

Keep in touch

The role of cohesin and CTCF in organizing the human genome

Jessica Zuin

ISBN: 978-94-6182-289-5

Printing: Off Page, www.offpage.nl

Copyright © 2013 Jessica Zuin. All right reserved. No part of this thesis may be reproduced or transmitted in any forms by any means, electronically or mechanical, including photocopying, recording or any information storage or retrieval system, without permission in writing from the author.

Printing of this thesis was kindly supported by J.E. Jurriaanse Stichting and the Erasmus University of Rotterdam.

Cover designed by Thomas Clapes and Jessica Zuin representing a way to "*Keep in touch*". Postcards from Venice, Padova, Tokyo, Paris, Toulouse, New York, London, Toronto.

Keep in touch

The role of cohesin and CTCF in organizing the human genome

Houd contact

De rol van cohesin en CTCF in het
organiseren van het humane genoom

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. H.G. Schmidt

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

dinsdag 25 Juni 2013 om 15.30 uur

door

Jessica Zuin

Geboren te Padova, Italy



Promotiecommissie:

Promotor: Prof.dr. F.G. Grosveld

Overige leden: Dr.ir. W.M. Baarends
Dr.ir. N.J. Galjart
Prof.dr. J.H. Gribnau

Copromotor: Dr. K.S. Wendt

Table of contents

Chapter 1 - General Introduction	p. 7
Chapter 2 - Multiplexed Chromosome Conformation Capture Sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions	p. 37
Chapter 3 - Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells in human cells	p. 57
Chapter 4 - A cohesin-independent role for NIPBL at the promoters provide insights in Cornelia de Lange Syndrome	p.105
Chapter 5 - General Discussion	p.155
Abbreviations	p.169
Summary/Samenvatting	p.170
Curriculum Vitae	p.172
PhD Portfolio	p.174
Acknowledgements	p.176

Chapter 1

General introduction

Nuclear architecture

Packaging of DNA into chromatin

The genome of higher eukaryotes is organized in a complex structure consisting of DNA and proteins. In human cells the DNA has a length of two meters, divided over 46 chromosomes. Since the nuclear diameter is only 10 μ m, DNA is folded and compacted by protein, forming the chromatin fiber. The basic unit of the chromatin is the nucleosome that is constituted by DNA and positively charged proteins called histones. The nucleosome core consists of 146bp DNA wrapped around a histone octamer formed by two molecules of each H2A, H2B, H3, and H4 histone proteins (Luger, Mader et al., 1997). Each of these histones has an N-terminal amino acid tail which extends out from the DNA-histone core and represents a site for numerous post-translational modifications, including acetylation, methylation, phosphorylation, ubiquitination and sumoylation. These modifications regulate different aspects of the chromatin structure, such as accessibility of the DNA to regulatory proteins like transcription factors. Core nucleosomes are linked by a small stretch of DNA known as the linker. This arrangement can be visualized by electron microscope as “beads on a string” (Oudet, Gross-Bellard et al., 1975) and is known as the 10nm chromatin fiber (based on its approximate diameter). Nucleosomes are stabilized by the binding of histone H1 to the DNA linker.

The 10nm chromatin fiber has to be further compacted to fit in the nucleus. For this reason it was proposed that this fiber might be additionally packed into a 30nm chromatin fiber. Studies in vitro and in vivo are still trying to prove the existence of the 30nm chromatin fiber (Fussner, Ching et al., 2011).

Concept of euchromatin and heterochromatin

Light microscopy studies revealed that in interphase the chromatin exists in two compaction states called *euchromatin* and *heterochromatin*.

Euchromatin comprises decondensed and in general more genetically active regions while heterochromatin is formed by more condensed and in general transcriptionally inactive areas. Heterochromatin was first defined as the structure that does not change its compaction level during the cell cycle and can be divided into constitutive heterochromatin, formed by highly repetitive regions like telomere and centromere, and facultative heterochromatin involving chromatin regions that can adopt the structural and the functional characteristics of a heterochromatic state in certain cells and tissues. Each of these chromatin types can be associated with specific histone tail modifications that keep the active or repressive state by regulating the accessibility of regulatory proteins to the chromatin (Kouzarides, 2007).

Accessibility to the DNA to regulate genome functions

The packaging of the DNA into the chromatin is necessary to fit the entire genome in the small nuclear volume and it is therefore very important to precisely regulate the accessibility to the DNA for various protein involved in gene regulation. Despite the high degree of compaction, chromatin structure has to be highly dynamic to allow access to the DNA by the complex protein machineries involved in gene transcription, DNA replication and DNA repair (reviewed in Luger, Dechassa et al., 2012). This process is carried out by different modifications which cause conformational changes that regulate accessibility to the DNA.

Chromatin remodeling complexes can change the structure of the nucleosome temporarily by loosening the DNA from the histones or displacing a nucleosome, thereby making the DNA accessible for other factors.

Modifications of the histone tails play an important role in altering the chromatin state and accessibility of the DNA by loosening the histone-DNA interaction and enabling binding of specific regulatory proteins (Kouzarides, 2007; Ernst, Kheradpour et al., 2011). Histone acetylation is mainly associated with transcription activation, for instance histone 3 lysine 9 acetylation (H3K9ac) and lysine 27 acetylation (H3K27ac) are associated with active regulatory regions in the DNA. Histone methylation can be implicated in both transcriptional activation and repression depending on the residue that is modified; in particular histone 3 lysine 4 methylation (H3K4me1) correlates with enhancers; histone 3 lysine 4 di-methylation (H3K4me2) with enhancer and promoter regions, histone 3 lysine 4 tri-methylation (H3K4me3) with promoter regions and histone 3 lysine 27 tri-methylation (H3K27me3) with transcriptionally repressed regions. Phosphorylation of histones mainly takes place on serine, threonine and tyrosine residues. This modification adds a significant negative charge to the histones and in general this is linked to transcriptional activation. Histones tails can be also ubiquitylated and sumoylated. Ubiquitylation has been found on histone 2A lysine 119 (H2AK119) in association with transcriptional repression and on histone 2B lysine 20 (H2BK20) in association with transcriptional activation. Sumoylation has been shown to take place on all four core histones (reviewed in Kouzarides, 2007).

Modifications of the DNA itself can also modulate expression. One classical example is the methylation of cytosines in CpG islands that causes repression of gene transcription by preventing the transcription machinery to associate with the DNA.

Transcription and regulatory elements

The achievement of different biological processes depends on the correct spatial and temporal expression of genes. Gene transcription is performed by RNA polymerase and it is a process tightly controlled by transcription factors and transcriptional regulatory elements such as promoters, enhancers, silencers, Locus Control Regions (LCR) and insulators (reviewed in Maston, Evans et al., 2006).

Transcription factors are proteins that bind to specific regions of the DNA and can modulate gene regulation by acting on transcriptional regulatory regions.

Promoters are regions located at the gene start that serve as docking and assembly site for the transcriptional machinery (general and specific transcription factors) and specific activator proteins.

Enhancers are regions on the genome that markedly increase the transcription of a gene by acting on the respective promoters. They function independently of orientation and distance from the target promoter and can occasionally act over long distances. Enhancers provide the binding platforms for several transcription factors that work cooperatively to enhance transcription.

Silencers are sequence-specific elements that confer negative effects on the transcription of a target gene. Typically they function independently of orientation and distance from the target promoter and they provide the binding sites for transcription factors with a repressive function. Locus Control Regions (LCR) are regulatory elements involved in regulating entire loci or gene clusters. In general they are composed of several regulatory elements comprising enhancers, silencers or insulators that can be bound by general or specific transcription factors with activating or repressing function, thereby controlling transcription. Although LCRs are typically located upstream of their target genes, they can also be found within the intron of their target gene or even in the introns of neighboring genes.

Insulators are regulatory elements that can be bound by specific proteins, called insulator proteins, and can have enhancer-blocking activity by preventing an enhancer to interact with a specific promoter, or chromatin-barrier activity by preventing the spreading of heterochromatic marks into active chromatin. Much of the work that defined the properties of an insulator was performed in *Drosophila*, where different insulator regions have been described. Each of these insulators consists of a DNA sequence and specific insulator proteins that interact with this sequence (reviewed in Gurudatta and Corces, 2009). In *Drosophila*, six different insulator-associated proteins have been identified: dCTCF, CP190, BEAF-32, Su(Hw), Mod(mdg4) and GAF. In vertebrates the only protein identified so far as chromatin insulator is CTCF, which will be explained later in this chapter.

Functional organization of the genome in higher-order chromatin structure

Based on the observation that DNA regulatory elements can be found quite distant from their target promoters, several models have been proposed to explain how insulator proteins or transcription factors can function to facilitate the communication between regulatory elements and promoters.

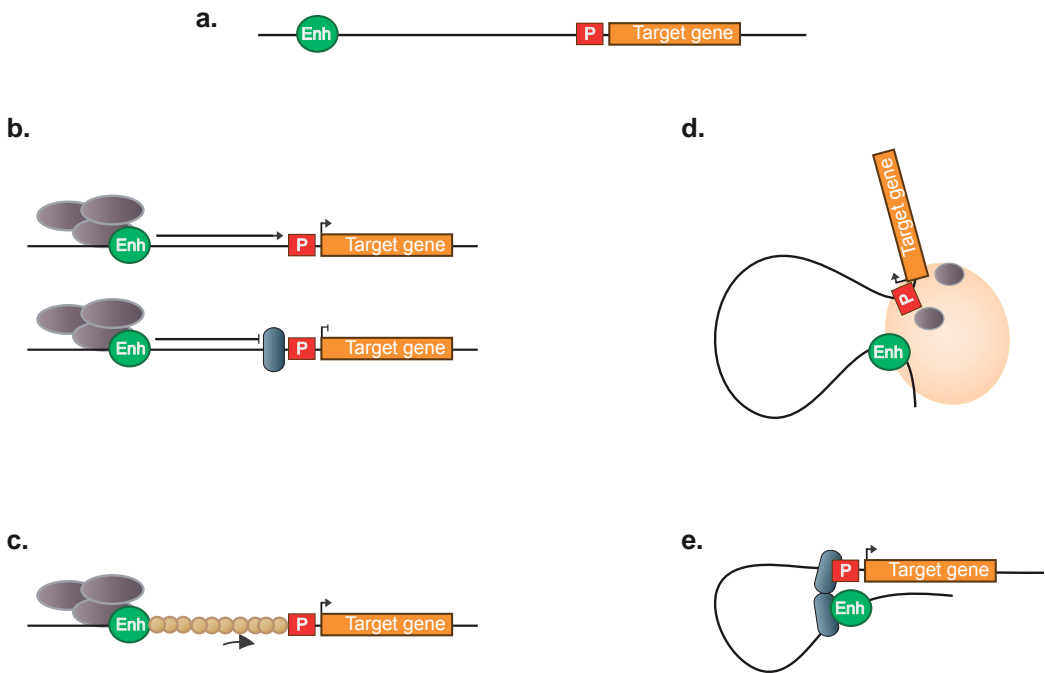


Figure 1: Models proposed to explain the formation of long-range chromosomal interaction for gene regulation.

a. Linear representation of the genomic distance between an enhancer (green oval) and a promoter (orange square). **b.** Tracking model in which the “tracking complex” (brown ovals) bound to the enhancer can slide along the chromatin fiber towards the target promoter activating the transcription (upper model). The “tracking complex” might be also trapped by an insulator protein (grey oval) which leads to transcriptional repression (lower model). **c.** Linking model in which DNA binding proteins (brown ovals) bind to an enhancer and drive the polymerization of other proteins (orange ovals) towards the target promoter. **d.** Relocation model in which a particular gene can relocate into a nuclear compartment (orange oval) containing transcription factors and active polymerase (brown ovals) that favour the interaction between the enhancer and the target promoter. **e.** DNA looping model in which the regulatory proteins (gray ovals) promote the formation of a DNA loop, thereby bringing the enhancer in close proximity to a specific promoter.

In the *tracking model*, DNA binding proteins bind to the enhancer, forming a “tracking complex” that can slide along the chromatin fiber and reach a target promoter where it activates transcription. One example supporting this model was the transcription of the late genes of the bacteriophage T4 where the activator gp45 is loaded on the DNA and can reach the target promoter by sliding along the DNA (Kolesky, Ouhammouch et al., 2002). The tracking model could also be used to explain gene silencing mediated by chromatin insulators. The insulator protein would block the “tracking complex” before it reaches the target promoter, but so far this model is not supported by evidence (**Figure 1b**). In the *linking model*, DNA binding proteins can bind to an enhancer and drive the sequential binding of other proteins towards the target promoter. This model was used to explain the action of the *Drosophila* Chip protein that interacts with several other factors to facilitate enhancer-promoter communication (Bulger and Groudine, 1999) (**Figure 1c**). The tracking and the linking models assume that

enhancers could emit some signals that travel along the chromatin fiber towards the target gene to modulate the transcription. In the *relocation model*, a particular gene can relocate to a nuclear compartment in which the interaction between the enhancer and the promoter of the gene is favoured (Cook, 2003). This model is based on observations that active RNA polymerase can be found in focal sites within the nucleus (Jackson, Hassan et al., 1993) and that during transcription active genes are associated with this focal site (Osborne, Chakalova et al., 2004) (**Figure 1d**). In the *DNA looping model*, regulatory proteins can promote the formation of DNA loops which bring the enhancer in close proximity to a specific promoter and regulate transcription (**Figure 1e**). This model is the favoured one because it assumes flexibility of the chromatin fiber. Evidence for this type of interactions was found using chromosome conformation capture methods, for instance at the mouse β -globin locus the *locus control region* (LCR) was found to interact with the distal β -globin promoter (Tolhuis, Palstra et al., 2002).

The effect of gene position

The evidences that DNA folds into configurations that bring regulatory elements in close proximity to control gene expression, raised the question on how the 3D nuclear position of the chromosomes could influence such functional long-range interactions. In fact, the genome is not randomly positioned within the nucleus and interphase chromosomes occupy distinct territories in the nuclear space, termed as *chromosome territories* (CT) (Stack, Brown et al., 1977; Cremer, Cremer et al., 1982). CTs positions correlate with gene density; high gene density chromosomes are located in the interior of the nucleus and chromosomes with low gene density are found at the nuclear periphery (Boyle, Gilchrist et al., 2001). Furthermore, heterochromatin and silent genes mainly localize at the nuclear lamina while euchromatin and active genes are found at the nuclear interior, suggesting that the nuclear periphery is a repressive environment for transcription and the nuclear interior a compartment for transcriptional activation (Schneider and Grosschedl, 2007). Initial microscopy studies regarding the arrangement of the CTs suggested an organization in loop domains of on average 1Mb size (Manuelidis, 1985; Sachs, van den Engh et al., 1995; Munkel, Eils et al., 1999; Bolzer, Kreth et al., 2005; Cremer, Grasser et al., 2008; Cremer and Cremer, 2010; Markaki, Smeets et al., 2012). Based on these observations, two main models were proposed to explain how the chromatin fiber folds within chromosome territories and how the organization of the nuclear architecture influences genome functions: the *chromosome territory-interchromatin compartment* (CT-IC) model and the *interchromatin network* (ICN) model (Cremer and Cremer, 2010).

According to the CT-IC model, the nucleus is composed of two main compartments: the CTs, constituted by chromatin domains, and a DNA free compartment carrying splicing speckles and nuclear bodies expanding in between these chromatin domains called *interchromatin compartment* (IC). The width of the IC space is highly variable depending on movements of the chromatin domains and allowing transient contacts of domains surface in *cis* or *trans*. Both compartments were first observed by light and electron microscopy (Visser, Jaunin et al., 2000;

Rouquette, Genoud et al., 2009). The IC is further separated from the more condensed interior of the CTs by a layer of relatively decondensed chromatin located at the domain periphery, called *perichromatin region* (PR) which might represent the main nuclear compartment for transcription. Electron microscopy evidence has supported the view of the PR area (Fakan and van Driel, 2007). During ongoing transcription genes are at least partially decondensed into the PR generating *perichromatin fibrils* (PF) that carry the nascent transcripts. PF are then processed in nearby splicing factories located at the IC (**Figure 2a**).

The ICN model predicts that loops of a particular CT can contact loops located in the same CT (*cis* interactions) or loops of a neighbouring CTs (*trans* interactions). These interactions between the chromatin fibers of CTs were first observed by FISH experiments in which different labelled chromosome painting probes revealed zones of colour overlap between the CTs (Branco and Pombo, 2006). In this interchromatin network, active genes are located on decondensed chromatin loops that extend outside of the chromosome territory and can localize with areas of focal concentration of RNA polymerase II, called transcription factories, leading to regulation of their expression (Fraser and Bickmore, 2007) (**Figure 2b**).

Both models are mainly based on microscopy studies which have some limitations in analyzing the overall spatial organization of chromosomes (Rajapakse and Groudine, 2011) and in fact they provide only few mechanistic details about the relationship between higher-order chromatin structure and the genome.

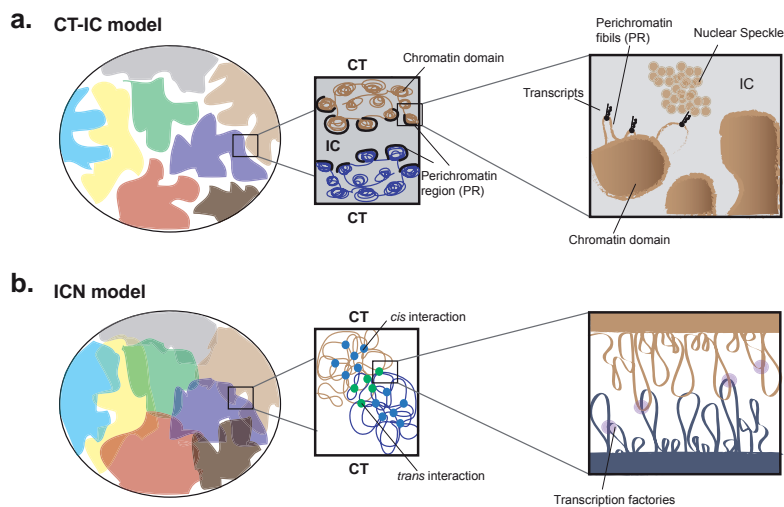


Figure 2: Schematic representation of the chromosome territory-interchromatin compartment (CT-IC) model (a) and the interchromatin network (ICN) model (b) (modified from Cremer and Cremer, 2010).

a. CT-IC model in which CTs are constituted by chromatin domains and a DNA free nuclear compartment carrying splicing speckles and nuclear body called IC. Perichromatin region (PR) located at the domain periphery is also indicated. During transcription, genes in the PR are decondensed generating perichromatin fibrils (PF) that carry the nascent transcripts which are then processed in nearby splicing factories. Movement of chromatin domains allows transient contacts of domains surface in *cis* or *trans*. **b.** ICN model in which loops from the same CT as well as from neighbouring CTs can make contacts in *cis* (blue dots) or *trans* (green dots). Transcription factories (pink ovals) can recruit genes on decondensed chromatin loops extending outside of the chromosome territory to regulate transcription (model suggested by Fraser and Bickmore, 2007).

Insights into nuclear organization by 3C technologies

In the past decade the development of *chromosome conformation capturing* (3C) techniques has made it possible to study the nuclear organization at a high resolution level. These technologies can elucidate how the chromatin fibers are folded into the higher-order chromatin organizations that form the CTs and which factors are involved in promoting these contacts. All these methods are based on fixation of the chromatin to stabilize the 3D structure of the DNA, digestion with a restriction enzyme and ligation of the cross-linked DNA fragments (**Figure 3**). In this way, DNA fragments located in close proximity in the nuclear space are ligated to each other. The subsequent detection of these fragments, which varies in method between the different 3C technologies, provides information about the 3D spatial organization of the genome.

Chromosome Conformation Capture (3C): The detection of the ligation product is done by PCR (semi-quantitative) or qPCR (quantitative) measuring the number of the ligation events between the selected DNA fragment (anchor or viewpoint) and the interacting fragments. PCR primers are designed near the restriction ends of the anchor fragment and in all the restriction fragments of interest (one-to-one).

The primer design requires some knowledge of potential interactions (Dekker, Rippe et al., 2002; Hagege, Klous et al., 2007).

Chromosome Conformation Capture-on-chip (4C): The ligated 3C template is further digested and ligated to form small DNA circles. By using PCR primers designed in the fragment of interest (viewpoint), it is possible by inverse PCR to amplify all unknown fragments ligated with the viewpoint. The amplified fragments can be identified by microarray (one-to-all) (Simonis, Klous et al., 2006).

Chromosome Conformation Capture sequence (3C-seq or multiplexed 3C-seq): Based on the 4C method, but the detection of the amplified fragments is done using next generation sequencing technology. This allows to analyze one viewpoint (3C-seq; one-to-all) or, by using a multiplexed approach, several viewpoints per sequencing lane (multiplexed 3C; many-to-all) (Soler, Andrieu-Soler et al., 2010; Stadhouders, Thongjuea et al., 2012; Stadhouders, Kolovos et al., 2013).

