fission yeast cohesin is removed from chromosomes during mitotic anaphase, a population remains associated throughout mitosis. This implies the existence of cohesin sub-populations responsible for (i) sister chromatid cohesion, (ii) transcriptional regulation, and (ii) homologous recombination and DNA damage repair.

The results described in this chapter provide evidence for the existence of two functionally different populations of cohesin in *S. pombe.* Cohesin present at the loading sites is stably associated with chromosomes, while cohesin away from such sites represents less-functional cohesin. As *S. pombe* is the main experimental

system, the fission yeast nomenclature for cohesin complexes is followed throughout this chapter (and also chapter V). The fission yeast analogs for cohesin subunits, Smc3, Smc1 and Rad21 are known as Psm3, Psm1 and Rad21 (Tomonaga et al. 2000). The fourth subunit (Scc3 in budding yeast) and (SA/STAG in mammals) is known as Psc3. Likewise, in fission yeast, the cohesin loading/kollerin complex consists of Mis4/Ssl3, analogs of Scc2/Scc4 (*S. cerevisiae*) and NIPBL/MAU2 (mammals).

IV.1 VALIDATION OF Mis4/Ssl3 and COHESIN BINDING PATTERN IN *S. pombe*

First, I confirmed the cohesin and Mis4/Ssl3 ChIP-chip profiles along chromosome 2 by ChIP analysis. I employed strains where Rad21 and Mis4 are endogenously tagged with GFP (Rad21-GFP and Mis4-GFP) and immunoprecipitated the crosslinked complexes with an anti-GFP antibody (Abcam). qPCR primers were specifically designed for the intergenic regions between convergent gene pairs, as ~94% of the cohesin peaks and 41.7% of the Mis4 peaks overlap with convergent gene pairs (Schmidt et al. 2009).

Figure IV.1A shows a schematic of the cohesin and Mis4 binding pattern identified by ChIP-chip analysis. The orange arrows represent gene orientation in the sense and antisense directions. The Rad21 binding pattern is shown as blue ovals and Mis4/Ssl3 is shown as yellow circles. The Rad21 and Mis4/Ssl3 binding pattern spans multiple open reading frames, with maximum signal intensity at intergenic regions. In cycling cells (G_2), Rad21 and Mis4 ChIP signals (Figures IV.1B and 1C) were detected at intergenic regions (dis2, srp1, ubp9, mrs2 and hrf1) that have been shown to bind Rad21 and Mis4/Ssl3 (Schmidt et al. 2009). Furthermore, Rad21 and Mis4/Ssl3 were absent from a region previously shown to be devoid of the two proteins and hence, was used as a negative control in subsequent experiments.

IV.2 “COHESIVE” versus “LESS-COHESIVE” COHESIN

ChIP-chip analysis has identified 228 cohesin sites but only 72 Mis4 sites across chromosome 2 (Schmidt et al. 2009). Out of 228, only 44 cohesin sites overlap with a Mis4 binding site. It has been hypothesized that cohesin at its loading site may maintain a dynamic state of binding, useful for promoting reversible interactions between regulatory regions (Schmidt et al. 2009). In contrast, cohesin

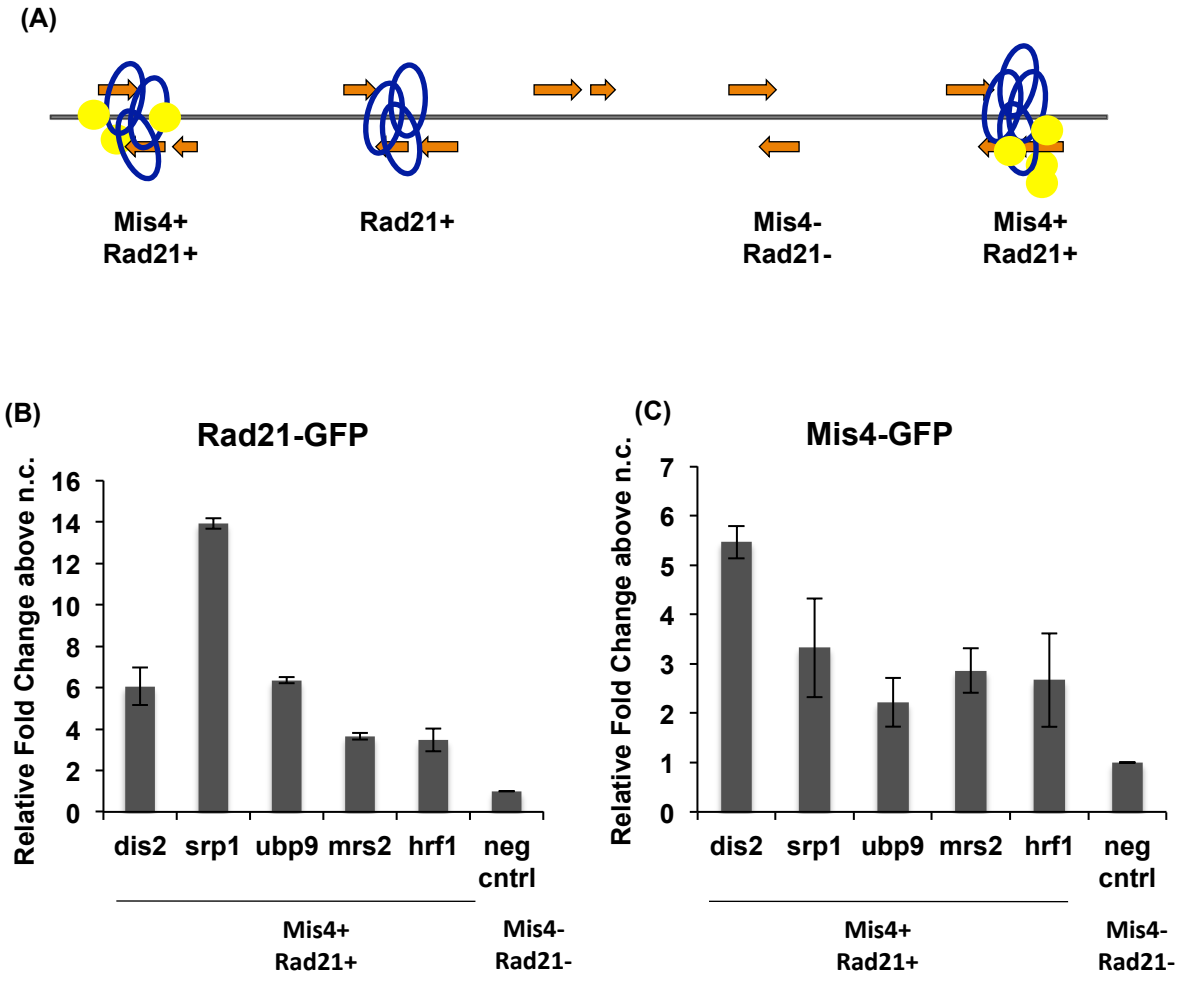


Figure IV.1 Validation of cohesin and Mis4 binding along chromosome 2

(A) Schematic representation cohesin (blue oval) and Mis4 (yellow circles) binding profile along chromosome 2 in *S. pombe.* The orange arrows represent genes transcribing either in the sense (above black line) or antisense (below black line) directions. Arrows pointing towards each other represent convergent gene pairs and those in the same direction are tandem genes. Binding sites with cohesin and Mis4 colocalization (Mis4+/Rad21+) or cohesin alone (Rad21+) or without Mis4 and cohesin (Mis4-/Rad21-) are shown.

(B) Rad21-GFP occupancy in G_2 cells detected by ChIP analysis using a Rad21-GFP strain. The grey bars represent Rad21 enrichment relative to the negative control (Mis4-/Rad21-). The probes (dis2, srp1, ubp9, mrs2 and hrf1) are positioned in the intergenic regions of corresponding convergent gene pairs where cohesin and Mis4 binding has been previously detected. Error bars represent standard deviation from three independent experiments.

(C) Mis4-GFP occupancy in G_2 cells detected by ChIP analysis using a Mis4-GFP strain, at the same loci as mentioned in (B). Error bars represent standard deviation from three independent experiments.

translocated away from the loading site, may function in sister chromatid cohesion. To investigate such functional disparity between the Mis4/Rad21 and Rad21 only sites, I performed single locus DNA-Fluorescent *in situ* hybridization (FISH) analysis.

Five sites with Mis4/Rad21 (Chapter II, Table II.9a and II.9b) and Rad21 only (Chapter II, Table II.9c) were randomly selected for FISH analysis. Overlapping PCRs were performed to amplify ~10 kbp regions inclusive and around the convergent regions. For each locus, the PCR products were mixed in equimolar amounts and labeled with the FISH-Tag Kit (Molecular probes, Invitrogen), according to the manufactures' protocol. Cycling cells (G_2) were fixed with para-formaldehyde, spheroplasted with zymolyaseT100 (100 units) and treated with RNase A (1 µg/ml), to prepare them for hybridization. 100 ng of labeled probe (1 ng/µl) was hybridized to fixed cells overnight. The following day, cells were washed, stained with DAPI (0.2 µg/ml) and transferred to poly-L-lysine coated slides. Imaging was performed using the DeltaVision microscope with a 100X-1.3 NA objective lens (Department of Biochemistry, Oxford). Z-sections were taken at 200 nm intervals and raw images were deconvolved using Softworx (Applied Precisions). Images with single or double FISH signals were identified and quantified manually. The final images representing maximum intensity plots for the z-sections were obtained using the Fiji/ImageJ program (NIH).

It was hypothesized that if two sister chromatids are stably held together within the 40 nm diameter of the cohesin ring, they would appear as a single dot. On the other hand, unstable cohesion would result in separation of sister chromatids, which would appear as two dots, one for each chromatid. Interestingly, single dots (Figure IV.2B) were observed for Mis4/Rad21 regions, implying stable sister chromatid cohesion. In over 200 cells counted, dis2, srp1, tif11, ubp9 and rpl3801 demonstrated 100% cohesion. The centromeric and per-centric regions are major cohesin binding sites (Peters JM 2008). Therefore, the centromeric degenerate *(dg)* repeats were used as a positive control and 100% cohesion was observed at the *dg*. The centromeric probe targets the degenerate repeats within the outer region. The *dg* repeats are present in centromeres 1, 2 and 3. As the *S. pombe* genome is highly compact, it is difficult to resolve the three centromeres as separate entities.

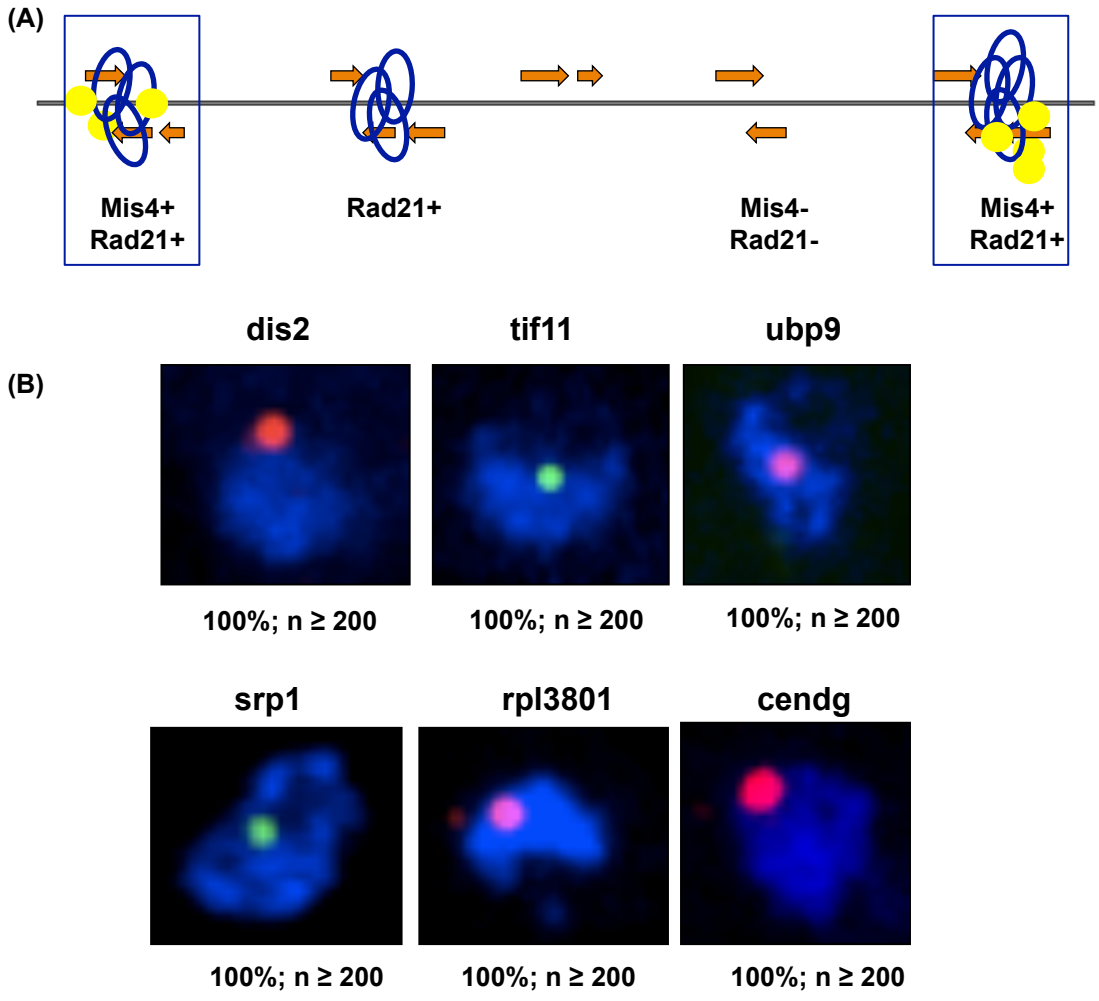


Figure IV.2 Stable sister chromatid cohesion at Mis4+/Rad21+ sites.

(A) Schematic representation cohesin (blue oval) and Mis4 (yellow circles) binding profile along chromosome 2 in *S. pombe.* The orange arrows represent genes transcribing either in the sense (above black line) or antisense (below black line) directions. Arrows pointing towards each other represent convergent gene pairs and in the same direction are tandem genes. Binding sites with cohesin and Mis4 colocalization (Mis4+/Rad21+) or cohesin alone (Rad21+) are shown. The blue rectangles represent Mis4+/Rad21+regions selected for FISH analysis in (B).

(B) Single locus FISH analysis of regions along chromosome 2. Nearly 10 kbp long overlapping PCR products that span the Mis4+/Rad21+ convergent gene loci were pooled in equimolar quantities and labeled according to the FISH-Tag kit (Invitrogen). The blue regions depict DAPI stained DNA and the single dot within the nucleus is the DNA-FISH signal. Probes labeled with Alexa-488 are green, while those labeled with Alexa-555 are red. For each experiment n ≥ 200 cells were counted. The images represent maximum intensity projection of z-sections taken at 200 nm intervals. Clearly, Mis4+/Rad21+ regions exhibit 100% cohesion, similar to the centromere *dg* positive control.

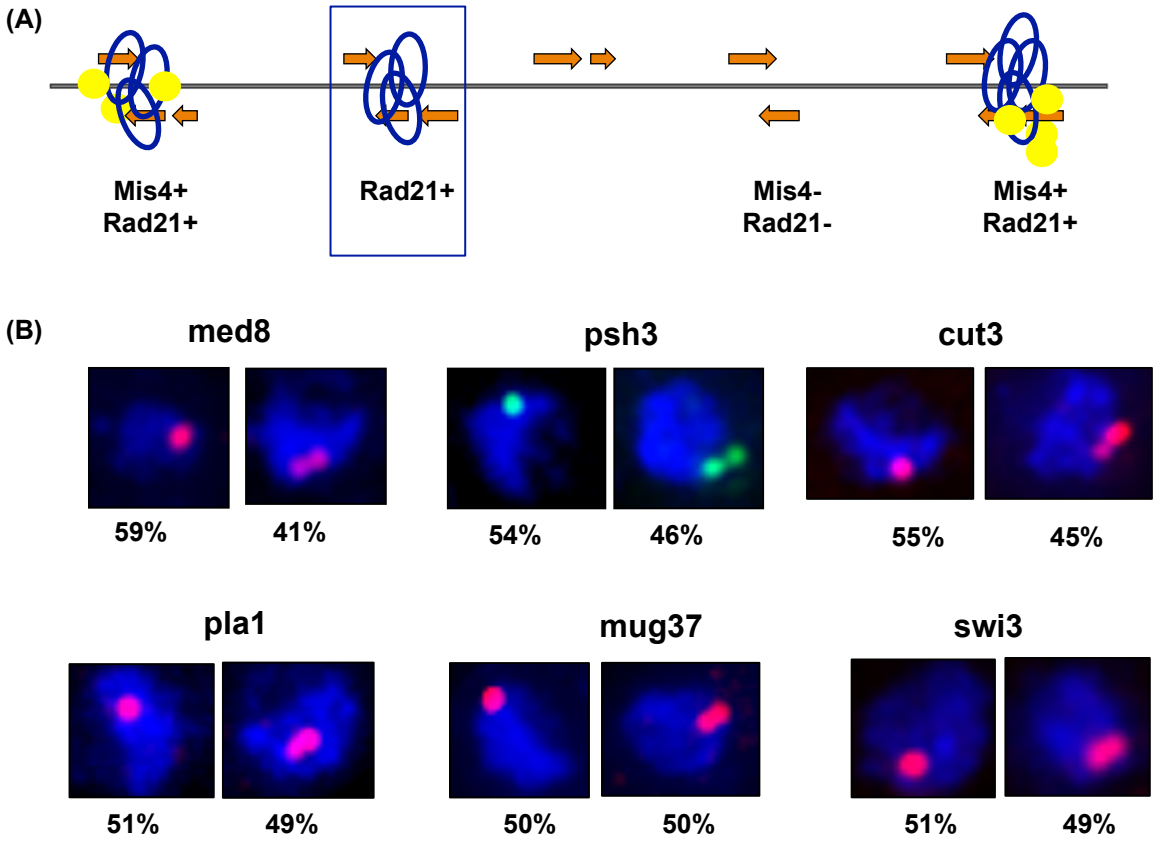


Figure IV.3 Unstable sister chromatid cohesion at Rad21+ sites

(A) Schematic representation cohesin (blue oval) and Mis4 (yellow circles) binding profile along chromosome 2 in *S. pombe.* The orange arrows represent genes transcribing either in the sense (above black line) or antisense (below black line) directions. Arrows pointing towards each other represent convergent gene pairs and in the same direction are tandem genes. Binding sites with cohesin and Mis4 colocalization (Mis4+/Rad21+) or cohesin alone (Rad21+) are shown. The blue rectangle represents the cohesin only (Rad21+) regions analyzed by FISH analysis in (B).

(B) Single locus FISH analysis of Rad21+ only regions along chromosome 2. Nearly 10 kbp long overlapping PCR products that span the Rad21+ convergent gene loci were pooled in equimolar quantities and labeled according to the FISH-Tag kit (Invitrogen). The blue regions depict DAPI stained DNA and the single/double dots within the nucleus is the DNA-FISH signal. Probes labeled with Alexa-488 appear as green, while those labeled with Alexa-555 appear red. For each experiment n ≥ 100 cells were counted. The images represent maximum intensity projection of z-sections taken at 200 nm intervals. Clearly, the Rad21+ only regions exhibit unstable, ~50% cohesion.

On the other hand, each Rad21 only regions (med8, psh3, cut3, pla1, mug37 and swi3) showed ~50% single and ~50% double dots (Figure IV.3B). This suggests that Rad21 regions away from the loading site are unable to maintain stable sister chromatid cohesion. Thus, *S. pombe* cohesin can be divided into two populations, (i) "cohesive" cohesin that is present at its loading site in along with Mis4 (Mis4+/Rad21+) and (ii) "less-cohesive" cohesin, i.e., cohesin present by itself (Rad21+ only).

IV.3 LOSS OF SISTER CHROMATID COHESION IN *rad21ts*

To confirm that sister chromatids at "cohesive" sites are indeed held together by Rad21, FISH analysis (as above) was performed in a Rad21 temperature sensitive strain. In a preliminary experiment, the *rad21-K1ts* mutant (Tatebayashi et al. 1998) and wild type (972) cells were synchronized in G_1 by nitrogen starvation. In response to nitrogen depletion, *S. pombe* cells become smaller and exit the cell cycle in a stationary phase, accumulating with 1C DNA content. Upon re-feeding with complete media, cells re-enter the cell cycle. Synchronization is complete only after four generation times (Luche and Forsburg 2009) and therefore, cells were grown in essential minimal medium without any nitrogen source, for 16 h at 25°C. Subsequently, the cultures were transferred to rich medium and shifted to 37°C (restrictive temperature). Aliquots were taken at hourly intervals till 8 h and DNA content was measured by FACS analysis (Figure IV.4B). Both, the wild type (left panel) and *rad21-K1ts* (right panel) progressed from 1C to 2C DNA content ~4 h after the shift to 37°C. Such a delay is observed because nitrogen-starved cells exhibit a 2 h lag period before re-entering the cell cycle (Luche and Forsburg 2009). Thereafter, the DNA content in *rad21-K1ts* gradually became heterogeneous, predominantly between 6-8 h. This result complies with previous studies that indicate functional replication but defective chromosome segregation in *rad21-K1ts* mutant (Tatebayashi et al. 1998).

FACS analysis indicated that replication occurred ~4 h after temperature shift (Figure IV.4B). Therefore, DNA-FISH was performed in wild type and *rad21-K1ts* cells at t = 4, 6 and 8 h after the temperature shift. Wild type cells (Figure IV.4A) maintained stable cohesion at chromosomal arm (tif11) and centromeric *dg* repeats throughout. Centromeric and arm cohesion was intact in *rad21-K1ts* at t = 4 h. While, centromeric cohesion was still stable at t =6 h, cohesion at the arm (tif11) was

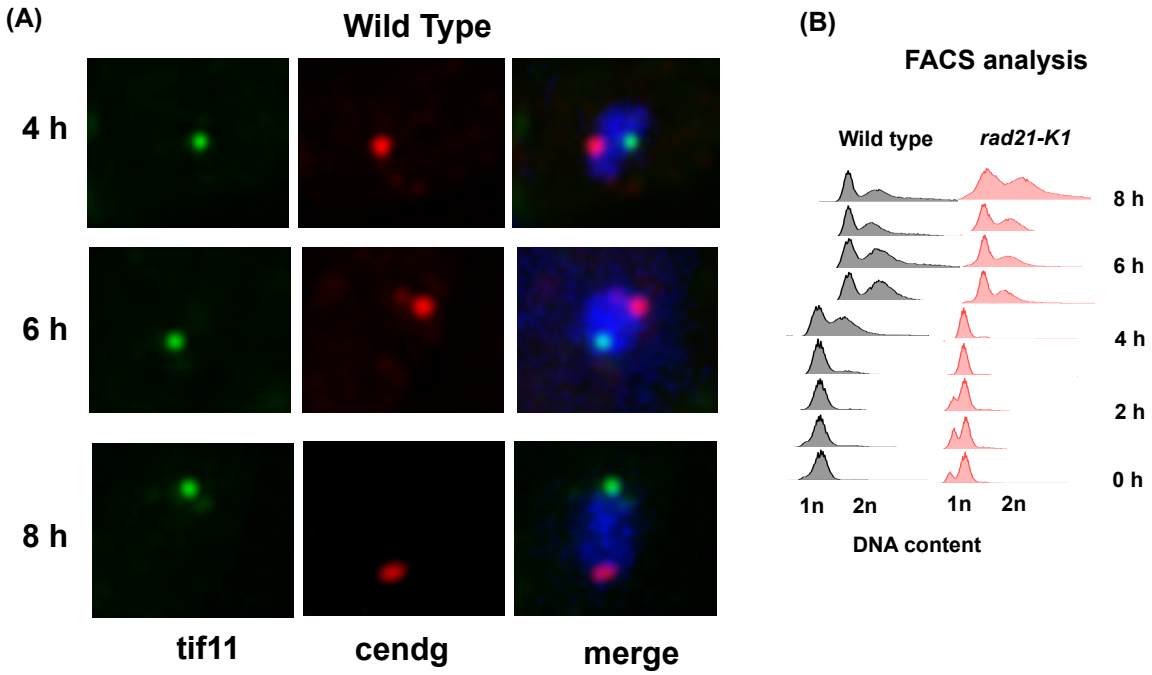


Figure IV.4 Sister chromatid cohesion at "cohesive" sites is mediated by Rad21 and Mis4.

(A) Single locus FISH analysis of Mis4+/Rad21+ regions along chromosome 2 in a wild type strain. The blue regions depict DAPI stained DNA, while the single dot within the nucleus is the DNA-FISH signal. Tif11 is a FISH probe targeting Mis4+/Rad21+ loci along the chromosome arms (as shown in Figure IV.2). cendg is the FISH probe specific to the centromere *dg* repeat (as shown in Figure IV.2). At all time points, the arm and centromeric cohesion remains intact. Images represent maximum intensity projection of z-sections taken at 200 nm intervals (n = 50-100 cells).

(B) Wild type and *rad21-K1* cells were synchronized in G_1 by nitrogen starvation for 15 h at 25°C. Subsequently, cells were transferred to rich medium and shifted to 37°C. Aliquots were taken at hourly intervals till 8 h for FACS analysis. Cells were fixed in cold 70% ethanol, washed with sodium citrate, RNase treated and stained with propidium iodide overnight at 4°C. FACS analysis was performed using the FACS Calibur. Both cell types completed replication ~4 h after shift to 37°C. Subsequently, the WT progressed normally to the second cell cycle but *rad21-K1* exhibited chromosome segregation defects.