Chromosome Conformation Capture Carbon copy (5C): The 3C template library is hybridized with a mix of oligos covering a particular region of interest and designed in the ligation junctions of the 3C fragments. After PCR the amplified fragments are detected by microarray or next generation sequencing technology (many-to-many) (Dostie, Richmond et al., 2006).

By using this technology, the spatial organization of the mouse regulatory *X-inactivation*

centre (Xic) was analyzed demonstrating that the region was partitioned in a series of 200Kb to 1Mb *topologically associated domains (TAD)* (Nora, Lajoie et al., 2012). The TADs spatially segregated the promoters of the non-translated RNA genes *Xist/Tsix* sense/antisense which lay in adjacent domain pointing to their role in shaping regulatory landscape.

Hi-C: In this case the procedure to obtain the template is different: before ligation of the crosslinked DNA fragments, the ligation ends are filled with biotinylated-nucleotides. The ligation occurred then via the blunt-ends and the interacting fragments are enriched by a biotin pull-down. The fragments are amplified by PCR and detected using next generation sequencing technology (all-to-all) (Lieberman-Aiden, van Berkum et al., 2009; van Berkum, Lieberman-Aiden et al., 2010).

This technology was used to study chromatin structure in mammalian cells and revealed that the genome is partitioned into large mega-based self-associated domains. These so called topological domains are conserved between tissues and to some extent also between mouse and human (Dixon, Selvaraj et al., 2012). The different topological domains are linked by narrow segments, called boundaries, which delimit the transition between domains.

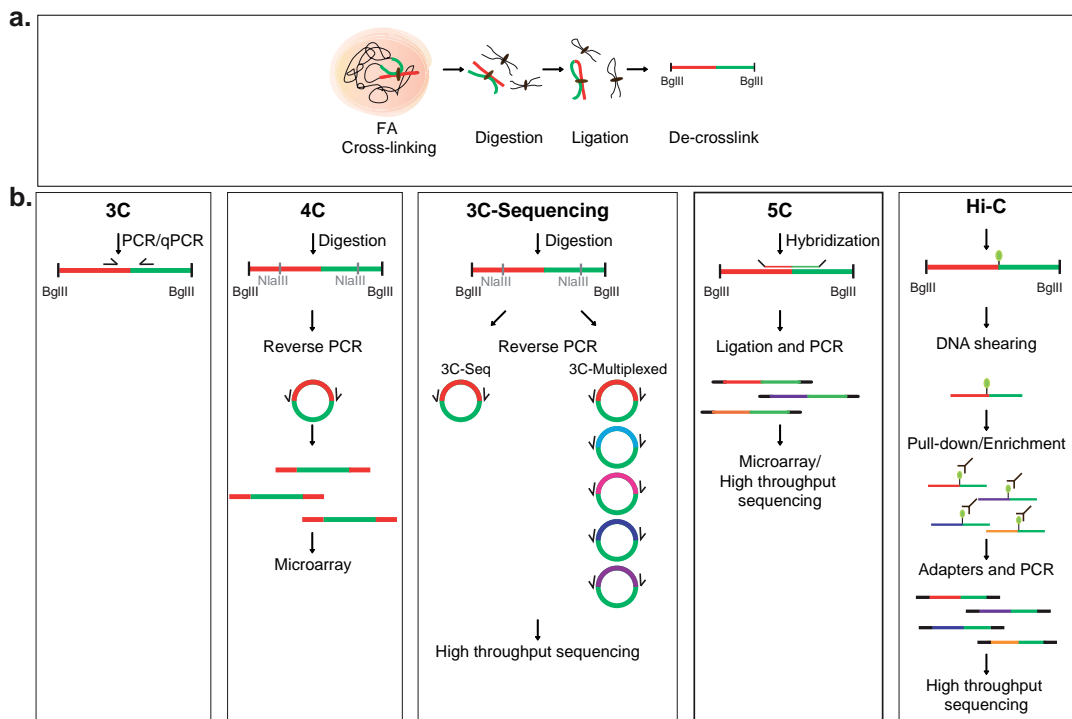


Figure 3: Schematic overview of chromosome conformation capturing methods.

a. Common steps between "C" methods. **b.** Specific step for the different "C" methods.

Factors promoting higher-order chromatin structure

Several studies on the 3D organization of the genome by using the “C” methods have highlighted important factors that play a central role in promoting higher-order chromatin structure.

In these sections I will focus on the cohesin complex and CTCF (CCCTC-binding factor).

The cohesin complex

Structure of the cohesin core complex

Cohesin is a chromosome-associated multisubunit protein complex which is highly conserved in eukaryotes. In somatic vertebrate cells, the core cohesin complex consists of four subunits: SMC1, SMC3, RAD21 and SA1 or SA2 (**Table 1** for nomenclature in different species).

The two core subunits of the complex, SMC1 and SMC3, are members of the Structural Maintenance of Chromosomes (SMC) family. The polypeptide chain of SMC proteins folds back on it-self forming a long anti-parallel coiled-coil structure of 45 nm with the hinge domain at one end and the N- and C-terminal domains assembling into a functional ATPase (ATPase head).

SMC1 and SMC3 in the cohesin complex interact via the hinge domain and the ATPase heads are connected by the alpha-kleisin subunit, RAD21, resulting in the formation of a tripartite molecular ring which can be visualized by electron microscopy (Anderson, Losada et al., 2002; Haering, Lowe et al., 2002).

The N-terminus of RAD21 binds to the SMC1 subunit while the C-terminus binds to SMC3. A fourth subunit, either SA1 or SA2 (stromalin antigens 1 and 2), binds to RAD21 (**Figure 4a**).

	<i>S. Cerevisiae</i>	<i>S. pombe</i>	<i>D. melanogaster</i>	<i>X. leavis</i>	<i>M. Musculus</i>	<i>H. Sapiens</i>
SMC	Smc1	Psm1	Smc1	Smc1	Smc1	SMC1
	Smc3	Psm3	Smc3	Smc3	Smc3	SMC3
Kleisin	Scc1/Mcd1	Rad21	Rad21	Rad21	Scc1	RAD21
Kleisin interacting subunits	Scc3	Psc3	SA	SA1/stag1	SA1/Stag1	SA1/STAG1
				SA2/stag2	SA2/Stag2	SA2/STAG2
Cohesin loading factor	Scc4	Mis4	Nipped-B	Scc2	Scc2	NIPBL
		Ssl3	Scc4	Scc4	Scc4	Mau2
Regulatory proteins	Pds5	Pds5	Pds5	Pds5A, Pds5B	APRIN	PDS5A, PDS5B
	Rad61/Wapl	Wapl	Wapl	Wapl	Wapal	WAPAL
	-	-	Dalmatian	Sororin/cdca5-b	Sororin/Cdca5	Sororin/CDCA5
Acetyl-transferase	Eco/Ctf7	Eso1	Deco, San	Esco1, Esco2	Esco1, Esco2	ESCO1, ESCO2
Deacetylases	Hos1	-	-	-	-	HDAC8

Table 1: Cohesin complex subunits and cohesin-regulatory proteins in different species.

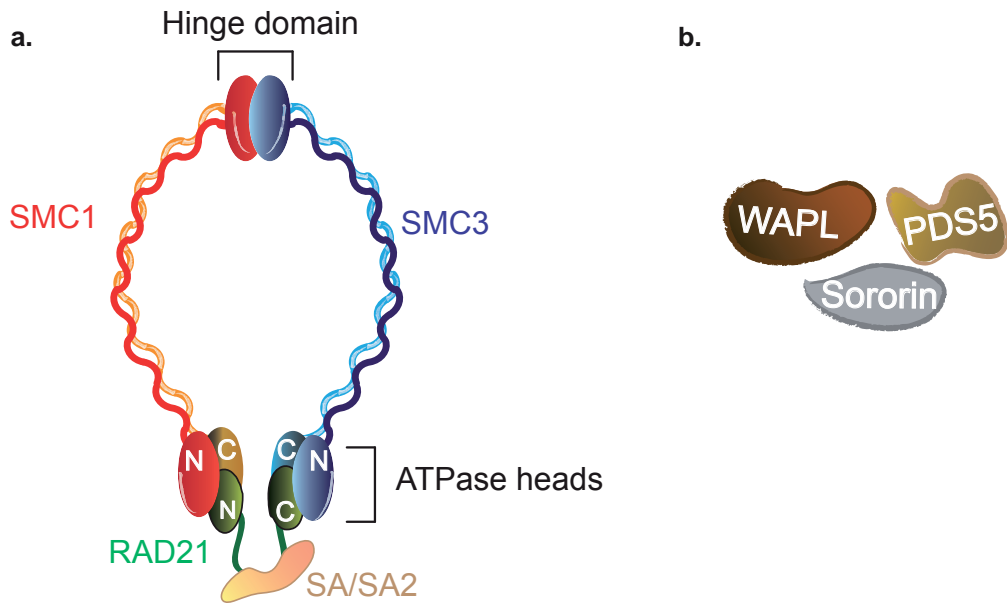


Figure 4: Schematic representation of the cohesin complex (a) and the transiently associated regulatory proteins (b).

- a. SMC3 and SMC1 are bound via the hinge domain and the ATPase heads are connected by RAD21 which in turn can bind either SA1 or SA2.
- b. Cohesin regulatory protein PDS5, WAPL and Sororin: PDS5/Sororin act during S phase to stabilize cohesion while PDS5/WAPL act during prophase to facilitate cohesin dissociation from chromatin arms.

Cohesin and transiently-associated regulatory proteins

RAD21 and SA subunits are transiently associated with three main regulatory proteins: PDS5 (“*Precocious Dissociation of Sisters*”), WAPL (“*Wings Apart-like*”) and Sororin (**Figure 4b**). PDS5, which in vertebrates exists in two isoforms PDS5A and PDS5B, is a protein that associates with SA subunit of the cohesin complex and is characterized by several HEAT repeats.

WAPL contains at the N-terminal region three FGF motifs that promote interaction with PDS5, and a HEAT repeat that might favour the interaction with SMC3 or SA subunits. The complex PDS5/WAPL is very important during prophase to facilitate the dissociation of cohesin from chromatin arms.

Sororin contains FGF motifs that bind to PDS5. In particular during S phase Sororin can displace WAPL from PDS5, thereby avoiding the dissociation of cohesin from chromatin and stabilizing the cohesion between the sister chromatid (Nishiyama, Ladurner et al., 2010).

Functions of the cohesin complex

All cohesin functions were initially studied in yeast and subsequently in vertebrates. The cohesin complex was first discovered by a genetic screen in yeast for its role in mediating cohesion between sister chromatid during cell division (Michaelis, Ciosk et al., 1997) and later on this function was also studied in *Xenopus* and human (Losada, Hirano et al., 1998; Sumara, Vorlaufer et al., 2000). Cohesin was also found to be essential for efficient DNA double strand break repair (Sjogren and Nasmyth, 2001; Kim, Krasieva et al., 2002) (reviewed in Watrin and Peters, 2006).

Recently it has become increasingly clear that the complex has roles in regulating gene expression, mediating chromatin insulation and promoting the formation of higher-order chromatin structure (reviewed in Wendt and Peters, 2009; Remeseiro and Losada, 2013).

Canonical function of cohesin: sister chromatid cohesion

To ensure genome stability and accurate inheritance of genetic information, DNA has to be replicated and a correct and complete set of chromosomes has to be distributed onto the daughter cells. The cohesin complex plays an important role in this process by holding the replicated sister chromatids together until they are properly attached to the mitotic spindle.

Loading of cohesin

In vertebrates, cohesin is loaded onto DNA in telophase after the formation of the nuclear envelope. This process depends on the cohesin-loading factors, NIPBL and MAU-2 (**Figure 5a**). It is not clear how cohesin is exactly loaded and bound onto DNA but observations in yeast suggest that the binding depends on ATP hydrolysis by SMC1 and SMC3 globular ATPase heads that is somehow linked to the opening of the hinge domain, allowing the passage of the chromatin strand into the cohesin ring (Gruber, Arumugam et al., 2006).

Establishment and maintenance of cohesion

The cohesion between the sister chromatid is established during DNA replication in S phase and it requires the acetylation of two lysine residues located at the ATPase globular domain of the SMC3 subunit by an acetyltransferase enzyme called ESCO, a protein that exists as two isoforms in human cells, ESCO1 and ESCO2 (Unal, Heidinger-Pauli et al., 2008).

Cohesion between the sister chromatid is maintain by a protein called Sororin (Schmitz, Watrin et al., 2007). In particular, SMC3 acetylation and the DNA replication recruit Sororin (Lafont, Song et al., 2010). Sororin binds to PDS5 and it displaces WAPL from PDS5, thus stabilizing cohesin binding and maintaining stable cohesion (Nishiyama, Ladurner et al., 2010) (**Figure 5b**).

Removal of cohesin: prophase pathway

In vertebrates, the bulk of cohesin at the chromosome arms is removed during prophase, while it is preserved at the centromeres (Waizenegger, Hauf et al., 2000). Cohesin removal requires Polo-like kinase 1 (Plk1) that is activated in early mitosis and is translocated from the cytoplasm into the nucleus at the end of prophase. In mitosis, the RAD21 and SA1/SA2 subunits of cohesin are phosphorylated and, based on observations in mitotic *Xenopus* egg extracts, Plk1 is required for these modifications (Sumara, Vorlaufer et al., 2002). Another protein required for the dissociation of cohesin from chromosome arms is the mitotic kinase Aurora B (Losada, Hirano et al., 2002). Since this protein is unable to phosphorylate cohesin subunits *in vitro*, it is possible that Aurora B can influence cohesin indirectly by phosphorylating other regulators. One possible candidate is Condensin I whose association with chromosomes is depending on Aurora B (Lipp, Hirota et al., 2007) and cohesin dissociation (Hirota, Gerlich et al., 2004) (**Figure 5c**). The phosphorylation of the cohesin complex is required but not sufficient to dissociate cohesin from chromosome arms. Several studies identified WAPL as a key regulator of cohesin dissociation. In fact, depletion of WAPL prevents removal of cohesin from chromosome arms in prophase (Kueg, Hegemann et al., 2006).

At this stage, phosphorylation of Sororin by cyclin-dependent kinase1 (Cdk1) (Dreier, Bekier et al., 2011) allows the binding of WAPL to PDS5, promoting the dissociation of cohesin from chromosome arms (Nishiyama, Ladurner et al., 2010) (**Figure 5d**). At this phase, centromeric cohesin is protected by binding of shugoshin (Sgo1) which in turn recruits the protein phosphatase 2A complex (PP2A). It was proposed that the Sgo1/PP2A complex might dephosphorylate centromeric cohesin, thereby preventing its removal by the prophase pathway (Kitajima, Sakuno et al., 2006; Tang, Shu et al., 2006) (**Figure 5d**).

Removal of cohesin: cleavage by separase

Sister chromatid cohesion at the centromere persists until the chromatids are properly attached to the opposite poles of the mitotic spindle. At this point the remaining cohesin is removed by proteolytic cleavage of its RAD21 subunit by the protease separase. The proper timing of these events is ensured by the ubiquitin ligase anaphase-promoting complex/cyclosome (APC/C) which activates separase by ubiquitylation of the separase inhibitor cycline B/ securine, thereby targeting it for degradation by the proteasome. Cleavage of RAD21 leads to the release of cohesin from chromatin and segregation of the sister chromatid (**Figure 5e**).

Re-cycle of cohesin

In order to be reused in the next cell cycle, the cohesin SMC3 subunit is deacetylated by Hos1 in yeast (Borges, Lehane et al., 2010) and HDAC8 in human cells (Deardorff, Bando et al., 2012) (**Figure 5f**). Failure of deacetylation prevents the disassembly of the cleaved cohesin complex and reduces the level of cohesin available for the next cell cycle. This acetylated cohesin complex can be loaded onto DNA but chromatin immunoprecipitation sequencing analysis has shown a decreased enrichment of cohesin at the its sites (Deardorff, Bando et al., 2012).

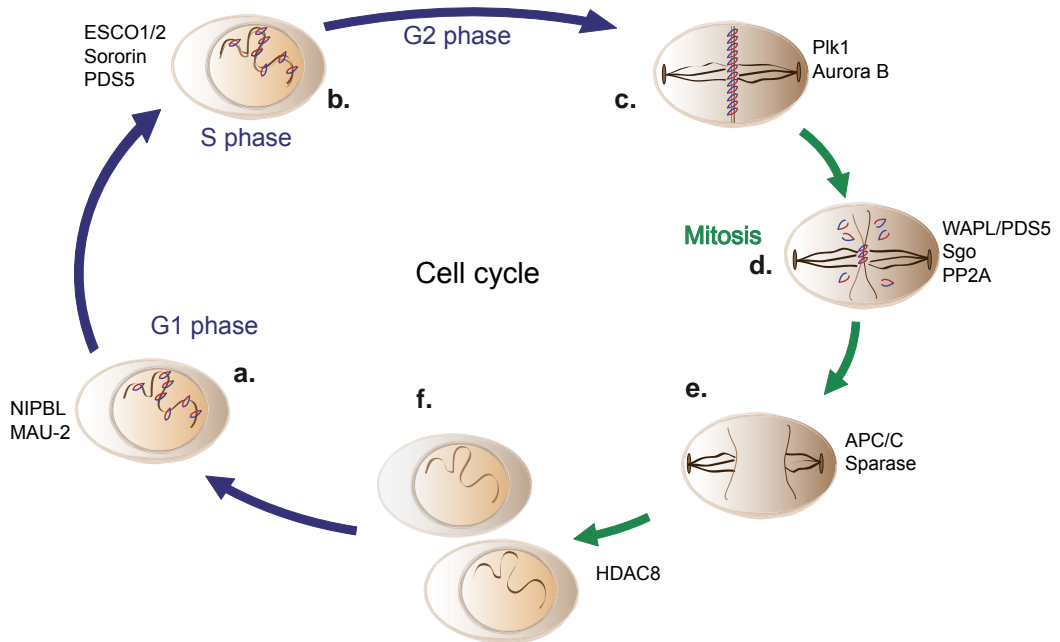


Figure 5: Schematic representation of the canonical function of the cohesin complex.

a. Loading of cohesin by NIPBL/MAU-2. b. Establishment of sister chromatid cohesion via acetylation by ESCO and maintenance of stable cohesion by Sororin/PDS5. c-d. Removal of cohesin from chromatid arms by PIK1, Aurora B and WAPL/PDS5 and protection of centromeric cohesin by Sgo/PP2A. e. Cleavage of centromeric cohesin by Separase and chromatin segregation. f. De-acetylation of cohesin by HDAC8.

How does cohesin mediate sister chromatid cohesion?

Several models have been proposed to explain how cohesin can tether sister chromatids but I will only focus on the two most popular ones: the ring model and the hand-cuff model.

The *Ring model* proposes that cohesion between the sister chromatid is mediated by the ring-like structure of the cohesin complex that can topologically entrap the replicated DNA strands (Haering, Lowe et al., 2002; Gruber, Haering et al., 2003) involving a minimal contact between the DNA and the SMC subunits (**Figure 6a**). The observations that proteolytic cleavage of Scc1 or Smc3 is sufficient to release cohesin from chromosomes (Waizenegger, Hauf et al., 2000; Gruber, Haering et al., 2003) and that linearization of yeast minichromosomes leads to loss of cohesin (Ivanov and Nasmyth, 2005) support this model.

In the *Hand-cuff model*, cohesion of the sister chromatid is established by interaction of two cohesin complexes via RAD21 subunits, enforced by either SA1 or SA2 (**Figure 6b**). Based on this model, single-ring cohesin complexes are loaded onto the chromosomes and during DNA replication the rings are distributed onto the sister chromatids. The handcuff is established when two RAD21 molecules are paired and tethered either by SA1 or SA2 (Zhang and Pati, 2009). This model is supported by one study using a fluorescence protein complement assay (PCA) and a yeast two-hybrid assay showing that two RAD21 molecules can interact and that

the depletion of SA1 and SA2 prevents this RAD21–RAD21 interaction causing loss of sister chromatid cohesion (Zhang and Pati, 2009).

Experimental observations support both models and it is still unclear how exactly cohesin mediates cohesion.

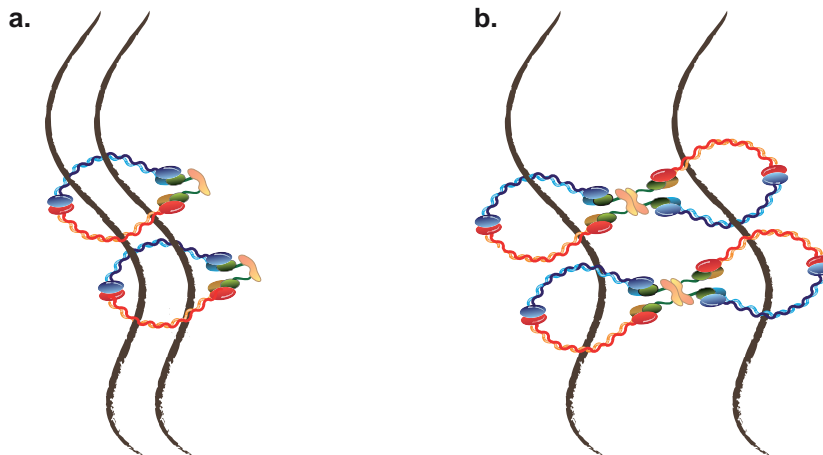


Figure 6: Possible mechanisms how cohesin can hold sister chromatids.

a. Ring model in which the tripartite ring-like structure of the cohesin complex topologically embraces the sister chromatids. **b.** Hand-cuff model in which cohesion is established by two cohesin complexes that interact via their RAD21 subunits tethered by SA1 or SA2. Each ring entraps one chromatid fiber.

Non-canonical functions of cohesin

Cohesin in Double-Strand Break (DSB) repair

In eukaryotic cells there are two distinct mechanisms for DNA double-strand break (DSB) repair: the non-homologous end-joining (NHEJ) pathway and the homologous recombination (HR) pathway. The NHEJ pathway re-ligates the two free ends of DNA created at the DSB while the HR pathway uses the intact sister chromatid to repair the damaged DNA. The function of the HR pathway depends on the ability of cohesin to tether sister chromatids. In yeast, *de novo* loading of cohesin by Scc2/Scc4 occurs after DNA damage. In human, the recruitment of cohesin to the damage sites depends on SMC5/SMC6, a protein complex required for efficient HR repair (Potts, Porteus et al., 2006). Recently it was also demonstrated that the cohesin loading factor NIPBL is recruited to DSB sites, leading to the hypothesis that NIPBL could promote *de novo* loading of cohesin at the damage sites, as observed in yeast (Oka, Suzuki et al., 2011).

In vertebrates it was shown that after a IR DNA-damage, the cohesin subunits SMC1 and SMC3 are phosphorylated by several kinases, such as ATM kinase, crucial for the initiation of the DNA-repair signalling pathway (Yazdi, Wang et al., 2002). Different studies suggested that these phosphorylations might be important to activate the DNA damage checkpoint to inhibit DNA replication (Kim, Krasieva et al., 2002; Yazdi, Wang et al., 2002). However, how SMC1

and SMC3 phosphorylation functions in the DNA damage response is still to be clarified. Based on the fact that the phosphorylations occur in a region of the coiled-coil next to the globular ATPase domains, it was proposed that these modifications could alter the ATPase activity and the ability to open and close the ring or to laterally move along the DNA fiber (reviewer by Watrin and Peters, 2006).

Cohesin in gene regulation

Several studies demonstrated that cohesin might have a function in gene regulation in addition to its roles in chromosomal cohesion and DNA damage repair. The initial evidence came from the observation in yeast that cohesin can block the spreading of the Silent Information Regulatory complex (SIR) that represses transcription by occupying the boundaries of the HRM silent mating-type locus (Donze, Adams et al., 1999). Studies in *Drosophila* demonstrated that cohesin and the cohesin loading factor Nipped-B are required for proper transcription via long-range contacts of the homeobox genes *cut* and *Ultrabithorax*, but with opposite effects: cohesin decreases *cut* expression (Rollins, Korom et al., 2004) while Nipped-B activates both genes (Rollins, Morcillo et al., 1999). Furthermore, mutations of cohesin subunits led to defects during development; for instance, in Zebrafish, mutation of the Rad21 gene was found to reduce the expression of *Runx* genes during embryonic development (Horsfield, Anagnostou et al., 2007) and in *Drosophila*, mutations of Smc1 and SA led to defects in axon pruning of the mushroom body neurons (Schuldiner, Berdnik et al., 2008). These defects cannot be explained by a sister chromatid cohesion defect but point to a role of cohesin in regulation of gene expression.

Moreover, the observations that cohesin binding to DNA persists during the cell cycle and that cohesin is expressed in different tissues (Monnich, Banks et al., 2009) including post-mitotic cells like neurons where chromatid cohesion is not required (Wendt, Yoshida et al., 2008), support the idea of other roles of cohesin besides cohesion.

Cohesinopathies

Different studies have shown that mutations in cohesin regulatory proteins or cohesin subunits are associated with severe developmental disorders also termed *Cohesinopathies* such as Roberts/SC phocomelia syndrome (RBS) and Cornelia de Lange syndrome (CdLS).

Roberts/SC phocomelia syndrome (OMIM #268300 and #269000) is an autosomal recessive disorder that is linked to homozygous mutations (in general truncation and missense mutation) of the acetyltransferase enzyme ESCO2. This syndrome is characterized by severe developmental defects, such as limb and facial abnormalities, mental and growth retardation as well as defects in centromeric cohesion and other heterochromatic areas but not along the euchromatic arms (Vega, Waisfisz et al., 2005). Similar defects were also observed in *Drosophila* after mutation of the acetyltransferase *deco*, the fly homolog of ESCO

(Williams, Garrett-Engle et al., 2003). In this mutant the association of the Rad21 subunit with centromeric regions was found to be reduced in comparison to the wild-type. Thus it was hypothesized that changes in the level or mode of cohesin binding in heterochromatic regions could interfere with expression of heterochromatic genes and contribute to the developmental deficit (reviewed in Dorsett, 2007). Furthermore, a defect in the mitotic mechanism may also be involved in the pathogenesis of RBS/SC. It was observed that the lack of cohesion at heterochromatic regions results in improper biorientation of the chromosomes. This might activate the mitotic spindle checkpoint leading the cells to divide very slowly or not divide at all thus contributing to the malformations associated with the syndrome (Jabs, Tuck-Muller et al., 1991; Vega, Waisfisz et al., 2005).

Cornelia de Lange syndrome (OMIM #122470) is a disorder characterized by typical facial features, growth and mental retardation, upper limb malformations, hirsutism, gastrointestinal and other visceral system defects. In the majority of the cases (60%) this syndrome is caused by heterozygous mutation of the cohesin loading factor NIPBL (Krantz, McCallum et al., 2004) and in a smaller fraction of patients by mutations of the cohesin core subunits SMC1A (5%) or SMC3 (<1%) (Musio, Selicorni et al., 2006; Deardorff, Kaur et al., 2007) or by mutations in the cohesin-deacetylation protein HDAC8 (Deardorff, Bando et al., 2012). NIPBL-related CdLS and SMC3-related CdLS are inherited in an autosomal dominant manner; SMC1A-related CdLS is inherited in an X-linked manner.

NIPBL mutations (in general truncation, deletion or missense mutations) are linked to a severe phenotype and mutations of the cohesin core subunits (in general missense or small deletion) to the mild phenotype. Interestingly, sister chromatid cohesion is not affected in cells derived from CdLS patients, likely due to a partial dosage compensation for NIPBL. This observation led to the hypothesis that a defect in the gene regulation function of cohesin might be the cause of the phenotype (reviewed in Dorsett and Krantz, 2009). Comparison of the transcriptomes of *lymphoblastoid cells lines* (LCLs) derived from CdLS patients with NIPBL mutation and controls identified a group of specific genes which are differentially expressed and involved in different cellular processes (embryonic and tissue development, hematological and immune system development and functions, cell death, cell proliferation and cell cycle regulation). In the control cell lines, some of these genes have cohesin binding sites next to the *transcription start sites* (TSS) (Liu, Zhang et al., 2009). These sites are lost in LCLs with NIPBL mutation suggesting that the loss of cohesin due to NIPBL mutation might affect transcriptional regulation at specific loci and might contribute to the CdLS phenotype.

Recently, mutations in the *histones deacetylase 8* (HDAC8) were found in a small cohort of CdLS patients (Deardorff, Bando et al., 2012). These patients present similar clinical features, transcriptional misregulation and loss of cohesin binding sites as seen in CdLS with NIPBL mutations, pointing to the idea that loss of cohesin results in both cellular and clinical features of CdLS. It remains possible that mutations in genes encoding other cohesin subunits or cohesin-regulatory proteins could contribute to the disorder, but to date no such mutations have been found in CdLS probands. Different mouse models suggest that mutations of

other cohesin subunits and cohesin-regulatory proteins give a phenotype reminiscent of CdLS. For instance, mice with heterozygous mutations of *Pds5B* and *Pds5A* exhibit CdLS like developmental abnormalities with no obvious cohesion defect (Zhang, Jain et al., 2007; Zhang, Chang et al., 2009). Furthermore, it was observed that the absence of the Stag1/SA1 cohesin subunit in mouse embryonic fibroblast (MEF) cells results in a relocation of the cohesin complex which alters the expression of genes involved in biological processes related to CdLS (Remeseiro, Cuadrado et al., 2012).

Co-localization of cohesin with transcription factors

Cohesin and CCCTC-binding factor (CTCF)

In mammalian cells, genome-wide mapping of cohesin binding sites revealed that cohesin largely colocalizes with the DNA binding protein *CCCTC-binding factor* (CTCF) (Parelho, Hadjur et al., 2008; Wendt, Yoshida et al., 2008).

CCCTC-binding factor (CTCF)

CTCF is a highly conserved and ubiquitously expressed DNA binding protein consisting of eleven zinc fingers. It has been implicated in several functions including transcriptional repression and activation, imprinting and chromatin insulation. CTCF was first described as a transcription factor that repressed the chicken *c-myc* gene (Lobanenkov, Nicolas et al., 1990; Klenova, Nicolas et al., 1993). Later studies demonstrated that CTCF is not only involved in gene silencing but also in activation, as shown for the *Amyloid β -Protein Precursor* (APP) (Vostrov and Quitschke, 1997).

In vertebrates, it is so far the only protein with a chromatin insulator function. Its binding to DNA can block enhancer-promoter interactions and can also provide a chromatin barrier function, as demonstrated for the chicken β -globin locus (Bell, West et al., 1999; Recillas-Targa, Pikaart et al., 2002). The CTCF insulation function plays also a critical role in the establishment and maintenance of imprinting expression patterns, as shown for the mouse and the human imprinted H19/Igf2 genes. These genes are expressed in a parent-specific manner with H19 transcribed from the maternal allele and Igf2 from the paternal allele. Both genes are under the control of the same enhancer located in close proximity to the H19 gene. CTCF was found to occupy the *imprinting control region* (ICR) of the locus only at the unmethylated maternal allele where it blocks the enhancer from activating the Igf2 gene (Bell and Felsenfeld, 2000; Hark, Schoenherr et al., 2000; Kanduri, Pant et al., 2000; Tost, Jammes et al., 2006).

Role of CTCF in promoting higher-order chromatin structure

The fact that CTCF can bind to distal regulatory elements to modulate gene transcription suggested that it might promote the formation of higher-order chromatin structures to bring

these elements in close proximity. An important study on the function of CTCF in promoting long-range chromosomal interactions was performed using the mouse H19/Igf2 locus (Murrell, Heeson et al., 2004). Chromosome conformation capture methods revealed that in mouse cells the binding of CTCF to the maternal ICR inhibits a direct communication between the enhancer and the Igf2 promoter, leading to the formation of a silent chromatin loop and to the transcriptional inactivation of Igf2. By contrast, at the paternal ICR the absence of CTCF facilitates the interactions between the enhancer and the Igf2 promoter forming an active chromatin loop and leading to Igf2 transcription (Murrell, Heeson et al., 2004; Kurukuti, Tiwari et al., 2006). CTCF was also found to be involved in organizing long-range chromosomal interactions at different other loci including the mouse β -globin locus in association with the erythroid specific factors EKLF, GATA-1, FOG1 (Drissen, Palstra et al., 2004; Vakoc, Letting et al., 2005; Splinter, Heath et al., 2006), the major histocompatibility complex II (MHCII) with the class II transactivator (CIITA) (Majumder, Gomez et al., 2006; Majumder, Gomez et al., 2008), the mouse interferon gamma ($Inf\gamma$) locus with the transcription factor T-bet (Sekimata, Perez-Melgosa et al., 2009) and the mouse Myb locus together with the Ldb1 complex (Stadhouders, Thongjuea et al., 2012). Interestingly, genome-wide studies focused on elucidating the 3D structure of the mouse and human genome identified large, megabase-sized local chromatin interaction domains, termed topological domains, in which the boundaries are enriched for CTCF sites, indicating that the presence of CTCF might be important in establishing the overall structure of the genome (Dixon, Selvaraj et al., 2012). These observations led to the initial thought of CTCF as the “*master weaver*” of the genome (Phillips and Corces, 2009) but how CTCF promotes long-range chromosomal interactions is still unclear.

Role of cohesin and CTCF in promoting higher-order chromatin structure

The finding that cohesin co-localizes with CTCF raised the question whether and how cohesin and CTCF functionally interact. CTCF was found to be important to position cohesin at CTCF sites as depletion of CTCF causes reduced cohesin enrichment at the shared sites but the general chromatin-binding of cohesin was not reduced after CTCF depletion. Conversely, cohesin is required to maintain CTCF-dependent chromatin-insulation. For instance, at the human IGF2/H19 imprinted gene, cohesin binds to the maternal ICR, as CTCF, and depletion of cohesin lead to loss of CTCF insulation, indicating that cohesin is also required for the imprinting regulation of the locus (Wendt, Yoshida et al., 2008). These results led to the idea that also the cohesin complex might be involved in promoting long-range interactions together with CTCF. Chromosome conformation capture studies at the human imprinted IGF2/H19 locus have demonstrated that both factors are helping in organizing the structure of the locus which results in different promoter-enhancer interaction on both alleles. In particular, for both alleles CTCF and cohesin were found to co-localize downstream of the enhancers (CTCF DS), in the Centrally Conserved DNase I hypersensitive domain (CCD) between H19 and IGF2 and in the region adjacent to the IGF2 promoter (CTCFAD/DMR0) (**Figure 7a**). As already shown, the binding at the ICR was found only on the maternal allele.

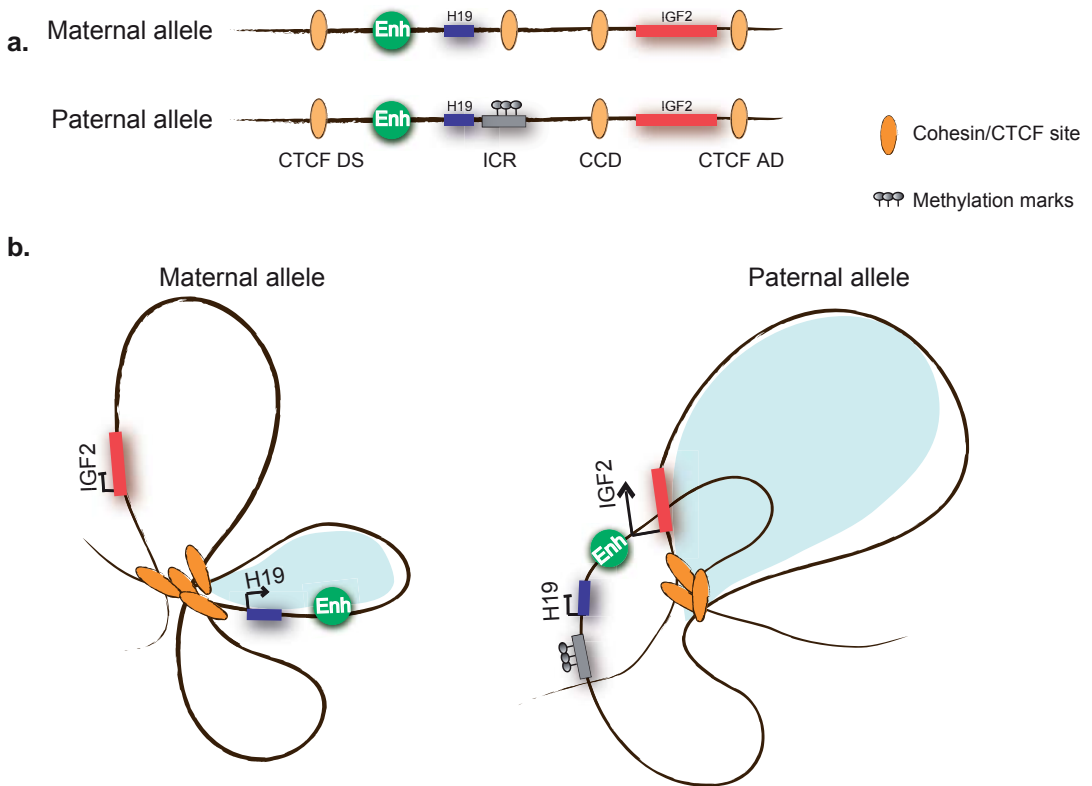


Figure 7: Model of cohesin and CTCF long-range chromosomal interactions at the human IGF2/H19 imprinted locus (modified from Nativio, Wendt et al., 2009).

a. Schematic representation of the different cohesin and CTCF binding (orange ovals) sites at the maternal and the paternal allele. **b.** Proposed loop-model for the human IGF2/H19 imprinted locus: on the maternal allele all cohesin and CTCF sites strongly interact forming a structure in which the enhancer (green circle) cannot interact with the IGF2 promoter, leading to the expression of only the H19 gene; on the paternal allele, the methylated ICR prevents cohesin and CTCF binding, thus it is excluded from the interacting region leading to an interaction between the enhancer to and the IGF2 promoter.

The chromosome conformation capturing data indicated that all cohesin and CTCF sites in the locus interact strongly with each other, while the ICR and the enhancer have limited allele specific interactions. These data suggested a model in which cohesin and CTCF binding sites on the paternal allele are coming together forming an interacting region in which the ICR is excluded, thus enabling the IGF2 gene promoters and enhancer region to interact. By contrast, the ICR on the maternal allele can interact with the other cohesin and CTCF sites redefining the locus structure and preventing the contact between the enhancer and the IGF2 promoter (**Figure 7b**). Moreover, cohesin has an important role in maintaining the higher-order chromatin structure at this locus since the depletion of cohesin destabilizes the chromatin conformation and this coincides with changes in gene expression (Nativio, Wendt et al., 2009). These observations indicated that both cohesin and CTCF are required for proper imprinted gene expression via formation of higher-order chromatin structure. Thus

it was hypothesized that CTCF might recruit cohesin to the shared sites and cohesin might establish the loop structure entrapping the chromatin strands, as it was suggested for its role in sister chromatid cohesion. In addition to the imprinted IGF2/H19 locus an “architectural role” for cohesin and CTCF was also found in other loci. Cohesin and CTCF were found to promote long-range chromosomal interactions at the human *Interferon gamma* (*Infγ*) locus and at the human *apolipoprotein gene cluster* (*APO*) and depletion of either cohesin or CTCF led to disruption of the respective locus organization and altered gene expression (Hadjur, Williams et al., 2009; Mishiro, Ishihara et al., 2009).

Cohesin and tissue-specific transcription factors

Several genome-wide studies have shown that a fraction of cohesin sites colocalizes with tissue-specific transcription factors independently from CTCF, suggesting a role for cohesin in tissue-specific gene expression (Kagey, Newman et al., 2010; Schmidt, Schwalie et al., 2010; Faure, Schmidt et al., 2012).

In *mouse embryonic stem cells* (mES), CTCF-independent cohesin sites occupy the enhancer and the core promoter of several active genes together with the transcriptional activator complex *mediator*. In *mouse embryonic fibroblasts* (MEFs) it was shown that cohesin and mediator bound to a different set of promoters when compared to mES, demonstrating the cell-type specificity of the binding (Kagey, Newman et al., 2010). Furthermore, it was demonstrated that cohesin and mediator promote cell-type-specific DNA loops favoring the interactions between distal enhancers and specific promoters. For instance, in mES enhancer-promoter interactions mediated by cohesin/mediator were detected for the active genes *Nanog*, *Phc1*, *Oct4* and *Lefty1*. These interactions were not observed in MEFs where *Nanog*, *Phc1*, *Oct4* and *Lefty1* are silent and not occupied by mediator and cohesin (Kagey, Newman et al., 2010). In human breast cancer cells (MCF-7) cohesin was found to co-bind with the *estrogen-receptor alpha* ($ER\alpha$) and in hepatocellular carcinoma cells (HepG2) with the two liver-specific transcription factors *CEBPA* and *HNF4A* independently from CTCF, indicating also in this case a tissue-specific binding pattern of the complex (Schmidt, Schwalie et al., 2010). In the breast cancer cell line MCF-7 cohesin was found enriched at $ER\alpha$ binding sites that correlate with estrogen-regulated genes, suggesting a role of cohesin in mediating estrogen-dependent transcriptional responses (Schmidt, Schwalie et al., 2010). ER was already found to be involved in long-range interactions for proper transcription of estrogen-regulated genes (Carroll, Liu et al., 2005; Fullwood, Liu et al., 2009) and interestingly, cohesin was significantly enriched at ER binding sites involved in chromatin interactions compared to ER binding events not implicated in long-range contacts. In contrast, CTCF binding was not observed to be involved in these interactions (Schmidt, Schwalie et al., 2010).