Figure IV.4 continued……

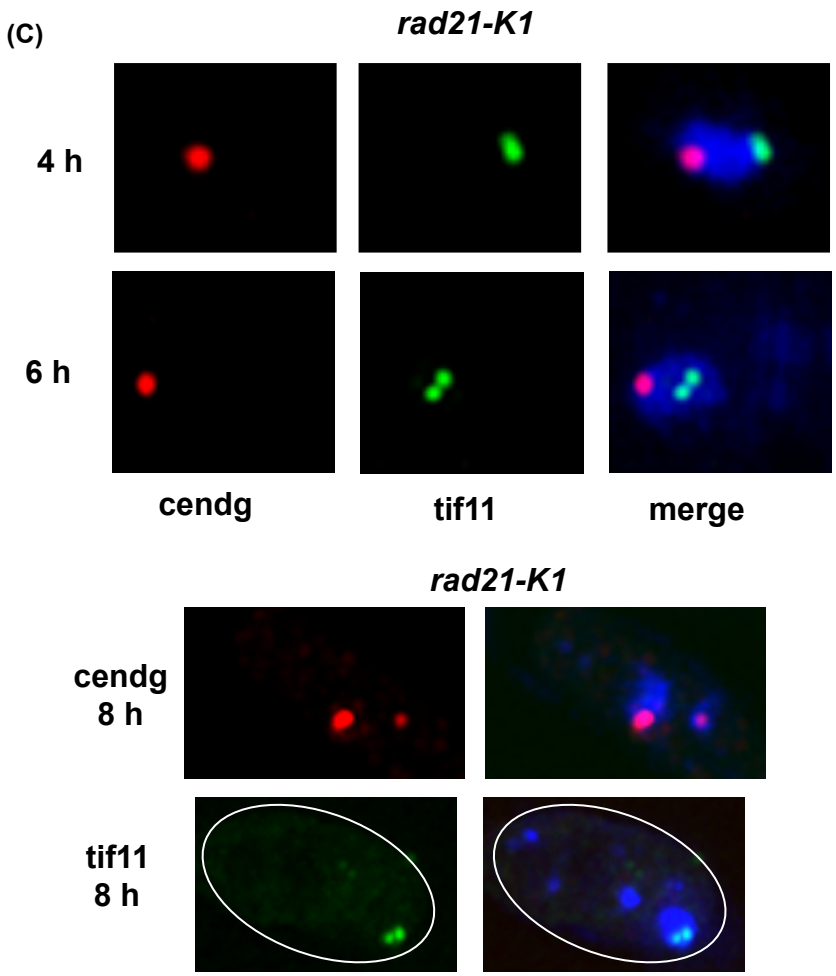


Figure IV.4 Sister chromatid cohesion at "cohesive" sites is mediated by Rad21 and Mis4.

(C) Single locus FISH analysis of Mis4+/Rad21+ regions along chromosome 2 in *rad21-K1ts* mutant. The blue regions depict DAPI stained DNA and the dot within the nucleus is the DNA-FISH signal. Tif11 targets Mis4+/Rad21+ loci along the chromosomal arms, while cendg is specific to the centromere *dg* repeats. At t = 4 h after the shift to 37°C, arm and centromeric cohesion was intact in the mutant. However, at t = 6 h, arm cohesion destabilized but centromeric cohesion was still intact. Furthermore, at t = 8 h arm and centromeric cohesion was destabilized, resulting in chromosome segregation defects as observed in the FACS analysis (IV.4B). Images represent maximum intensity projection of z-sections taken at 200 nm intervals. The white oval outlines the cell boundary (n = 50-100 cells).

Figure IV.4 continued……

destabilized, indicating inactivation of the Rad21 protein. Impaired arm cohesion persisted at t = 8 h and centromeric cohesion was visibly destabilized (Figure IV.4C). This emphasizes the fact that single dots observed at the “cohesive” sites (Figure IV.2) are indeed “cohesed” sister chromatids that separate upon inactivation of Rad21.

IV.4 LOSS OF SISTER CHROMATID COHESION IN *mis4ts*

A similar experiment (as above) was performed in a Mis4 temperature sensitive strain, *mis4-367ts* (Bernard et al. 2006). At t = 4, 6 and 8 h, wild type cells exhibited stable cohesion at centromeric *dg* repeats and the chromosomal arm, tif11 (data not shown). In *mis4-367ts*, apparent defects in arm cohesion were absent at t = 4 h, in accordance with previous reports that replication is unaffected upon Mis4 inactivation (Furuya et al. 1998). Defects in arm cohesion were observed at t = 6 h, continuing till t = 8 h (Figure IV.4D). Centromeric cohesion was intact till t = 6 h but destabilized at t = 8 h. This reiterates the fact that “cohesive” sites represent “cohesed” sister chromatids that are held together by the cohesin complex, which is in turn loaded by Mis4.

IV.5 STOICHIOMETRIC DIFFERENCES IN COHESIN OCCUPANCY AT “COHESIVE” versus “NON-COHESIVE” SITES

From a total of 228 cohesin peaks identified on chromosome 2, only 43 overlap with Mis4/Ssl3. Interestingly, the height of cohesin peaks with or without Mis4 is remarkably similar (compare peak height and length between peaks listed in Chapter II, Table II.9.a, 9.b and 9.c) (Schmidt et al. 2009). However, DNA FISH analysis (Figure IV.2B and IV.3B) suggests functional differences between the two regions. To address this discrepancy, ChIP analysis was performed using a strain where cohesin subunit Psm3 is endogenously tagged with GFP (Psm3-GFP).

Surprisingly, Psm3 levels at the “cohesive” sites (Figure IV.5B) (dis2 and srp1) were ~3 fold higher than the “less-cohesive” sites (med8, plh1 and 776). This implies that functional differences between the “cohesive” and “less-cohesive” sites possibly arise due to stoichiometric variation in the amount of cohesin associated along chromosome arms. In mammalian cells, FRAP data suggests only one-third nuclear cohesin is stably associated with chromatin, while the remaining two-third is

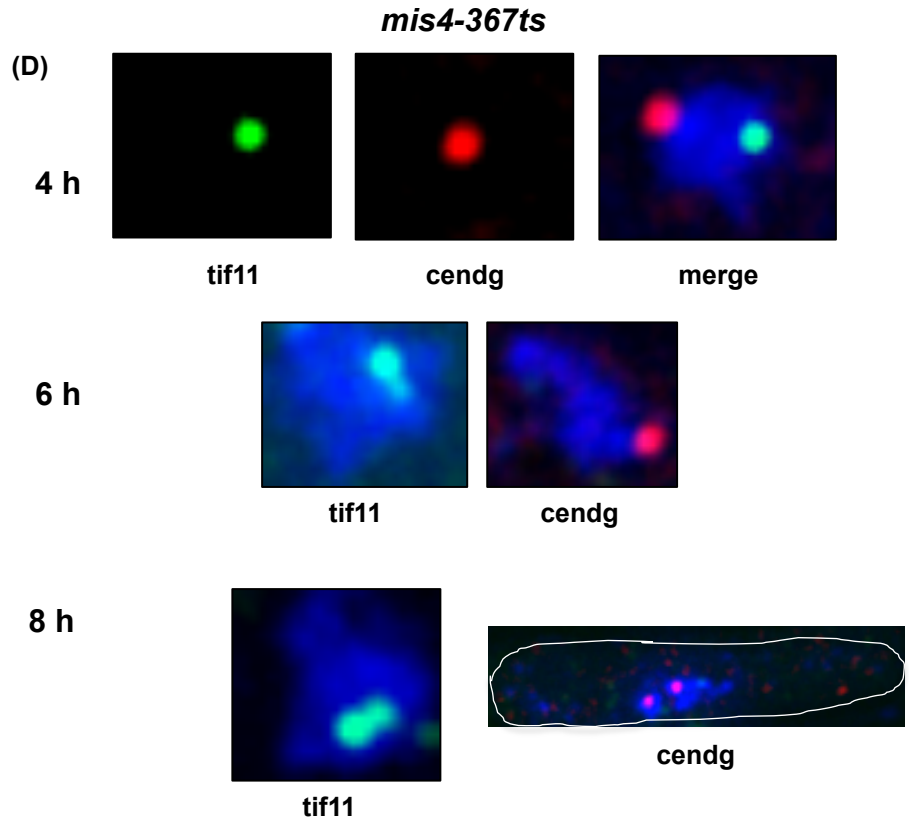


Figure IV.4 Sister chromatid cohesion at "cohesive" sites is mediated by Rad21 and Mis4.

(D) Single locus FISH analysis of Mis4+/Rad21+ regions along chromosome 2 in *mis4-367ts* mutant. The blue regions depict DAPI stained DNA and the dot within the nucleus is the DNA-FISH signal. Tif11 targets Mis4+/Rad21+ loci along the chromosomal arms, while cendg is specific to the centromere *dg* repeats. At t = 4 h after the shift to 37°C, arm and centromeric cohesion was intact in the mutant. However, at t = 6 h, arm cohesion destabilized but centromeric cohesion was still intact. Furthermore, at t = 8 h, arm and centromeric cohesion destabilized, resulting in chromosome segregation defects as observed in the FACS analysis (IV.4B). Images represent maximum intensity projection of z-sections taken at 200nm intervals. The white line outlines the cell boundary (n = 50-100 cells).

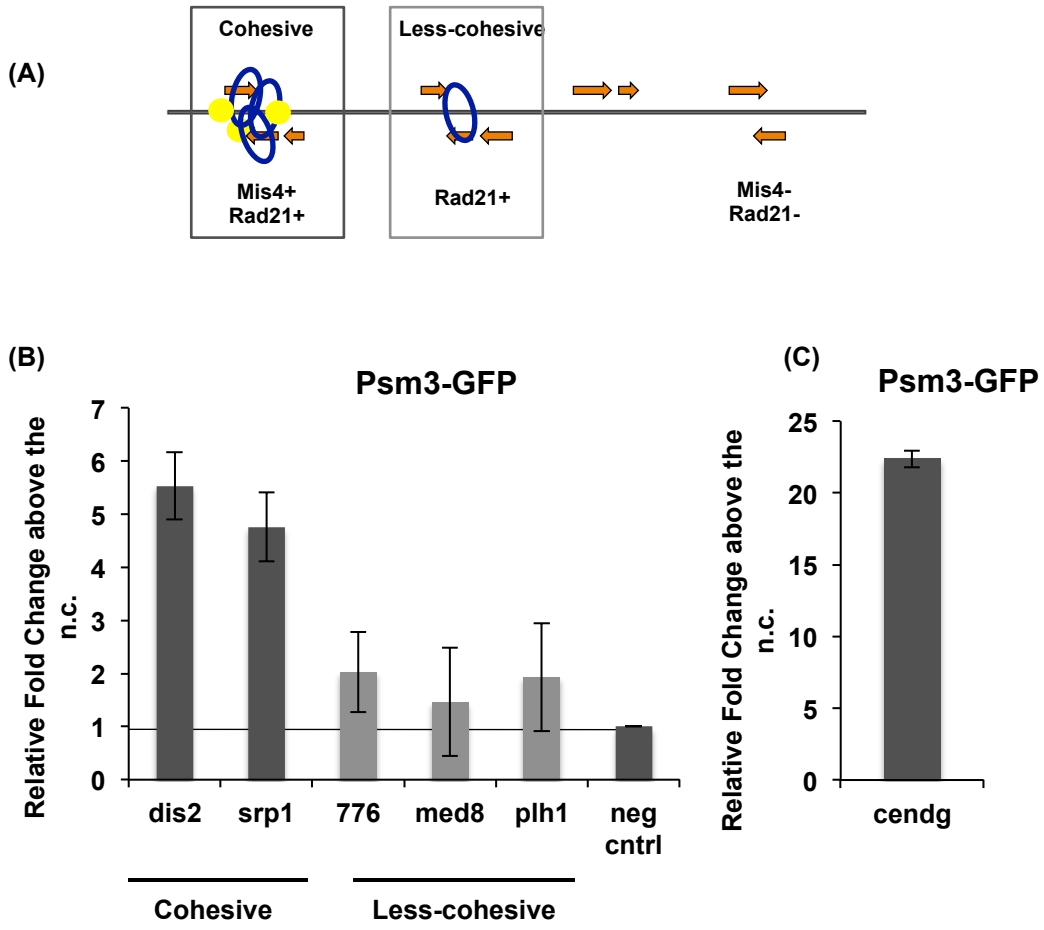


Figure IV.5 Cohesin occupancy at "cohesive" versus "less-cohesive" sites.

(A) Schematic representation cohesin (blue oval) and Mis4 (yellow circles) binding profile along chromosome 2 in *S. pombe.* The orange arrows represent genes transcribing either in the sense (above black line) or antisense (below black line) directions. Arrows pointing towards each other represent convergent gene pairs and those in the same direction are tandem genes. Regions with cohesin and Mis4 colocalization are the cohesive sites. Regions with cohesin only are less-cohesive. Regions without Mis4 and cohesin represent the negative control.

(B) Psm3 occupancy at the cohesive and less-cohesive sites. The cohesive sites (dis2 and srp1) demonstrate ~3 fold high levels of cohesin subunit Psm3 in comparison to less-cohesive sites (776, med8 and plh1). Psm3 enrichment is plotted relative to the negative control (n.c.) set at 1 (black horizontal line). Error bars represent SEM from two independent experiments.

(C) Psm3-GFP occupancy at the centromeric dg repeats used as a positive control.

dynamic (Gerlich et al. 2006). In such a scenario, low levels of cohesin at less-cohesive sites represent the dynamic pool, in contrast to stable cohesin at cohesive sites.

Although, the ChIP-chip dataset identifies Mis4/Rad21 peaks that span ≥2 kbp of DNA sequence, locus-specific ChIP-qPCR indicates otherwise. For example, cohesin peak number 158 (Chapter II, Table II.9.a and 9.b) spans ~4 kbp across the convergent gene locus. However, ChIP-qPCR using primers positioned at the respective gene promoters and intergenic region indicates massive accumulation of cohesin at the intergenic region only and not elsewhere along the 4 kbp ChIP-chip signal. Additionally, experimental variation arising due to differences in protocols or the use of different strains can be ruled out because cohesin occupancy at Mis4/Rad21 sites is highly reproducible. As a positive control, a strong signal of Psm3 was detected at the centromeric *dg* repeats, as centromeres harbor the bulk of cellular cohesin (Peters JM 2008) (Figure IV.5C).

IV.6 Eso1 DIFFERENTIATES BETWEEN COHESIVE AND LESS-COHESIVE SITES

Stoichiometric differences between the cohesin distribution at “cohesive” and “less-cohesive” sites are in accordance with FISH analysis. A plausible explanation stems from the requirement of Eso1 mediated stabilization of cohesin. To identify the Eso1 binding profile for cohesive and less-cohesive sites, Eso1 ChIP analysis was performed using a strain where Eso1 is endogenously tagged with GFP (Eso1-GFP). As shown in Figure IV.6B, Eso1 was enriched at the “cohesive” sites (cendg, dis2, srp1 and rpl3801), while the signals at “less-cohesive” sites (cut3, thi5, pla1 and med8) were similar or lower than the background. The negative control represents a region where no cohesin occupancy has been shown.

Taken together, these results demonstrate the existence of two, functionally different populations of cohesin during the G_2 phase of the cell cycle. While differences between Eso1 enrichment are evident, the overall signal was relatively low. In *S. cerevisiae*, Eco1 (Eso1 homolog) is phosphorylated by Clb2-Cdk1 after the S phase, followed by degradation via the proteasome pathway (Lyon et al. 2011). As a result, newly synthesized cohesin complexes are capable of loading onto chromosomes but are unable to establish sister chromatid cohesion. Such regulation of Eco1 levels restricts cohesion establishment to S phase and ensures a timely

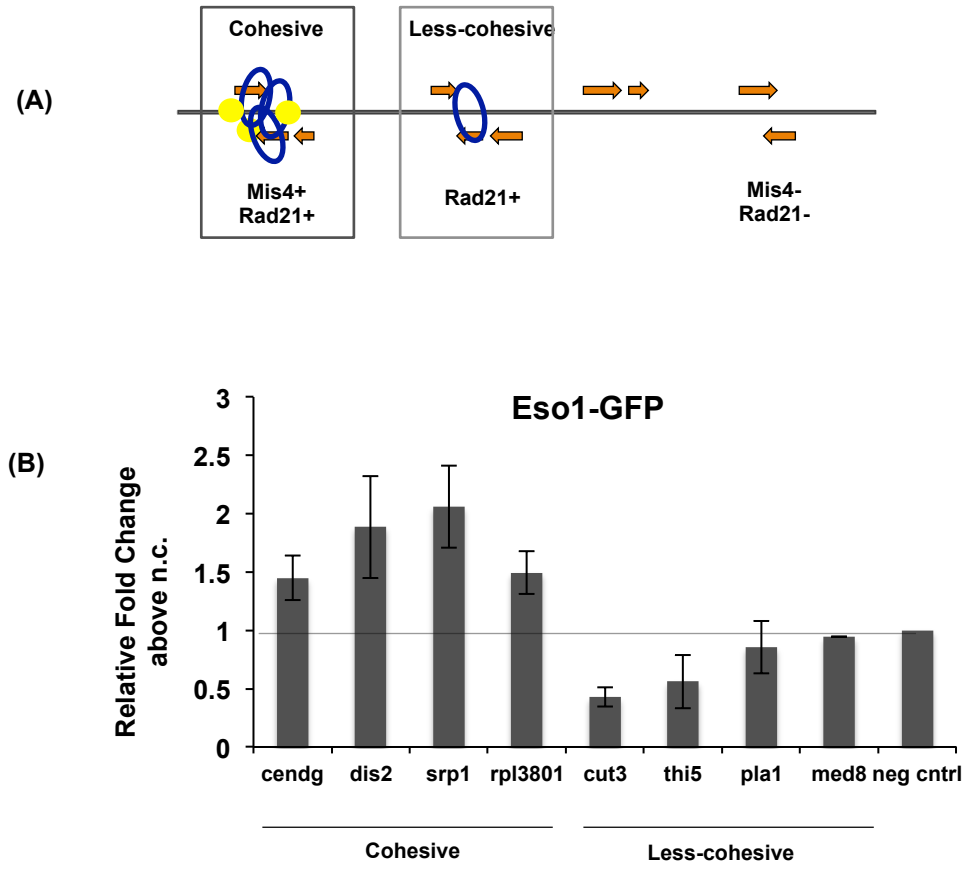


Figure IV.6 Eso1 enrichment at "cohesive" vs "less-cohesive" sites.

(A) Schematic representation cohesin (blue oval) and Mis4 (yellow circles) binding profile along chromosome 2 in *S. pombe*. The orange arrows represent genes transcribing either in the sense (above black line) or antisense (below black line) directions. Arrows pointing towards each other represent convergent gene pairs and in the same direction are tandem genes. Cohesive regions represent cohesin and Mis4 colocalization sites, while less-cohesive are bound by cohesin alone.

(B) Eso1 occupancy at the cohesive and less-cohesive sites. Eso1 ChIP analysis (G_2) was performed in a strain expressing Eso1 endogenously tagged with GFP (Eso1-GFP). The cohesive sites (cendg, dis2, srp1 and rpl3801) demonstrate slight enrichment for Eso1 binding in comparison to less-cohesive sites (cut3, thi5, pla1, med8). Eso1 signal is plotted relative to the negative control. The grey line indicates an arbitrary threshold comparing the the background signal as detected at the negative control. Error bars represent SEM from three independent experiments.

Figure IV.6 continued……

(C) **Eso1 is a Mis4-interacting protein**

PROTEIN	DESCRIPTION	%COVERAGE	Mis4-TAP	WT	Mis4-TAP/ WT
sp|Q02061| RPB2_SCHPO	RNAPII subunit Rpb2	5.30	3	3	1
sp|O42917| ESO1_SCHPO	**N-acetyltransferase-Eso1**	3.10	1	1	1
sp|O74825| RPB4_SCHPO	RNAPII subunit Rpb4	25.2	1	-	-
sp|P36594| RPB1_SCHPO	RNAPII subunit Rpb1	5.10	7	3	2.33

Figure IV.6 Eso1 enrichment at "cohesive" vs "less-cohesive" sites.

(C) Mass Spectrometry (MS) analysis of Mis4 interacting partners. MS analysis was performed using exponentially growing G_2 cells in a Mis4-TAP and a no TAP control strain. The potential interacting partners for Mis4 include RNAPII subunits Rpb1, Rpb2, Rpb3, and the cohesin acetyltransferase Eso1.

anaphase. Experimental evidence for Eso1 degradation in *S. pombe* is currently absent but this possibility cannot be discounted. Therefore, a rather prudent approach would be to analyze the binding pattern for Eso1 in S phase, where it is pivotal for cohesion establishment.

An obvious difference between "cohesive" and "less-cohesive" sites is the presence of Mis4/Ssl3. The Mis4/Ssl3 complex is analogous to the Scc2/Scc4 in *S .cerevisiae* and NIPBL/Mau2 in mammals (Nasmyth and Haering 2009). Scc2 is a large protein containing multiple HEAT repeats within its C-terminal domain and an Scc4 binding N-terminal extension (Neuwald et al. 2000; Panizza et al. 2000). In contrast, Scc4 is composed of TPR repeats (Watrin et al 2006). Both, HEAT and TPR repeats can act as binding platforms to facilitate multiple protein-protein interactions. To identify potential interacting partners for Mis4 and Ssl3, Mass Spectrometry (MS) analysis was performed using strains expressing Mis4-TAP or Ssl3-TAP, along with a wild type (no tag) control.

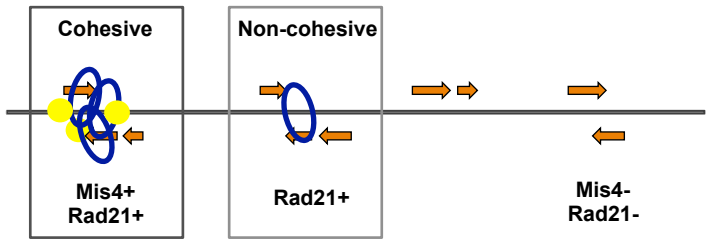
Interestingly, MS identified a novel interaction between Mis4 and the acetyltransferase Eso1 (Figure IV.6C). Eso1 is recruited at DNA replication fork assembly sites via its interaction with PCNA (Nasmyth and Haering 2009). These interactions were validated in two independent MS datasets. This implies that Mis4 performs two cellular functions, (i) loading cohesin complex onto chromosomes throughout the cell cycle and (ii) recruitment of Eso1 to chromosomal loading sites, which then stabilizes cohesin.

IV.7 "COHESIVE" SITES PREFERENTIALLY ASSOCIATE WITH ACTIVELY TRANSCRIBED GENES

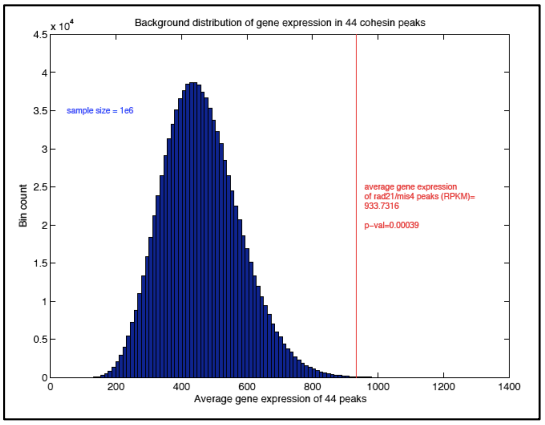
The bioinformatics analysis described in this section was performed by a fellow PhD student, Margarita Schlakow, Department of Mathematics at the University of Oxford.

The results described so far emphasize that cohesin at the loading sites is stably associated cohesin, while cohesin on its own is unstable and might be rapidly turned over. A characteristic feature of cohesin loading sites is their significant overlap with highly transcribed genes (Schmidt et al. 2009). Additionally, genes flanking cohesin-bound convergent sites are expressed at higher levels than genes

(A)



(B)



(C)

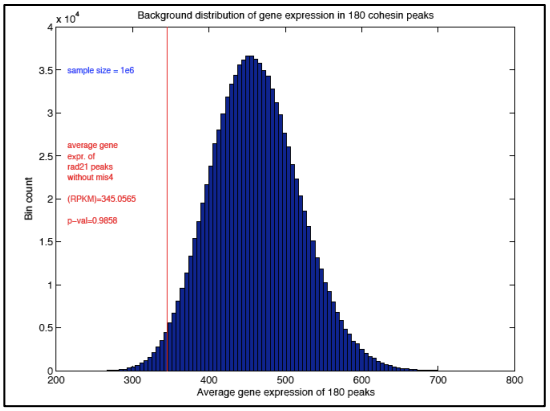


Figure IV.7 Cohesive sites (Mis4+/Rad21+) significantly overlap highly expressed genes.

(A) Schematic representation cohesin (blue oval) and Mis4 (yellow circles) binding profile along chromosome 2 in *S. pombe*. The orange arrows represent genes transcribing either in the sense (above black line) or antisense (below black line) directions. Arrows pointing towards each other represent convergent gene pairs and in the same direction are tandem genes. Cohesive regions represent cohesin and Mis4 colocalization sites and less-cohesive are bound by cohesin alone.

(B) A distribution comparing the average expression levels of genes overlapping with cohesive sites (Mis4+/Rad21+). The average value = 933.716 (p-value = 0.00039).

(C) A distribution comparing the average expression levels of genes overlapping with less-cohesive sites (Rad21+ only). The average value = 345.0568 (p-value=0.9858).