These observations indicate that the CTCF-independent cohesin sites overlap with the binding of tissue-specific transcription factors suggesting an important role of cohesin also in cell-type-specific gene regulation.

Aim of this thesis

The chromatin fiber is organized in topological domains. How the chromatin fiber is folded within these domains and which proteins are responsible for the establishment of these interactions is poorly understood as well as how these structures influence genes.

The cohesin complex and the insulator protein CCCTC-binding factor (CTCF) are factors involved in the formation of interactions but even though a role for both proteins in promoting higher-order chromatin interactions has been established at individual loci, their genome-wide function is still to be clarified. The aim of this thesis is to elucidate the role for cohesin and CTCF in organizing the architecture of the human genome and their contribution to gene regulation.

Chapter 1 provides a general overview of the current knowledge on nuclear organization and how gene regulation might work via formation of higher-order chromatin structure. It also summarizes the currently available methods to study the folding of the chromatin and describes in more details two of the factors involved in long-range interactions, the cohesin complex and CTCF.

Chapter 2 and 3 of this thesis focus on how cohesin and CTCF co-operate to organize the architecture of the genome. We developed a technique to assess long-range interactions of multiple regions of interest (viewpoints) in parallel. This method is explained in detail in *Chapter 2*. In *Chapter 3* the use of this technique as an approach to unravel the cohesin's role in establishing interactions at the human chromosome 11 p15.5 region is described. In addition, the global functions of both cohesin and CTCF in organizing the structure of the human genome is revealed using the genome-wide method, Hi-C.

The fact that mutations in the cohesin-regulatory protein NIPBL and in cohesin subunits SMC1A and SMC3 are the causes of the developmental syndrome Cornelia de Lange reveal an important role of the cohesin pathway during development. In *Chapter 4*, the role of NIPBL as (co)transcription factor will be described. We propose that NIPBL can regulate gene expression directly on the gene where it is bound and indirectly via loading of cohesin. These observations suggest that NIPBL mutation might alter the transcription of developmental genes contributing to the CdLS phenotype.

References

- Anderson D.E., A. Losada, et al., 2002. Condensin and cohesin display different arm conformations with characteristic hinge angles. *J Cell Biol* 156, 3, 419-424.
- Bell A.C. and G. Felsenfeld, 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* 405, 6785, 482-485.
- Bell A.C., A.G. West, et al., 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98, 3, 387-396.
- Bolzer A., G. Kreth, et al., 2005. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol* 3, 5, e157.
- Borges V., C. Lehane, et al., 2010. Hos1 deacetylates Smc3 to close the cohesin acetylation cycle. *Mol Cell* 39, 5, 677-688.
- Boyle S., S. Gilchrist, et al., 2001. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum Mol Genet* 10, 3, 211-219.
- Branco M.R. and A. Pombo, 2006. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol* 4, 5, e138.
- Bulger M. and M. Groudine, 1999. Looping versus linking: toward a model for long-distance gene activation. *Genes Dev* 13, 19, 2465-2477.
- Carroll J.S., X.S. Liu, et al., 2005. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* 122, 1, 33-43.
- Cook P.R., 2003. Nongenic transcription, gene regulation and action at a distance. *J Cell Sci* 116, Pt 22, 4483-4491.
- Cremer M., F. Grasser, et al., 2008. Multicolor 3D fluorescence in situ hybridization for imaging interphase chromosomes. *Methods in molecular biology (Clifton, N.J)* 463, 205-239.
- Cremer T., C. Cremer, et al., 1982. Analysis of chromosome positions in the interphase nucleus of Chinese hamster cells by laser-UV-microirradiation experiments. *Hum Genet* 62, 3, 201-209.
- Cremer T. and M. Cremer, 2010. Chromosome territories. *Cold Spring Harb Perspect Biol* 2, 3, a003889.
- Deardorff M.A., M. Bando, et al., 2012. HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. *Nature* 489, 7415, 313-317.
- Deardorff M.A., M. Kaur, et al., 2007. Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of cornelia de Lange syndrome with predominant mental retardation. *Am J Hum Genet* 80, 3, 485-494.
- Dekker J., K. Rippe, et al., 2002. Capturing chromosome conformation. *Science* 295, 5558, 1306-1311.
- Dixon J.R., S. Selvaraj, et al., 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 7398, 376-380.

- Donze D., C.R. Adams, et al., 1999. The boundaries of the silenced HMR domain in *Saccharomyces cerevisiae*. *Genes Dev* 13, 6, 698-708.
- Dorsett D., 2007. Roles of the sister chromatid cohesion apparatus in gene expression, development, and human syndromes. *Chromosoma* 116, 1, 1-13.
- Dorsett D. and I.D. Krantz, 2009. On the molecular etiology of Cornelia de Lange syndrome. *Ann N Y Acad Sci* 1151, 22-37.
- Dostie J., T.A. Richmond, et al., 2006. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res* 16, 10, 1299-1309.
- Dreier M.R., M.E. Bekier, 2nd, et al., 2011. Regulation of sororin by Cdk1-mediated phosphorylation. *J Cell Sci* 124, Pt 17, 2976-2987.
- Drissen R., R.J. Palstra, et al., 2004. The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev* 18, 20, 2485-2490.
- Ernst J., P. Kheradpour, et al., 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473, 7345, 43-49.
- Fakan S. and R. van Driel, 2007. The perichromatin region: a functional compartment in the nucleus that determines large-scale chromatin folding. *Semin Cell Dev Biol* 18, 5, 676-681.
- Faure A.J., D. Schmidt, et al., 2012. Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules. *Genome Res* 22, 11, 2163-2175.
- Fraser P. and W. Bickmore, 2007. Nuclear organization of the genome and the potential for gene regulation. *Nature* 447, 7143, 413-417.
- Fullwood M.J., M.H. Liu, et al., 2009. An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* 462, 7269, 58-64.
- Fussner E., R.W. Ching, et al., 2011. Living without 30nm chromatin fibers. *Trends Biochem Sci* 36, 1, 1-6.
- Gruber S., P. Arumugam, et al., 2006. Evidence that loading of cohesin onto chromosomes involves opening of its SMC hinge. *Cell* 127, 3, 523-537.
- Gruber S., C.H. Haering, et al., 2003. Chromosomal cohesin forms a ring. *Cell* 112, 6, 765-777.
- Gurudatta B.V. and V.G. Corces, 2009. Chromatin insulators: lessons from the fly. *Brief Funct Genomic Proteomic* 8, 4, 276-282.
- Hadjur S., L.M. Williams, et al., 2009. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* 460, 7253, 410-413.
- Haering C.H., J. Lowe, et al., 2002. Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol Cell* 9, 4, 773-788.
- Hagege H., P. Klous, et al., 2007. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* 2, 7, 1722-1733.
- Hark A.T., C.J. Schoenherr, et al., 2000. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* 405, 6785, 486-489.

- Hirota T., D. Gerlich, et al., 2004. Distinct functions of condensin I and II in mitotic chromosome assembly. *J Cell Sci* 117, Pt 26, 6435-6445.
- Horsfield J.A., S.H. Anagnostou, et al., 2007. Cohesin-dependent regulation of Runx genes. *Development* 134, 14, 2639-2649.
- Ivanov D. and K. Nasmyth, 2005. A topological interaction between cohesin rings and a circular minichromosome. *Cell* 122, 6, 849-860.
- Jabs E.W., C.M. Tuck-Muller, et al., 1991. Studies of mitotic and centromeric abnormalities in Roberts syndrome: implications for a defect in the mitotic mechanism. *Chromosoma* 100, 4, 251-261.
- Jackson D.A., A.B. Hassan, et al., 1993. Visualization of focal sites of transcription within human nuclei. *Embo J* 12, 3, 1059-1065.
- Kagey M.H., J.J. Newman, et al., 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 7314, 430-435.
- Kanduri C., V. Pant, et al., 2000. Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr Biol* 10, 14, 853-856.
- Kim J.S., T.B. Krasieva, et al., 2002. Specific recruitment of human cohesin to laser-induced DNA damage. *J Biol Chem* 277, 47, 45149-45153.
- Kitajima T.S., T. Sakuno, et al., 2006. Shugoshin collaborates with protein phosphatase 2A to protect cohesin. *Nature* 441, 7089, 46-52.
- Klenova E.M., R.H. Nicolas, et al., 1993. CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Mol Cell Biol* 13, 12, 7612-7624.
- Kolesky S.E., M. Ouhammouch, et al., 2002. The mechanism of transcriptional activation by the topologically DNA-linked sliding clamp of bacteriophage T4. *J Mol Biol* 321, 5, 767-784.
- Kouzarides T., 2007. Chromatin modifications and their function. *Cell* 128, 4, 693-705.
- Krantz I.D., J. McCallum, et al., 2004. Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of *Drosophila melanogaster* Nipped-B. *Nat Genet* 36, 6, 631-635.
- Kueng S., B. Hegemann, et al., 2006. Wapl controls the dynamic association of cohesin with chromatin. *Cell* 127, 5, 955-967.
- Kurukuti S., V.K. Tiwari, et al., 2006. CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc Natl Acad Sci U S A* 103, 28, 10684-10689.
- Lafont A.L., J. Song, et al., 2010. Sororin cooperates with the acetyltransferase Eco2 to ensure DNA replication-dependent sister chromatid cohesion. *Proc Natl Acad Sci U S A* 107, 47, 20364-20369.
- Lieberman-Aiden E., N.L. van Berkum, et al., 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 5950, 289-293.

- Lipp J.J., T. Hirota, et al., 2007. Aurora B controls the association of condensin I but not condensin II with mitotic chromosomes. *J Cell Sci* 120, Pt 7, 1245-1255.
- Liu J., Z. Zhang, et al., 2009. Transcriptional dysregulation in NIPBL and cohesin mutant human cells. *PLoS Biol* 7, 5, e1000119.
- Lobanenkov V.V., R.H. Nicolas, et al., 1990. A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. *Oncogene* 5, 12, 1743-1753.
- Losada A., M. Hirano, et al., 1998. Identification of Xenopus SMC protein complexes required for sister chromatid cohesion. *Genes Dev* 12, 13, 1986-1997.
- Losada A., M. Hirano, et al., 2002. Cohesin release is required for sister chromatid resolution, but not for condensin-mediated compaction, at the onset of mitosis. *Genes Dev* 16, 23, 3004-3016.
- Luger K., M.L. Dechassa, et al., 2012. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat Rev Mol Cell Biol* 13, 7, 436-447.
- Luger K., A.W. Mader, et al., 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 6648, 251-260.
- Majumder P., J.A. Gomez, et al., 2006. The human major histocompatibility complex class II HLA-DRB1 and HLA-DQA1 genes are separated by a CTCF-binding enhancer-blocking element. *J Biol Chem* 281, 27, 18435-18443.
- Majumder P., J.A. Gomez, et al., 2008. The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. *J Exp Med* 205, 4, 785-798.
- Manuelidis L., 1985. Individual interphase chromosome domains revealed by in situ hybridization. *Hum Genet* 71, 4, 288-293.
- Markaki Y., D. Smeets, et al., 2012. The potential of 3D-FISH and super-resolution structured illumination microscopy for studies of 3D nuclear architecture: 3D structured illumination microscopy of defined chromosomal structures visualized by 3D (immuno)-FISH opens new perspectives for studies of nuclear architecture. *Bioessays* 34, 5, 412-426.
- Maston G.A., S.K. Evans, et al., 2006. Transcriptional regulatory elements in the human genome. *Annual review of genomics and human genetics* 7, 29-59.
- Michaelis C., R. Ciosk, et al., 1997. Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell* 91, 1, 35-45.
- Mishiro T., K. Ishihara, et al., 2009. Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *Embo J* 28, 9, 1234-1245.
- Monnich M., S. Banks, et al., 2009. Expression of cohesin and condensin genes during zebrafish development supports a non-proliferative role for cohesin. *Gene Expr Patterns*.
- Munkel C., R. Eils, et al., 1999. Compartmentalization of interphase chromosomes observed in simulation and experiment. *J Mol Biol* 285, 3, 1053-1065.

- Murrell A., S. Heeson, et al., 2004. Interaction between differentially methylated regions partitions the imprinted genes *Igf2* and *H19* into parent-specific chromatin loops. *Nat Genet* 36, 8, 889-893.
- Musio A., A. Selicorni, et al., 2006. X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. *Nat Genet* 38, 5, 528-530.
- Nativio R., K.S. Wendt, et al., 2009. Cohesin is required for higher-order chromatin conformation at the imprinted *IGF2-H19* locus. *PLoS Genet* 5, 11, e1000739.
- Nishiyama T., R. Ladurner, et al., 2010. Sororin mediates sister chromatid cohesion by antagonizing Wapl. *Cell* 143, 5, 737-749.
- Nora E.P., B.R. Lajoie, et al., 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 7398, 381-385.
- Oka Y., K. Suzuki, et al., 2011. Recruitment of the cohesin loading factor NIPBL to DNA double-strand breaks depends on MDC1, RNF168 and HP1gamma in human cells. *Biochem Biophys Res Commun* 411, 4, 762-767.
- Osborne C.S., L. Chakalova, et al., 2004. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet* 36, 10, 1065-1071.
- Oudet P., M. Gross-Bellard, et al., 1975. Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell* 4, 4, 281-300.
- Parelho V., S. Hadjur, et al., 2008. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 132, 3, 422-433.
- Phillips J.E. and V.G. Corces, 2009. CTCF: master weaver of the genome. *Cell* 137, 7, 1194-1211.
- Potts P.R., M.H. Porteus, et al., 2006. Human SMC5/6 complex promotes sister chromatid homologous recombination by recruiting the SMC1/3 cohesin complex to double-strand breaks. *Embo J* 25, 14, 3377-3388.
- Rajapakse I. and M. Groudine, 2011. On emerging nuclear order. *J Cell Biol* 192, 5, 711-721.
- Recillas-Targa F., M.J. Pikaart, et al., 2002. Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc Natl Acad Sci U S A* 99, 10, 6883-6888.
- Remeseiro S., A. Cuadrado, et al., 2012. A unique role of cohesin-SA1 in gene regulation and development. *Embo J* 31, 9, 2090-2102.
- Remeseiro S. and A. Losada, 2013. Cohesin, a chromatin engagement ring. *Curr Opin Cell Biol* 25, 1, 63-71.
- Rollins R.A., M. Korom, et al., 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. *Mol Cell Biol* 24, 8, 3100-3111.
- Rollins R.A., P. Morcillo, et al., 1999. Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. *Genetics* 152, 2, 577-593.

- Rouquette J., C. Genoud, et al., 2009. Revealing the high-resolution three-dimensional network of chromatin and interchromatin space: a novel electron-microscopic approach to reconstructing nuclear architecture. *Chromosome Res* 17, 6, 801-810.
- Sachs R.K., G. van den Engh, et al., 1995. A random-walk/giant-loop model for interphase chromosomes. *Proc Natl Acad Sci U S A* 92, 7, 2710-2714.
- Schmidt D., P.C. Schwalie, et al., 2010. A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res* 20, 5, 578-588.
- Schmitz J., E. Watrin, et al., 2007. Sororin is required for stable binding of cohesin to chromatin and for sister chromatid cohesion in interphase. *Curr Biol* 17, 7, 630-636.
- Schneider R. and R. Grosschedl, 2007. Dynamics and interplay of nuclear architecture, genome organization, and gene expression. *Genes Dev* 21, 23, 3027-3043.
- Schuldiner O., D. Berdnik, et al., 2008. piggyBac-Based Mosaic Screen Identifies a Postmitotic Function for Cohesin in Regulating Developmental Axon Pruning. *Dev Cell* 14, 2, 227-238.
- Sekimata M., M. Perez-Melgosa, et al., 2009. CCCTC-binding factor and the transcription factor T-bet orchestrate T helper 1 cell-specific structure and function at the interferon-gamma locus. *Immunity* 31, 4, 551-564.
- Simonis M., P. Klous, et al., 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* 38, 11, 1348-1354.
- Sjogren C. and K. Nasmyth, 2001. Sister chromatid cohesion is required for postreplicative double-strand break repair in *Saccharomyces cerevisiae*. *Curr Biol* 11, 12, 991-995.
- Soler E., C. Andrieu-Soler, et al., 2010. The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes Dev* 24, 3, 277-289.
- Splinter E., H. Heath, et al., 2006. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev* 20, 17, 2349-2354.
- Stack S.M., D.B. Brown, et al., 1977. Visualization of interphase chromosomes. *J Cell Sci* 26, 281-299.
- Stadhouders R., P. Kolovos, et al., 2013. Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nat Protoc* 8, 3, 509-524.
- Stadhouders R., S. Thongjuea, et al., 2012. Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development. *Embo J* 31, 4, 986-999.
- Sumara I., E. Vorlaufer, et al., 2000. Characterization of vertebrate cohesin complexes and their regulation in prophase. *J Cell Biol* 151, 4, 749-762.
- Sumara I., E. Vorlaufer, et al., 2002. The dissociation of cohesin from chromosomes in prophase is regulated by Polo-like kinase. *Mol Cell* 9, 3, 515-525.
- Tang Z., H. Shu, et al., 2006. PP2A is required for centromeric localization of Sgo1 and proper chromosome segregation. *Dev Cell* 10, 5, 575-585.

- Tolhuis B., R.J. Palstra, et al., 2002. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* 10, 6, 1453-1465.
- Tost J., H. Jammes, et al., 2006. Non-random, individual-specific methylation profiles are present at the sixth CTCF binding site in the human H19/IGF2 imprinting control region. *Nucleic Acids Res* 34, 19, 5438-5448.
- Unal E., J.M. Heidinger-Pauli, et al., 2008. A molecular determinant for the establishment of sister chromatid cohesion. *Science* 321, 5888, 566-569.
- Vakoc C.R., D.L. Letting, et al., 2005. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol Cell* 17, 3, 453-462.
- van Berkum N.L., E. Lieberman-Aiden, et al., 2010. Hi-C: a method to study the three-dimensional architecture of genomes. *J Vis Exp*, 39.
- Vega H., Q. Waisfisz, et al., 2005. Roberts syndrome is caused by mutations in ESCO2, a human homolog of yeast ECO1 that is essential for the establishment of sister chromatid cohesion. *Nat Genet* 37, 5, 468-470.
- Visser A.E., F. Jaunin, et al., 2000. High resolution analysis of interphase chromosome domains. *J Cell Sci* 113 (Pt 14), 2585-2593.
- Vostrov A.A. and W.W. Quitschke, 1997. The zinc finger protein CTCF binds to the APBbeta domain of the amyloid beta-protein precursor promoter. Evidence for a role in transcriptional activation. *J Biol Chem* 272, 52, 33353-33359.
- Waizenegger I.C., S. Hauf, et al., 2000. Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase. *Cell* 103, 3, 399-410.
- Watrin E. and J.M. Peters, 2006. Cohesin and DNA damage repair. *Exp Cell Res* 312, 14, 2687-2693.
- Wendt K.S. and J.M. Peters, 2009. How cohesin and CTCF cooperate in regulating gene expression. *Chromosome Res* 17, 2, 201-214.
- Wendt K.S., K. Yoshida, et al., 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451, 7180, 796-801.
- Williams B.C., C.M. Garrett-Engele, et al., 2003. Two putative acetyltransferases, san and deco, are required for establishing sister chromatid cohesion in *Drosophila*. *Curr Biol* 13, 23, 2025-2036.
- Yazdi P.T., Y. Wang, et al., 2002. SMC1 is a downstream effector in the ATM/NBS1 branch of the human S-phase checkpoint. *Genes Dev* 16, 5, 571-582.
- Zhang B., J. Chang, et al., 2009. Dosage effects of cohesin regulatory factor PDS5 on mammalian development: implications for cohesinopathies. *PLoS ONE* 4, 5, e5232.
- Zhang B., S. Jain, et al., 2007. Mice lacking sister chromatid cohesion protein PDS5B exhibit developmental abnormalities reminiscent of Cornelia de Lange syndrome. *Development* 134, 17, 3191-3201.
- Zhang N. and D. Pati, 2009. Handcuff for sisters: a new model for sister chromatid cohesion. *Cell Cycle* 8, 3, 399-402.

Chapter 2



Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions

Ralph Stadhouders, Petros Kolovos, Rutger Brouwer, Jessica Zuin, Anita van den Heuvel, Christel Kockx, Robert-Jan Palstra, Kerstin Wendt, Frank Grosveld, Wilfred van Ijcken & Eric Soler.

Nature Protocols 8, 3, 509-524 (2013)

Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions

Ralph Stadhouders^{1,6}, Petros Kolovos^{1,6}, Rutger Brouwer^{2,3,6}, Jessica Zuin¹, Anita van den Heuvel¹, Christel Kockx², Robert-Jan Palstra¹, Kerstin S Wendt¹, Frank Grosveld^{1,4}, Wilfred van Ijcken² & Eric Soler^{1,4,5}

¹Department of Cell Biology, Erasmus Medical Center, Rotterdam, The Netherlands. ²Center for Biomics, Erasmus Medical Center, Rotterdam, The Netherlands.