Figure IV.7 continued……

(D)

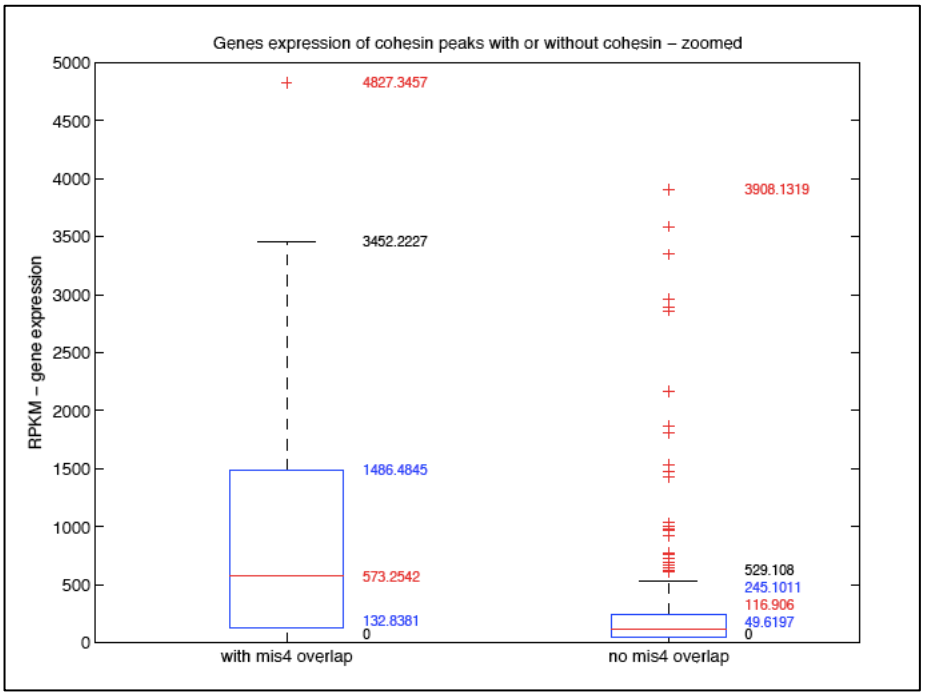


Figure IV.7 Cohesive sites (Mis4+/Rad21+) significantly overlap highly expressed genes.

(D) A box plot comparing the average gene expression of genes overlapping cohesive sites with those of less-cohesive sites. The median value (red horizontal line) of genes overlapping with Mis4+/Rad21+ is 573.2542, significantly higher than of genes overlapping with Rad21+ only, 116.906.

flanking cohesin-free convergent sites (Schmidt et al. 2009). However, a comparison between the expression level of genes that overlap with “cohesive” and “less-cohesive” sites has not been done so far.

Based on the ChIP-chip dataset, 44 out of 224 (excluding centromeric and telomeric regions) cohesin peaks overlap with Mis4 binding sites. Gene expression data that represents an average RPKM (reads per kilobase per million) score was obtained from two-strand specific RNA-Seq analysis of exponentially growing cells (Marguerat et al. 2012). Assuming M1 as the mean gene expression of “cohesive” sites and M2 as the mean of “less-cohesive” sites, the 224 peaks were randomly distributed into groups of 44 and 180 and the means for each group were computed. This was repeated $1x10^6$ times to obtain a distribution of the means. A p-value was obtained by dividing the number of means ≥ to M1 (or M2) with $1x10^6$.

As shown in Figure (IV), the average gene expression score for the “cohesive” sites, M1 = 933.736 (Figure IV.7B), is higher than the average gene expression for “less-cohesive” sites, M2 = 345.0565 (Figure IV.7C). This was further validated by a permutation test that calculates the difference between the two sites. The average expression for cohesive was higher than less-cohesive by a value of 588.6751 (p-value=$2.9e^{-05}$). Furthermore, a comparison of the median gene expression values (Figure IV.7D, red horizontal line) demonstrates a significantly high expression level for genes overlapping the cohesive sites. Note, the average and median values for a distribution are not the same. Thus, we can conclude that “cohesive” sites significantly overlap (p-value= 0.00039) with highly transcribed genes in comparison to “less-cohesive” sites (p-value=0.9858).

A curious question that requires further investigation is, “what defines a cohesin loading site? What attracts Mis4 to highly transcribed genes?” Is it the open chromatin structure or protein-protein interactions between cohesin complexes and the transcription machinery or underlying sequence specificity? Results described in chapter V shed some light upon the physical association of cohesin complexes with RNAPII, wherein the latter appears to be an essential player in the cohesin loading process.

IV.8 DISCUSSION

Several studies have highlighted the fact that cohesin exists in sub-sets, each with variable dwelling time on the chromosomes. In mammals, FRAP analysis indicates that cohesin maintains a dynamic state of association and dissociation in G_1, with a mean residence time <30 min. As cells enter G_2, the cohesin pool gets divided into soluble or weakly bound cohesin ($t_{1/2}$ <30 min) and stable cohesin, associating with chromosomes for >6 h (Gerlich et al. 2006; McNairn and Gerton 2009; Gause et al. 2010; Onn and Koshland 2011). Undoubtedly, it is the stable cohesin that persists until anaphase onset and is responsible for maintaining cohesion. In contrast, the soluble/dynamic cohesin might execute cohesion-independent functions, such as regulation of gene expression or act as a reservoir that can be rapidly stabilized in event of DNA damage or loss of chromosome biorientation. Whether colocalization with the kollerin complex is the underlying molecular determinant differentiating between soluble and stable cohesin has not been established so far.

However, for the very first time, using single-locus FISH analysis, I provide direct evidence that Mis4 is not only required for cohesin loading but it also facilitates cohesion establishment. Regions across chromosomal arms, where cohesin colocalizes with Mis4 (dis2, tif11, ubp9, srp1 and rpl3801) maintain stable sister chromatid cohesion, and hence “cohesive” cohesin. In contrast, cohesin without Mis4 (swi3, cut3, pla1, mug37, med8 and psh3) is unable to hold sister chromatids together and therefore, is “less-cohesive” cohesin. The Mis4/Ssl3 complex facilitates association of cohesin with DNA during late G_1 in budding yeast or end of telophase in vertebrate cells (Furuya et al. 1998; Ciosk et al. 2000; Rollins et al. 2004; Seitan et al. 2006; Bernard et al. 2006; Kogut et al. 2009). Unless otherwise stated, the cohesin-loading complex is now referred as “kollerin” (Nasmyth 2011). In *Xenopus*, kollerin is recruited to chromatin via the Cdc7/Drf1 kinase (DDK) component of the pre-replication complex (Gillespie and Hirano 2004; Takahashi et al. 2004). In *S. cerevisiae,* kollerin is enriched at centromeric and pericentromeric regions via kinetochore proteins. In *S. pombe*, kollerin is recruited at centromeric regions via heterochromatin binding protein Swi6/HP1 (Bernard et al. 2001; Nonaka et al. 2002). Interestingly, overlapping transcription at convergent gene loci results in double strand RNA (dsRNA) formation during G_1. The dsRNA is processed by the RNAi machinery, followed by RITS-induced heterochromatin formation and Swi6 recruitment. In turn, Swi6 facilitates cohesin

recruitment, which prevents read-through transcription and ensures transcription termination (Gullerova and Proudfoot 2008). Such conspicuous series of events provide a straightforward answer to the molecular distinction between cohesive and less-cohesive sites and are being currently investigated.

I also present bioinformatics data suggesting that cohesive sites overlap highly expressed RNAPII transcribed genes (Figure IV.7). The nucleosome-free promoter regions might appear as attractive sites for cohesin loading. Alternatively, the transcription machinery itself might facilitate recruitment of the kollerin and/or cohesin complexes to such loci. Results described in Chapter V shed light upon the connection between transcription and cohesion.

Both subunits of the kollerin complex are composed of repetitive protein sequences such as HEAT and TPR repeats. HEAT repeats have also been identified in *Xenopus* condensin and *Drosophila* Nipped-B (Neuwald and Hirano 2000). Additionally, HEAT repeats are present in TBP-associated protein TIP120 and Mot1 component of the SWI/SNF2 remodeling complex. HEAT repeat proteins may provide a scaffold for assembly of the machinery involved in chromosome dynamics, such as the SMC complexes. Similarly, TPR-containing proteins are involved in a diverse array of cellular processes, such as cell cycle control, transcription, splicing, protein transport and protein folding (Blatch and Lassle 1999). Therefore, the kollerin complex is inherently capable of interacting with nucleosomes, chromatin remodeling complexes, transcription factors and most definitely the cohesin and condensin complexes. The importance of kollerin is also highlighted by the fact that mutations in the human homolog (*NIPBL*) are a leading cause of Cornelia deLange Syndrome (CdLS) (Liu and Krantz 2008). CdLS is a genetic dominant disorder, with an occurrence of 1:10,000. Patients show growth delay and neuro-developmental retardation. Extreme cases develop upper extremity malformations, hirsutism, dysmorphism and other dysfunctions that affect cardiac, renal and gastroesophageal systems. Pathogenic mutations in *NIPBL* account for ~60% cases of classical CdLS, while ~5% cases have mutations in the *SMC1A* and *SMC3* subunits (Deardorff et al. 2007; Wierzba et al. 2011; Liu et al. 2009; Strachan T 2005). Surprisingly, CdLS patients do not show severe cohesion defects, pointing towards a yet uncharacterized function of kollerin in development.

Given the critical importance of kollerin, I performed MS analysis to identify potential interacting partners of Mis4. Duplicate MS datasets detected Mis4 in

association with Eso1 (Eco1 in *S. cerevisiae*, ESCO1/2 in humans). Eso1 is crucial for replication-coupled establishment of sister chromatid cohesion. Eso1 interacts with DNA replication components such as PCNA, RF-C, DNA helicase Chl1 and the Okazaki fragment maturation flap endonuclease I (Fen1) (Skibbens et al. 1999; Mayer et al. 2004; Skibbens et al. 2004; Petronczki et al. 2004; Rudra and Skibbens 2012). However, cohesion establishment can also occur independent of replication in events of DNA damage (Heidinger-Pauli et al. 2008). During replication, Eso1 acetylates K112 and K113 residues of the Smc3 subunit; as a result, the binding of cohesin to DNA becomes resistant to the destabilizing effect of the cohesin-associated Wapl protein (Rudra and Skibbens 2013). In vertebrates, Smc3 acetylation promotes the recruitment of soronin to the cohesin complex, which is thought to antagonize Wapl (Nishiyama et al. 2010). However, Smc3 acetylation can also facilitate the recruitment of accessory factors required for chromatid tethering (Rudra and Skibbens 2013).

The timing of Mis4-Eso1 interaction during the cell cycle is of some importance. While the Mis4 binding pattern is identical during G_1 and G_2, a genome-wide Eso1 binding profile has not been established so far (Schmidt et al. 2009). Given the essential nature of Eso1 during S-phase, it is possible that Mis4 and Eso1 interact during S-phase but how? Does Eso1 preferentially associate with a population of Mis4, thereby stabilizing a sub-set of cohesin? Alternatively, this interaction might reflect proximity between Mis4 and Eso1 chromosomal binding sites. In *S. cerevisiae*, genome-wide analysis identified 96 sites with characteristic enrichment of DNA-polymerase that overlap with the open reading frames of highly transcribed genes (Azvolinsky et al. 2009; Tuduri et al. 2009). Furthermore, genome-wide mapping in metazoans shows that transcribed regions are replicated early in S-phase (Schubeler et al. 2002; White et al. 2004; Woodfine et al. 2004). The open chromatin structure associated with active transcription probably facilitates replication-fork assembly as replication origins are usually enriched in promoter regions and DNaseI hypersensitive sites (Cardoret et al. 2008; Cyrou et al. 2011; Martin et al. 2011). As mentioned earlier, bioinformatics analysis shows that cohesive peaks significantly overlap with highly transcribed genes. Taken together, these observations suggest a chronological series of events involved in cohesin loading. During G_1, Mis4 associates with highly transcribed gene loci, where the nucleosome-free architecture facilitates replication fork assembly. Eso1 is recruited to the replication fork either via interacting partners described above or via Mis4 itself. Eso1 then acetylates Smc3, stabilizing the

association of cohesin onto chromatin till anaphase. A complete understanding of the timing and contribution of Mis4-Eso1 interaction requires further investigation.

Another interesting question that requires further study is, “what is the function of soluble cohesin?” Apart from cohesion, cohesin is also involved in transcription regulation and DNA damage repair via homologous recombination. Experiments outlined below can help delineate whether less-cohesive cohesin indeed performs any of the above functions. Chromosome spreads demonstrate that about half of chromosomal cohesin is cleaved and removed during anaphase, while a sub-pool persists (Schmidt et al. 2009). ChIP analysis of cells synchronized in mitosis by using the cold sensitive β-tubulin *nda3-KM311* mutant (Hiraoka et al. 1984), followed by release into anaphase, would help identify whether less-cohesive cohesin represents the sub-pool that remains after anaphase. Additionally, inducing DNA damage during G_2 by ultraviolet radiation or methyl methanesulphonate (MMS) or bleomycin, could reinforce genome-wide cohesion establishment. If damage-induced cohesin coincides with the less-cohesive sites (analyzed by ChIP and FISH), this would argue for a DNA repair function of less-cohesive cohesin. Damage induced (DI) cohesin loading can also be detected by ChIP analysis for Rad21 phosphorylation and SUMOlyation (subject to antibody availability), a known mark of DI cohesion, unlike Smc3 acetylation that coincides with S-phase cohesion (Heidinger-Pauli et al. 2008; Rolef Ben-Shahar et al. 2008; Unal et al. 2008; Heidinger-Pauli et al. 2009). Furthermore, DI cohesion at less-cohesive sites would be enriched for γH2AX and repair proteins such as Nse2 (Rudra and Skibbens 2013).

Further support for a repair/recombination-related function for less-cohesive cohesin comes from ChIP-chip analysis. Cohesin exhibits an ordered arrangement across chromosome 2. The mean distance between cohesin binding sites along chromosome 2 is ~13.1 kbp and the median distance is 16.5 kbp (Schmidt et al. 2009). This suggests an underlying principle for the distribution of cohesin along chromosomes, possibly to maintain proximity between sister chromatids in order to facilitate efficient double strand break repair via homologous recombination. In fact, Rad21 was first identified as playing a role in DNA repair when fission yeast Rad21 mutants were found to be sensitive to γ-radiation and defective in repair (Birkenbhil and Subramani 1992).

To summarize, I described experiments that characterize the functional difference between cohesin population associated with or without Mis4. My results uncover two different features of the cohesin complex. First, cohesin at its loading site is stable cohesin. Second, cohesin at less-cohesive sites might have a cohesion-independent role in DNA damage response. Additionally, cohesive sites preferentially overlap with highly transcribed genes, implicating either the transcription process or the transcription machinery in the cohesin cycle. I provide new insight into transcription and cohesion in the following Chapter V.

CHAPTER V

COHESIN AND TRANSCRIPTION: A REQUIREMENT OF RNAPII FOR CHROMOSOMAL ASSOCIATION OF COHESIN

While cohesin is well known for its role in sister chromatid cohesion, current evidence argues for a cohesion-independent function of cohesin in the regulation of gene expression. Mutations in the *Smc1* and *SA* genes block ecdysone steroid signaling, axon pruning and dendrite formation in *Drosophila* neurons (Pauli et al. 2010; Schuldiner et al. 2008). Cohesin depletion in non-dividing mouse thymocytes results in reduced transcription and rearrangements at the T cell receptor α locus, thereby affecting thymocyte differentiation (Seitan et al. 2011). Mutations in *NIPBL* are the leading cause of CdLS, where hundreds of genes are dysregulated in patient cell lines without any cohesion defects (Muto et al. 2011; Krantz et al. 2004; Liu et al. 2009). In zebrafish, cohesin is a positive regulator of *runx* expression that functions in cellular differentiation (Horsfield et al. 2007).

In mammals, cohesin (along with CTCF) is considered to be a regulator of gene expression primarily as an organizer of interphase chromatin. Cohesin facilitates long-range contacts between enhancer and promoter regions of pluripotency genes, in association with NIPBL and the mediator complex (Kagey et al. 2010). Direct evidence from knockdown experiments implicates CTCF and cohesin in maintaining long-range interactions at a variety of loci. These include the imprinted *Igf2/H19* locus (Nativio et al. 2009), Interferon gamma locus *IFNG* (Hadjur et al. 2009), *β-globin* locus (Chien et al. 2011), *APO* gene cluster (Mishiro et al. 2009) and *MHC* class II cluster (Maunder et al. 2011). In a CTCF-independent manner, cohesin also localizes with tissue-specific transcription factors (Schmidt et al. 2010). How cohesin mediates such contacts is unclear, but it presumably entraps regulatory elements in *cis*, in a manner related to the way it holds together sister chromatids in *trans.*

A global organization for the *S. pombe* genome has been suggested recently (Tanizawa et al. 2010). 3C-based methods demonstrate significant association among highly transcribed genes that frequently contain the same DNA motifs at their promoter regions. The authors propose the existence of “transcription factories” in yeast, analogous to transcription factories in mammals. While a CTCF-like protein has not been identified in yeast, cohesin and kollerin might perform similar functions. On the other hand, cohesin physically interacts with transcription factor NFκB and stimulates target gene expression (Lara-Pezzi et al. 2004). Furthermore, cohesin functions in transcription termination via heterochromatin formation (Gullerova and Proudfoot 2008). In this chapter, I describe experiments aimed at understanding

whether cohesin and kollerin directly influence localized transcription initiation or elongation. I have gone on to investigate the influence of transcription on the chromosomal association of cohesin and kollerin. While previous studies indicate that active transcription acts to push cohesin on chromosomes, I have now identified a specific requirement of RNAPII for the association of kollerin and cohesin on chromosome arms. I begin this chapter by focusing on the effect of cohesin and kollerin on RNAPII transcription and then follow on by investigating the affect of transcription inhibition on cohesin.

V.1 RNAPII OCCUPANCY IN *rad21ts* AND *mis4ts.*

To identify a function for cohesin and accessory proteins in transcriptional regulation, I employed temperature sensitive alleles of the cohesin subunit Rad21 and cohesin loading protein Mis4. *rad21-K1ts* (Tatebayashi et al. 1998) and *mis4-367ts* (Bernard et al. 2006) mutant strains were grown overnight at the permissive temperature (25°C) and then shifted to the restrictive temperature (37°C) for 2 h. Subsequently, cells were cross linked and processed for ChIP to assess RNAPII chromatin occupancy. Previous studies have investigated the effect of cohesin on transcription in G_1 synchronized cells using Rad21 or Mis4 *ts* mutants (Dheur et al. 2011). However, Rad21 levels in G_1 are low because of proteasome-mediated degradation of Rad21 during anaphase (Nasmyth 2011). Moreover, the G_1 phase of the fission yeast cell cycle is very short (Luche and Forsburg 2009). Additionally, Rad21 and Mis4 are essential during the S phase. Therefore, I chose exponentially growing cells present in the G_2 phase of the cell cycle. Simultaneously, aliquots were fixed in 70% ethanol, DAPI stained and examined microscopically to identify premature cohesion defects due to the temperature shift. qRT-PCR primers were designed specifically in the promoter regions of highly transcribed genes that associate with (i) both cohesin and Mis4, these include the 60S ribosomal subunit encoding gene *rpl29*, protein phosphatase *dis2* and polypeptide-associated complex alpha subunit *egd2*; (ii) only Mis4, 60S ribosomal subunit gene *rpl1701* and (iii) only cohesin, DNA-directed RNA polymerase I complex large subunit *nuc1* (Schmidt et al. 2009). A primer positioned in a non-transcribed region was used as the negative control.

After shifting to 37°C, *rad21-K1ts* cells exhibited abnormal chromosome structures (Figure V.1A, right panel) such as condensed chromosomes, unequal

chromosome segregation and lagging chromosomes (white arrowheads). Similarly, the *mis4-367ts* cells were also defective in chromosome segregation (Figure V.2A, right panel). In contrast, the wild type cells showed punctate nuclei without any cohesion defects. In wild type and *ts* mutants, genes encoding ribosomal proteins displayed high levels of RNAPII occupancy in comparison to other mRNA genes (dark bars in Figure V.1B & V.1C; and Figure V.2B &V.2C). In both *ts* mutants, RNAPII occupancy increased between 1.2-2.5 fold, with maximum change associated with the highly transcribed ribosomal protein coding genes *rpl29* and *rpl1701* (Figure V.1C and V.2C). No significant change was associated with RNAPII occupancy at *egd2* and *nuc1* promoters. However, *dis2* promoter showed a ~1.6 fold increase in RNAPII occupancy in *rad21-K1ts* and ~2.5 fold increase in *mis4-367ts* mutant background. This marginal increase may suggest that cohesin depletion results in RNAPII pausing at highly transcribed gene loci. However, post-shift to 37°C, RNAPII occupancy was also increased in the wild type strain (Figures V.1B and V.2B). This negates a role for cohesin in RNAPII pausing at promoter regions. The increase in RNAPII occupancy at gene promoters across all the three strains most definitely indicates a high rate of RNAPII transcription at 37°C in comparison to 25°C and therefore, high levels of RNAPII occupancy at gene promoters.

Histone H3 binding was analyzed as an internal control in wild type cells and the *ts* mutants. The H3 profile was remarkably similar in the wild type (Figure V.1D and V.2D), *rad21-K1ts* (Figure V.1E) and *mis4-367ts* (Figure V.2E) between the permissive (25°C) and restrictive (37°C) temperatures.

V.2 LOSS OF COHESIN AFTER INHIBITION OF TRANSCRIPTION.

In a reciprocal experiment, I analyzed the cohesin occupancy profile after inhibition of RNAPII transcription with thiolutin. Thiolutin is an antibiotic derived from *Streptomyces luteoreticuli* (Calmer and Solomon 1955), which inhibits all three RNA polymerases, primarily at the level of initiation (Tipper DJ 1973). In a preliminary experiment (Figure V.3A), exponentially growing wild type cells were treated with thiolutin at a concentration of 20 µg/ml (Tipper DJ 1973, Tang-Wong et al. 2012) and total cellular RNA was isolated at 15 min intervals for 1 h. qRT-PCR was performed using gene-specific primers for *dis2* and *act1*. Raw qRT-PCR values were first normalized to the mitochondrial *cox* transcript, as the mitochondrial RNA polymerase is not targeted by thiolutin. Subsequently, the values were plotted relative to

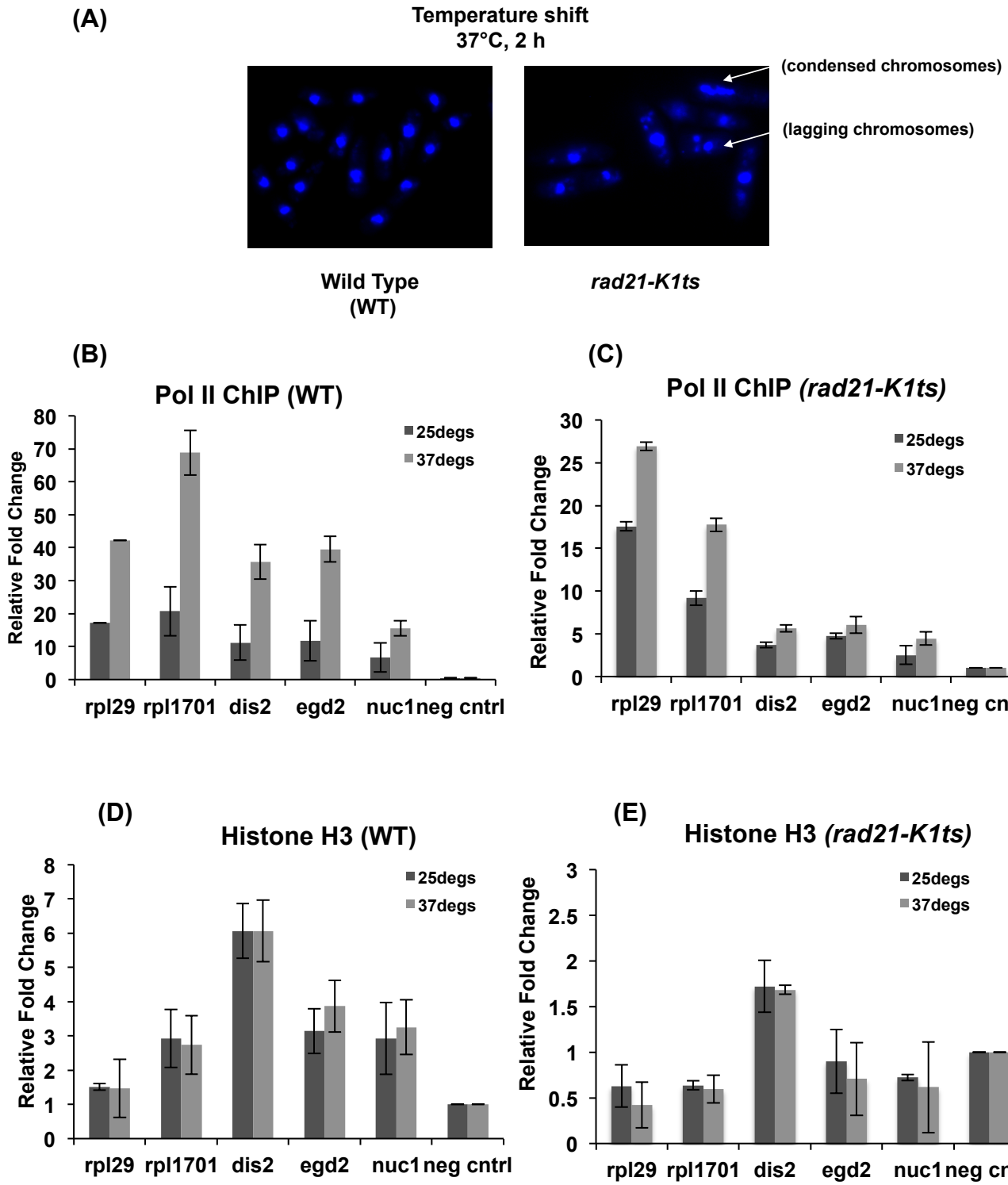

(A)
Temperature shift
37°C, 2 h
(condensed chromosomes)
(lagging chromosomes)
Wild Type
(WT)
rad21-K1ts
(B)
Pol II ChIP (WT)
Relative Fold Change
25degs
37degs
rpl29
rpl1701
dis2
egd2
nuc1
neg cntrl
(C)
Pol II ChIP (rad21-K1ts)
(D)
Histone H3 (WT)
(E)
Histone H3 (rad21-K1ts)

Figure V.1 RNAPII occupancy in wild type and *rad21-K1ts.*

(A) DAPI stained nuclei depicting wild type (left) and *rad21-K1ts* (right) cells after a shift from 25°C to 37°C for 2 h. WT nuclei are punctate but *rad21-K1ts* shows chromosome segregation defects (white arrows), indicating inactivation of Rad21.