³Netherlands Bioinformatics Centre (NBIC), Nijmegen, The Netherlands. ⁴Cancer Genomics Center, Erasmus Medical Center, Rotterdam, The Netherlands. ⁵Laboratory of Hematopoiesis and Leukemic Stem Cells (LSHL), French Alternative Energies and Atomic Energy Commission (CEA)/Institut National de la Santé et de la Recherche Médicale (INSERM) U967, Fontenay-aux-Roses, France. ⁶These authors contributed equally to this work. Correspondence should be addressed to W.v.I. (w.vanijcken@erasmusmc.nl) or E.S. (eric.soler@cea.fr).

Published online 14 February 2013; doi:10.1038/nprot.2013.018

Chromosome conformation capture (3C) technology is a powerful and increasingly popular tool for analyzing the spatial organization of genomes. Several 3C variants have been developed (e.g., 4C, 5C, ChIA-PET, Hi-C), allowing large-scale mapping of long-range genomic interactions. Here we describe multiplexed 3C sequencing (3C-seq), a 4C variant coupled to next-generation sequencing, allowing genome-scale detection of long-range interactions with candidate regions. Compared with several other available techniques, 3C-seq offers a superior resolution (typically single restriction fragment resolution; approximately 1–8 kb on average) and can be applied in a semi-high-throughput fashion. It allows the assessment of long-range interactions of up to 192 genes or regions of interest in parallel by multiplexing library sequencing. This renders multiplexed 3C-seq an inexpensive, quick (total hands-on time of 2 weeks) and efficient method that is ideal for the in-depth analysis of complex genetic loci. The preparation of multiplexed 3C-seq libraries can be performed by any investigator with basic skills in molecular biology techniques. Data analysis requires basic expertise in bioinformatics and in Linux and Python environments. The protocol describes all materials, critical steps and bioinformatics tools required for successful application of 3C-seq technology.

INTRODUCTION

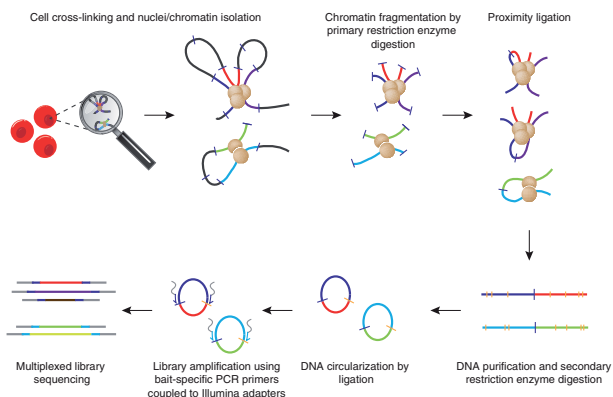
In recent years, it has become evident that the 3D organization of genomes is not random. Numerous studies have implicated long-range chromosomal interactions in several crucial cellular processes, including the regulation of gene expression^{1–4}. Indeed, chromatin coassociations mediated by chromatin looping provide a means by which distal enhancers communicate with their target genes and stimulate transcription^{5–7}. Accordingly, methods providing efficient and sensitive detection of chromatin looping events with high resolution are becoming increasingly popular. The development of 3C technology has revolutionized the analysis of spatial genomic organization by allowing the detection of chromatin coassociations with a resolution far beyond that provided by light microscopy-based studies⁸. 3C relies on the ability of distal DNA fragments to be ligated together when positioned in close proximity in the nuclear space. Over the past decade, several 3C variants have been developed, offering the possibility of analyzing chromatin looping events on a genome-wide scale (e.g., 4C^{9–12}, 5C¹³, ChIA-PET¹⁴, Hi-C¹⁵). We describe here in detail multiplexed 3C-seq, a 3C variant coupled to high-throughput sequencing that we recently developed^{16,17}. Multiplexed 3C-seq allows genome-scale simultaneous detection of long-range chromatin interactions of numerous genomic elements in parallel and can be applied to low numbers of cells (from 1×10^6 cells¹⁸ to as low as 300,000 cells (P.K. and E.S., unpublished data)). We recently used this technique to analyze the spatial organization of several loci, including the mouse β -globin (*Hbb*), myeloblastosis oncogene (*Myb*) and IgK loci (*Igk*), revealing crucial enhancer-gene communications^{16–18}.

Overview of the procedure

All 3C-based procedures use formaldehyde fixation of living cells or fresh tissues to preserve genomic architecture in its native state before fragmentation by restriction enzyme digestion. The digested cross-linked chromatin is subjected to a ligation reaction under dilute conditions, favoring intramolecular ligation events over intermolecular ligation events (proximity ligation). This step yields a 3C library composed of chimeric DNA molecules resulting from the ligation of (distal) chromatin fragments that were in physical proximity in the nuclear space (Fig. 1). The subsequent steps differ depending on the type of assay used. The 3C library can be directly analyzed by probing for specific interactions by PCR^{19,20} or further processed for more global analyses using bait-specific primers (e.g., promoter-specific primer pair^{9–12,16–18}) or whole-genome looping assays as in Hi-C¹⁵. In the 3C-seq procedure, the 3C library is subjected to a second restriction enzyme digestion using a frequent cutter, and fragments are circularized before an inverse PCR step using bait-specific primers (Fig. 1), similar to the original microarray-based 4C protocol¹¹. This second restriction digest is necessary to decrease the size of the DNA circles, resulting in fragments that can be PCR-amplified efficiently. The inverse PCR products contain the DNA elements that were captured (i.e., ligated) by the bait sequence and thereby represent its native chromatin environment in the nucleus. The 3C-seq library is then directly sequenced on an Illumina HiSeq2000 platform, with the possibility of multiplexing sample sequencing by pooling up to 12 different bait-specific 3C-seq libraries in a single lane of a HiSeq2000 flow cell, providing marked cost reduction and increased throughput.

PROTOCOL

Figure 1 | Overview of the multiplexed 3C-seq procedure. Nuclei from cross-linked cells are digested (primary restriction enzyme) and ligated under dilute conditions to physically link *in vivo* interacting DNA fragments. After a secondary digestion (secondary restriction enzyme) and ligation, inverse PCR is performed using bait-specific primers containing Illumina sequencing adapters to amplify unknown fragments interacting with the bait. PCR samples generated with different primer sets are then pooled and subjected to multiplexed library sequencing.



Other sequencing platforms are, in principle, compatible with multiplexed 3C-seq, but the multiplexing/de-multiplexing steps and associated informatics tools described here may need further optimization and adjustments.

Comparison of 3C-seq with other 3C-based methods

The choice between 3C and the different derivatives strongly depends on the biological question under consideration (Table 1). Although 3C-qPCR is particularly suited to quantitatively probe for specific interactions and interrogate a restricted number of chosen chromatin coassociations, it rapidly becomes technically demanding when large chromosomal domains are under investigation or when numerous interactions need to be analyzed in parallel for *de novo* detection of chromatin looping events. In the latter cases, high-throughput 3C derivatives such as 4C, 5C, 3C-seq or Hi-C technologies will be preferred. The 4C approach^{10,11} consists of a large-scale analysis of chromatin interactions with a chosen bait sequence by probing the 4C library on DNA microarrays. It produces chromatin interaction maps of a single bait, with the coverage depending on the array used. 4C has the advantage of allowing unbiased detection of unknown bait-specific interactions, but is limited by the number of arrays needed to

achieve genome-wide coverage and by the saturation of signals around the bait sequence, preventing the detection of medium- to close-range interactions (up to 200 kb away). The 5C variant¹³ overcomes this limitation and offers the possibility of exploring every potential chromatin coassociation in large subchromosomal domains by using primer sets covering all possible interactions. It is, however, difficult to reach genome-wide coverage using 5C, as it requires extremely large numbers of primers for all possible intrachromosomal and interchromosomal interactions. HiC, in contrast, provides a global genome-wide analysis of all possible chromatin associations by coupling a modified 3C procedure to high-throughput sequencing¹⁵. Although it is extremely powerful, Hi-C requires substantial computational resources, and the number of sequence reads needed to obtain high coverage of mammalian genomes renders it very expensive and, as a consequence, unaffordable for a large number of academic laboratories.

TABLE 1 | Comparison between different 3C variants.

3C-based method	Applications	Advantages	Limitations
3C-(q)PCR ^{19,20}	One-to-one	Relatively simple analysis (no bioinformatics required)	Laborious, knowledge of locus required, proper controls are essential
3C-on-chip (4C) ^{9–11}	One-to-all	Relatively simple data analysis	Poor signal-to-noise ratio, difficult to obtain genome-wide coverage
3C sequencing (3C-seq or 4C-seq) ^{12,16}	One-to-all	Genome-wide coverage, high resolution, good signal-to-noise ratio, allows multiplexing for high-throughput	Restricted to a single view point per experiment (except when multiplexing), analysis requires some bioinformatics expertise
Multiplexed 3C-seq ^{17,18}	Many-to-all		
3C carbon copy (5C) ¹³	Many-to-many	Explores interactions between many individual fragments simultaneously (instead of using a single viewpoint)	No genome-wide coverage, primer design can be challenging
Hi-C ¹⁵	All-to-all	Explores the genome-wide interactions between all individual fragments simultaneously	Obtaining high resolution requires a massive sequencing effort; expensive, complicated analysis

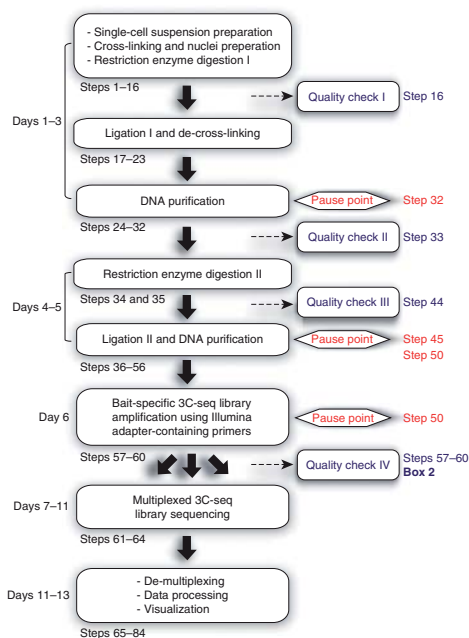


Figure 2 | Flowchart of multiplexed 3C-seq data generation and processing. Steps involved in the multiplexed 3C-seq procedure are shown in blue rectangles. Time needed to complete these steps is depicted on the left. Pause points are indicated together with the timing of the different quality checkpoints: I, primary digestion efficiency (Step 16); II, ligation efficiency (Step 33); III, secondary digestion efficiency (Step 44); IV, 3C-seq PCR performance (Steps 57–60 and **Box 2**).

© 2013 Nature America, Inc. All rights reserved.



3C-seq provides a fast and affordable genome-scale 3C alternative (**Fig. 2**). The use of high-throughput sequencing eliminates the problems of limited coverage and saturating signals associated with microarray technology and markedly increases resolution and signal-to-noise ratios. A disadvantage of 3C-seq is that, as in 4C, the analysis is restricted to a single bait sequence and does not provide deep characterization of chromatin coassociations of several regulatory elements in parallel. The multiplexed 3C-seq protocol presented here (**Figs. 1 and 2**) addresses this limitation and shows that, by efficiently multiplexing bait-specific library sequencing, genome-scale interactions of up to 192 different genomic elements can be assessed in parallel on an Illumina HiSeq2000 platform, thereby markedly increasing the throughput of the technique and decreasing sequencing costs. Moreover, 3C-seq data analysis is facilitated by the availability of bioinformatics tools. We provide here a dedicated analysis pipeline facilitating the entire data handling process, including de-multiplexing, alignment and visualization. Together, this renders multiplexed 3C-seq an inexpensive and efficient method for in-depth analysis of complex genetic loci and genomic regulatory regions.

Applications of the method

3C-seq can be applied to any nonrepetitive region of a genome. It is generally used to unravel medium- to long-range interactions (i.e., few kb to hundreds of kb) of a genomic element of interest. It is usually applied to detect interactions between promoter elements and the surrounding regions, or to connect distal enhancers to their target gene(s). With the recent developments in high-throughput chromatin occupancy profiling²¹, large numbers of transcription factor binding and chromatin modification data sets are becoming available. Combined with this knowledge, 3C-seq can be used to analyze the functional relationships existing between regulatory elements, sites of active transcription, gene deserts or boundary elements where transitions in chromatin structure or transcription are observed (e.g., insulator elements or initiation sites for productive transcription elongation).

Limitations of 3C-seq

Similar to all 3C-based procedures, 3C-seq only provides topological information. The control experiments discussed in Experimental design will help validate and ensure the specificity of the observed interactions. Even so, it is recommended to combine 3C-seq data with results from complementary experiments (e.g., fluorescence *in situ* hybridization (FISH), gene expression analysis, chromatin immunoprecipitation (ChIP)^{7,17,22} or, even better, with functional experiments, before drawing conclusions on the functional impact of chromatin coassociations.

Experimental design

Fixing cells. Cell fixation, which represents the starting point of the procedure, provides the template for the essential proximity ligation step used to capture DNA-DNA interactions. Fixation conditions need to be standardized for increased reproducibility and efficient comparison between samples. In our hands, formaldehyde fixation conditions used in ChIP experiments (1–2% (vol/vol) formaldehyde, 10 min at room temperature (18–22 °C)) work well for 3C-seq^{16–18}. More extensive fixation protocols have been reported to improve signal-to-noise ratios in the distance range of a few kb (ref. 23), although this protocol utilizes more frequently cutting restriction enzymes to obtain such resolution and might therefore be difficult to compare with our protocol.

Starting material. We have used many human and mouse cell or tissue types in 3C-seq experiments (**Table 2**), although certain cell or tissue types (e.g., fibroblasts) can be more difficult to handle. The use of single-cell suspensions is essential when performing 3C-seq (and other 3C-based protocols, for that matter). When working with tissues that are difficult to dissociate (e.g., brain, heart, lung), consider treating them with collagenase before formaldehyde fixation (see PROCEDURE Step 1 and TROUBLESHOOTING section). Previously published 3C (and derivate) protocols describe using 10⁶ cells or more per experiment. We, however, have successfully applied 3C-seq on much smaller numbers of cells (i.e., FACS-sorted cell populations, using < 10⁶ cells), further extending its applicability (P.K. and E.S., unpublished data, and ref. 18).

Restriction enzyme choice. The resolution of a 3C-seq experiment depends on the first restriction enzyme used. Ideally, the restriction pattern given by the enzyme should provide evenly distributed

PROTOCOL

TABLE 2 | Performance of different cell types and tissues successfully used for 3C-seq.

Cell or tissue type	Performance in 3C-seq	Special requirements
Hematopoietic cell types: mouse and human erythroid cells (FACS sorted and cultured), mouse B and T lymphocytes (FACS sorted and cultured), mouse erythroleukemia cell lines (MEL, I11) Hematopoietic tissue (mouse fetal liver E12.5-15.5, human fetal liver) Mouse ES cells (IB10), ES-derived Flk1 ⁺ cells (magnetic-activated cell sorting (MACS)-sorted) HeLa cells	Excellent	None
Other mouse tissues (Mouse fetal brain E12.5-15.5) Rat tissues (liver, heart and lung)	Good	Use a collagenase treatment (PROCEDURE Step 1) to obtain a single-cell suspension for efficient cross-linking
Human primary melanocytes ³³ Fibroblast cells: cell lines (NIH3T3) and primary cells (mouse dermal fibroblasts, mouse and human lung fibroblasts) HEK/293T cells K562 cells HUVEC cells Human ES cells (H9)	Poor: extensive nuclei aggregation resulting in poor digestion efficiencies	Ensure gentle handling of the cells and nuclei. Preferentially collect adherent cells with a scraper instead of trypsin. In case of aggregation, see Table 3 for additional troubleshooting. Melanin produced by melanocytes is a potent PCR inhibitor and can be removed using a suitable column purification step ³³

© 2013 Nature America, Inc. All rights reserved.

fragments, separating the different regulatory elements of interest (e.g., promoter, enhancers). When possible, check for the presence of regulatory elements, transcription factor binding sites and histone modification patterns relevant for the tissue to be analyzed using publicly accessible databases such as ENCODE (<http://genome.ucsc.edu/ENCODE/>) in order to determine the most appropriate enzyme for the region of interest. We suggest using 6-base-recognizing enzymes (referred to as a 'six-cutter') such as EcoRI, HindIII, BglII, BamHI and XhoI, which perform well on cross-linked chromatin.

The enzymes should be insensitive to mammalian DNA methylation in order to prevent introducing digestion biases. We observed that the use of a six-cutter yields better reproducibility at the single restriction fragment level than enzymes that cut more frequently (e.g., 4-base-recognizing enzymes, referred to as a 'four-cutter'). The latter generate many more fragments per kb, which may lead to a poorer signal-to-noise ratio owing to more frequent intermolecular ligations. This could result in interaction signals being spread over several restriction fragments, thereby yielding



Box 1 | 3C-seq primer design

Two primers, a P5 primer and a P7 primer, need to be designed for each bait fragment of interest:

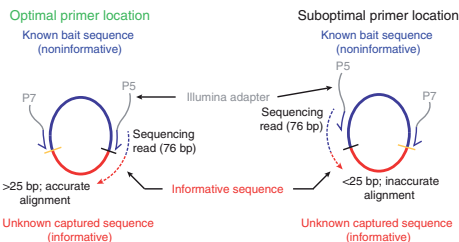
The P5 primer must be located as close as possible to the primary restriction enzyme site (usually the six-cutter). As only the sequence located after the restriction site is informative for identifying interacting fragments, the distance between the primary restriction enzyme primer and the restriction site itself should be minimized to ensure unambiguous alignment and identification of the interacting fragments (**Fig. 3**). This primer contains the P5 Illumina adapter sequence (5'-AATGATACGGGACCACCGAACACTCTTCCCTACACGACCTCTCCGATCT-3') to be placed upstream of the annealing sequence; **Fig. 3** from which library sequencing will be initiated. The sequencing reaction starts from the bait fragment, reads through the annealing primer sequence and extends into the unknown captured fragment. To allow more flexibility for primer design and to ensure optimal alignment of the sequences, we use a 76-bp sequencing read length (Step 64).

The second primer, located near the secondary restriction enzyme site (the four-cutter), contains the P7 Illumina adapter sequence (5'-CAAGCAGAAGCGGCATACGA-3', **Fig. 3**), and although it is required for the inverse PCR and the Illumina sequencing chemistry it is not sequenced (in contrast to paired-end sequencing, for which a different adapter is required). Therefore, the location of the P7 primer with regard to the secondary restriction site is more flexible (within 100 bp of the restriction site).

Actual primer requirements are similar to those used in standard PCR reactions. Oligo length is kept between 17 and 24 nt to facilitate efficient amplification and annealing temperatures are generally chosen between 54 and 59 °C. We regularly use primer design software (DNAMAN 5.0) to check these parameters and to ensure that primers are not prone to form dimers.

Note: Oligonucleotide sequences are copyright 2007–2012 Illumina. All rights reserved. Derivative works created by Illumina customers are authorized for use with Illumina instruments and products only. All other uses are strictly prohibited.

Figure 3 | 3C-seq primer design and positioning. Schematic drawing of the location of the inverse PCR primers used to amplify a 3C-seq library. The ring represents a circular DNA molecule composed of the bait fragment (blue) ligated to an unknown captured fragment (red). The two PCR primers are located on the bait fragment next to the restriction sites, with adapters shown as gray overhangs. The P5 primer is located next to the primary restriction site (black dash), and the P7 primer is located next to the secondary restriction site (yellow dash). Illumina sequencing is initiated from the P5 primer and extends into the unknown fragment (dashed arrow). If the P5 primer is located right next to the primary restriction site (within 50 bp), sequence reads generated will be long enough for highly accurate alignment (>25 bp, left). If the distance between the P5 primer and the primary restriction site becomes too large (>50 bp, right), accurate alignment might be compromised.



interaction profiles that are sometimes more difficult to interpret. For instance, enhancer-promoter communication might be difficult to analyze using a small four-cutter bait fragment encompassing the transcription start site, as in some cases enhancers tend to associate with slightly more downstream or upstream sequences, which may not be encompassed by the four-cutter fragment used in the analysis^{7,17,24}. We suggest using a four-cutter as the primary restriction enzyme only when you are refining interactions initially detected by a six-cutter or if interactions have to be investigated within a narrow genomic region. For the secondary restriction enzyme, any four-cutter insensitive to mammalian DNA methylation and with good re-ligation efficiencies can, in principle, be used. We have performed successful 3C-seq experiments using NlaIII, DpnII, HaeIII and MseI. The final combination of primary and secondary restriction enzymes will ultimately depend on their compatibility in terms of generating a suitable bait fragment for the inverse PCR primer design (see below and **Box 1**). To maximize efficient circularization in the second ligation step, the final bait fragment should be at least ~250 bp (ref. 25), although we have succeeded in obtaining good interaction profiles with bait fragments as small as 120–180 bp (ref. 18; P.K. and E.S., unpublished data). Please note that for some potential interacting fragments both restriction enzyme sites will be very close (<50 bp). When such a fragment ligates to the bait, the resulting sequencing reads might be problematic to align (see TROUBLESHOOTING section). Such a read is not a combination of the bait sequence and a single interacting fragment, as it will also contain sequences from the other side of the bait fragment. By trimming the 3' end of the reads (PROCEDURE Step 75), a large portion of these fragments can be rehabilitated.

Primer design. The 3C-seq library is amplified using primers annealing to the bait sequence, facing outward. Proper design of both primers for the inverse PCR is crucial in the 3C-seq procedure (**Box 1** and **Fig. 3**). Efficiency and reproducibility of the PCR primers are first tested without the addition of the Illumina adapters (**Box 2**). If performing well, oligonucleotides containing appropriate Illumina adapters are then tested again before being used in the final library amplification PCR before sequencing. For multiplexing purposes, the bait-specific primer sequence itself is used as a bar code to identify reads originating from each individual 3C-seq library. If identical bait-specific libraries need to be sequenced in parallel (e.g., the same promoter for different biological conditions), small bar codes (2–6 nt) may be added to the primers (PROCEDURE Step 62; **Box 3**).

Controls. 3C-seq data need to be interpreted carefully, as high interaction signals are not necessarily indicators of functionally relevant chromatin coassociations (also see the 'Limitations' section). Furthermore, the PCR amplification step may introduce biases owing to differences in fragment length and GC content, which can affect amplification efficiencies. To ensure proper data interpretation, consider including several control experiments²⁶. Whether an interaction is specific for a certain tissue/cell type or whether it correlates with the activity of a specific gene can be tested by analyzing different tissues/cell types or non-expressing cells, respectively. For example, we generally use embryonic stem (ES) cells, cell lines, tissues or FACS-sorted cells that do not express the gene under investigation as controls when investigating promoter-enhancer interactions of an active gene. In addition, using a captured interaction site of interest as bait in a 'reverse experiment' can provide excellent validation of the interaction.

Box 2 | 3C-seq PCR setup and optimization

As 3C-seq library fragments differ in length and abundance, we use the Expand long template system to minimize any biases resulting from these differences¹¹. Bait-specific primers (without adapters) are first tested for proper linearity and efficiency.

1. Test the increasing amounts of 3C-seq library DNA (up to 200 ng) using a 50- μ l PCR. Reaction components and conditions are described in PROCEDURE Step 57.
2. Analyze PCR products on a 1.5% (wt/vol) agarose gel, where they should appear as a reproducible smear of DNA fragments, usually showing two prominent bands¹¹. These prominent bands are the result of recircularization of the bait fragment in the first ligation step, and of detection of the neighboring fragment owing to incomplete digestion of the primary restriction site on the bait fragment¹¹.
3. Assess the linear range of the individual primer pairs by quantifying prominent bands in each reaction of the dilution range.
4. Order versions of the primer pairs that perform well, including the P5 and P7 Illumina adapter sequences (**Box 1**). Test these new primers as described in steps 1–3 of **Box 2**.
5. Use successful P5 and P7 primers to prepare 3C-seq samples for sequencing (PROCEDURE Steps 57–60).

Box 3 | 3C-seq pooling guidelines

The Illumina sequencers use the first four sequenced bases to locate the DNA clusters on the flow cell. When too little variation is present in these first bases, the DNA clusters will not be correctly recognized and base calling will be compromised. The following pooling guidelines are used to ensure that the sequencing process proceeds correctly.

1. Pool *at least* six samples together in a single lane for multiplexing. As one sample can be sequenced in multiple lanes, there is no physical limit as to how many samples can be pooled. We have regularly pooled up to 12 samples in one lane.
2. Ensure that at least one adenine and one thymine base are present in each of the first four cycles of a sample pool. The cycles with the highest intensity of the adenine and thymine bases are used for cluster recognition by the sequencer. Without these specific nucleotides in the first four bases, base calling will be compromised and the sequencing run will fail.
3. Do not pool samples generated with the same bait-specific PCR primer, as sequences derived from these samples cannot be discriminated in the downstream analysis. If pooling of such samples is desired, short bar-code sequences (2–6 nt) will have to be added to the adapter-containing bait-specific primers in the final PCRs (Step 57).

MATERIALS

- Freshly collected tissues, sorted populations of cells and/or cell lines
- **CAUTION** Approved governmental and institutional regulations must be followed and adhered to.
- FCS (Sigma-Aldrich, cat. no. A4781)
- DMEM (Gibco, cat. no. 41966)
- Glycine (1 M in PBS; Sigma-Aldrich, cat. no. G7126) **▲ CRITICAL** Glycine stocks should be stored at 4 °C and used cold. They can be stored for a maximum of 6 months.
- PBS (Sigma-Aldrich, cat. no. P4417)
- FCS/PBS (10% (vol/vol))
- Lysis buffer (see Reagent Setup)
- Sodium chloride (NaCl; Sigma-Aldrich, cat. no. S7653)
- Nonidet P-40 substitute (NP-40, Sigma-Aldrich, cat. no. 74385)
- Complete protease inhibitor, EDTA free (Roche, cat. no. 11873580001, see Reagent Setup)
- Milli-Q H₂O
- Collagenase, 2.5% (wt/vol) (Sigma-Aldrich, cat. no. C1639), in PBS
- Formaldehyde, 37% (vol/vol) (Merck, cat. no. 1039992500)
- **CAUTION** Formaldehyde is toxic.
- Restriction enzymes with 6-bp and 4-bp recognition sites and their corresponding buffers (see INTRODUCTION; Roche or New England Biolabs)
- SDS (20% (wt/vol); Sigma-Aldrich, cat. no. 05030)
- Triton X-100 (20% (vol/vol); Sigma-Aldrich, cat. no. T8787)
- T4 DNA ligation buffer (Roche, cat. no. 10799009001)
- T4 DNA ligase, high concentration (Roche, cat. no. 10799009001)
- Proteinase K (10 mg ml⁻¹, Sigma-Aldrich, cat. no. P2308)
- RNase (10 mg ml⁻¹, Sigma-Aldrich, cat. no. R6513)
- Phenol/chloroform/isoamyl alcohol (25:24:1 (vol/vol/vol); pH 8; Sigma-Aldrich, cat. no. 77617) **CAUTION** Phenol/chloroform is toxic.
- Glycogen (20 mg ml⁻¹, Roche, cat. no. 10901393001)
- Ethanol (100% (vol/vol) or 70% (vol/vol); Sigma-Aldrich, cat. no. 459844)
- Sodium acetate (2 M, pH 5.6; Sigma-Aldrich, cat. no. S2889)
- Tris-HCl (10 mM, pH 7.5, or 1 M, pH 8.0)
- Liquid N₂
- Agarose electrophoresis gels (0.6% and 1.5% (wt/vol))
- Expand long template system 10× buffer 1 (Roche, cat. no. 11759060001)

- dNTPs (10 mM each)
- Expand long template system DNA polymerase (Roche, cat. no. 11759060001)
- PCR primers (see INTRODUCTION)
- QIAquick gel extraction kit (Qiagen, cat. no. 28706)
- TruSeq SR cluster kit v3-cBot-HS (Illumina, cat. no. GD-401-3001)
- TruSeq SBS kit v3-HS (50 cycles) (Illumina, cat. no. FC-401-3002)
- Python 2.6 (<http://www.python.org/>)
- Illumina offline base calling software (http://support.illumina.com/sequencing/sequencing_software/offline_basecaller_olb.ilmn)
- NARWHAL (<https://trac.nbic.nl/narwhal/>)
- Pysam (<http://code.google.com/p/pysam/>)
- Supplementary analysis scripts (see **Supplementary Data**; the scripts `findSequence.py`, `regionsBetween.py`, `alignCounter.py` and `libutil.py` should be extracted to the same directory)

EQUIPMENT

- Cell strainer, 40 μm (BD Falcon, cat. no. 352340)
- Polypropylene centrifugation tubes (Greiner bio-one, cat. no. 188271)
- Safe-Lock 1.5-ml centrifugation tubes (Eppendorf, cat. no. 0030120.086)
- Thermomixer (Eppendorf, cat. no. EF4283)
- Water bath
- Microcentrifuge (Eppendorf, cat. no. 5417R)
- PCR thermocycler (MJ Research, cat. no. PTC-200)
- Spectrophotometer (NanoDrop 2000c, Thermo Scientific)
- Agilent 2100 Bioanalyzer (Agilent Technologies, cat. no. G2938C) with the 7500 DNA chip (cat. no. 5067-1506)
- Illumina HiSeq2000 high-throughput sequencing machine (Illumina)
- Excel spreadsheet software (Microsoft)
- Computer with a minimum of 8 Gb RAM and 1.5 Tb attached storage running a Linux distribution and the software listed above

REAGENT SETUP

Complete protease inhibitor, EDTA free Dissolve one tablet in 1 ml of PBS to create a 50× working solution. Store the solution at –20 °C for up to 2–3 months; avoid repeated freeze-thaw cycles.

Lysis buffer Prepare the following solution in Milli-Q H₂O: 10 mM Tris-HCl (pH 8.0), 10 mM NaCl, 0.2% (vol/vol) NP-40 and 1× protease inhibitor solution. **▲ CRITICAL** Because protease inhibitors degrade quickly in solution, use freshly prepared lysis buffer for each new experiment.

PROCEDURE

Single-cell preparation and cross-linking ● TIMING 1–2 h

1| Obtain single-cell preparations from fresh tissue, FACS-sorted cells or cell lines in 10% (vol/vol) FCS/PBS (see **Table 2** for cell types successfully used by us in 3C-seq experiments). Tissues rich in extracellular matrix (e.g., brain) can be treated with collagenase (0.125% (wt/vol) in PBS; incubate the tissues for 30–60 min at 37 °C) first. Filter tissue-harvested cell preparations through a 40-μm cell strainer to obtain single-cell suspensions (see ref. 19). Determine cell concentrations and

dilute 0.3×10^6 to 10×10^6 cells (10×10^6 is preferred but substantially fewer starting cells can be used) in 12 ml of culture medium (e.g., DMEM) or 10% (vol/vol) FCS/PBS (15-ml polypropylene tube).

▲ CRITICAL STEP Cell preparations need to be single-cell suspensions in order for proper formaldehyde cross-linking to be achieved.

2| Add 649 μ l of 37% (vol/vol) formaldehyde to each 15-ml tube (2% (vol/vol) final formaldehyde concentration), and incubate it for 10 min at room temperature while tumbling.

▲ CRITICAL STEP 1% (vol/vol) formaldehyde can also be used, especially if digestion efficiencies are suboptimal.

3| Transfer the tubes to ice and add 1.6 ml of cold 1 M glycine (0.125 M final concentration). Immediately proceed with Step 4.

Cell lysis, nuclei preparation and first restriction enzyme digestion ● TIMING 18–20 h

4| Centrifuge the mixture for 8 min at 340g (4 °C) and remove all of the supernatant.

5| Carefully add ice-cold PBS to a volume of 14 ml and resuspend the pellet.

6| Pellet the cells again as in Step 4. Remove all of the supernatant.

7| Carefully resuspend the pellet in 1 ml of cold lysis buffer and add another 4 ml of lysis buffer to obtain a total volume of 5 ml for each tube. Incubate the mixture for 10 min on ice.

8| Centrifuge the mixture for 5 min at 650g (4 °C) to pellet the nuclei.

■ PAUSE POINT The pelleted nuclei can be washed with PBS, snap-frozen in liquid N₂ and stored at –80 °C for several months.

9| Resuspend the nuclei in 0.5 ml of 1.2× restriction buffer and transfer them to a 1.5-ml Safe-Lock microcentrifuge tube.

10| Place the tubes at 37 °C in a thermomixer and add 7.5 μ l of 20% (wt/vol) SDS (final: 0.3% SDS).

? TROUBLESHOOTING

11| Incubate the mixture at 37 °C for 1 h while shaking (900 r.p.m.).

12| Add 50 μ l of 20% (vol/vol) Triton X-100 (final: 2% Triton X-100).

13| Incubate the mixture at 37 °C for 1 h while shaking (900 r.p.m.).

14| Take a 5- μ l aliquot (undigested control sample) of each sample and store it at –20 °C until analysis of digestion efficiency is required (see Step 16).

15| Add 400 U of the selected six-cutter restriction enzyme to the remaining samples and incubate them overnight at 37 °C while shaking (900 r.p.m.).

▲ CRITICAL STEP More unconventional primary restriction enzymes with optimal temperatures of 38–50 °C (e.g., ApoI) are also used at 37 °C to avoid partial de-cross-linking of the sample. Prolonged incubation times and/or addition of more enzyme might be required in these cases.

16| Take a 5- μ l aliquot (digested control sample) of each sample. At this point, digestion efficiencies can be analyzed by purifying the genomic DNA from the control samples using a standard phenol/chloroform extraction and running it on a 0.6% (wt/vol) agarose gel (see ref. 19). A successful six-cutter restriction enzyme digestion results in a DNA smear with the majority of fragments located between 5 and 10 kb (Fig. 4a).

? TROUBLESHOOTING

Preparation of the 3C library: first ligation and de-cross-linking ● TIMING 20–22 h

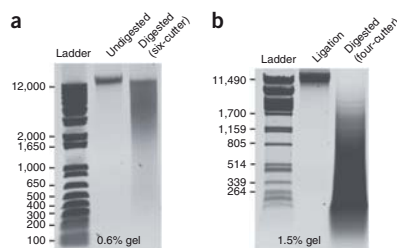
17| Add 40 μ l of 20% (wt/vol) SDS (final: 1.6% SDS) to the remaining sample from Step 15.

18| Incubate the mixture for 20–25 min at 65 °C while shaking (900 r.p.m.).



PROTOCOL

Figure 4 | Examples of successful digestion and ligation efficiencies. (a) Agarose gel (0.6%, wt/vol) on which an aliquot of undigested (left lane) and digested (right lane) sample (primary restriction digestion, Step 16) was run. A six-cutter was used, showing a typical smear of DNA fragments (a majority of DNA fragments residing between the 12 kb and 4 kb marker bands). (b) After ligation (left lane, Step 33), the DNA smear has returned to a sharp band (~12 kb). Secondary enzyme digestion (four-cutter) of the ligated 3C library typically results in a DNA smear of 2–0.1-kb fragments (1.5% (wt/vol) agarose gel).



19| Transfer the digested nuclei to 50-ml centrifugation tubes and add 6.125 ml of 1.15× ligation buffer.

20| Add 375 µl of 20% (vol/vol) Triton X-100 (final: 1% Triton X-100).

21| Incubate the mixture for 1 h at 37 °C in a water bath while shaking gently.

22| Add 100 U of T4 DNA ligase (20 µl of a high-concentration stock) and incubate it at 16 °C for 4 h.

■ **PAUSE POINT** The samples can be kept overnight at 16 °C if necessary.

23| Add 30 µl of 10 mg ml⁻¹ proteinase K (300 µg in total) and incubate it overnight at 65 °C to de-cross-link the samples.

Preparation of the 3C library (DNA purification) ● **TIMING 7–8 h**

24| Add 30 µl of 10 mg ml⁻¹ RNase (300 µg in total) and incubate the mixture for 30–45 min at 37 °C.

25| Briefly cool the samples to room temperature and add 7 ml of phenol/chloroform/isoamyl alcohol (25:24:1) and shake the samples vigorously.

26| Centrifuge the samples for 15 min at 3,200g (room temperature).

27| Transfer the upper aqueous phase into a new tube and add 7 ml of Milli-Q H₂O. Add 1.5 ml of 2 M sodium acetate (pH 5.6), and then add 35 ml of 100% ethanol.

28| Mix the tubes thoroughly and place them at –80 °C for 2–3 h until the liquid is frozen solid.

29| Directly centrifuge the frozen samples for 45 min at 3,200g (4 °C).

30| Remove the supernatant and add 10 ml of 70% ethanol.

31| Centrifuge the mixture for 15 min at 3,200g (4 °C).

32| Remove the supernatant, air-dry the pellet for ~20 min at room temperature and dissolve the pellet in 150 µl of 10 mM Tris-HCl (pH 7.5) by incubating it for 30 min at 37 °C.

■ **PAUSE POINT** This material is referred to as the '3C library' and can be stored at –20 °C for several months.

33| To determine ligation efficiency, run 0.5–1.0 µl of 3C material on a 0.6% (wt/vol) agarose gel. A successful ligation of six-cutter-digested 3C material should result in a single band, running at a similar height as the undigested control sample from Step 14 (Fig. 4b).

Preparation of the 3C-seq library (determination of DNA concentration and secondary digestion of 3C material)

● **TIMING 16–18 h**

34| If primary digestion and ligation were successful, the 3C library (Step 32) can either be used for 3C-qPCR experiments (see Hagege *et al.*¹⁹ for a detailed protocol) or be used to prepare the 3C-seq library as described here. First, run an aliquot (e.g., 1 µl) of 3C library DNA alongside a reference sample of species-matched genomic DNA to estimate DNA concentrations. To obtain sharp bands suitable for accurate gel densitometry quantification, a 1.5–2% (wt/vol) agarose gel is used. Optical density (OD) measurements do not provide an accurate estimation of DNA concentrations in 3C library samples.

35| Digest a preferred amount of the 3C library overnight (generally 25–50 µg) with a 4-base recognition restriction enzyme of choice (the four-cutter), at a DNA concentration of 100 ng µl⁻¹, using 1 U of enzyme per µg of DNA. Use buffers and incubation temperatures as recommended in the manufacturer's instructions.

Preparation of the 3C-seq library (Second ligation and DNA purification) ● TIMING 12–13 h

36| Transfer the sample to a 1.5-ml Safe-Lock tube. Add an equal amount of phenol/chloroform/isoamyl alcohol (25:24:1) and mix it vigorously.

37| Centrifuge the mixture for 15 min at 15,800g (room temperature).

38| Transfer the upper phase to a new tube and add 2 µl of 20 mg ml⁻¹ glycogen. Add a one-tenth volume of 2 M sodium acetate (pH 5.6), mix the contents and add 850 µl of 100% ethanol.

39| Mix the tubes thoroughly and snap-freeze them in liquid N₂.

40| Directly centrifuge the frozen tubes for 20 min at 15,800g (4 °C).

41| Remove the supernatant carefully and add 1 ml of 70% (vol/vol) ethanol.

42| Centrifuge the mixture for 5 min at 15,800g (4 °C).

43| Remove the supernatant carefully, air-dry the pellet for ~15 min and dissolve the pellet in 100 µl of Milli-Q H₂O by incubating it for 15 min at 37 °C.

44| Analyze 5 µl of the digested DNA on a 1.5% (wt/vol) agarose gel to check digestion efficiency. The resulting type of smear depends on the enzyme used, but the majority of fragments should be <1 kb and are usually between 300 and 500 bp (Fig. 4b).

45| Transfer the remaining sample to a 50-ml centrifugation tube. Add the components tabulated below and incubate the mixture at 16 °C for 4 h.

Component	Amount per reaction	Final
10× ligation buffer	1.4 ml	1×
T4 DNA ligase (5 U µl ⁻¹)	40 µl	200 U
Milli-Q H ₂ O	Up to 14 ml	

■ **PAUSE POINT** The samples can be kept overnight at 16 °C if necessary.

46| Add 14 ml of phenol/chloroform/isoamyl alcohol (25:24:1) and shake the mixture vigorously.

47| Centrifuge the mixture for 10 min at 3,200g (room temperature).

48| Split the upper phase into two new 50-ml tubes. Add an equal amount of Milli-Q H₂O to each tube and add 1 µl of 20 mg ml⁻¹ glycogen per ml.

▲ **CRITICAL STEP** Increasing the volume before precipitation will greatly reduce the amount of coprecipitating DTT.

49| Add a one-tenth volume of 2 M sodium acetate (pH 5.6), mix the contents and add two volumes of 100% ethanol.

50| Place the tubes at –80 °C for 2–3 h until the liquid is frozen solid.

■ **PAUSE POINT** The samples can be kept at –80 °C for several days.

51| Directly centrifuge the frozen tubes for 45 min at 3,200g (4 °C).