(B) RNAPII ChIP analysis in exponentially growing wild type cells at 25°C (dark bars) and 37°C (light bars) using anti-RNAPII 8WG16 antibody. ChIP enrichment is plotted relative to a non transcribed region as the negative control. The horizontal axis denotes the promoter regions of genes where RNAPII occupancy was analyzed. Error bars = SEM from two independent experiments.

(C) RNAPII ChIP analysis in exponentially growing *rad21-K1ts* cells at 25°C (dark bars) and 37°C (light bars) using anti-RNAPII 8WG16 antibody. ChIP enrichment is plotted relative to a non transcribed region as the negative control. The horizontal axis denotes the promoter regions of genes where RNAPII occupancy was analyzed. Error bars = SEM from two independent experiments.

(D) Histone H3 ChIP analysis in exponentially growing wild type cells at 25°C (dark bars) and 37°C (light bars) using anti-H3 (Abcam) antibody. ChIP enrichment is plotted relative to a non transcribed region as the negative control. The horizontal axis denotes the promoter regions of genes where RNAPII occupancy was analyzed. Error bars = SEM from two independent experiments. The negative control (neg cntrl) is just to compare H3 occupancy before and after shift. It does not imply a region negative for H3 binding.

(E) Histone H3 ChIP analysis in exponentially growing *rad21-K1ts* at 25°C (dark bars) and 37°C (light bars) using anti-H3 (Abcam) antibody. ChIP enrichment is plotted relative to a non transcribed region as the negative control. The horizontal axis denotes the promoter regions of genes where RNAPII occupancy was analyzed. Error bars = SEM from two independent experiments. The negative control (neg cntrl) is just to compare H3 occupancy before and after shift. It does not imply a region negative for H3 binding.

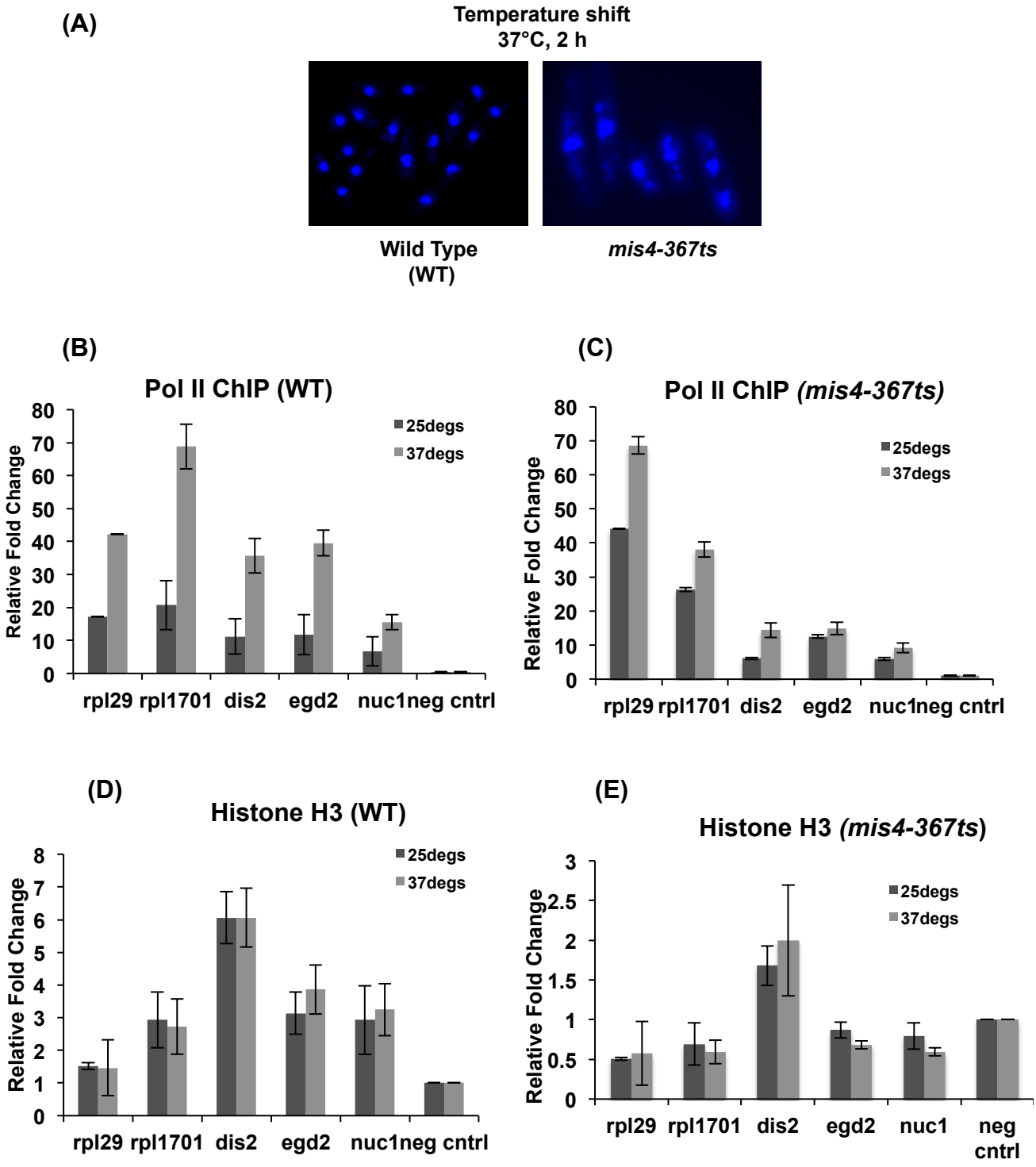

(A)
Temperature shift
37°C, 2 h
Wild Type
(WT)
mis4-367ts
(B)
Pol II ChIP (WT)
Relative Fold Change
25degs
37degs
rpl29 rpl1701 dis2 egd2 nuc1neg cntrl
(C)
Pol II ChIP (mis4-367ts)
Relative Fold Change
25degs
37degs
rpl29 rpl1701 dis2 egd2 nuc1neg cntrl
(D)
Histone H3 (WT)
Relative Fold Change
25degs
37degs
rpl29 rpl1701 dis2 egd2 nuc1neg cntrl
(E)
Histone H3 (mis4-367ts)
Relative Fold Change
25degs
37degs
rpl29 rpl1701 dis2 egd2 nuc1 neg cntrl

Figure V.2 RNAPII occupancy in wild type and *mis4-367ts.*

(A) DAPI stained nuclei depicting wild type (left) and *mis4-367ts* (right) cells after a shift from 25°C to 37°C for 2 h. WT nuclei are punctate but *mis4-367ts* shows chromosome segregation defects (white arrows), indicating inactivation of Mis4.

(B) RNAPII ChIP analysis in exponentially growing wild type cells at 25°C (dark bars) and 37°C (light bars) using anti-RNAPII 8WG16 antibody. ChIP enrichment is plotted relative to a non transcribed region as the negative control. The horizontal axis denotes the promoter regions of genes where RNAPII occupancy was analyzed. Error bars = SEM from two independent experiments.

(C) RNAPII ChIP analysis in exponentially growing *mis4-367ts* cells at 25°C (dark bars) and 37°C (light bars) using anti-RNAPII 8WG16 antibody. ChIP enrichment is plotted relative to a non transcribed region as the negative control. The horizontal axis denotes the promoter regions of genes where RNAPII occupancy was analyzed. Error bars = SEM from two independent experiments.

(D) Histone H3 ChIP analysis in exponentially growing wild type cells at 25°C (dark bars) and 37°C (light bars) using anti-H3 (Abcam) antibody. ChIP enrichment is plotted relative to a non transcribed region as the negative control. The horizontal axis denotes the promoter regions of genes where RNAPII occupancy was analyzed. Error bars = SEM from two independent experiments. The negative control (neg cntrl) is just to compare H3 occupancy before and after shift. It does not imply a region negative for H3 binding.

(E) Histone H3 ChIP analysis in exponentially growing *mis4-367ts* at 25°C (dark bars) and 37°C (light bars) using anti-H3 (Abcam) antibody. ChIP enrichment is plotted relative to a non transcribed region as the negative control. The horizontal axis denotes the promoter regions of genes where RNAPII occupancy was analyzed. Error bars = SEM from two independent experiments. The negative control (neg cntrl) is just to compare H3 occupancy before and after shift. It does not imply a region negative for H3 binding.

untreated cells. Nearly 50% reduction in *dis2* mRNA was visible at t = 30 min (Figure V.3A), while *act1* levels dropped only after t = 45 min. Therefore, 45 min was selected as the time point for subsequent experiments.

Thiolutin has also been suggested to inhibit translation (Jimenez et al. 1973, Tipper DJ 1973). Therefore, total RNAPII levels were assessed by western blot analysis with anti-RNAPII (8WG16) antibody. As shown in Figure V.3B, RNAPII levels remain unchanged based on a histone H3 normalization control. Furthermore, GFP fluorescence intensity was unaffected between untreated and treated cells (Figure V.3C), in a strain where the cohesin subunit Rad21 is endogenously tagged with GFP (Rad21-GFP strain).

I next investigated the binding profiles of Rad21 at different positions along chromosome 2 after thiolutin treatment. Exponentially growing (G_2) cells expressing endogenously tagged Rad21-GFP were treated with thiolutin for 45 min and processed for ChIP. As shown in Figure V.4B, thiolutin resulted in a substantial drop in RNAPII binding at the *rpl29* and *egd2* promoters, indicating efficient inhibition of transcription. Cohesin binding was analyzed at five regions overlapping a “cohesive” site (Mis4+/Rad21+) along chromosome 2, (i) centromeric degenerate repeats (dg), (ii) sub-telomeric region (telo), (iii) intergenic region between a convergent gene pair (srp1), (iv) *rpl29* promoter and (v) *egd2* promoter. While only background levels were detected at promoter regions, significant enrichment of Rad21 was observed at the centromeric *dg* repeats and telomeric region (Figure V.4A). In most organisms, bulk cohesin is indeed concentrated at the pericentromeric regions and telomeres (Lengronne et al. 2004; Schmidt et al. 2009; Hu et al. 2011). In agreement with the ChIP-chip data (Schmidt et al. 2009), Rad21 accumulated at the intergenic regions between convergent gene pairs (srp1). Surprisingly, transcription inhibition resulted in ~4 fold reduction in cohesin binding along the chromosome arm (srp1), with only a slight decrease (~1.5 fold) at the centromeric and telomeric regions (Figure V.4A). Note that only one-third of cellular cohesin is required for cohesion and this population is stabilized during S-phase (Heidinger-Pauli et al. 2010). Therefore, the drop in cohesin must represent the dynamic cohesin, which is continually being loaded by Mis4 at these loci in G_2. Secondary effects due to chemical treatment were ruled out since the histone H3 occupancy (Figure V.4C) remained unchanged between treated and untreated cells.

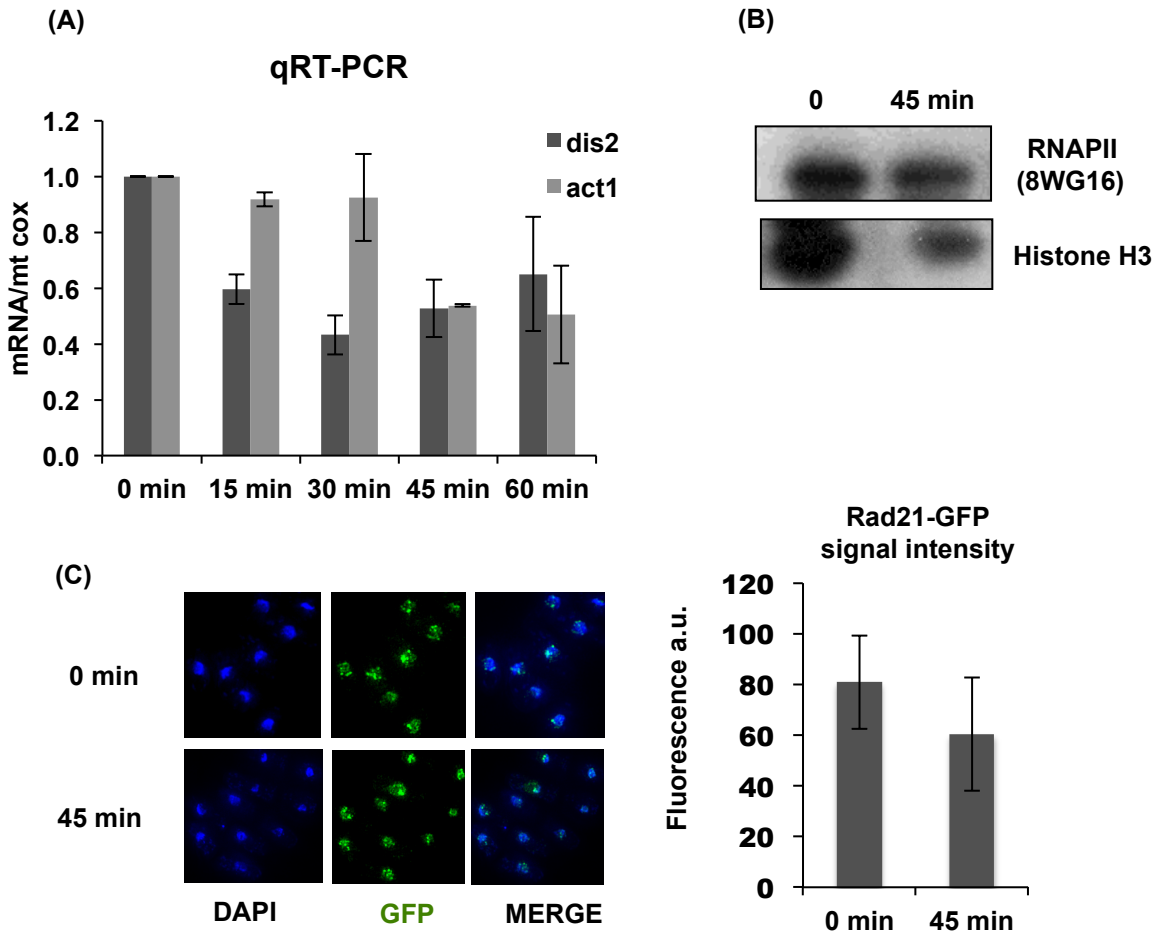


Figure V.3 RNAPII transcription inhibition by thiolutin.

(A) Reverse transcription and qPCR analysis of *dis2* and *act1* mRNA levels after thiolutin treatment. G_2 cells were treated with 20 µg/ml of thiolutin and samples were collected at 15 min intervals till 1 h. RNA was isolated using hot-phenol, DNaseI (Roche) treated and reverse transcribed using Superscript-SIII (Invitrogen). First strand cDNA was quantified using qPCR. The raw signals were first normalized to the mitochondrial *cox* transcript and then plotted relative to t = 0 min. While *dis2* RNA dropped to ~50% levels at t = 30 min, *act1* levels reduced at t = 45 min. Therefore, 45 min was selected as the suitable time for thiolutin treatment for subsequent experiments.

(B) Western Blot to confirm that thiolutin does not result in protein degradation. Whole cell extracts were prepared from wild type G_2 cells, boiled with SDS-gel loading dye and separated on a 4-15% bis-tris precast gel. The proteins were transferred to a nitrocellulose membrane and RNAPII levels were detected with anti-RNAPII antibody 8WG16. Histone H3 (Abcam) was used as a loading control. No change was observed in the protein levels between 0 and 45 min after thiolutin treatment.

(C) Microscopic visualization of Rad21-GFP intensity after thiolutin treatment. Cycling cells (G_2) were fixed in 20% methanol for 20 min. Nuclei were stained with DAPI (0.2 µg/ml) and cells were visualized using the DeltaVision microscope. Images were captured using Softworx, deconvoluted and subsequently, processed and quantified using ImageJ (NIH). The adjacent graph depicts that Rad21-GFP fluorescence was unaffected within the experimental set up.

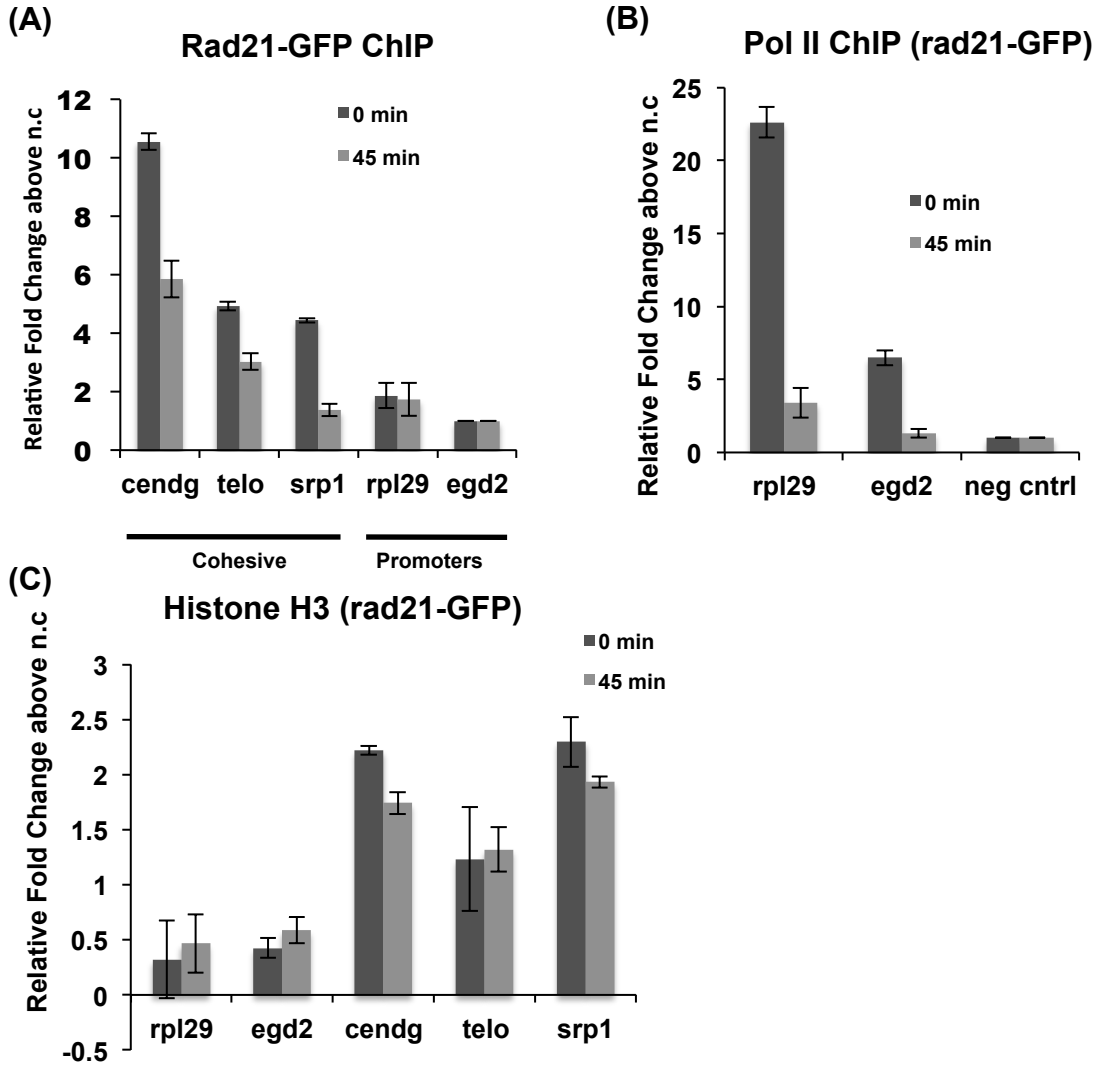


Figure V.4 Transcription inhibition results in reduction in cohesin levels.

(A) Cycling cells (G_2) were treated with thiolutin and processed for ChIP analysis. Rad21-GFP levels were detected with anti-GFP antibody (Abcam). Rad21 occupancy was analyzed at centromeric *dg* repeats, sub-telomeric regions, chromosomal arm srp1 and gene promoters *rpl29* and *egd2*. Centromeric and telomeric cohesin reduced ~1 fold, while cohesin on the arms reduced by ~4 fold. Promoter regions showed very low levels of cohesin which remained unaffected after transcription inhibition. Error Bars = SEM from two independent replicates.

(B) Transcription inhibition was assessed by RNAPII occupancy (8WG16). RNAPII levels significantly reduced at *rpl29* and *egd2* (t = 45 min) promoter regions coinciding with thiolutin treatment. Error Bars = SEM from two independent replicates.

(C) H3 levels (Abcam) were measured as a control between 0 and 45 min. Error Bars = SEM from two independent replicates. The negative control (neg cntrl) is just to compare the H3 occupancy before and after thiolutin treatment. It does not imply a region negative for H3 binding.

The above data suggests a striking interplay between RNAPII transcription and the association of cohesin with chromosomes. Based on genome-wide mapping of cohesin complexes, the view has been that convergent transcription repositions cohesin from its loading site to intergenic regions (Lengronne et al. 2004; Glynn et al. 2004; Schmidt et al. 2009). In such a scenario, a decrease in polymerase should coincide with accumulation of cohesin at the promoter regions. However, Rad21 signal at the *rpl29* and *egd2* promoters (Figure V.4A) did not change between the two time points, arguing instead for a mechanistic connection between cohesin loading and RNAPII.

V.3 LOSS OF MIS4 BINDING AFTER TRANSCRIPTION INHIBITION.

The association of cohesin with chromatin depends on a separate loading complex known as kollerin (Nasmyth 2011). In fission yeast, kollerin consists of two essential subunits, Mis4 and Ssl3 (Bernard et al. 2006). Kollerin binding and ATP hydrolysis at the NBDs is thought to introduce conformational changes in the cohesin ring that result in transient opening of the hinge domain, thereby forming a DNA "entry gate" (Gruber et al. 2006; Mishra et al. 2010; Kurze et al. 2011; Hu et al. 2011). Mis4/Ssl3 have been shown to bind highly transcribed genes (Schmidt et al. 2009). Since my results imply that cohesin loading is transcription-dependent, I next investigated whether Mis4/Ssl3 association with chromosomes also requires RNAPII. To address this, I performed ChIP analysis in a strain where Mis4 was endogenously tagged with GFP (Mis4-GFP) after thiolutin treatment.

Mis4 binding was analyzed at five regions overlapping a "cohesive" site (Mis4+/Rad21+) along chromosome 2, (i) centromeric degenerate repeats (dg), (ii) sub-telomeric region (telo), (iii) intergenic region between a convergent gene pair (srp1), (iv) *rpl29* promoter, and (v) *egd2* promoter. In an unexpected manner, Mis4 binding onto chromatin (Figure V.5.A) was dramatically reduced after RNAPII inhibition (V.5B). Mis4 signal at the promoter regions (*rpl29* and *egd2*) was relatively low at t = 0 min in comparison to centromeric *dg*, telomeric and chromosomal (srp1) regions (Figure V.5A, dark bars). At t = 45 min, Mis4 levels at the promoter were still low but Mis4 chromatin occupancy reduced drastically at the centromere, telomeric and arm regions (Figure V.5A, light bars). So far, there is no evidence towards a requirement of RNAPII for the association of Mis4 and cohesin onto chromatin. Thiolutin inhibits transcription initiation, suggesting that RNAPII recruits Mis4 alone or

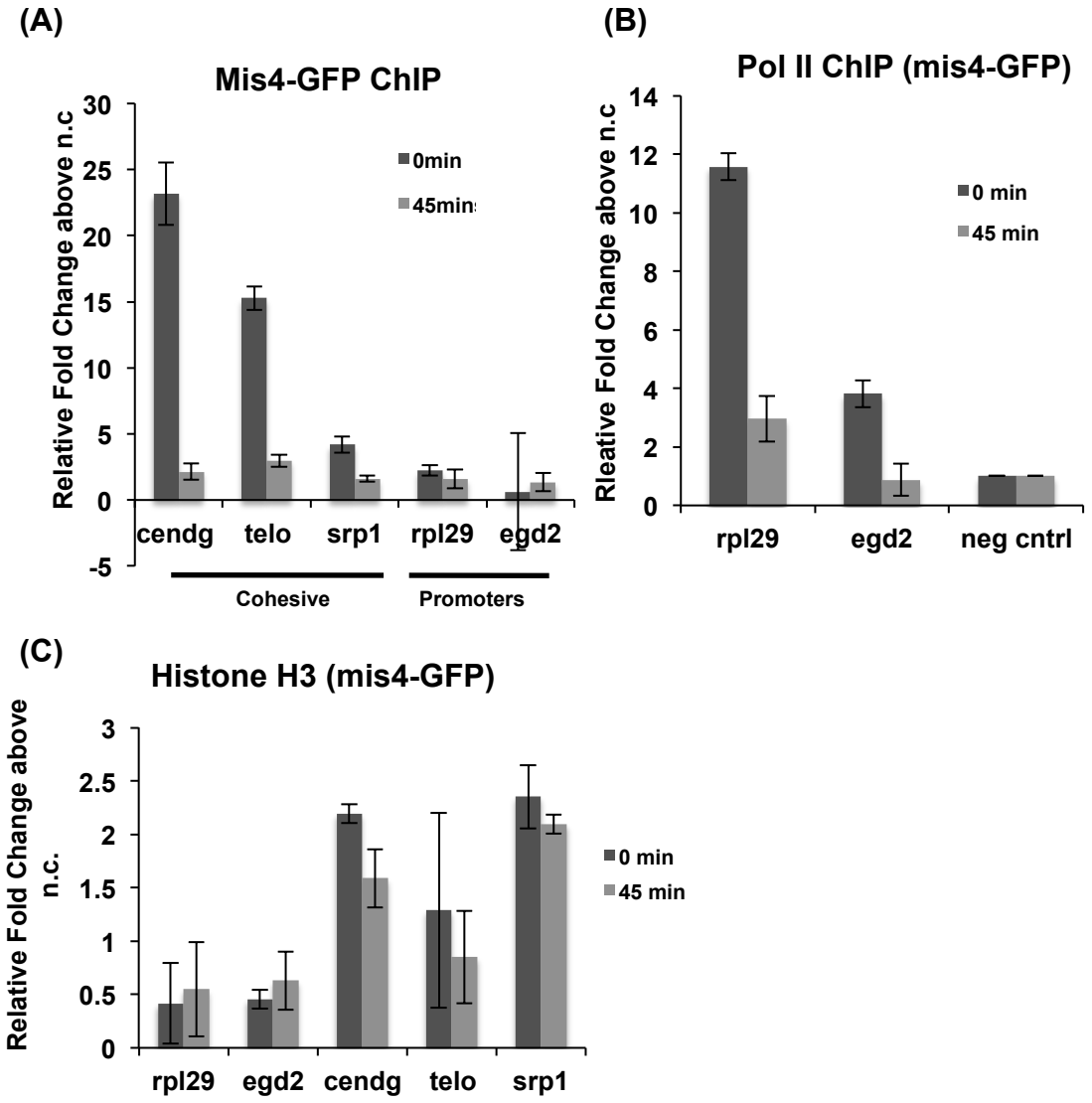


Figure V.5 Transcription inhibition results in reduction in Mis4 levels.