PROTOCOL

- 52| Remove the supernatant and add 15 ml of 70% (vol/vol) ethanol.
- 53| Centrifuge the mixture for 15 min at 3,200g (4 °C).
- 54| Remove the supernatant, air-dry the pellet for ~20 min and dissolve it in 75 µl of 10 mM Tris-HCl (pH 7.5 (per pellet)) by incubating it for 30 min at 37 °C. Thereafter, samples divided over two tubes can be recombined into a single tube.
- 55| Purify the DNA using the QIAquick gel purification kit according to the manufacturer's recommendations for direct cleanup from enzymatic reactions. Other DNA purification kits can be used, but we have obtained excellent purities with the QIAquick kit. **▲ CRITICAL STEP** One column can bind a maximum of 10 µg of DNA: use enough columns to avoid overloading and a subsequent loss of material.
- 56| Determine the DNA concentration of the resulting 3C-seq library using NanoDrop OD measurements.

3C-seq inverse PCR (preparing the sample for Illumina sequencing) ● TIMING 5–6 h

57| Perform several PCR reactions (we generally amplify the equivalent of 500–1,000 ng input DNA per bait fragment) using the primers containing the P5/P7 Illumina adapters as overhang using the PCR reaction setup and program tabulated below. The amount of input 3C-seq library DNA used should be the maximum amount for which the PCR reaction is still linear and reproducible (see tables below and Step 58), not exceeding 200 ng per reaction.

Component	Amount per reaction	Final
10× buffer I	5 µl	1×
10 mM dNTPs	1 µl	0.2 mM
25 pmol µl ⁻¹ forward primer	1 µl	25 pmol
25 pmol µl ⁻¹ reverse primer	1 µl	25 pmol
Polymerase mix (5 U µl ⁻¹)	0.75 µl	3.75 U
3C-seq library DNA	Depends on concentration	25–200 ng
Milli-Q H ₂ O	Add up to 50 µl	

Cycle number	Denature	Anneal	Extend
1	94 °C, 2 min		
2–31	94 °C, 15 s	Primer-specific, 1 min	68 °C, 3 min
32			68 °C, 7 min

▲ CRITICAL STEP Inverse PCR primers first have to be tested for linearity and reproducibility as described in **Box 2** (also see ref. 11), first without and then with the P5/P7 Illumina sequencing adapters attached.

? TROUBLESHOOTING

- 58| Verify PCR success by running small aliquots (10 µl) of each reaction on a 1.5% (wt/vol) agarose gel.
- 59| Pool all successful reactions from the same bait fragment and purify the DNA using 2 QIAquick gel purification columns. Elute the columns with 40 µl of Milli-Q H₂O and combine the samples.
- 60| Verify the purification procedure success by running an aliquot (5–10 µl) on a 1.5% (wt/vol) agarose gel. The sample is now ready to be used for Illumina high-throughput sequencing. **■ PAUSE POINT** The samples can be kept at –20 °C for several months.

3C-seq sample pooling and Illumina high-throughput sequencing ● **TIMING 4 d**

61| Quantify the DNA molarity of the individual samples on an Agilent Bioanalyzer with the DNA 7500 chip cartridge according to the manufacturer's instructions. Perform a 'smear analysis' quantification using the Bioanalyzer software.
▲ CRITICAL STEP Make sure to use the DNA 7500 chip cartridge, as 3C material contains large (1–5 kb) DNA fragments that will influence DNA molarity and may not be detected using other DNA chip cartridges.

62| Design a pool of 3C-seq samples to be sequenced together in a single lane on the flow cell using the guidelines described in **Box 3**.

63| Pool the selected samples in equal molarities in a single tube.

64| Proceed with the sequencing procedure as described by the manufacturer in the Illumina TruSeq SR cluster kit and TruSeq SBS manuals. The sequencing procedure can be outsourced to a sequence service provider. We generally use 76-bp single-read sequencing; paired-end sequencing is not required for 3C-seq.

▲ CRITICAL STEP When loading the flow cell, aim for a cluster density of 750,000–850,000 clusters per mm². In our case, this is usually achieved with a final template DNA concentration of 9 pM.

▲ CRITICAL STEP Ensure that the total number of sequencing cycles exceeds the sum of the bait-specific sequence length and a minimum of 36 bases for optimal alignment of the unknown interacting fragments.

Initial data processing ● **TIMING 1–2 d**

65| Copy the whole run folder generated by the Illumina sequencer to the storage on the Linux computer.

66| Open a terminal on the Linux computer and enter the commands described after the > signs.

67| Convert the binary output from the sequencer to text files in the Qseq format by using the BclToQseq scripts included in the Illumina Offline Basecaller (available at the Illumina website <http://www.illumina.com/>):

```
> cd Illumina_Run_Folder/Data/Intensities/BaseCalls
> /path_to_OLB/bin/setupBclToQseq.py --in-place -b.
> make -j 6
```

68| Determine the bait-specific sequences for de-multiplexing. Note that this also includes the primer, the primary restriction site and any sequence in between. To obtain the highest yield while still retaining high specificity, de-multiplexing is performed using only 6 bases instead of the entire bait-specific sequence. The first set of 6 bases that differ for 2 or more bases from the other bait sequences are used for de-multiplexing.

▲ CRITICAL STEP Record the unique 6-bp bait-specific sequences (6-bp-bait) and their positions (6 bp-bait-pos) in the bait for each sample.

69| Determine the number of bases to trim from the 5' and the 3' ends of the reads as described in Steps 70–75. This procedure is performed in Microsoft Excel.

▲ CRITICAL STEP The 5' trimming is crucial, as the remaining bait-specific sequences will prevent the read from aligning to the reference sequence (**Fig. 3**). The 3' trimming prevents the loss of short interacting fragments (see Experimental design).

70| First, extend the bait-specific primer sequence with the genomic sequence up to and including the primary restriction site.

71| Extend the bait-specific primer sequence with the genomic sequence up to and including the primary restriction site.

72| Subtract the forward Illumina P5 adapter sequence from the 5' end of this sequence (**Box 1**).

73| Count the number of bases in the resulting sequence using the *len()* function to obtain the number of bases to trim from the 5' end of the read (*n5trim*).

74| Subtract *n5trim* from the read length.

PROTOCOL

75| Subtract 36 bases from the result of Step 74 to obtain the number of bases to trim from the 3' end (*n3trim*).

76| Create a NARWHAL²⁷ sample sheet (**Supplementary Table 1**) for the lanes that contain the 3C-seq samples. In this sample sheet, use any profile that runs BOWTIE²⁸ with the `--best` option. To de-multiplex, several options need to be set in the sample sheet: the bar code-read field is set to 1; the bar code-start field is set to the 6-bp-bait-pos; the bar code field is set to the 6-bp-bait sequence. For the trimming, the following options are added to the options field of the sample sheet to trim the sequences: `--trim5=n5trim,--trim3=n3trim`.

77| Copy the NARWHAL sample sheet to the Linux computer.

78| (Optional) When the flow cell does not exclusively contain 3C-seq samples, it might be necessary to analyze only specific lanes. This can be achieved by setting up a directory with only the Qseq files for the specific lanes to be analyzed. This can be performed as follows, with *i* as the lanes to be analyzed:

```
> mkdir MyLanes/  
  
> ln -s /full_path_to_qseq_folder/s_[i]_1*_qseq.txt MyLanes/
```

79| Run NARWHAL using the following command:

```
> narwhal.sh -s samplesheet.txt Qseq_folder output_folder
```

After the alignment, NARWHAL will generate a PDF reporting the total number of reads generated, the percentage successfully aligned reads, the read distribution across the chromosomes, edit rates and duplication rates²⁷. Successful 3C-seq experiments should have high duplication rates (>95%), with a majority of reads (>50%) mapped to the chromosome on which the bait is located.

? TROUBLESHOOTING

Bioinformatics and initial data visualization ● TIMING 2 h

80| After the initial data processing, a restriction map of the genome needs to be generated as described in Steps 80–82. First, Search the genome for restriction sites using the `findSequence.py` script (**Supplementary Data**). This script will generate a BED file containing all the occurrences of a given sequence in the genome.

```
> python findSequence.py -f genome.fasta -s  
primary_restriction_sequence -b occurrences.bed
```

81| Create a BED file containing the regions between the restriction sites by using the `regionsBetween.py` script (**Supplementary Data**):

```
> python regionsBetween.py -i occurrences.bed -s chromsizes.txt -o regions.bed
```

82| Sort the regions with the BEDtools²⁹ `sort` command:

```
> bedtools sort -i regions.bed > sorted_regions.bed
```

83| Count the reads per target fragment using the `alignCounter.py` tool (**Supplementary Data**). The count result is a table that can be loaded into other tools such as R.

```
> python alignCounter.py -b aln.srt.bam -r sorted_regions.bed -o output_table.txt
```

84| Convert the read count tables to BED files using the command below. These BED files can be loaded into a variety of genome browsers including the UCSC Genome Browser (<http://genome.ucsc.edu/>).

```
> gawk '/^[#]/{ if($4 > 0){print $1 "\t" $2 "\t" $3 "\t" $4 ;} ; }'  
output_table.txt > output_table.bed
```

? TROUBLESHOOTING

? TROUBLESHOOTING

Multiplexed 3C-seq success primarily depends on digestion efficiencies, 3C-seq PCR setup (**Boxes 1 and 2**) and Illumina sequencing. **Table 3** contains 3C-seq troubleshooting advice, mainly concerning these steps. Digestion efficiencies are also highly dependent on the cell or tissue type used. **Table 2** provides additional cell type-specific troubleshooting information. Other published protocols have also provided detailed troubleshooting for the 3C procedure^{19,30}.

TABLE 3 | Troubleshooting table.

Step	Problem	Possible reason	Solution
10	Formation of aggregates after addition of SDS to the restriction buffer	Too many nuclei are used or the nuclei are of poor quality	Dilute the material 2–4 times in 1.2× restriction buffer containing 0.3% (wt/vol) SDS. For future experiments, ensure gentle handling of the cells and nuclei. A more stringent lysis buffer and/or Douncing step can also be beneficial. If persistent, consider starting with fewer cells in future experiments
16	Poor primary digestion efficiency	Formaldehyde concentrations used are too high for the enzyme; the enzyme is not compatible with the 3C protocol and/or extensive nuclei aggregation	Lower formaldehyde concentrations (e.g., 1% instead of 2% (vol/vol)) or increase Triton X-100 concentration in Step 12. Alternatively, consider changing to a different enzyme. If nuclei are forming large aggregates, see Step 10 troubleshooting for advice
57	Poor PCR linearity, reproducibility or PCR failure	PCR conditions or design are suboptimal	Ensure that the correct primer T_m is used. Further optimizing the T_m using a gradient can be beneficial. Often, simply redesigning the 3C-seq primers will greatly improve PCR success
	Primer dimer formation	PCR conditions or design are suboptimal	See above. If primer dimer formation specifically occurs after addition of the P5/P7 adaptors, DNA purification kits with a >100-bp cutoff can be used to remove dimers before sequencing
79	Fewer than expected sequence yield for a particular sample	Unanticipated bait-specific sequence	Compare the list of expected barcodes to the most abundant sequences. To generate a list with the most abundant barcode sequences from a FastQ file, the following Linux command-line code can be used: <pre>> grep '^[ACTGN]\+\$' in.fastq sed 's/^\(.{6}\).*\/1/g' sort uniq -c sort -nr head -n 30</pre> Cross-reference unexpected highly abundant sequences with the expected primers and if possible assign these reads to a sample. Re-do de-multiplexing with the updated barcodes
	Low mapping percentage after sequencing	Primer dimers present in 3C-seq sample or the secondary restriction site occurs directly after the primary restriction site in the most abundant target fragments	Obtain all the non-aligning sequences from the BAM file: <pre>> samtools view aln.srt.bam grep -P '^S+\t\d+\t\.*\$' > not_aligned.aln</pre> Check these sequences for subsequences of the primers used in the amplification. Determine whether these sequences contain the restriction site for the secondary restriction enzyme. This issue occurs more frequently with increasing read-length. For this reason, we strongly recommend using the 3' trimming procedure from Steps 70–75. If after trimming the target sequence is shorter than 25 bp, the secondary restriction enzyme needs to be changed in order for the read to be aligned properly

(continued)



TABLE 3 | Troubleshooting table (continued).

Step	Problem	Possible reason	Solution
84	Complete absence of reads at expected sites of interaction	The fragment expected to interact with the bait is <36 bp	Further extend the 3' trimming procedure or use a different six-cutter/four-cutter combination
		The genome assembly has changed (updated)	Reanalyze older data sets using the proper version of the genome assembly. This may be crucial when recent data sets need to be compared with older ones
	Weak 3C-seq interaction signals	Poor signal-to-noise ratio	Consider using a double cross-linking procedure by using ethylene glycol bis-succinimidylsuccinate treatment before formaldehyde as described in Lin <i>et al.</i> ³⁴

● TIMING

Steps 1–3, single-cell preparation and cross-linking: 1–2 h
 Steps 4–16, cell lysis, nuclei preparation and first restriction enzyme digestion: 18–20 h
 Steps 17–23, preparation of the 3C library: first ligation and de-cross-linking: 20–22 h
 Steps 24–33, preparation of the 3C library: DNA purification: 7–8 h
 Steps 34 and 35, preparation of the 3C-seq library: determination of DNA concentration and secondary digestion of 3C material: 16–18 h
 Steps 36–56, Preparation of the 3C-seq library: second ligation and DNA purification: 12–13 h
 Steps 57–60, 3C-seq inverse PCR: preparing the sample for Illumina sequencing: 5–6 h
 Steps 61–64, 3C-seq sample pooling and Illumina high-throughput sequencing: 4 d
 Steps 65–79, initial data processing: 1–2 d
 Steps 80–84, bioinformatics and initial data visualization: 2 h

ANTICIPATED RESULTS

After sequencing and data processing, the resulting BED files (Step 84) can be visualized in a genome browser (e.g., UCSC genome browser, <http://genome.ucsc.edu/>). Careful attention should be given to the particular version of the genome that is used for analysis, especially when different experiments are compared. Several simple but important checks can provide information on whether the 3C-seq experiment was successful, which are automatically provided during initial data processing (Steps 65–79) by the NARWAL software²⁷. The PDF file provided contains statistics on the chromosomal location of the aligned reads and the duplication percentage. These are important metrics for the initial validation of a 3C-seq experiment: the vast majority (>50%) of reads are usually found *in cis* (i.e., on the same chromosome), and as 3C-seq profiles consist of stacked reads the duplication percentage should be >95%. Typical alignment percentages are above 70%, although this can vary considerably between different primer sets. Lower percentages are often caused by the sequencing of primer dimers present in the PCR samples or failure to align reads coming from the (in general) most abundant interactions (the bait fragment itself and the neighboring fragment, see **Box 2** and **Table 3**). However, low alignment percentages can still provide informative data, as long as the total number of aligned reads is high enough (>1 million reads³⁰) and read distribution is as expected (see below and **Fig. 5**). After uploading the BED output file (Step 84) in a genome browser, interactions with the chosen bait fragments can be observed. Signals are represented as bars (**Fig. 5**), the width of which is determined by the size of the actual restriction fragment. The height of the bars represents the number of reads found on the fragment and is a measurement of the frequency of interaction with the bait fragment. The highest signal density is always found around the viewpoint (typically ~40% of all reads are located within 1 Mb of the bait), with the two most abundant interactions being the bait and its neighboring fragment (**Box 2**). Signal intensity tends to rapidly decline with increasing genomic distance from the bait (a classic characteristic of 3C and its derivatives, see refs. 11,26), resembling a bell-shaped distribution around the bait (**Fig. 5a**). The majority (>75%) of *cis* interactions are normally found within a 1-Mb window around the bait, although bait fragments within highly complex genomic structures (e.g., immunoglobulin loci) can produce profiles that deviate from this general picture¹⁸. Interactions found *in trans* (generally about 40–50% of the reads) often show low interaction frequencies and appear to be randomly scattered around the genome. *Trans*-interaction signals therefore need to be interpreted with caution, as their reproducibility may appear questionable in a number of cases. However, several studies have begun to probe their functional relevance in specific cases, in particular in light of chromosomal translocations, and showed correlation between physical proximity and sites of recombination, indicating that physical proximity *in trans* may be relevant^{31,32}.

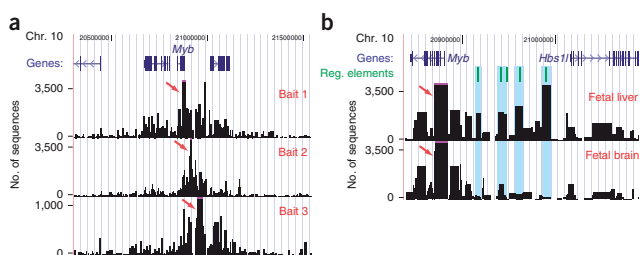
© 2013 Nature America, Inc. All rights reserved. mpj

Figure 5 | Typical interaction profiles obtained from a multiplexed 3C-seq experiment.

(a) 3C-seq interaction profiles in mouse fetal liver cells shown for three bait fragments in the *Myb* locus¹⁷ (1.2-Mb region shown). Bait signals are depicted by an arrow.

(b) 3C-seq interaction profiles generated from both mouse fetal liver and brain using the *Myb* promoter as bait (shown is an ~250-kb region encompassing the *Hbs1l*-like (*Hbs1l*) neighboring gene). *Myb* is highly expressed in fetal liver cells, but expression is much lower in fetal brain cells. Several fetal liver-specific interactions are located within an intergenic region containing several regulatory (Reg.) elements (green lines and blue shading)¹⁷. Bait signals are depicted by an arrow.

Data were visualized using the UCSC genome browser. All animal work was approved by the Netherlands Animal Experimental Committee (DEC) and the Institutional Ethical Review Board of Erasmus Medical Center, and was carried out according to institutional and national guidelines.



Multiplexing 3C-seq samples greatly increases the technique's throughput and results in a substantial cost reduction. Even though the total number of reads is lower in a multiplexed sample compared with a nonmultiplexed sample, interaction patterns remain almost identical (Fig. 6). Thus, multiplexing 3C samples seems to have little effect on the resulting interaction profiles (Fig. 6).

Further validation of detected interactions can be obtained by complementary experiments (e.g., 3C-qPCR, FISH) or by performing new 3C-seq experiments with these interactions as bait (a 'reverse experiment', see 'Controls' section of INTRODUCTION). Functional interpretation of 3C-seq profiles is often desired and requires correlation with other data sets, usually transcription factor binding and/or histone modification patterns for the locus of interest. When using 3C-seq to explore the regulatory elements in close proximity to a gene, strong interaction signals can often be positively correlated to the binding of transcription factors and the presence of specific histone modifications¹⁷. Performing 3C-seq experiments in different cell or tissue types can further provide valuable information on the tissue specificity of interactions and whether their presence can be correlated to differences in gene expression or protein binding (Fig. 5b). The 3C-seq data can also be

further processed using dedicated tools and scripts (S.Thongjuea, R.S., F.G., E.S. and B. Lenhard, unpublished data, and ref. 12) for more in-depth analysis.

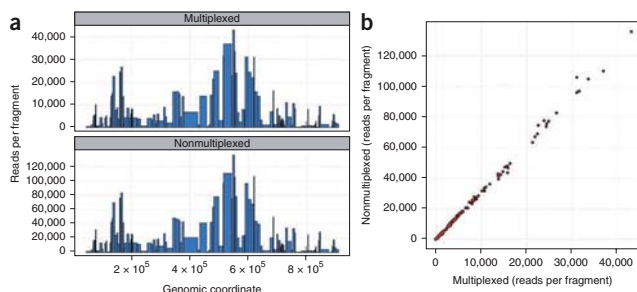


Figure 6 | Comparison of interactions detected for the same 3C-seq sample after single or multiplexed library sequencing. (a) Interaction profiles around the bait fragment for a 3C-seq sample after multiplexed (top) or nonmultiplexed (bottom) library sequencing, showing highly similar profiles. (b) Scatter plot comparing read counts for 146 fragments around the bait fragment between nonmultiplexed and multiplexed data sets.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS We thank A. van der Sloot, Z. Ozgur, E. Oole, M. van den Hout, F. Sleutels, S.Thongjuea and B. Lenhard for their help in sample processing, bioinformatics pipeline development and data analysis. R.S. received support from the Royal Netherlands Academy of Arts and Sciences (KNAW). P.K. was supported by grants from ERASysBio+/FP7 (project no. 93511024). E.S. was supported by grants from the Dutch Cancer Genomics Center, the Netherlands Genomics Initiative (project no. 40-41009-98-9082) and the French Alternative Energies and Atomic Energy Commission (CEA). This work was supported by the EU-FP7 Euracc consortium.