(A) Cycling cells (G2) were treated with thiolutin and processed for ChIP analysis. Mis4-GFP levels were detected with anti-GFP antibody (Abcam). Mis4 occupancy was analyzed at centromeric *dg* repeats, sub-telomeric regions, chromosomal arm srp1 and gene promoters *rpl29* and *egd2*. Centromeric and telomeric cohesin reduced ~10 fold, while cohesin on the arms reduced by ~5 fold. Promoter regions showed very low levels of cohesin, which was unaffected after transcription inhibition. Error bars = SEM two independent replicates.

(B) Transcription inhibition was assessed by RNAPII occupancy (8WG16). RNAPII levels significantly reduced at *rpl29* and *egd2* (t = 45 min) promoter regions, coinciding with thiolutin treatment. Error bars = SEM two independent replicates.

(C) H3 levels (Abcam) were measured as a control between 0 and 45 min. The negative control (neg cntrl) is just to compare the H3 occupancy before and after thiolutin treatment. It does not imply a region negative for H3 binding.

the Mis4-cohesin complex onto chromatin. In such a scenario, RNAPII may exhibit physical interaction with Mis4 and cohesin. Further results support this hypothesis and are discussed in the following sections. Additionally, Histone H3 occupancy remained unaffected (Figure V.5C).

V.4 TRANSCRIPTION INHIBITION IMPAIRS COHESIN BINDING IN A *Δwapl* BACKGROUND.

During G_1 and G_2 in mammals, a fraction of cohesin associates with chromosomes in a transient manner with a mean residence time of ~25 min (Gerlich et al. 2006). This instability is attributed to the "anti-establishment" activity of the *Wpl* gene product (Sutani et al. 2009). Wapl exerts a destabilizing effect on cohesin association throughout the cell cycle, which is counteracted by Eso1 mediated acetylation of Psm3 (in S phase), resulting in the establishment of stable sister chromatid cohesion from G_2 till anaphase (Nasmyth K 2011). Wapl deletion bypasses Eso1 requirement and therefore, cohesin binding onto chromatin during G_1, S and G_2 in a *Δwapl* strain should inherently be stable. In such a scenario, if transcription were indeed a requirement for loading cohesin onto chromosomes, then inhibition of transcription would still result in reduction of cohesin occupancy along the chromosomes. RNAPII and H3 occupancy was also analyzed (data not shown) and was similar to that in Figures V4 and V5.

The association of cohesin after thiolutin treatment was compared between the wild type and *Δwapl* strain, both expressing cohesin subunit Psm3 endogenously tagged with GFP (Psm3-GFP). Cohesin occupancy was assessed at cohesive sites along chromosome 2, (i) centromeric *dg* repeats, (ii) telomeric region, (iii) intergenic region between convergent genes, dis2, (iv) intergenic region between convergent gene, srp1, (v) negative control with no Mis4 or cohesin binding.

As observed with Rad21-GFP (Figure V.4B), Psm3-GFP binding at intergenic regions along chromosome 2 arms (dis2 and srp1) was reduced in both wild type (Figure V.6A) and *Δwapl* strains (Figure V.6B). There was a slight decrease in Psm3-GFP occupancy at the centromeric *dg* repeats and telomeres in the wild type, in contrast to no change at the telomeres and an increase at the centromeric *dg* in the *Δwapl* strain. Focusing on the *rpl29-dis2* convergent gene pair (Figure V.7A), Psm3 did not accumulate at promoter regions (probes A and C) at t =45 min (Figure

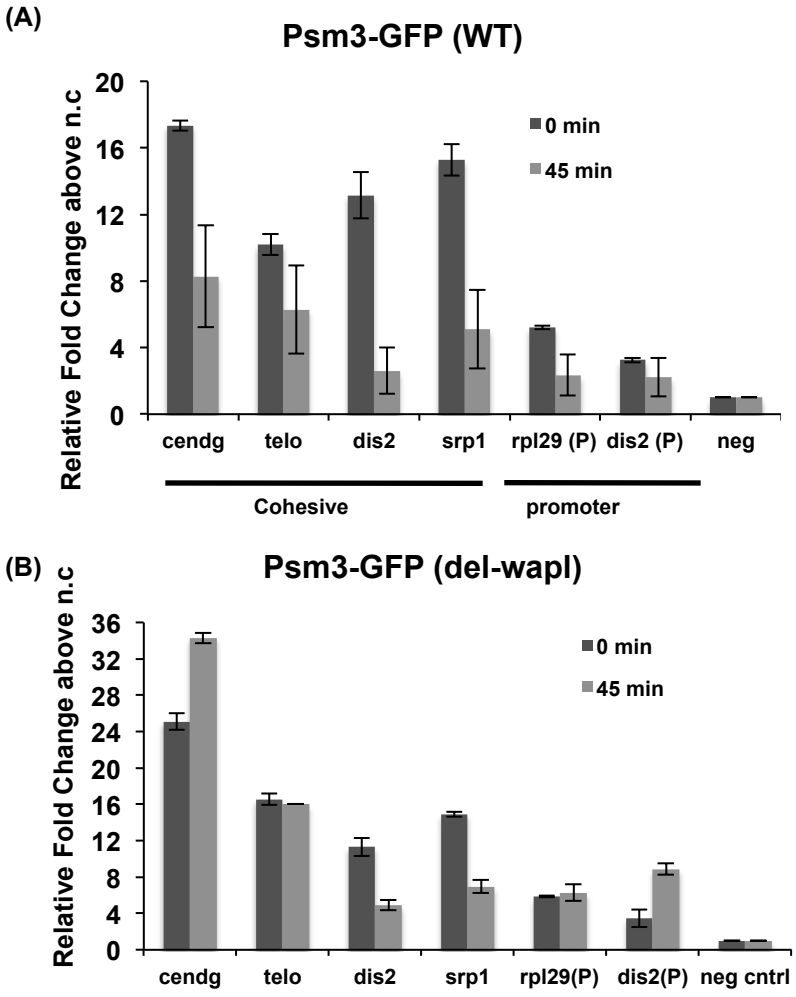


Figure V.6 Loss of transcription reduces inherently stable cohesin in a *Δwapl* background.

(A) Psm3 ChIP analysis in exponentially growing wild type cells after thiolutin treatment. As shown previously, cohesin binding at chromosome arms was evidently reduced at t = 45 min. A nominal difference was observed at the *dg* and telomeric regions. Furthermore, only background Psm3 levels were detected at the *rpl29* and *dis2* promoters, which remained unchanged between t = 0 and 45 min. Error bars = SEM from two independent replicates.

(B) Psm3 ChIP analysis in exponentially growing *Δwapl* cells after thiolutin treatment. Cohesin binding at the centromeric and telomeric regions was stable but cohesin on the arms dropped. A significant change at the promoter regions rpl29(P) and dis2(P) was absent. This implies that transcription is crucial for chromosomal association of cohesin. Error bars = SEM from two independent replicates.

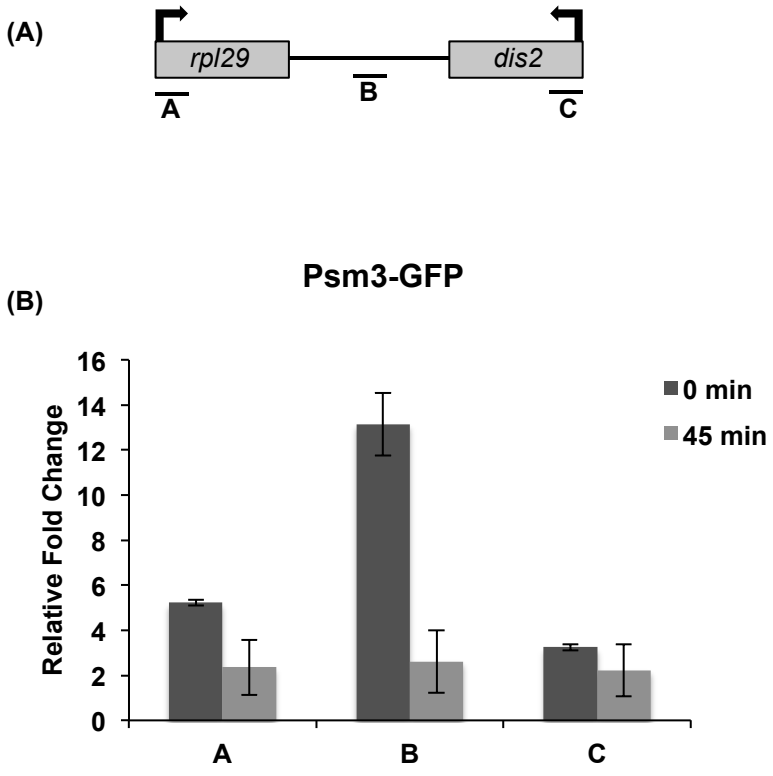


Figure V.7 Loss of transcription does not increase cohesin at promoter regions.

(A) Schematic representation of a convergent gene pair. *rpl29* and *dis2* are transcribed in sense and antisense direction, respectively (black arrows). A, B and C depict probes used for ChIP-qPCR analysis to determine cohesin levels after transcription inhibition.

(B) RNAPII ChIP analysis in exponentially growing cells after thiolutin treatment. While Psm3 signal dropped in the intergenic region (probe B), it remained unchanged at the promoters (A and C). This suggests that cohesin reduction after RNAPII inhibition is not a mere consequence of less transcription and therefore, less accumulation of cohesin at intergenic sites. Error bars = SEM from two independent replicates.

V.7B) but bulk of cohesin present at the intergenic region (probe B) was reduced at t = 45 min. This negates the hypothesis that transcription is required only to reposition cohesin. Additionally, reduction in cohesin levels after transcription inhibition in *Δwapl* suggests that RNAPII is required for cohesin loading onto chromosomes.

V.5 LOSS OF COHESIN BINDING AFTER INHIBITION OF TRANSCRIPTION BY 1,10-PHENANTHROLINE.

Thiolutin has been suggested to illicit stress responses (Adams and Gross et al. 1991, Grigull et al. 2004) and inhibit translation (Jimenez et al. 1973; Tipper DJ 1973), probably because it inhibits transcription by all yeast polymerases (RNAPI, II and III). Therefore, to further confirm the dependency between cohesin loading and RNAPII transcription, cycling cells were treated with an alternative transcription inhibitor called 1,10-phenanthroline. 1,10-phenanthroline is a metal chelator with some specificity for RNAPII (Parker et al. 1991; Gallagher et al. 1996). Cycling cells expressing cohesin subunit Psm3 endogenously tagged with GFP (Psm3-GFP), were grown in minimal medium, treated with 100 µg/ml of 1,10-phenanthroline for 30 min (Lackner et al. 2007) and processed for ChIP.

In addition to analyzing cohesin occupancy at cohesive sites, the cohesin binding pattern at less-cohesive sites was also investigated. Except centromeric and telomeric regions, all other probes target intergenic regions between convergent gene pairs (Lengronne et al. 2006; Schmidt et al. 2009). The cohesive regions along chromosome 2 include, (i) centromeric *dg* repeats, (ii) telomeric region, (iii) dis2, and (iv) srp1. The less-cohesive regions along chromosome 2 include intergenic regions between convergent genes (i) 776, (ii) med8 and (iii) plh1. The negative control represents a region with no Mis4 or cohesin binding.

Reduction in RNAPII initiation (Figure V.8C) coincided with a simultaneous decrease in the occupancy of Psm3 along chromosome arms (dis2 and srp1). While dis2 and srp1 represent “cohesive-cohesin” sites (Figure V.8A), transcription inhibition had little or no effect on the binding of Psm3 at the “less-cohesive sites”, i.e., 776, med8 and plh1. This can be attributed to the relatively low levels of functional cohesin present at the less-cohesive sites. In contrast to thiolutin treatment, no change was observed at the centromeric *dg* repeats (Figure V.8B), suggesting that thiolutin probably results in some stress response. The possibility of secondary

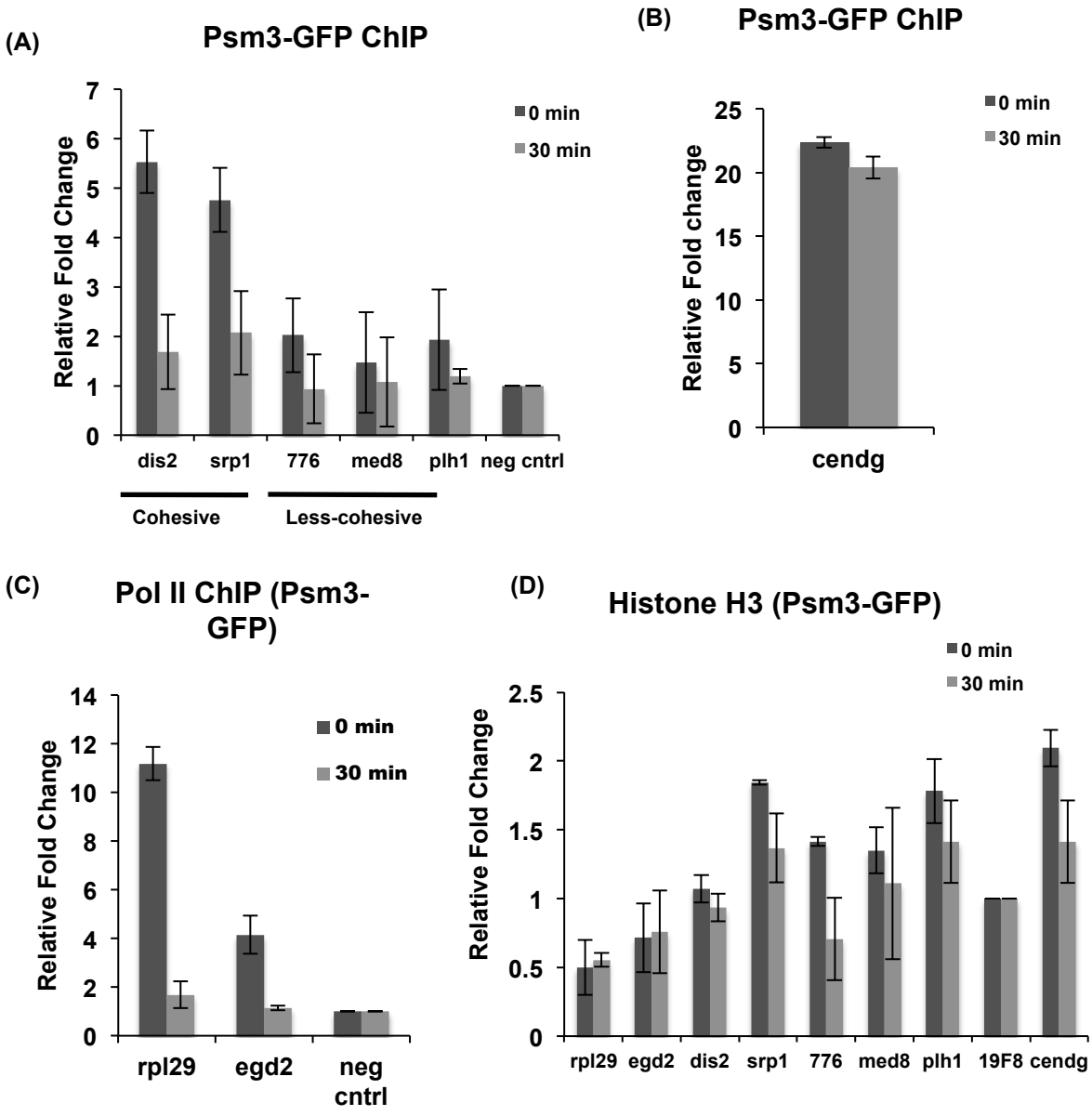


Figure V.8 Transcription inhibition by 1,10-phenanthroline also results in reduction in cohesin levels.

(A) Cycling cells (G_2) were treated with 1,10-phenanthroline and processed for ChIP analysis. Psm3-GFP levels were detected with anti-GFP antibody (Abcam). Psm3 occupancy was analyzed across chromosome arms at cohesive sites dis2 and srp1 and less-cohesive sites, 776, med8 and plh1. Cohesin at cohesive sites reduced ~2.5-3 fold, whereas cohesin at less-cohesive sites was unaffected. Error bars = SEM from two independent replicates.

(B) Centromeric cohesion was largely unaffected.

Figure V.8 continued……..

Figure V8 continued……..

Figure V.8 Transcription inhibition by 1,10-phenanthroline also results in reduction in cohesin levels.

(C) Transcription inhibition was assessed by RNAPII occupancy (8WG16). RNAPII levels significantly reduced at *rpl29* and *egd2* (t = 30 min) promoter regions, coinciding with 1,10-phenanthroline treatment.

(D) H3 levels (Abcam) were measured as a control between 0 and 30 min and remained similar between the time points.

effects arising due to chemical treatment was discounted since the H3 profile was consistent between t = 0 and 30 min (Figure V.8D)

V.6 COHESIN INTERACTS WITH RNAPII.

Genome-wide mapping analysis in yeast has shown that the position of cohesin occupancy correlates with transcriptional activity, such that transcription appears to reposition cohesin from its loading sites to intergenic regions between convergent gene pairs (Glynn et a. 2004; Lengronne et al. 2006; Schmidt et al. 2009). In such a scenario, the observed interplay between transcription and cohesin would simply mean that less polymerase is initiating and therefore, less cohesin is being pushed to intergenic regions. On the contrary, if RNAPII is actually required for the association of cohesin onto chromosomes, then the two complexes should physically interact. One might suspect that RNAPII and the cohesin complexes (cohesin and Mis4) exist in physical proximity but surprisingly, an interaction between them has not been reported. A cohesin interactome in MOLT-4 T-lymphocytes revealed novel interactions between RAD21 and regulators of chromosome dynamics, replication, apoptosis, DNA-damage response, RNA-processing and cytoskeleton functions (Panigrahi et al. 2012). However, an interaction between cohesin and RNAPII was not observed.

To address this critical issue, co-immunoprecipitation experiments were performed using TAP-tagged strains of core cohesin subunits (Psm3, Rad21 and Psc3), cohesin loading complex (Mis4 and Ssl3) and cohesin maintenance proteins (Pds5 and Wapl). An untagged strain (972) was used as a negative control. Whole cell extracts (WCE) were prepared from cycling cells. 10% of the WCE was kept as the input (IN) and the remaining material was used for immunoprecipitation (IP). Overnight IP was performed with anti-phosphoSer5 RNAPII antibody (Abcam). Phosphorylation of the RNAPII CTD at Ser 5 is a hallmark of transcription initiation and my results (Figures V.4, V.5, V6 and V.8) suggest that inhibition of transcription initiation results in loss of cohesin binding. The input and IP complexes were boiled with SDS-PAGE loading dye, separated on a 4-15% bistris precast gel, transferred to a nitrocellulose membrane and probed with antibodies for TAP-tag (Pierce Antibodies) and phosphoSer5. First, the immunoprecipitation (IP) efficiency for all the strains was analyzed using an anti-Ser5 (Figure V.9A, top panels) antibody. In comparison to the input, a strong enrichment was observed in the IP lanes, which for

(A) Phosho-Ser5 RNAPII pull down-TAP western

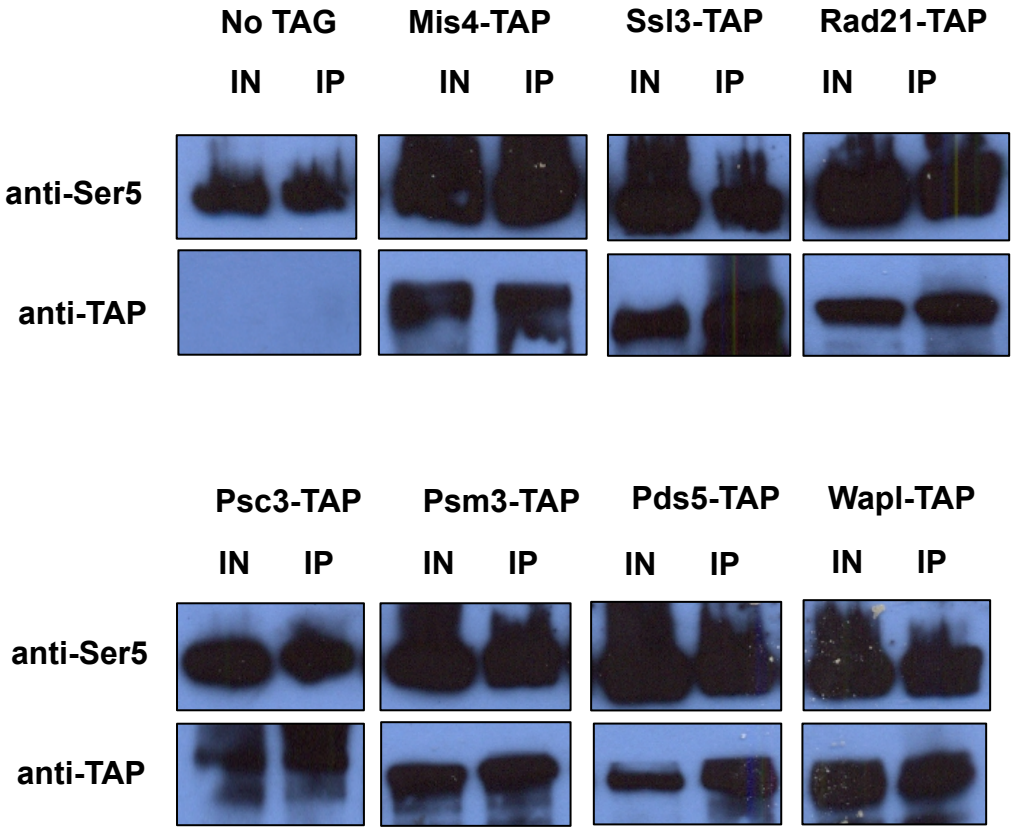


Figure V.9 RNAPII interacts with cohesin complex.

(A) Whole cell extracts were prepared from cycling cells (G_2) with no tag, *mis4-TAP, ssl3-TAP, rad21-TAP*, *psc3-TAP, psm3-TAP, pds5-TAP* and *wapl-TAP* strains. Overnight immunoprecipitation was performed with anti-phosphoSer5 antibody (Abcam) targeting RNAPII CTD and western blot performed with anti-TAP antibody (Pierce Antibodies). "IN" and "IP" represent total input and immunoprecipitated samples. The strong phosphoSer5 signal in the IPs indicates efficient protein pull down. All subunits of the cohesin complexes were found to interact with RNAPII. No TAP signal was observed in IP from the no tag control.

(B) TAP pull down-anti RNAPII 8WG16 western

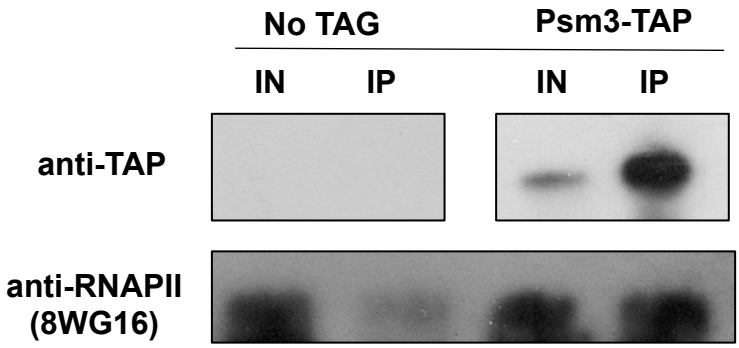


Figure V.9 RNAPII interacts with cohesin complex.

(B) Whole cell extracts were prepared from cycling cells (G_2) with no tag and *psm3-TAP* strains. Overnight immunoprecipitation was performed with IgG sepharose beads (GE healthcare) and western blot with anti-phosphoSer5 antibody (Abcam). “IN” and “IP” represent total input and immunoprecipitated samples. The strong TAP signal in the IP indicates efficient protein pull down. RNAPII was indeed found to co-precipitate with cohesin subunit Psm3. No TAP signal was observed in IP from the no tag control.

the TAP-tagged strains was absent in 972. In a reciprocal experiment, WCEs from 972 and Psm3-TAP strains were immunoprecipitated with anti-TAP antibody (Pierce Antibodies) and probed with anti-TAP and anti-RNAPII (8WG16) antibodies (Figure V.9B). Enrichment in the IP over input demonstrated high IP efficiency (Figure V.9B, top panel). Surprisingly, RNAPII was immunoprecipitated in the Psm3 IP (Figure V.9b, bottom panel) but not in the IP from the untagged wild type control. This confirms an interaction between RNAPII and the core cohesin complex, and between RNAPII and Mis4. In an unexpected result, even Pds5 and Wapl interacted with RNAPII. This either implies a hitherto unidentified “regulatory” interaction, or it is a reflection of proximity in the nuclear space.

The above results lend support to the connection observed between transcription initiation and chromosomal association of cohesin and Mis4.

V.7 MASS SPECTROMETRY CONFIRMS THE INTERACTION BETWEEN COHESIN COMPLEXES AND RNAPII.

To further validate the above interaction between RNAPII and cohesin, mass spectrometry (MS) analysis was performed using TAP-tagged strains of cohesin subunits and the untagged control. WCEs were prepared as above and immunoprecipitated overnight with IgG-sepharose beads (GE Healthcare). The protein-IgG-beads complexes were washed, boiled with SDS-PAGE loading dye, and 25 µl of each IP was loaded on a 4-15% bistris precast gel (BioRAD) and run for a few minutes at 180V. The run was stopped once the sample had descended ~5 mm into the wells and each band was cut, chopped into small pieces and sent to the mass spectrometry facility (Sir William Dunn School of Pathology). The MS data is presented in a tabular form (Tables V1-V7). In agreement with the co-IP data (Figure V.9A), the core cohesin subunits (Psm3, Rad21 and Psc3), loading proteins (Mis4 and Ssl3) and accessory proteins (Pds5 and Wapl) interacted with RNAPII subunits Rpb1 and Rpb2. Additionally, Mis4 also interacted with Rpb4 subunit of the RNAPII holoenzyme. Moreover, Mis4 and Wapl associated with the cohesin acetyltransferase Eso1. Surprisingly, Psm3, Rad21 and Pds5 also exhibited an interaction with the transcription elongation factor Spt5.

Table V1 **Psm3 MS Analysis**

Protein	Description	% Coverage	TAP tag	WT	TAP tag/ WT
PSM3-TAP					
sp|O13936| SPT5_SCHPO	Transcription elongation factor spt5 OS=Schizosaccharomyces pombe	4.70	1.00	1.00	1.00
sp|Q02061| RPB2_SCHPO	DNA-directed RNA polymerase II subunit RPB2 OS=Schizosaccharomyces pombe	4.70	5.00	3.00	1.67
sp|O42649| SMC3_SCHPO	Structural maintenance of chromosomes protein 3 OS=Schizosaccharomyces pombe	12.50	10.00		
sp|O94383| SMC1_SCHPO	Structural maintenance of chromosomes protein 1 OS=Schizosaccharomyces pombe	14.10	12.00		
sp|P36594| RPB1_SCHPO	DNA-directed RNA polymerase II subunit rpb1 OS=Schizosaccharomyces pombe	5.70	5.00	3.00	1.67

Table V2 **Rad21 MS Analysis**

Protein	**Description**	**% Coverage**	**TAP tag**	**WT**	**TAP tag/ WT**
RAD21-TAP					
sp|Q02061| RPB2_SCHPO	DNA-directed RNA polymerase II subunit RPB2 OS=Schizosaccharomyces pombe	3.60	2.00	3.00	0.67
sp|O42649| SMC3_SCHPO	Structural maintenance of chromosomes protein 3 OS=Schizosaccharomyces pombe	6.30	5.00		
sp|P36594| RPB1_SCHPO	DNA-directed RNA polymerase II subunit rpb1 OS=Schizosaccharomyces pombe	4.10	4.00	3.00	1.33
sp|O94383| SMC1_SCHPO	Structural maintenance of chromosomes protein 1 OS=Schizosaccharomyces pombe	3.30	2.00		
sp|O13936| SPT5_SCHPO	Transcription elongation factor spt5 OS=Schizosaccharomyces pombe	4.60	2.00	1.00	2.00

Table V3 **Psc3 MS Analysis**

Protein	Description	% Coverage	TAP tag	WT	TAP tag/ WT
Psc3TAP					
sp|P36594| RPB1_SCHPO	DNA-directed RNA polymerase II subunit rpb1 OS=Schizosaccharomyces pombe	2.50	2.00	3.00	0.67
sp|Q02061| RPB2_SCHPO	DNA-directed RNA polymerase II subunit RPB2 OS=Schizosaccharomyces pombe	4.80	5.00	3.00	1.67
sp|O13816| SCC3_SCHPO	Cohesin subunit psc3 OS=Schizosaccharomyces pombe	3.50	2.00		

Table V4 **Mis4 MS Analysis**

Protein	Description	% Coverage	TAP tag	WT	TAP tag/ WT
MIS4-TAP					
sp|Q02061| RPB2_SCHPO	DNA-directed RNA polymerase II subunit RPB2 OS=Schizosaccharomyces pombe	5.30	3.00	3.00	1.00
sp|O42917| ESO1_SCHPO	N-acetyltransferase eso1 OS=Schizosaccharomyces pombe	3.10	1.00	1.00	1.00
sp|O74825| RPB4_SCHPO	DNA-directed RNA polymerase II subunit rpb4 OS=Schizosaccharomyces pombe	25.20	1.00		
sp|P36594| RPB1_SCHPO	DNA-directed RNA polymerase II subunit rpb1 OS=Schizosaccharomyces pombe	5.10	7.00	3.00	2.33

Table V5 **Ssl3 MS Analysis**

Protein	Description	% Coverage	TAP tag	WT	TAP tag/ WT
SSL3-TAP					
sp|P36594| RPB1_SCHPO	DNA-directed RNA polymerase II subunit rpb1 OS=Schizosaccharomyces pombe	6.10	6.00	3.00	2.00
sp|Q02061| RPB2_SCHPO	DNA-directed RNA polymerase II subunit RPB2 OS=Schizosaccharomyces pombe	7.00	5.00	3.00	1.67

Table V6 **Pds5 MS Analysis**

Protein	Description	% Coverage	TAP tag	WT	TAP tag/ WT
PDS5-TAP					
sp|O13936| SPT5_SCHPO	Transcription elongation factor spt5 OS=Schizosaccharomyces pombe	3.70	1.00	1.00	1.00
sp|Q02061| RPB2_SCHPO	DNA-directed RNA polymerase II subunit RPB2 OS=Schizosaccharomyces pombe	4.20	4.00	3.00	1.33
sp|P36594| RPB1_SCHPO	DNA-directed RNA polymerase II subunit rpb1 OS=Schizosaccharomyces pombe	4.80	5.00	3.00	1.67

Table V7 **Wapl MS Analysis**

Protein	Description	% Coverage	TAP tag	WT	TAP tag/ WT
WAPL-TAP					
sp|Q02061| RPB2_SCHPO	DNA-directed RNA polymerase II subunit RPB2 OS=Schizosaccharomyces pombe	4.10	3.00	3.00	1.00
sp|O42917| ESO1_SCHPO	N-acetyltransferase eso1 OS=Schizosaccharomyces pombe	3.10	2.00	2.00	1.00
sp|P36594| RPB1_SCHPO	DNA-directed RNA polymerase II subunit rpb1 OS=Schizosaccharomyces pombe	2.50	1.00	3.00	0.33

V.8 COHESIN INTERACTS WITH TRANSCRIPTION ELONGATION FACTOR Spt5.

At the tRNA gene loci, Mis4/Ssl3 (and partially Rad21) binding sites overlap the binding profile of Sfc6, a subunit of RNAPIII transcription factor TFIIIC (Schmidt et al. 2009). Similarly, at the genes encoding ribosomal proteins, Mis4/Ssl3 colocalize with the forkhead-domain containing protein Fhl1 (Schmidt et al. 2009). However, whether Sfc6 or Fhl1 are indeed required for loading cohesin lacks experimental evidence. Transcription elongation factor Spt5 (in association with Spt4) is a ubiquitously expressed regulator of transcriptional elongation for most RNAPII genes (Hartzog et al. 2013). The MS data suggested an interaction between Spt5 and core cohesin subunits Psm3 and Rad21, and cohesion maintenance protein Pds5. This argues for a possible involvement of Spt5 in loading cohesin at actively transcribed genes. However, definitive evidence requires further experimentation.

V.9 DISCUSSION.

Previous studies have shown that yeast Scc2 (Mis4 homolog) mutants mimic CdLS mutation by showing altered *GAL* gene expression and chromosome organization (Gard et al. 2009). However, in budding yeast, Rad21 inactivation results in inactivation of ~29 (out of 6300) genes with inter-related functions and positional clustering (Skibbens et al. 2011). In an indirect approach, inactivation of the cohesin loader in *S. pombe* (*mis4-367ts*) also results in misregulation of ~25 genes (Dheur et al. 2011). While the above studies address cohesin and transcription function in G_1-synchronized cells, I have investigated transcriptional changes associated with Rad21 and Mis4 inactivation in G_2. Moreover, microarray analyses primarily investigate steady state transcripts; therefore, I focused on ChIP-qPCR to identify immediate changes in RNAPII occupancy at promoter regions.

Rad21 or Mis4 inactivation did not result in any decrease in RNAPII occupancy. Interestingly, I did observe a ~1-2 fold increase in RNAPII occupancy in *rad21-K1ts* and *mis4-367ts*. However, similar increase was also observed in the wild type strain, demonstrating increased efficiency of formaldehyde cross-linking at high temperatures.

Genome-wide mapping shows that the position of cohesin complexes on chromosome arms, in both *S. cerevisiae* and *S. pombe,* correlates with transcriptional activity. This appears to reposition cohesin from the loading sites to intergenic regions of convergent gene pairs. Although, clear definition of a loading site is missing, preferential binding of Mis4 to highly transcribed genes suggests that cohesin associates with chromosomes at gene promoters and is subsequently pushed by RNAPII. In such a scenario, inhibition of transcription initiation would impede cohesin repositioning and cohesin would now accumulate at gene promoters. This question was addressed by treating *S. pombe* cells with thiolutin, a potent inhibitor of transcription by RNAPI, II and III (Tipper DJ 1973). Surprisingly, inhibition of transcription initiation resulted in loss of cohesin binding at intergenic regions, without any increase at upstream promoter regions. Furthermore, transcription inhibition also led to reduction in Mis4 occupancy along chromosome 2. These observations are inconsistent with the view that cohesin loading is independent of transcription initiation. Indeed, my results directly implicate RNAPII in chromosomal association of cohesin, possibly via Mis4.

Centromeres are hubs of heterochromatin, regions that are inherently inaccessible to the transcription machinery. However, loss of transcription resulted in reduction of cohesin and Mis4 even at the centromeric *dg* repeats. The *S. pombe* centromeres are divided into inner most (*imr*) and outer most (*otr*) repeats, wherein the *dg* resides within the *otr* (Chan et al. 2012). Cryptic promoters at the outer repeats transcribe overlapping *dg/dh* transcripts, maximally in the S-phase, which are then processed into siRNAs by Dicer. The siRNAs are amplified by RNA-directed RNA polymerase (Rdrp) and incorporated into the RNA-induced transcriptional silencing (RITS) complex (Djupedal et al. 2005; Djupedal et al. 2009). RITS recruits the Clr4 methyltransferase that deposits the H3K9me mark (Rea S et al. 2000), crucial for recruitment of heterochromatic protein Swi6 (Lachner et al. 2001) and heterochromatin establishment (Zhang et al. 2008). Interestingly, similar to Dicer, Clr4 and Rdrp mutants, Rpb2 and Rpb7 mutants (RNAPII holoenzyme subunits) also exhibit defects in transcriptional silencing and heterochromatin formation (Kato et al. 2005). As thiolutin ChIP-qPCR experiments were performed in unsynchronized cells, loss of cohesin and Mis4 at the *dg* could be attributed to RNAPII transcription occurring in S-phase.

As previously stated, establishment of stable sister chromatid cohesion occurs during S-phase, when Eso1 acetylates K112 and K113 residues of Smc3. How acetylation stabilizes cohesin is currently unknown. At the molecular level, histone acetylation sequesters the positive charge of lysine residues and thereby disrupts histone-DNA interaction (Polevoda and Sherman 2002). Additionally, histone acetylation can serve as a docking site for bromodomain transcription activators (Sanchez and Zhou 2009). Furthermore, acetylation of lysine residues near enzyme active sites can inhibit enzymatic activity (Miller et al. 2010). On similar lines, the prevalent idea is that Smc3 acetylation reduces the affinity of cohesin for the "anti-establishment" activity of Wapl (Chan et al. 2012; Rowland et al. 2009; Sherwood et al. 2010; Terret et al. 2009). In support for this hypothesis, Wapl and Pds5 deletion mutants bypass the requirement of Eso1 for cell viability and chromosomal association of cohesin is sensitive to Wapl protein levels (Chan et al. 2012). Furthermore, Wapl depleted fission yeast cells inherently stabilize cohesin (Sutani et al. 2009). In such a scenario, if transcription were indeed a requirement for cohesin loading, then inhibiting transcription would reduce cohesin occupancy even in a *Δwapl* genetic background. Indeed, cohesin occupancy was reduced in the *Δwapl* strain.

Results presented in Chapter IV describe two sub-sets of nuclear cohesin, "cohesive cohesin" and "less-cohesive" cohesin. The cohesive cohesin overlaps with highly transcribed gene loci due to preferential binding of Mis4 at such loci. In contrast, the less-cohesive cohesin sites lack Mis4 and colocalize with poorly transcribed genes. Further comparison was made between the two sites by the use of 1,10-phenanthroline in a wild type background. 1,10-phenanthroline is a metal chelator with some specificity for RNAPII (Parker et al. 1991; Gallagher et al. 1996). Similar effects were observed with 1,10-phenanthroline and thiolutin. While the association of cohesive cohesin (dis2 and srp1) with chromatin decreased following transcription inhibition (Figure V.8A, cohesive), the less-cohesive cohesin (Figure V.8A, less-cohesive) remained unchanged. Furthermore, FISH analysis targeting cohesive sites (dis2 and tif11) after thiolutin treatment demonstrates separation of sister chromatids at t = 45 min, while the less-cohesive sites (pla1 and swi3) are indistinguishable between 0 and 45 min (data not shown).

The indirect connection between RNAPII and cohesin complexes described above was further supported by co-immunoprecipitation and MS analysis. Core

cohesin subunits (Psm3, Rad21 and Psc3), cohesin loading protein (Mis4 and Ssl3), and accessory proteins (Pds5 and Wapl), all interacted with the initiating form of RNAPII. The association of Pds5 and Wapl with RNAPII can either be due to physical proximity of the entire complex to RNAPII or might have a regulatory role. While MS identified Spt5 as a cohesin interacting partner, co-IP did not provide definitive evidence due to lack of antibody specificity. Nonetheless, by analogy to TFIIIC, which colocalizes with cohesin at tRNA genes (Schmidt et al. 2009), the ubiquitous factor Spt5 might perform a similar function for RNAPII genes.

The experiments discussed so far provide clear evidence that RNAPII is required for the chromosomal association of Mis4 and subsequently cohesin. However, despite a lack of change in the histone H3 profiles between several experimental repeats, it can still be argued that the above effects are due to chemical treatment. To address this concern I will use the auxin-inducible degron (AID) system for RNAPII depletion (Nishimura et al. 2009). Briefly, the strategy employs a plant-specific mechanism that relies on auxin induced proteasome pathway. In plants, auxin binds to transport inhibitor response 1 (TIR1) protein and promotes binding of the SCF^{TIR1} to Aux/IAA transcription repressors. The Aux/IAA proteins are polyubiquitinated and degraded by the proteasome. Except TIR1, components of the proteasome pathway are conserved among eukaryotes and thus, employing AID in non-plant systems is seemingly possible. In a strain expressing the TIR1 complex, any protein of interest can be tagged with the AID domain and degraded via the proteasome pathway after auxin treatment. This system has been applied to address the connection between Pds5 and cohesin acetylation in *S. cerevisiae* (Chan et al. 2013) and the MCM replication complex in *S. pombe* (Kane et al. 2011). This system would be a powerful way to perform cell-cycle stage experiments and gain further mechanistic insights into transcription and cohesin association.

A complete picture of the likely series of events that govern the association of cohesin with chromosomes is elaborated in the final discussion chapter (Chapter VI).

CHAPTER VI

GENERAL DISCUSSION

During my undergraduate studies, textbooks such as, “GENES VIII” (Benjamin Lewin), “Molecular Biology of The Cell” (Bruce Alberts), “Molecular Biology of the Gene” (James D. Watson) and “Molecular Cell Biology” (Harvey Lodish) gave the impression that gene expression is a streamlined manifestation of the central dogma of cell and molecular biology. The linear DNA sequence is converted into RNA (transcription) and the RNA serves as a template for protein synthesis (translation). Proteins, in turn, execute a multitude of regulatory functions, tailoring the cellular phenotype.

My doctoral studies led me to appreciate that gene expression is in part a streamlined manifestation of the central dogma, but is not straightforward. The genome doesn’t operate on a linear scale and the cellular phenotype is an intricately controlled orchestration of the three-dimensional organization of the genome. The compaction of DNA into arrays of nucleosomes, formation of higher-order chromatin structure and their positioning within nuclear space are collectively essential in establishing the gene expression signatures during differentiation and development. The importance of three-dimensional regulation of the genome is further emphasized by the fact that a breakdown in the above processes results in a variety of diseases and syndromes, such as cancer, autoimmunity, neurological disorders and cardiovascular diseases.

Microscopic and 3C-based methods have clearly established compartmentalization of the interphase nucleus into chromosome territories (van Steensel and Dekker 2010). Gene loci within a territory tend to communicate with each other but inter-territorial contacts have also been observed. Furthermore, co-regulated genes tend to associate at higher frequencies, presumably sharing the same cellular pool of transcription factors and activators. However, a twist in the tale comes from observations that co-regulated genes are not co-transcribed. Genes are expressed in “bursts” and most transcription activators and coactivators exhibit short binding times on chromatin templates (Misteli 2007). Despite of this, the cell must mount a rapid transcriptional response at the right moment. Several studies have suggested that this is in part achieved by organization of genes into “loop” structures. Gene-loops are formed by the juxtaposition of promoter and terminator regions and, in most cases, are dependent on active transcription. In higher eukaryotes, extensive literature supports the existence of long-range “loop” interactions between distant

regulatory elements. However, single gene-loops have only been studied at a few loci.

Our understanding of gene-loops is largely based on studies performed in budding yeast, where gene-loops are formed after the first round of transcription. Once established, loops may facilitate recycling of RNAPII from the terminator to promoter regions. Gene-loops are also linked to transcription memory and maintenance of promoter directionality (Tan-Wong et al. 2009; Laine et al. 2009; Tan-Wong et al. 2012). When I initiated my doctorate studies, a detailed analysis of gene-loop formation was absent in mammalian systems. It was surprising that, even though gene-loops are dependent on transcription and RNA processing factors, not all transcription units appear associated with gene loops. Additionally, 3C-based evidence for gene-loops represents an average analysis of a heterogeneous cell population. A single allele analysis by nuclear imaging will be required to resolve the generality of gene looping.

To address the above concerns, I undertook gene-loop analysis at a constitutively expressed gene, mouse serum albumin *(Alb)*, and an inducible gene, the nuclear factor of kappa polypeptide *(NFκB1)*. Chapter III provides a detailed description of the analysis. To summarize, using 3C, a gene loop was detected at *Alb*. The loop was dependent on active transcription, specifically phospho-Ser2 RNAPII, while indirect evidence also implicated phospho-Ser5 RNAPII. In contrast to *Alb*, *NFκB1* 3C analysis provided insights into chromatin conformation changes in response to extracellular stimuli. In response to tumor necrosis factor α (TNFα), *NFκB1* appears to form a transient gene-loop. This was followed by clover-leaf shaped contacts between promoter and possible regulatory intronic sequences (similar to *Alb*). Subsequently, activation of the *IκB*-mediated feedback pathway translocated NFκB to the cytoplasm, ceasing further transcription and concomitant disappearance of 3C contacts. The length of *NFκB1* (>100 kbp) prompted us to employ it as a model system to perform kinetic analysis of gene-loop formation. RNA-FISH was reasoned as the method of choice because multiple RNA molecules make signal detection easier in comparison to DNA-FISH. However, RNA-FISH did not support a gene-loop at *NFκB1*. Whilst 3C represents average population data, limited understanding of RNA residency time at transcription initiation and processing sites compromised the RNA-FISH results. Perhaps, the method of choice is DNA-FISH. However, gene-loop signatures such as TFIIB and phospho-Ser5 RNAPII are

present only at a subset of actively transcribed genes, predominantly in a cell-type specific manner (Grosso et al. 2012). Most highly expressed genes are small in length and hence, unsuitable for microscopic analysis. One of my future goals is to identify an appropriate model gene that overcomes the above limitations and to look for single cell evidence supporting the existence of gene-loops.

Several architectural proteins act as key mediators of chromatin organization, of which cohesin is a prime example. The multi-subunit cohesin complex is highly conserved across species, executing cohesion, transcription and repair/recombination functions. Interestingly, cohesin occupancy along chromosomes is highly variable among different organisms. In mammals, cohesin primarily colocalizes with CTCF, while in *Drosophila* it is present near highly transcribed genes. In *S. cerevisiae*, cohesin concentrates at intergenic regions separated from loading sites. Surprisingly, in *S. pombe*, cohesin is divided into two classes, one colocalizes with the loading complex Mis4/Ssl3, while the other accumulates at intergenic regions of convergent gene pairs.

I decided to investigate whether the two populations of cohesin in fission yeast manifest any functional differences. As described in Chapter IV, cellular cohesin demonstrated a clear “division of labor”. Cohesin colocalizing with Mis4 is the stable/cohesive cohesin. In contrast, cohesin without Mis4 represents soluble/dynamic cohesin, with a shorter residency time on chromosomes. The distinction between the sub-sets is perhaps established during S-phase, when Eso1 acetylates cohesin colocalizing with Mis4, as it appears that Eso1 physically interacts with Mis4. In Chapter V, I study the stable cohesin form at cohesin loading sites and observe a previously unexpected connection between cohesin loading onto chromosomes and RNAPII. Inhibition of transcription initiation results in reduced cohesin occupancy on chromatin. Furthermore, inhibition of transcription also leads to loss of Mis4 chromatin binding. These results are supported by preferential colocalization of Mis4 and cohesin (at cohesive sites) with highly transcribed genes. Additionally, both Mis4 and cohesin physically interact with RNAPII. In simple terms, my data suggests a stepwise mechanism for cohesin loading. However, prior to describing the mechanism, I first outline a few key points about DNA replication.

In eukaryotes, DNA synthesis initiates at origins of replication throughout the genome. During S-phase, each cell utilizes a subset of origins that fire at

characteristic times, implying a programmed choice of origin usage (Aladjem 2007). Like most cellular processes, replication initiation is a highly coordinated assembly of many components. First, the origin recognition complex (ORC) selects sites for initiating replication. Next, the pre-replicative complex (pre-RC) is assembled by the recruitment of Cdc6, Cdt1 and MCM helicase in G_1. Finally, Cdc45 is necessary for loading DNA polymerase, forming the pre-initiation complex (pre-IC) and initiating replication (Bell and Dutta 2002). Genome-wide analysis in fission yeast has identified ~401 strong and 503 putative weak origins (Heichinger et al. 2006). The strong origins are spaced from 3-114 kbp, with a mean inter-origin distance of 31 kbp, while the overall average spacing between origins is ~14 kbp, similar to *Drosophila* and *Xenopus* cells (Hyrien et al. 2003). Additionally, it has also been proposed that early firing origins maximally associate with ORCs at the exit of mitosis, which enables them to quickly assemble the pre-RC and initiate DNA synthesis. On the other hand, replication at late origins maybe inefficient due to limiting amounts of replication factors in the cellular pool. The *S. pombe* ORCs directly bind AT-rich intergenic sequences that have a destabilizing effect on nucleosome assembly, suggesting that origins are nucleosome-depleted (Xu et al. 2012). In support of the above observation, an increase in histone acetylation by the deletion of Rpd3 histone deacetylase results in earlier firing of several origins (Aparicio et al. 2004). Finally, in *S. pombe*, origins are activated by Swi6-mediated recruitment of Dbf4-dependent kinase (DDK) and Cdc7 to heterochromatin foci, which are replicated early (Kim et al. 2003). In higher eukaryotes, replication timing has also been linked to transcription, such that active genes and gene-rich regions are replicated earlier than inactive, gene-poor regions (Danis et al. 2004; Schubuler et al. 2002; Hiratani et al. 2009; Gilbert 2002).

Based on my results and the documented literature, I suggest a model for cohesin loading onto chromosomes (Figure VI). First, RNAPII recruits Mis4 (or Mis4 co-bound with cohesin) to highly transcribed gene loci during G_1. In late G_1, synthesis of Rad21 results in reassembly of the tripartite cohesin ring. Mis4 now facilitates the association of cohesin with chromatin. However, stabilization of cohesin is challenged due to the absence of Eso1. Highly transcribed genes are known to exhibit open chromatin conformation and therefore, can act as platforms for the assembly of the replication initiation complexes during S-phase. The replication fork components recruit Eso1, which can now acetylate cohesin at Mis4 and cohesin overlapping sites. As a result, stable cohesion is established at the cohesive sites

COHESIN LOADING DURING THE CELL CYCLE

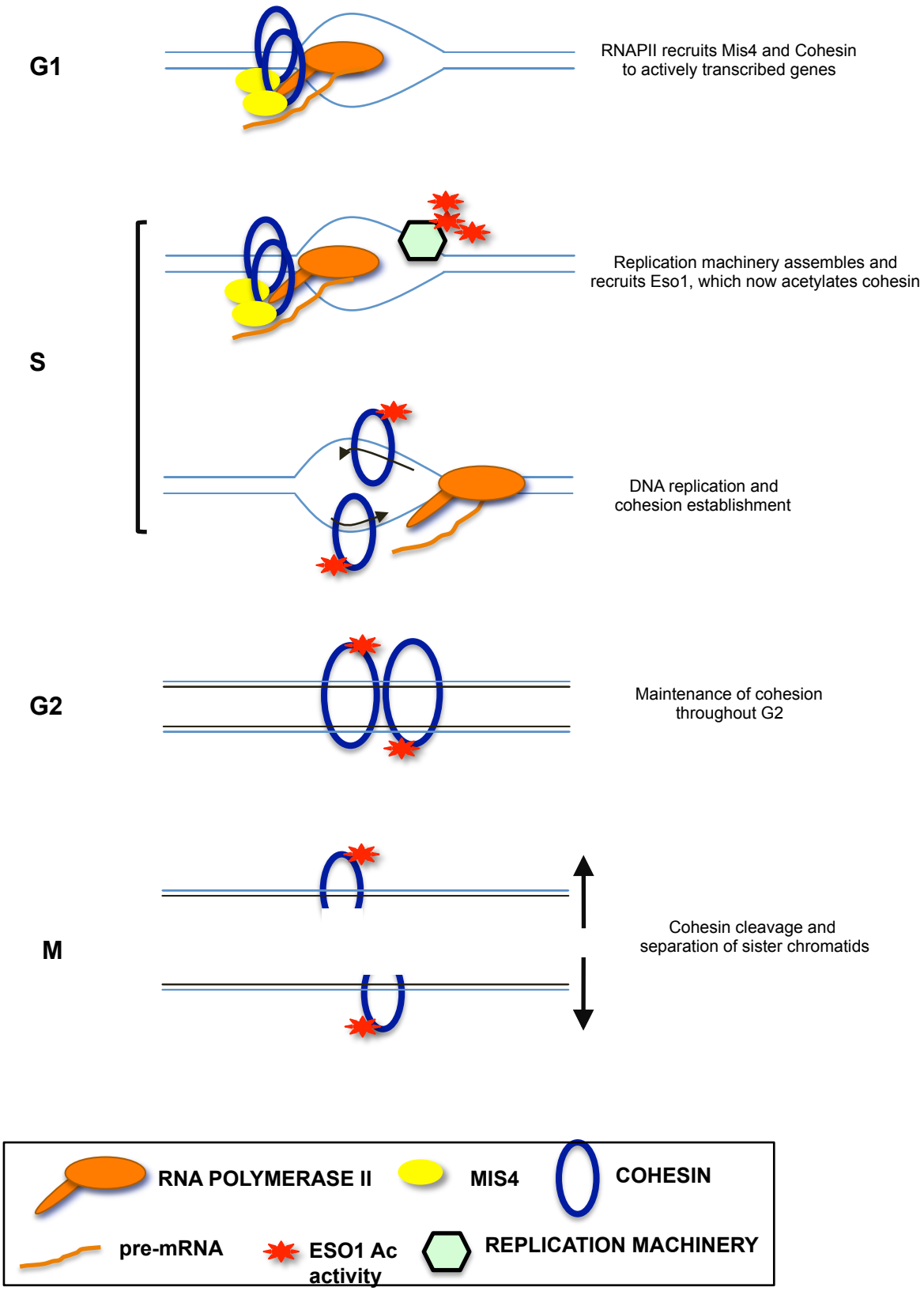


Figure VI. A model for cohesin loading onto chromosomes during the fission yeast cell cycle.

During G1, RNAPII recruits Mis4 and cohesin to actively transcribed gene loci. In S, the open chromatin conformation of highly transcribed genes, provides a easy platform for assembly of the replication machinery. Replication fork proteins recruit Eso1 to the Mis4/cohesin sites. Continuing in S phase, Eso1 acetylates cohesin, thereby resulting in stabilization cohesin and establishment of sister chromatid cohesion. Thereafter, cohesion is maintained at “cohesive sites” throughout G2. During mitosis, cohesin subunit Rad21 is cleaved , resulting in dismantling of the cohesin complex and dissolution of sister chromatids.

(Mis4/Rad21) during S-phase, continuing into G_2. During mitosis, separase-mediated cleavage of Rad21 dismantles the cohesin ring, followed by dissociation of cohesive cohesin from chromosomes and separation of sister chromatids to the daughter nuclei. In contrast, at poorly expressed genes, reduced levels of RNAPII result in low levels of chromatin bound Mis4 and cohesin. In these regions, cohesin maintains transient association and dissociation kinetics, failing to stabilize due to the absence of Eso1. However, in events of DNA damage, the less-cohesive cohesin may stabilize and facilitate recombination/repair.

I end with a few words highlighting the strong association of cohesin and associated proteins in disease and cancer. Mutations in human cohesin subunits, *SMC1A, SMC3, STAG2, RAD21*, *NIPBL* and *STAG3* have been identified in patients with acute adult *de novo* myeloid leukemia (The Cancer Genome Atlas 2013; Dolnik et al. 2012). SMC1A has also been implicated in the pathogenesis of gliomas, where SMC1A expression is up-regulated and loss of SMC1A suppresses glioma cell growth *in vitro* (Ma et al. 2013). Several studies have documented aberrant expression of RAD21 in cancer cells. Thus, RAD21 is over-expressed in undifferentiated cancers of the breast, lung, bladder, brain and ovaries (Rhodes et al. 2004) In contrast, RAD21 is downregulated in oral squamous cell carcinoma and cell lines with high metastatic potential (Yamamoto et al. 2006). Interestingly, studies in zebrafish have revealed *rad21* to be a regulator of *runx1* (Horsefield 2007). RUNX1/AML1 (runt-related transcription factor 1/ acute myeloid leukemia 1) belongs to the RUNX transcription factor family, which regulates transcription of genes involved in cell differentiation, growth and survival. *RAD21* and *RUNX* mRNA levels are known to be deregulated and co-dependent in endometrial cancers (Supernat et al. 2012). Glioblastoma cell lines, melanomas and Ewing's sarcomas also display cohesion defects, lagging chromosomes and anaphase bridges (classic signs of aneuploidy) due to mutations in the X-linked *STAG2* gene, which has been proposed to function as a "caretaker" tumor suppressor gene (Solomon et al. 2011). Somatic mutations in *SMC1A, SMC3, SA3,* and *NIPBL* have also been documented in colorectal cancers characterized by almost 100-fold higher CIN than normal cells (Barber et al. 2008). Conversely, downregulation of *SMC1A, STAG3, STAG2* and *RAD21* has been found in ovarian cancer, acute myeloid leukemia and in chronic myelomonocytic leukemia (Yamamoto et al. 2006).

Additionally, several reports link cohesin regulatory/accessory proteins to tumorigenesis. Over-expression of WAPL and ESCO2 is associated with tumor progression in cervical cancer and melanomas, respectively (Oikawa et al. 2008; Ryu et al. 2007). Separase upregulation is sufficient to induce tumorigenesis in mammary epithelial cells in a *p53* mutant background (Dorsett 2007). Interestingly, cohesin-associated Pds5 paralog APRIN (Pds5B) is also downregulated in a variety of cancers and loss of APRIN affects stem cell differentiation by disrupting Oct4, Nanog and SOX2 patterns (Denes et al. 2010). Furthermore, mutations in *NIPBL* are the leading cause of developmental disorders such as Cornelia deLange Syndrome (Dorsett 2007). CdLS patients display abnormal physical and mental development, including facial abnormalities, oesophageal defects, cardiac malformations and mental retardation (Dorsett 2007). Moreover, mutations in *ESCO2* are the cause of another developmental disorder, Roberts Syndrome. (Dorsett 2007).

It is abundantly clear that the cohesin pathway is a common target in a variety of cancer and developmental disorders. Its functions in cohesion, gene expression, DNA repair and recombination are certainly central to the maintenance of a healthy biological system. Further research on cohesin and its functions will help us understand the pathophysiology of cohesinopathies and cancer.

CHAPTER VII

REFERENCES

Anderson DE, Losada A, Erickson HP, Hirano T. 2002. Condensin and cohesin display different arm conformations with characteristic hinge angles. *The Journal of cell biology* **156**: 419-424.

Ansari A, Hampsey M. 2005. A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. *Genes & development* **19**: 2969-2978.

Aparicio JG, Viggiani CJ, Gibson DG, Aparicio OM. 2004. The Rpd3-Sin3 histone deacetylase regulates replication timing and enables intra-S origin control in Saccharomyces cerevisiae. *Molecular and cellular biology* **24**: 4769-4780.

Apostolou E, Thanos D. 2008. Virus Infection Induces NF-kappaB-dependent interchromosomal associations mediating monoallelic IFN-beta gene expression. *Cell* **134**: 85-96.

Arumugam P, Gruber S, Tanaka K, Haering CH, Mechtler K, Nasmyth K. 2003. ATP hydrolysis is required for cohesin's association with chromosomes. *Current biology : CB* **13**: 1941-1953.

Asp E, Sunnerhagen P. 2003. Mkp1 and Mkp2, two MAPKAP-kinase homologues in Schizosaccharomyces pombe, interact with the MAP kinase Sty1. *Molecular genetics and genomics : MGG* **268**: 585-597.

Augui S, Filion GJ, Huart S, Nora E, Guggiari M, Maresca M, Stewart AF, Heard E. 2007. Sensing X chromosome pairs before X inactivation via a novel X-pairing region of the Xic. *Science* **318**: 1632-1636.

Azvolinsky A, Giresi PG, Lieb JD, Zakian VA. 2009. Highly transcribed RNA polymerase II genes are impediments to replication fork progression in Saccharomyces cerevisiae. *Molecular cell* **34**: 722-734.

Bacher CP, Guggiari M, Brors B, Augui S, Clerc P, Avner P, Eils R, Heard E. 2006. Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. *Nature cell biology* **8**: 293-299.

Bantignies F, Cavalli G. 2011. Polycomb group proteins: repression in 3D. *Trends in genetics : TIG* **27**: 454-464.

Barber TD, McManus K, Yuen KW, 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. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 3443-3448.

Bau D, Sanyal A, Lajoie BR, Capriotti E, Byron M, Lawrence JB, Dekker J, Marti-Renom MA. 2011. The three-dimensional folding of the alpha-globin gene

domain reveals formation of chromatin globules. *Nature structural & molecular biology* **18**: 107-114.

Bell AC, West AG, Felsenfeld G. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98**: 387-396.

Bender MA, Mehaffey MG, Telling A, Hug B, Ley TJ, Groudine M, Fiering S. 2000. Independent formation of DnaseI hypersensitive sites in the murine beta-globin locus control region. *Blood* **95**: 3600-3604.

Bernard P, Drogat J, Maure JF, Dheur S, Vaur S, Genier S, Javerzat JP. 2006. A screen for cohesion mutants uncovers Ssl3, the fission yeast counterpart of the cohesin loading factor Scc4. *Current biology : CB* **16**: 875-881.

Bernard P, Maure JF, Partridge JF, Genier S, Javerzat JP, Allshire RC. 2001. Requirement of heterochromatin for cohesion at centromeres. *Science* **294**: 2539-2542.

Binnie A, Castelo-Branco P, Monks J, Proudfoot NJ. 2006. Homologous gene sequences mediate transcription-domain formation. *Journal of cell science* **119**: 3876-3887.

Blanton J, Gaszner M, Schedl P. 2003. Protein:protein interactions and the pairing of boundary elements in vivo. *Genes & development* **17**: 664-675.

Blatch GL, Lassle M. 1999. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *BioEssays : news and reviews in molecular, cellular and developmental biology* **21**: 932-939.

Borges V, Smith DJ, Whitehouse I, Uhlmann F. 2013. An Eco1-independent sister chromatid cohesion establishment pathway in S. cerevisiae. *Chromosoma* **122**: 121-134.

Branco MR, Pombo A. 2006. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS biology* **4**: e138.

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

Cai S, Lee CC, Kohwi-Shigematsu T. 2006. SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nature genetics* **38**: 1278-1288.

Calvo O, Manley JL. 2001. Evolutionarily conserved interaction between CstF-64 and PC4 links transcription, polyadenylation, and termination. *Molecular cell* **7**: 1013-1023.

Cancer Genome Atlas Research N. 2013. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *The New England journal of medicine* **368**: 2059-2074.

Canudas S, Smith S. 2009. Differential regulation of telomere and centromere cohesion by the Scc3 homologues SA1 and SA2, respectively, in human cells. *The Journal of cell biology* **187**: 165-173.

Carter D, Chakalova L, Osborne CS, Dai YF, Fraser P. 2002. Long-range chromatin regulatory interactions in vivo. *Nature genetics* **32**: 623-626.

Chan KL, Gligoris T, Upcher W, Kato Y, Shirahige K, Nasmyth K, Beckouet F. 2013. Pds5 promotes and protects cohesin acetylation. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 13020-13025.

Chan KL, Roig MB, Hu B, Beckouet F, Metson J, Nasmyth K. 2012. Cohesin's DNA exit gate is distinct from its entrance gate and is regulated by acetylation. *Cell* **150**: 961-974.

Chang TT, Hughes-Fulford M. 2009. Monolayer and spheroid culture of human liver hepatocellular carcinoma cell line cells demonstrate distinct global gene expression patterns and functional phenotypes. *Tissue engineering Part A* **15**: 559-567.

Chen Z, McCrosky S, Guo W, Li H, Gerton JL. 2012. A genetic screen to discover pathways affecting cohesin function in Schizosaccharomyces pombe identifies chromatin effectors. *G3* **2**: 1161-1168.

Chien R, Zeng W, Ball AR, Yokomori K. 2011. Cohesin: a critical chromatin organizer in mammalian gene regulation. *Biochemistry and cell biology = Biochimie et biologie cellulaire* **89**: 445-458.

Chubb JR, Trcek T, Shenoy SM, Singer RH. 2006. Transcriptional pulsing of a developmental gene. *Current biology : CB* **16**: 1018-1025.

Ciosk R, Shirayama M, Shevchenko A, Tanaka T, Toth A, Shevchenko A, Nasmyth K. 2000. Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. *Molecular cell* **5**: 243-254.

Cisse, II, Izeddin I, Causse SZ, Boudarene L, Senecal A, Muresan L, Dugast-Darzacq C, Hajj B, Dahan M, Darzacq X. 2013. Real-time dynamics of RNA polymerase II clustering in live human cells. *Science* **341**: 664-667.

Croft JA, Bridger JM, Boyle S, Perry P, Teague P, Bickmore WA. 1999. Differences in the localization and morphology of chromosomes in the human nucleus. *The Journal of cell biology* **145**: 1119-1131.

D'Ambrosio C, Schmidt CK, Katou Y, Kelly G, Itoh T, Shirahige K, Uhlmann F. 2008. Identification of cis-acting sites for condensin loading onto budding yeast chromosomes. *Genes & development* **22**: 2215-2227.

Dalby MJ, Gadegaard N, Herzyk P, Sutherland D, Agheli H, Wilkinson CD, Curtis AS. 2007. Nanomechanotransduction and interphase nuclear organization influence on genomic control. *Journal of cellular biochemistry* **102**: 1234-1244.

Danis E, Brodolin K, Menut S, Maiorano D, Girard-Reydet C, Mechali M. 2004. Specification of a DNA replication origin by a transcription complex. *Nature cell biology* **6**: 721-730.

Dantonel JC, Murthy KG, Manley JL, Tora L. 1997. Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. *Nature* **389**: 399-402.

Deardorff MA, Kaur M, Yaeger D, Rampuria A, Korolev S, Pie J, Gil-Rodriguez C, Arnedo M, Loeys B, Kline AD et al. 2007. Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of cornelia de Lange syndrome with predominant mental retardation. *American journal of human genetics* **80**: 485-494.

Denes V, Pilichowska M, Makarovskiy A, Carpinito G, Geck P. 2010. Loss of a cohesin-linked suppressor APRIN (Pds5b) disrupts stem cell programs in embryonal carcinoma: an emerging cohesin role in tumor suppression. *Oncogene* **29**: 3446-3452.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**: 376-380.

Dorsett D. 2007. Roles of the sister chromatid cohesion apparatus in gene expression, development, and human syndromes. *Chromosoma* **116**: 1-13.

-. 2009. Cohesin, gene expression and development: lessons from Drosophila. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **17**: 185-200.

Dorsett D, Eissenberg JC, Misulovin Z, Martens A, Redding B, McKim K. 2005. Effects of sister chromatid cohesion proteins on cut gene expression during wing development in Drosophila. *Development* **132**: 4743-4753.

Dorsett D, Merkenschlager M. 2013. Cohesin at active genes: a unifying theme for cohesin and gene expression from model organisms to humans. *Current opinion in cell biology* **25**: 327-333.

Dostie J, Zhan Y, Dekker J. 2007. Chromosome conformation capture carbon copy technology. *Current protocols in molecular biology / edited by Frederick M Ausubel [et al]* **Chapter 21**: Unit 21 14.

Drissen R, Palstra RJ, Gillemans N, Splinter E, Grosveld F, Philipsen S, de Laat W. 2004. The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes & development* **18**: 2485-2490.

Duan Z, Andronescu M, Schutz K, McIlwain S, Kim YJ, Lee C, Shendure J, Fields S, Blau CA, Noble WS. 2010. A three-dimensional model of the yeast genome. *Nature* **465**: 363-367.

Eichinger CS, Kurze A, Oliveira RA, Nasmyth K. 2013. Disengaging the Smc3/kleisin interface releases cohesin from Drosophila chromosomes during interphase and mitosis. *The EMBO journal* **32**: 656-665.

Errico A, Cosentino C, Rivera T, Losada A, Schwob E, Hunt T, Costanzo V. 2009. Tipin/Tim1/And1 protein complex promotes Pol alpha chromatin binding and sister chromatid cohesion. *The EMBO journal* **28**: 3681-3692.

Eskiw CH, Rapp A, Carter DR, Cook PR. 2008. RNA polymerase II activity is located on the surface of protein-rich transcription factories. *Journal of cell science* **121**: 1999-2007.

Fay A, Misulovin Z, Li J, Schaaf CA, Gause M, Gilmour DS, Dorsett D. 2011. Cohesin selectively binds and regulates genes with paused RNA polymerase. *Current biology : CB* **21**: 1624-1634.

Fernius J, Marston AL. 2009. Establishment of cohesion at the pericentromere by the Ctf19 kinetochore subcomplex and the replication fork-associated factor, Csm3. *PLoS genetics* **5**: e1000629.

Ferraiuolo MA, Rousseau M, Miyamoto C, Shenker S, Wang XQ, Nadler M, Blanchette M, Dostie J. 2010. The three-dimensional architecture of Hox cluster silencing. *Nucleic acids research* **38**: 7472-7484.

Fraser J, Rousseau M, Shenker S, Ferraiuolo MA, Hayashizaki Y, Blanchette M, Dostie J. 2009. Chromatin conformation signatures of cellular differentiation. *Genome biology* **10**: R37.

Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH et al. 2009. An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462**: 58-64.

Furuya K, Takahashi K, Yanagida M. 1998. Faithful anaphase is ensured by Mis4, a sister chromatid cohesion molecule required in S phase and not destroyed in G1 phase. *Genes & development* **12**: 3408-3418.

Gaszner M, Felsenfeld G. 2006. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nature reviews Genetics* **7**: 703-713.

Gause M, Misulovin Z, Bilyeu A, Dorsett D. 2010. Dosage-sensitive regulation of cohesin chromosome binding and dynamics by Nipped-B, Pds5, and Wapl. *Molecular and cellular biology* **30**: 4940-4951.

Gerlich D, Koch B, Dupeux F, Peters JM, Ellenberg J. 2006. Live-cell imaging reveals a stable cohesin-chromatin interaction after but not before DNA replication. *Current biology : CB* **16**: 1571-1578.

Ghamari A, van de Corput MP, Thongjuea S, van Cappellen WA, van Ijcken W, van Haren J, Soler E, Eick D, Lenhard B, Grosveld FG. 2013. In vivo live imaging of RNA polymerase II transcription factories in primary cells. *Genes & development* **27**: 767-777.

Gheldof N, Smith EM, Tabuchi TM, Koch CM, Dunham I, Stamatoyannopoulos JA, Dekker J. 2010. Cell-type-specific long-range looping interactions identify distant regulatory elements of the CFTR gene. *Nucleic acids research* **38**: 4325-4336.

Gilbert DM. 2002. Replication timing and transcriptional control: beyond cause and effect. *Current opinion in cell biology* **14**: 377-383.

Gillespie PJ, Hirano T. 2004. Scc2 couples replication licensing to sister chromatid cohesion in Xenopus egg extracts. *Current biology : CB* **14**: 1598-1603.

Glover-Cutter K, Kim S, Espinosa J, Bentley DL. 2008. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nature structural & molecular biology* **15**: 71-78.

Glynn EF, Megee PC, Yu HG, Mistrot C, Unal E, Koshland DE, DeRisi JL, Gerton JL. 2004. Genome-wide mapping of the cohesin complex in the yeast Saccharomyces cerevisiae. *PLoS biology* **2**: E259.

Grontved L, Hager GL. 2012. Impact of chromatin structure on PR signaling: transition from local to global analysis. *Molecular and cellular endocrinology* **357**: 30-36.

Grosveld F. 1999. Activation by locus control regions? *Current opinion in genetics & development* **9**: 152-157.

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

Gruber S, Haering CH, Nasmyth K. 2003. Chromosomal cohesin forms a ring. *Cell* **112**: 765-777.

Gullerova M, Proudfoot NJ. 2008. Cohesin complex promotes transcriptional termination between convergent genes in S. pombe. *Cell* **132**: 983-995.

Hadjur S, Williams LM, Ryan NK, Cobb BS, Sexton T, Fraser P, Fisher AG, Merkenschlager M. 2009. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* **460**: 410-413.

Haering CH. 2009. Foreword: the many fascinating functions of SMC protein complexes. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **17**: 127-129.

Haering CH, Farcas AM, Arumugam P, Metson J, Nasmyth K. 2008. The cohesin ring concatenates sister DNA molecules. *Nature* **454**: 297-301.

Haering CH, Schoffnegger D, Nishino T, Helmhart W, Nasmyth K, Lowe J. 2004. Structure and stability of cohesin's Smc1-kleisin interaction. *Molecular cell* **15**: 951-964.

Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, Lee CW, Ye C, Ping JL, Mulawadi F et al. 2011. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nature genetics* **43**: 630-638.

Hanna JS, Kroll ES, Lundblad V, Spencer FA. 2001. Saccharomyces cerevisiae CTF18 and CTF4 are required for sister chromatid cohesion. *Molecular and cellular biology* **21**: 3144-3158.

Hauf S, Roitinger E, Koch B, Dittrich CM, Mechtler K, Peters JM. 2005. Dissociation of cohesin from chromosome arms and loss of arm cohesion during early mitosis depends on phosphorylation of SA2. *PLoS biology* **3**: e69.

Hauf S, Waizenegger IC, Peters JM. 2001. Cohesin cleavage by separase required for anaphase and cytokinesis in human cells. *Science* **293**: 1320-1323.

Hayashi MT, Takahashi TS, Nakagawa T, Nakayama J, Masukata H. 2009. The heterochromatin protein Swi6/HP1 activates replication origins at the pericentromeric region and silent mating-type locus. *Nature cell biology* **11**: 357-362.

He X, Khan AU, Cheng H, Pappas DL, Jr., Hampsey M, Moore CL. 2003. Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1. *Genes & development* **17**: 1030-1042.

Hegemann B, Hutchins JR, Hudecz O, Novatchkova M, Rameseder J, Sykora MM, Liu S, Mazanek M, Lenart P, Heriche JK et al. 2011. Systematic

phosphorylation analysis of human mitotic protein complexes. *Science signaling* **4**: rs12.

Heichinger C, Penkett CJ, Bahler J, Nurse P. 2006. Genome-wide characterization of fission yeast DNA replication origins. *The EMBO journal* **25**: 5171-5179.

Heidinger-Pauli JM, Mert O, Davenport C, Guacci V, Koshland D. 2010a. Systematic reduction of cohesin differentially affects chromosome segregation, condensation, and DNA repair. *Current biology : CB* **20**: 957-963.

Heidinger-Pauli JM, Onn I, Koshland D. 2010b. Genetic evidence that the acetylation of the Smc3p subunit of cohesin modulates its ATP-bound state to promote cohesion establishment in Saccharomyces cerevisiae. *Genetics* **185**: 1249-1256.

Heidinger-Pauli JM, Unal E, Guacci V, Koshland D. 2008. The kleisin subunit of cohesin dictates damage-induced cohesion. *Molecular cell* **31**: 47-56.

Hieda M, Winstanley H, Maini P, Iborra FJ, Cook PR. 2005. Different populations of RNA polymerase II in living mammalian cells. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **13**: 135-144.

Hiratani I, Takebayashi S, Lu J, Gilbert DM. 2009. Replication timing and transcriptional control: beyond cause and effect--part II. *Current opinion in genetics & development* **19**: 142-149.

Horike S, Cai S, Miyano M, Cheng JF, Kohwi-Shigematsu T. 2005. Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nature genetics* **37**: 31-40.

Hu B, Itoh T, Mishra A, Katoh Y, Chan KL, Upcher W, Godlee C, Roig MB, Shirahige K, Nasmyth K. 2011. ATP hydrolysis is required for relocating cohesin from sites occupied by its Scc2/4 loading complex. *Current biology : CB* **21**: 12-24.

Hunt P, Hassold T. 2010. Female meiosis: coming unglued with age. *Current biology : CB* **20**: R699-702.

Hyrien O, Marheineke K, Goldar A. 2003. Paradoxes of eukaryotic DNA replication: MCM proteins and the random completion problem. *BioEssays : news and reviews in molecular, cellular and developmental biology* **25**: 116-125.

Iborra FJ, Pombo A, Jackson DA, Cook PR. 1996. Active RNA polymerases are localized within discrete transcription "factories' in human nuclei. *Journal of cell science* **109 (Pt 6)**: 1427-1436.

Ivanov D, Nasmyth K. 2005. A topological interaction between cohesin rings and a circular minichromosome. *Cell* **122**: 849-860.

Jackson DA, Cook PR. 1985. A general method for preparing chromatin containing intact DNA. *The EMBO journal* **4**: 913-918.

Jackson DA, Iborra FJ, Manders EM, Cook PR. 1998. Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Molecular biology of the cell* **9**: 1523-1536.

Kadener S, Fededa JP, Rosbash M, Kornblihtt AR. 2002. Regulation of alternative splicing by a transcriptional enhancer through RNA pol II elongation. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 8185-8190.

Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS et al. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**: 430-435.

Kaur M, DeScipio C, McCallum J, Yaeger D, Devoto M, Jackson LG, Spinner NB, Krantz ID. 2005. Precocious sister chromatid separation (PSCS) in Cornelia de Lange syndrome. *American journal of medical genetics Part A* **138**: 27-31.

Kim JM, Yamada M, Masai H. 2003. Functions of mammalian Cdc7 kinase in initiation/monitoring of DNA replication and development. *Mutation research* **532**: 29-40.

Kimura H, Tao Y, Roeder RG, Cook PR. 1999. Quantitation of RNA polymerase II and its transcription factors in an HeLa cell: little soluble holoenzyme but significant amounts of polymerases attached to the nuclear substructure. *Molecular and cellular biology* **19**: 5383-5392.

Knaus R, Pollock R, Guarente L. 1996. Yeast SUB1 is a suppressor of TFIIB mutations and has homology to the human co-activator PC4. *The EMBO journal* **15**: 1933-1940.

Kogut I, Wang J, Guacci V, Mistry RK, Megee PC. 2009. The Scc2/Scc4 cohesin loader determines the distribution of cohesin on budding yeast chromosomes. *Genes & development* **23**: 2345-2357.

Krantz ID, McCallum J, DeScipio C, Kaur M, Gillis LA, Yaeger D, Jukofsky L, Wasserman N, Bottani A, Morris CA et al. 2004. Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B. *Nature genetics* **36**: 631-635.

Kurukuti S, Tiwari VK, Tavoosidana G, Pugacheva E, Murrell A, Zhao Z, Lobanenkov V, Reik W, Ohlsson R. 2006. CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to

restrict enhancer access to Igf2. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 10684-10689.

Lafont AL, Song J, Rankin S. 2010. Sororin cooperates with the acetyltransferase Eco2 to ensure DNA replication-dependent sister chromatid cohesion. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 20364-20369.

Laine JP, Singh BN, Krishnamurthy S, Hampsey M. 2009. A physiological role for gene loops in yeast. *Genes & development* **23**: 2604-2609.

Larkin JD, Cook PR, Papantonis A. 2012. Dynamic reconfiguration of long human genes during one transcription cycle. *Molecular and cellular biology* **32**: 2738-2747.

Larkin JD, Papantonis A, Cook PR, Marenduzzo D. 2013. Space exploration by the promoter of a long human gene during one transcription cycle. *Nucleic acids research* **41**: 2216-2227.

Lee GR, Fields PE, Griffin TJ, Flavell RA. 2003. Regulation of the Th2 cytokine locus by a locus control region. *Immunity* **19**: 145-153.

Lengronne A, Katou Y, Mori S, Yokobayashi S, Kelly GP, Itoh T, Watanabe Y, Shirahige K, Uhlmann F. 2004. Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature* **430**: 573-578.

Lengronne A, McIntyre J, Katou Y, Kanoh Y, Hopfner KP, Shirahige K, Uhlmann F. 2006. Establishment of sister chromatid cohesion at the S. cerevisiae replication fork. *Molecular cell* **23**: 787-799.

Lewis A, Murrell A. 2004. Genomic imprinting: CTCF protects the boundaries. *Current biology : CB* **14**: R284-286.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**: 289-293.

Ling JQ, Li T, Hu JF, Vu TH, Chen HL, Qiu XW, Cherry AM, Hoffman AR. 2006. CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science* **312**: 269-272.

Liu J, Krantz ID. 2009. Cornelia de Lange syndrome, cohesin, and beyond. *Clinical genetics* **76**: 303-314.

Liu Z, Scannell DR, Eisen MB, Tjian R. 2011. Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. *Cell* **146**: 720-731.

Lopez-Serra L, Lengronne A, Borges V, Kelly G, Uhlmann F. 2013. Budding yeast Wapl controls sister chromatid cohesion maintenance and chromosome condensation. *Current biology : CB* **23**: 64-69.

Losada A, Yokochi T, Hirano T. 2005. Functional contribution of Pds5 to cohesin-mediated cohesion in human cells and Xenopus egg extracts. *Journal of cell science* **118**: 2133-2141.

Losada A, Yokochi T, Kobayashi R, Hirano T. 2000. Identification and characterization of SA/Scc3p subunits in the Xenopus and human cohesin complexes. *The Journal of cell biology* **150**: 405-416.

Lyons NA, Morgan DO. 2011. Cdk1-dependent destruction of Eco1 prevents cohesion establishment after S phase. *Molecular cell* **42**: 378-389.

Ma Z, Lin M, Li K, Fu Y, Liu X, Yang D, Zhao Y, Zheng J, Sun B. 2013. Knocking down SMC1A inhibits growth and leads to G2/M arrest in human glioma cells. *International journal of clinical and experimental pathology* **6**: 862-869.

Majumder P, Boss JM. 2011. Cohesin regulates MHC class II genes through interactions with MHC class II insulators. *Journal of immunology* **187**: 4236-4244.

Masui O, Bonnet I, Le Baccon P, Brito I, Pollex T, Murphy N, Hupe P, Barillot E, Belmont AS, Heard E. 2011. Live-cell chromosome dynamics and outcome of X chromosome pairing events during ES cell differentiation. *Cell* **145**: 447-458.

Mayer ML, Pot I, Chang M, Xu H, Aneliunas V, Kwok T, Newitt R, Aebersold R, Boone C, Brown GW et al. 2004. Identification of protein complexes required for efficient sister chromatid cohesion. *Molecular biology of the cell* **15**: 1736-1745.

Mc Intyre J, Muller EG, Weitzer S, Snydsman BE, Davis TN, Uhlmann F. 2007. In vivo analysis of cohesin architecture using FRET in the budding yeast Saccharomyces cerevisiae. *The EMBO journal* **26**: 3783-3793.

McGuinness BE, Hirota T, Kudo NR, Peters JM, Nasmyth K. 2005. Shugoshin prevents dissociation of cohesin from centromeres during mitosis in vertebrate cells. *PLoS biology* **3**: e86.

McNairn AJ, Gerton JL. 2009. Intersection of ChIP and FLIP, genomic methods to study the dynamics of the cohesin proteins. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **17**: 155-163.

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

Mishiro T, Ishihara K, Hino S, Tsutsumi S, Aburatani H, Shirahige K, Kinoshita Y, Nakao M. 2009. Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *The EMBO journal* **28**: 1234-1245.

Mishra A, Hu B, Kurze A, Beckouet F, Farcas AM, Dixon SE, Katou Y, Khalid S, Shirahige K, Nasmyth K. 2010. Both interaction surfaces within cohesin's hinge domain are essential for its stable chromosomal association. *Current biology : CB* **20**: 279-289.

Misteli T. 2007. Beyond the sequence: cellular organization of genome function. *Cell* **128**: 787-800.

Misulovin Z, Schwartz YB, Li XY, Kahn TG, Gause M, MacArthur S, Fay JC, Eisen MB, Pirrotta V, Biggin MD et al. 2008. Association of cohesin and Nipped-B with transcriptionally active regions of the Drosophila melanogaster genome. *Chromosoma* **117**: 89-102.

Mitchell JA, Fraser P. 2008. Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes & development* **22**: 20-25.

Moldovan GL, Pfander B, Jentsch S. 2006. PCNA controls establishment of sister chromatid cohesion during S phase. *Molecular cell* **23**: 723-732.

Moreno S, Klar A, Nurse P. 1991. Molecular genetic analysis of fission yeast Schizosaccharomyces pombe. *Methods in enzymology* **194**: 795-823.

Muto A, Calof AL, Lander AD, Schilling TF. 2011. Multifactorial origins of heart and gut defects in nipbl-deficient zebrafish, a model of Cornelia de Lange Syndrome. *PLoS biology* **9**: e1001181.

Nasmyth K. 2011. Cohesin: a catenase with separate entry and exit gates? *Nature cell biology* **13**: 1170-1177.

Nativio R, Wendt KS, Ito Y, Huddleston JE, Uribe-Lewis S, Woodfine K, Krueger C, Reik W, Peters JM, Murrell A. 2009. Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. *PLoS genetics* **5**: e1000739.

Neuwald AF, Hirano T. 2000. HEAT repeats associated with condensins, cohesins, and other complexes involved in chromosome-related functions. *Genome research* **10**: 1445-1452.

Nie Z, Hu G, Wei G, Cui K, Yamane A, Resch W, Wang R, Green DR, 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 AA, 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 EM, Hubner NC, 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.

Nonaka N, Kitajima T, Yokobayashi S, Xiao G, Yamamoto M, Grewal SI, Watanabe Y. 2002. Recruitment of cohesin to heterochromatic regions by Swi6/HP1 in fission yeast. *Nature cell biology* **4**: 89-93.

Noordermeer D, de Wit E, Klous P, van de Werken H, Simonis M, Lopez-Jones M, Eussen B, de Klein A, Singer RH, de Laat W. 2011a. Variegated gene expression caused by cell-specific long-range DNA interactions. *Nature cell biology* **13**: 944-951.

Noordermeer D, Leleu M, Splinter E, Rougemont J, De Laat W, Duboule D. 2011b. The dynamic architecture of Hox gene clusters. *Science* **334**: 222-225.

Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J et al. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**: 381-385.

O'Reilly D, Greaves DR. 2007. Cell-type-specific expression of the human CD68 gene is associated with changes in Pol II phosphorylation and short-range intrachromosomal gene looping. *Genomics* **90**: 407-415.

O'Sullivan JM, Tan-Wong SM, Morillon A, Lee B, Coles J, Mellor J, Proudfoot NJ. 2004. Gene loops juxtapose promoters and terminators in yeast. *Nature genetics* **36**: 1014-1018.

Oikawa K, Akiyoshi A, Tanaka M, Takanashi M, Nishi H, Isaka K, Kiseki H, Idei T, Tsukahara Y, Hashimura N et al. 2008. Expression of various types of alternatively spliced WAPL transcripts in human cervical epithelia. *Gene* **423**: 57-62.

Onn I, Koshland D. 2011. In vitro assembly of physiological cohesin/DNA complexes. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 12198-12205.

Osborne CS, Chakalova L, Brown KE, Carter D, Horton A, Debrand E, Goyenechea B, Mitchell JA, Lopes S, Reik W et al. 2004. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nature genetics* **36**: 1065-1071.

Osborne CS, Chakalova L, Mitchell JA, Horton A, Wood AL, Bolland DJ, Corcoran AE, Fraser P. 2007. Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS biology* **5**: e192.

Panigrahi AK, Zhang N, Otta SK, Pati D. 2012. A cohesin-RAD21 interactome. *The Biochemical journal* **442**: 661-670.

Panizza S, Tanaka T, Hochwagen A, Eisenhaber F, Nasmyth K. 2000. Pds5 cooperates with cohesin in maintaining sister chromatid cohesion. *Current biology : CB* **10**: 1557-1564.

Papantonis A, Cook PR. 2010. Genome architecture and the role of transcription. *Current opinion in cell biology* **22**: 271-276.

Papantonis A, Kohro T, Baboo S, Larkin JD, Deng B, Short P, Tsutsumi S, Taylor S, Kanki Y, Kobayashi M et al. 2012. TNFalpha signals through specialized factories where responsive coding and miRNA genes are transcribed. *The EMBO journal* **31**: 4404-4414.

Papantonis A, Larkin JD, Wada Y, Ohta Y, Ihara S, Kodama T, Cook PR. 2010. Active RNA polymerases: mobile or immobile molecular machines? *PLoS biology* **8**: e1000419.

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

Patrinos GP, de Krom M, de Boer E, Langeveld A, Imam AM, Strouboulis J, de Laat W, Grosveld FG. 2004. Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. *Genes & development* **18**: 1495-1509.

Pauli A, van Bemmel JG, Oliveira RA, Itoh T, Shirahige K, van Steensel B, Nasmyth K. 2010. A direct role for cohesin in gene regulation and ecdysone response in Drosophila salivary glands. *Current biology : CB* **20**: 1787-1798.

Perkins KJ, Lusic M, Mitar I, Giacca M, Proudfoot NJ. 2008. Transcription-dependent gene looping of the HIV-1 provirus is dictated by recognition of pre-mRNA processing signals. *Molecular cell* **29**: 56-68.

Peters JM, Tedeschi A, Schmitz J. 2008. The cohesin complex and its roles in chromosome biology. *Genes & development* **22**: 3089-3114.

Pombo A, Hollinshead M, Cook PR. 1999. Bridging the resolution gap: Imaging the same transcription factories in cryosections by light and electron microscopy. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* **47**: 471-480.

Pombo A, Jones E, Iborra FJ, Kimura H, Sugaya K, Cook PR, Jackson DA. 2000. Specialized transcription factories within mammalian nuclei. *Critical reviews in eukaryotic gene expression* **10**: 21-29.

Ragoczy T, Bender MA, Telling A, Byron R, Groudine M. 2006. The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation. *Genes & development* **20**: 1447-1457.

Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. 2006. Stochastic mRNA synthesis in mammalian cells. *PLoS biology* **4**: e309.

Rhodes DR, Chinnaiyan AM. 2004. Bioinformatics strategies for translating genome-wide expression analyses into clinically useful cancer markers. *Annals of the New York Academy of Sciences* **1020**: 32-40.

Rhodes JM, Bentley FK, Print CG, Dorsett D, Misulovin Z, Dickinson EJ, Crosier KE, Crosier PS, Horsfield JA. 2010. Positive regulation of c-Myc by cohesin is direct, and evolutionarily conserved. *Developmental biology* **344**: 637-649.

Riedel CG, Katis VL, Katou Y, Mori S, Itoh T, Helmhart W, Galova M, Petronczki M, Gregan J, Cetin B et al. 2006. Protein phosphatase 2A protects centromeric sister chromatid cohesion during meiosis I. *Nature* **441**: 53-61.

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

Rollins RA, Korom M, Aulner N, Martens A, Dorsett D. 2004. Drosophila nipped-B protein supports sister chromatid cohesion and opposes the stromalin/Scc3 cohesion factor to facilitate long-range activation of the cut gene. *Molecular and cellular biology* **24**: 3100-3111.

Rollins RA, Morcillo P, 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.

Rowland BD, Roig MB, Nishino T, Kurze A, Uluocak P, Mishra A, Beckouet F, Underwood P, Metson J, Imre R et al. 2009. Building sister chromatid cohesion: smc3 acetylation counteracts an antiestablishment activity. *Molecular cell* **33**: 763-774.

Rudra S, Skibbens RV. 2013. Cohesin codes - interpreting chromatin architecture and the many facets of cohesin function. *Journal of cell science* **126**: 31-41.

Ryu B, Kim DS, Deluca AM, Alani RM. 2007. Comprehensive expression profiling of tumor cell lines identifies molecular signatures of melanoma progression. *PloS one* **2**: e594.

Saitoh N, Bell AC, Recillas-Targa F, West AG, Simpson M, Pikaart M, Felsenfeld G. 2000. Structural and functional conservation at the boundaries of the chicken beta-globin domain. *The EMBO journal* **19**: 2315-2322.

Schaaf CA, Kwak H, Koenig A, Misulovin Z, Gohara DW, Watson A, Zhou Y, Lis JT, Dorsett D. 2013. Genome-wide control of RNA polymerase II activity by cohesin. *PLoS genetics* **9**: e1003382.

Schmidt CK, Brookes N, Uhlmann F. 2009. Conserved features of cohesin binding along fission yeast chromosomes. *Genome biology* **10**: R52.

Schmidt D, Schwalie PC, Ross-Innes CS, Hurtado A, Brown GD, Carroll JS, Flicek P, Odom DT. 2010. A CTCF-independent role for cohesin in tissue-specific transcription. *Genome research* **20**: 578-588.

Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S, Kurukuti S, Mitchell JA, Umlauf D, Dimitrova DS et al. 2010. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nature genetics* **42**: 53-61.

Schubeler D, Scalzo D, Kooperberg C, van Steensel B, Delrow J, Groudine M. 2002. Genome-wide DNA replication profile for Drosophila melanogaster: a link between transcription and replication timing. *Nature genetics* **32**: 438-442.

Schuldiner O, Berdnik D, Levy JM, Wu JS, Luginbuhl D, Gontang AC, Luo L. 2008. piggyBac-based mosaic screen identifies a postmitotic function for cohesin in regulating developmental axon pruning. *Developmental cell* **14**: 227-238.

Seitan VC, Banks P, Laval S, Majid NA, 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 biology* **4**: e242.

Seitan VC, Hao B, Tachibana-Konwalski K, Lavagnolli T, Mira-Bontenbal H, Brown KE, 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 VC, Merkenschlager M. 2012. Cohesin and chromatin organisation. *Current opinion in genetics & development* **22**: 93-100.

Shintomi K, Hirano T. 2009. Releasing cohesin from chromosome arms in early mitosis: opposing actions of Wapl-Pds5 and Sgo1. *Genes & development* **23**: 2224-2236.

Singh BN, Ansari A, Hampsey M. 2009. Detection of gene loops by 3C in yeast. *Methods* **48**: 361-367.

Skibbens RV. 2009. Establishment of sister chromatid cohesion. *Current biology : CB* **19**: R1126-1132.

Skibbens RV, Corson LB, Koshland D, Hieter P. 1999. Ctf7p is essential for sister chromatid cohesion and links mitotic chromosome structure to the DNA replication machinery. *Genes & development* **13**: 307-319.

Solomon DA, Kim T, Diaz-Martinez LA, Fair J, Elkahloun AG, Harris BT, Toretsky JA, Rosenberg SA, Shukla N, Ladanyi M et al. 2011. Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science* **333**: 1039-1043.

Song SH, Hou C, Dean A. 2007. A positive role for NLI/Ldb1 in long-range beta-globin locus control region function. *Molecular cell* **28**: 810-822.

Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA. 2005. Interchromosomal associations between alternatively expressed loci. *Nature* **435**: 637-645.

Stamatoyannopoulos G. 2005. Prospects for developing a molecular cure for thalassemia. *Hematology* **10 Suppl 1**: 255-257.

Strachan T. 2005. Cornelia de Lange Syndrome and the link between chromosomal function, DNA repair and developmental gene regulation. *Current opinion in genetics & development* **15**: 258-264.

Sun ZW, Hampsey M. 1996. Synthetic enhancement of a TFIIB defect by a mutation in SSU72, an essential yeast gene encoding a novel protein that affects transcription start site selection in vivo. *Molecular and cellular biology* **16**: 1557-1566.

Supernat A, Lapinska-Szumczyk S, Sawicki S, Wydra D, Biernat W, Zaczek AJ. 2012. Deregulation of RAD21 and RUNX1 expression in endometrial cancer. *Oncology letters* **4**: 727-732.

Sutani T, Kawaguchi T, Kanno R, Itoh T, Shirahige K. 2009. Budding yeast Wpl1(Rad61)-Pds5 complex counteracts sister chromatid cohesion-establishing reaction. *Current biology : CB* **19**: 492-497.

Takahashi TS, Yiu P, Chou MF, Gygi S, Walter JC. 2004. Recruitment of Xenopus Scc2 and cohesin to chromatin requires the pre-replication complex. *Nature cell biology* **6**: 991-996.

Tan-Wong SM, French JD, Proudfoot NJ, Brown MA. 2008. Dynamic interactions between the promoter and terminator regions of the mammalian BRCA1 gene. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 5160-5165.

Tan-Wong SM, Wijayatilake HD, Proudfoot NJ. 2009. Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex. *Genes & development* **23**: 2610-2624.

Tan-Wong SM, Zaugg JB, Camblong J, Xu Z, Zhang DW, Mischo HE, Ansari AZ, Luscombe NM, Steinmetz LM, Proudfoot NJ. 2012. Gene loops enhance transcriptional directionality. *Science* **338**: 671-675.

Tanaka K, Yonekawa T, Kawasaki Y, Kai M, Furuya K, Iwasaki M, Murakami H, Yanagida M, Okayama H. 2000. Fission yeast Eso1p is required for establishing sister chromatid cohesion during S phase. *Molecular and cellular biology* **20**: 3459-3469.

Tanizawa H, Iwasaki O, Tanaka A, Capizzi JR, Wickramasinghe P, Lee M, Fu Z, Noma K. 2010. Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. *Nucleic acids research* **38**: 8164-8177.

Tatebayashi K, Kato J, Ikeda H. 1998. Isolation of a Schizosaccharomyces pombe rad21ts mutant that is aberrant in chromosome segregation, microtubule function, DNA repair and sensitive to hydroxyurea: possible involvement of Rad21 in ubiquitin-mediated proteolysis. *Genetics* **148**: 49-57.

Terret ME, Sherwood R, Rahman S, Qin J, Jallepalli PV. 2009. Cohesin acetylation speeds the replication fork. *Nature* **462**: 231-234.

Tiwari VK, Baylin SB. 2009. Combined 3C-ChIP-cloning (6C) assay: a tool to unravel protein-mediated genome architecture. *Cold Spring Harbor protocols* **2009**: pdb prot5168.

Tolhuis B, Blom M, Kerkhoven RM, Pagie L, Teunissen H, Nieuwland M, Simonis M, de Laat W, van Lohuizen M, van Steensel B. 2011. Interactions among Polycomb domains are guided by chromosome architecture. *PLoS genetics* **7**: e1001343.

Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. 2002. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Molecular cell* **10**: 1453-1465.

Tomonaga T, Nagao K, Kawasaki Y, Furuya K, Murakami A, Morishita J, Yuasa T, Sutani T, Kearsey SE, Uhlmann F et al. 2000. Characterization of fission yeast cohesin: essential anaphase proteolysis of Rad21 phosphorylated in the S phase. *Genes & development* **14**: 2757-2770.

Tonkin ET, Wang TJ, Lisgo S, Bamshad MJ, Strachan T. 2004. NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly

Nipped-B, is mutated in Cornelia de Lange syndrome. *Nature genetics* **36**: 636-641.

Toth A, Ciosk R, Uhlmann F, Galova M, Schleiffer A, Nasmyth K. 1999. Yeast cohesin complex requires a conserved protein, Eco1p(Ctf7), to establish cohesion between sister chromatids during DNA replication. *Genes & development* **13**: 320-333.

Tuduri S, Crabbe L, Conti C, Tourriere H, Holtgreve-Grez H, Jauch A, Pantesco V, De Vos J, Thomas A, Theillet C et al. 2009. Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nature cell biology* **11**: 1315-1324.

Uhlmann F, Lottspeich F, Nasmyth K. 1999. Sister-chromatid separation at anaphase onset is promoted by cleavage of the cohesin subunit Scc1. *Nature* **400**: 37-42.

Unal E, Heidinger-Pauli JM, Kim W, Guacci V, Onn I, Gygi SP, Koshland DE. 2008. A molecular determinant for the establishment of sister chromatid cohesion. *Science* **321**: 566-569.

Vakoc CR, Letting DL, Gheldof N, Sawado T, Bender MA, Groudine M, Weiss MJ, Dekker J, Blobel GA. 2005. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Molecular cell* **17**: 453-462.

van Steensel B, Dekker J. 2010. Genomics tools for unraveling chromosome architecture. *Nature biotechnology* **28**: 1089-1095.

Vernimmen D, De Gobbi M, Sloane-Stanley JA, Wood WG, Higgs DR. 2007. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *The EMBO journal* **26**: 2041-2051.

Waizenegger IC, Hauf S, Meinke A, Peters JM. 2000. Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase. *Cell* **103**: 399-410.

Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA et al. 2011. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**: 120-124.

Watrin E, Peters JM. 2006. Cohesin and DNA damage repair. *Experimental cell research* **312**: 2687-2693.

Wendt KS, 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.

West S, Gromak N, Proudfoot NJ. 2004. Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432**: 522-525.

White EJ, Emanuelsson O, Scalzo D, Royce T, Kosak S, Oakeley EJ, Weissman S, Gerstein M, Groudine M, Snyder M et al. 2004. DNA replication-timing analysis of human chromosome 22 at high resolution and different developmental states. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 17771-17776.

Wierzba J, Gil-Rodriguez MC, Polucha A, Puisac B, Arnedo M, Teresa-Rodrigo ME, Winnicka D, Hegardt FG, Ramos FJ, Limon J et al. 2012. Cornelia de Lange syndrome with NIPBL mutation and mosaic Turner syndrome in the same individual. *BMC medical genetics* **13**: 43.

Wu WH, Pinto I, Chen BS, Hampsey M. 1999. Mutational analysis of yeast TFIIB. A functional relationship between Ssu72 and Sub1/Tsp1 defined by allele-specific interactions with TFIIB. *Genetics* **153**: 643-652.

Xu H, Beasley M, Verschoor S, Inselman A, Handel MA, McKay MJ. 2004. A new role for the mitotic RAD21/SCC1 cohesin in meiotic chromosome cohesion and segregation in the mouse. *EMBO reports* **5**: 378-384.

Xu M, Cook PR. 2008. The role of specialized transcription factories in chromosome pairing. *Biochimica et biophysica acta* **1783**: 2155-2160.

Xu N, Tsai CL, Lee JT. 2006. Transient homologous chromosome pairing marks the onset of X inactivation. *Science* **311**: 1149-1152.

Xu Z, Cetin B, Anger M, Cho US, Helmhart W, Nasmyth K, Xu W. 2009. Structure and function of the PP2A-shugoshin interaction. *Molecular cell* **35**: 426-441.

Yamamoto G, Irie T, Aida T, Nagoshi Y, Tsuchiya R, Tachikawa T. 2006. Correlation of invasion and metastasis of cancer cells, and expression of the RAD21 gene in oral squamous cell carcinoma. *Virchows Archiv : an international journal of pathology* **448**: 435-441.

Zhang J, Shi X, Li Y, Kim BJ, Jia J, Huang Z, Yang T, Fu X, Jung SY, Wang Y et al. 2008. Acetylation of Smc3 by Eco1 is required for S phase sister chromatid cohesion in both human and yeast. *Molecular cell* **31**: 143-151.