AUTHOR CONTRIBUTIONS R.S. and R.-J.P. adapted and optimized the protocol and library preparation for Illumina sequencing. R.S., P.K., A.v.d.H. and J.Z. used,

developed and troubleshot the technique. C.K. optimized procedures for library sequencing, and R.B. developed the informatics pipeline for data processing and analysis. W.v.I., F.G., K.S.W. and E.S. supervised the projects, and participated in technology design and discussions. R.S., P.K., R.B., W.v.I., F.G., K.S.W. and E.S. drafted the manuscript.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

Published online at <http://www.nature.com/doi/10.1038/nprot.2013.018>. Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. Dixon, J.R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
2. Nora, E.P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
3. Sanyal, A., Lajoie, B.R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113 (2012).
4. Splinter, E. & de Laat, W. The complex transcription regulatory landscape of our genome: control in three dimensions. *EMBO J.* **30**, 4345–4355 (2011).
5. Bulger, M. & Groudine, M. Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144**, 327–339 (2011).
6. Ong, C.T. & Corces, V.G. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.* **12**, 283–293 (2011).
7. Stadhouders, R. *et al.* Transcription regulation by distal enhancers: who's in the loop? *Transcription* **3**, 181–186 (2012).
8. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
9. Gondor, A., Rougier, C. & Ohlsson, R. High-resolution circular chromosome conformation capture assay. *Nat. Protoc.* **3**, 303–313 (2008).
10. Sexton, T. *et al.* Sensitive detection of chromatin coassociations using enhanced chromosome conformation capture on chip. *Nat. Protoc.* **7**, 1335–1350 (2012).
11. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* **38**, 1348–1354 (2006).
12. van de Werken, H.J. *et al.* Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat. Methods* **9**, 969–972 (2012).
13. Dostie, J. & Dekker, J. Mapping networks of physical interactions between genomic elements using 5C technology. *Nat. Protoc.* **2**, 988–1002 (2007).
14. Fullwood, M.J. *et al.* An oestrogen-receptor- α -bound human chromatin interactome. *Nature* **462**, 58–64 (2009).
15. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
16. Soler, E. *et al.* The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes Dev.* **24**, 277–289 (2010).
17. Stadhouders, R. *et al.* Dynamic long-range chromatin interactions control *Myb* proto-oncogene transcription during erythroid development. *EMBO J.* **31**, 986–999 (2012).
18. Ribeiro de Almeida, C. *et al.* The DNA-binding protein CTCF limits proximal V κ recombination and restricts κ enhancer interactions to the immunoglobulin κ light chain locus. *Immunity* **35**, 501–513 (2011).
19. Hagege, H. *et al.* Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat. Protoc.* **2**, 1722–1733 (2007).
20. Naumova, N., Smith, E.M., Zhan, Y. & Dekker, J. Analysis of long-range chromatin interactions using chromosome conformation capture. *Methods* (2012).
21. Ecker, J.R. *et al.* Genomics: ENCODE explained. *Nature* **489**, 52–55 (2012).
22. Dostie, J. & Bickmore, W.A. Chromosome organization in the nucleus—charting new territory across the Hi-Cs. *Curr. Opin. Genet. Dev.* **22**, 125–131 (2012).
23. Comet, I., Schuettengruber, B., Sexton, T. & Cavalli, G. A chromatin insulator driving three-dimensional Polycomb response element (PRE) contacts and Polycomb association with the chromatin fiber. *Proc. Natl. Acad. Sci. USA* **108**, 2294–2299 (2011).
24. Jing, H. *et al.* Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. *Mol. Cell* **29**, 232–242 (2008).
25. Rippe, K., von Hippel, P.H. & Langowski, J. Action at a distance: DNA-looping and initiation of transcription. *Trends Biochem. Sci.* **20**, 500–506 (1995).
26. Dekker, J. The three 'C's of chromosome conformation capture: controls, controls, controls. *Nat. Methods* **3**, 17–21 (2006).
27. Brouwer, R.W., van den Hout, M.C., Grosveld, F.G. & van Ijcken, W.F. NARWHAL, a primary analysis pipeline for NGS data. *Bioinformatics* **28**, 284–285 (2012).
28. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
29. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
30. van de Werken, H.J. *et al.* 4C technology: protocols and data analysis. *Methods Enzymol.* **513**, 89–112 (2012).
31. Hakim, O. *et al.* DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes. *Nature* **484**, 69–74 (2012).
32. Zhang, Y. *et al.* Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* **148**, 908–921 (2012).
33. Visser, M., Kayser, M. & Palstra, R.J. HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the *OCA2* promoter. *Genome Res.* **22**, 446–455 (2012).
34. Lin, Y.C. *et al.* Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nat. Immunol.* **13**, 1196–1204 (2012).



Supplementary table 1:

NARWHAL sample sheet. This file contains an example NARWHAL sample sheet that can be used in the primary data analysis. It serves to illustrate the specific fields that should be set in the primary data analysis procedure (Steps 65-79).

#sample_id	lane	data_reads	barcodes	multiplex_start	multiplex_read	reference_genome	PE	application	options
sample-06	1	1	ACAAAA	0	1	/data/Genomes/Bowtie/hg19	0	4C	--trim5=24,--trim3=17
sample-07	1	1	ATGGGG	0	1	/data/Genomes/Bowtie/hg19	0	4C	--trim5=20,--trim3=21
sample-02	1	1	CAACAG	0	1	/data/Genomes/Bowtie/hg19	0	4C	--trim5=27,--trim3=14
sample-05	1	1	CAGGGA	0	1	/data/Genomes/Bowtie/hg19	0	4C	--trim5=21,--trim3=20
sample-01	1	1	CTTTTT	0	1	/data/Genomes/Bowtie/hg19	0	4C	--trim5=24,--trim3=17
sample-09	1	1	GAAGCG	0	1	/data/Genomes/Bowtie/hg19	0	4C	--trim5=22,--trim3=19
sample-04	1	1	GATAAT	0	1	/data/Genomes/Bowtie/hg19	0	4C	--trim5=24,--trim3=17
sample-08	1	1	GCAAAC	0	1	/data/Genomes/Bowtie/hg19	0	4C	--trim5=32,--trim3=9
sample-10	1	1	GCGCTG	0	1	/data/Genomes/Bowtie/hg19	0	4C	--trim5=29,--trim3=12
sample-16	2	1	ACAAAA	0	1	/data/Genomes/Bowtie/hg19	0	4C	--trim5=24,--trim3=17
sample-17	2	1	ATGGGG	0	1	/data/Genomes/Bowtie/hg19	0	4C	--trim5=20,--trim3=21
sample-12	2	1	CAACAG	0	1	/data/Genomes/Bowtie/hg19	0	4C	--trim5=27,--trim3=14
sample-15	2	1	CAGGGA	0	1	/data/Genomes/Bowtie/hg19	0	4C	--trim5=21,--trim3=20
sample-11	2	1	CTTTTT	0	1	/data/Genomes/Bowtie/hg19	0	4C	--trim5=24,--trim3=17
sample-19	2	1	GAAGCG	0	1	/data/Genomes/Bowtie/hg19	0	4C	--trim5=22,--trim3=19
sample-14	2	1	GATAAT	0	1	/data/Genomes/Bowtie/hg19	0	4C	--trim5=24,--trim3=17
sample-18	2	1	GCAAAC	0	1	/data/Genomes/Bowtie/hg19	0	4C	--trim5=32,--trim3=9
sample-20	2	1	GCGCTG	0	1	/data/Genomes/Bowtie/hg19	0	4C	--trim5=29,--trim3=12

Supplementary data:

Scripts to analyze 3C-seq data. Within this supplementary archive, 4 python files are present that are used to analyze 3C-seq data. The findSequence.py and the regionsBetween.py files are used to generate a restriction map of the genome. The alignCounter.py script determines how many reads align to each of the restriction fragments. To run the alignCounter.py script the Pysam libraries should be installed on the system (see Materials). The libutils.py script is a shared library that should be placed in the same directory as the other py files. This library contains shared functionality between the 3 executable scripts. (Available at: <http://www.nature.com/nprot/journal/v8/n3/full/nprot.2013.018.html#supplementary-information>).

Chapter 3



Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells

Jessica Zuin, Jesse R. Dixon, Michael I.J.A. van der Reijden, Petros Kolovos, Zhen Ye, Rutger W.W. Brouwer, Mariëtte P. C. van de Corput, Wilfred F.J. van IJcken, Frank G. Grosveld, Bing Ren & Kerstin S. Wendt

Manuscript Submitted

Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells

Jessica Zuin^{1,9}, Jesse R. Dixon^{2,3,4,9}, Michael I.J.A. van der Reijden¹, Petros Kolovos¹, Zhen Ye^{2,8}, Rutger W.W. Brouwer^{5,6}, Mariëtte P. C. van de Corput¹, Wilfred F.J. van IJcken⁵, Frank G. Grosveld^{1,7}, Bing Ren^{2, 8} and Kerstin S. Wendt¹

¹ Department of Cell Biology, Erasmus Medical Center, Rotterdam, The Netherlands.

² Ludwig Institute for Cancer Research, 9500 Gilman Drive, La Jolla, California 92093, USA.

³ Medical Scientist Training Program, University of California, San Diego, La Jolla, California 92093, USA

⁴ Biomedical Sciences Graduate Program, University of California, San Diego, La Jolla, California 92093, USA.

⁵ Center for Biomics, Erasmus Medical Center, Rotterdam, The Netherlands.

⁶ Netherlands Bioinformatics Centre (NBIC), Nijmegen, The Netherlands.

⁷ Cancer Genomics Center, Erasmus Medical Center, Rotterdam, The Netherlands

⁸ University of California, San Diego School of Medicine, Department of Cellular and Molecular Medicine, Institute of Genomic Medicine, UCSD Moores Cancer Center, 9500 Gilman Drive, La Jolla, California 92093, USA.

⁹ These authors contributed equally to this work.

Recent studies of genome-wide chromatin interactions have revealed that the human genome is partitioned into many self-associating topological domains (Lieberman-Aiden, van Berkum et al., 2009; Dixon, Selvaraj et al., 2012; Nora, Lajoie et al., 2012; Sexton, Yaffe et al., 2012). The boundary sequences are enriched for binding sites of CTCF and the cohesin complex, implicating these two factors in the establishment or maintenance of topological domains (Splinter, Heath et al., 2006; Wendt, Yoshida et al., 2008; Hadjur, Williams et al., 2009; Nativio, Wendt et al., 2009; Kagey, Newman et al., 2010). To determine the role of cohesin and CTCF in higher-order chromatin architecture in human cells, we proteolytically cleaved the cohesin complex from interphase chromatin and examined changes in chromosomal organization as well as the transcriptome. We observed a general loss of local chromosomal interactions upon disruption of cohesin complex, but the topological domains remain intact. However, we found that depletion of CTCF by RNA interference in these cells not only reduced intra-domain interactions but also increased inter-domain interactions. Furthermore, distinct groups of genes become mis-regulated upon depletion of cohesin and CTCF. Taken together, these observations suggest that CTCF and cohesin contribute in different ways to chromatin organization and gene regulation.

Recent studies of the topological organization of the genome suggest that CTCF and cohesin might be involved in establishment or maintenance of topological domains in the mammalian genome, as their binding sites are enriched at the boundaries of these domains (Dixon, Selvaraj et al., 2012). CTCF is a transcription factor that can bind to transcriptional insulator sequences and prevent enhancers from inappropriately activating non-target genes (Bell, West et al., 1999; Phillips and Corces, 2009), however, the exact mechanism of CTCF's insulator function is not well understood. Several observations have led to the proposal that CTCF might act via recruitment of cohesin to facilitate long-range interactions as "master weaver" of the genome (Phillips and Corces, 2009). First, CTCF and the cohesin complex, consisting of the core subunits SMC3, SMC1, RAD21 and STAG1/SA1 or STAG2/SA2, were found to colocalize extensively throughout mammalian genomes (Parelho, Hadjur et al., 2008; Rubio, Reiss et al., 2008; Wendt, Yoshida et al., 2008). Second, both factors are involved in mediating long-range interactions (Kurukuti, Tiwari et al., 2006; Splinter, Heath et al., 2006; Hadjur, Williams et al., 2009; Mishiro, Ishihara et al., 2009; Nativio, Wendt et al., 2009; Hou, Dale et al., 2010). Finally, cohesin was shown to be important for CTCF's chromatin insulation function (Parelho, Hadjur et al., 2008; Rubio, Reiss et al., 2008; Wendt, Yoshida et al., 2008), while CTCF is necessary to recruit cohesin to the shared binding sites but not to chromatin (Wendt, Yoshida et al., 2008).

To understand the contribution of CTCF and cohesin to genome organization, we employed an engineered HEK293T cell line (original cell line derived from human embryonic kidney) in which we can rapidly remove the cohesin complex from interphase chromosomes by proteolytic cleavage of its RAD21 subunit (Uhlmann, Lottspeich et al., 1999). This cell line contains an episome-based vector that allows doxycycline-inducible expression of siRNA

targeting endogenous RAD21 and a RAD21-EGFP variant containing a recognition site for Human rhinovirus 3C (HRV) protease (RAD21cv) (Schockel, Mockel et al., 2011) (**Figures 1a,b**). Three days after doxycycline induction, RAD21cv completely replaces the endogenous RAD21 and is incorporated in the cohesin complex (**Supplementary Figure 1**). Subsequent transfection of the cells with a construct expressing HRV protease led to full cleavage of RAD21cv and release of cohesin from chromatin within 24 hours (**Figures 1c,d,e**).

Consistently, RAD21cv cells entering mitosis 24 hours after HRV transfection show increased defects in sister chromatid cohesion (**Supplementary Figure 2**). Cleavage of RAD21cv by HRV protease (RAD21cv/HRV) does not change the cell cycle distribution compared to transfection with a control protease from Tobacco Etch Virus (RAD21cv/TEV). Nevertheless, we noted that both transfected cell populations have more cells in G2 phase than untreated cells (RAD21cv) (**Supplementary Figure 3**). This rapid release of cohesin allows the study of the immediate effect of cohesin loss on chromatin structure, without interfering with cohesin function in cell division.

To test whether removal of cohesin from chromatin affects long-range chromatin interactions, we performed 3C-seq (a multiplexed 4C variant) (Stadhouders, Kolovos et al., 2013) in RAD21cv/TEV and RAD21cv/HRV cells. We examined the interior and the borders of one topological domain, based on previous domain assignment in the human IMR90 cells (Dixon, Selvaraj et al., 2012), at the well characterized chr11p15.5 region comprising H19, Igf2 and other imprinted genes (referred to here as H19/IGF2 domain, **Figures 1f-h**). We have previously used this region to establish the role of cohesin in chromatin insulation by CTCF (Wendt, Yoshida et al., 2008). In RAD21cv/TEV we observed, as reported before (Nativio, Wendt et al., 2009), that the IGF2 promoter region (VP1) and an intergenic region (VP2) interact strongly. Further contacts persist over a 500kb region until the proximal keratin cluster (KRTA5) marking the domain boundary. Viewpoints placed between H19 and IGF2 (VP2) and upstream of H19 (VP3) confirm these interactions (**Figure 1h**). A viewpoint in the neighbouring domain at the centromeric side (VP6) consistently shows interactions until the domain boundary. A viewpoint placed at the telomeric boundary (VP5) shows weak interactions with both domains. The CTSD gene residing in a cohesin-depleted region is remarkably excluded from interactions; although a 3C-seq viewpoint there (VP4) shows some of the interactions detected by the other viewpoints (**Figure 1h**). We observed similar interaction profiles in the breast endothelial cell line 1-7HB2 (abbreviated HB2) with normal karyotype, indicating their conservation between cell lines (**Supplementary Figure 4**).

Cleavage of RAD21 leads to a global loss of interactions across the entire domain at all viewpoints (**Figure 1h**). A control with a cell line lacking the HRV cleavage site in RAD21-EGFP (RAD21wt) did not show altered cohesin binding and long-range interactions after transfection with the cleavage protease (**Supplementary Figure 5**). These results strongly support that cohesin is required for the higher-order chromatin structure within this domain.

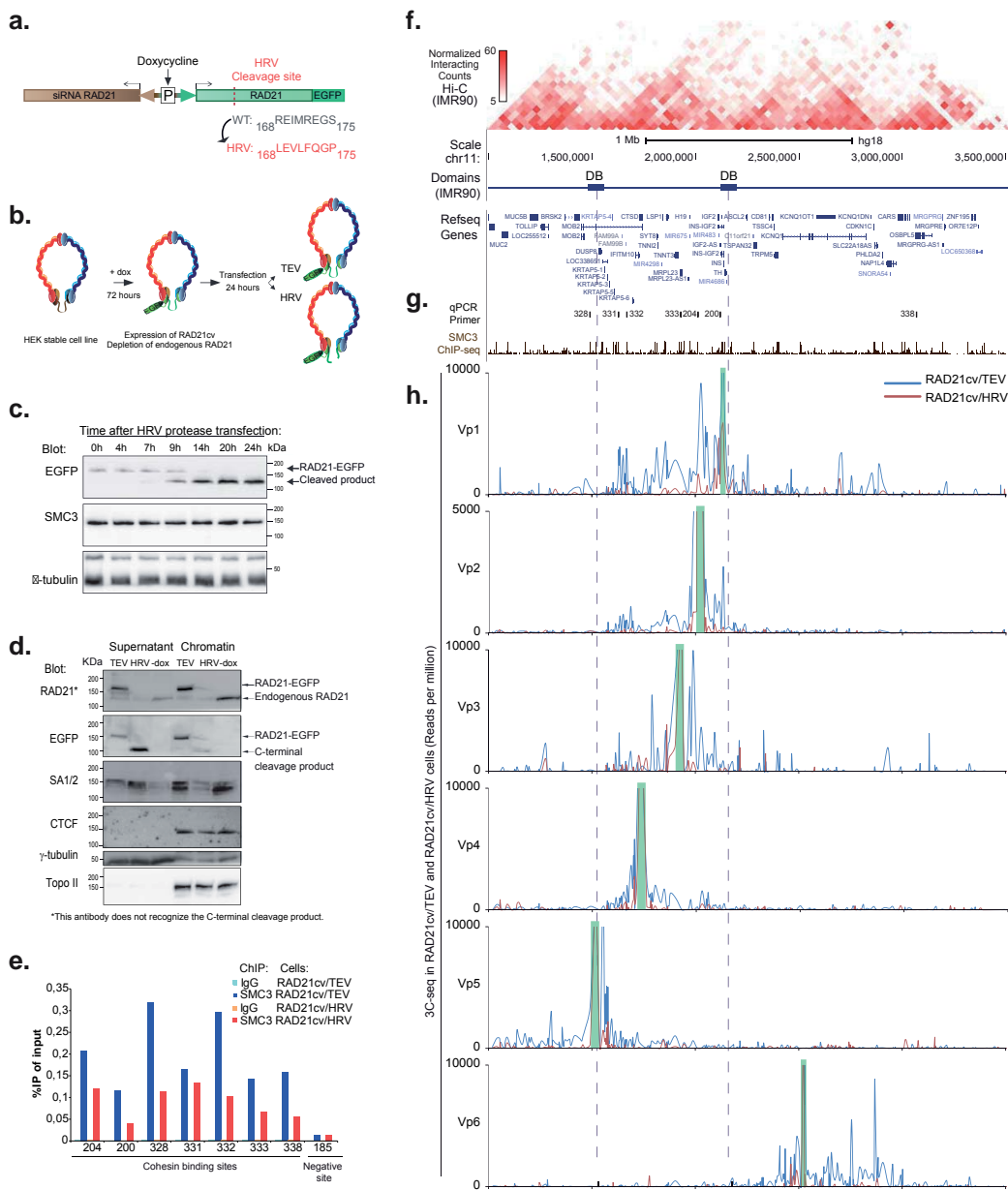


Figure 1: Cohesin cleavage reduces long-range interactions within the H19/IGF2 domain.

a. Endogenous RAD21 is replaced with RAD21cv using a doxycycline-inducible bidirectional promoter driving expression of RAD21cv and siRNA targeting the endogenous RAD21 from an episomal construct stably integrated in HEK293T cells. **b.** Outline of the experiment showing the replacement of RAD21 by RAD21cv and transfection of the protease HRV, which cleaves RAD21cv, and TEV, which does not. **c.** Time course showing full cleavage of RAD21cv after 24 hours by detecting the C-terminal EGFP-tag in RAD21cv. The shift is consistent with the loss of a 20 kD fragment from the N-terminus of RAD21cv. Note RAD21cv gets more abundant in the lysates due to its release from chromatin after cleavage. **d.** Fractionation of the lysates from RAD21cv/TEV (TEV), RAD21cv/HRV (HRV) and uninduced cells (-dox) in soluble (Supernatant) and chromatin-bound fraction (Chromatin). Similar levels of endogenous RAD21 (-dox) and RAD21cv (TEV control) are observed bound to chromatin. Blotting for RAD21 shows absence of endogenous RAD21 in TEV and HRV. Full cleavage of RAD21cv is shown by blotting for EGFP, due to the

release from chromatin the cleavage product appears now in the soluble fraction. Blotting with a bispecific antibody against the cohesin subunits STAG1/SA1 (SA1) and SATG2/SA2 (SA2) shows that these subunits are also released from chromatin after HRV cleavage. CTCF binding to chromatin is not affected. **e**, Chromatin immunoprecipitation (ChIP) with anti-SMC3 from RAD21cv/TEV and RAD21cv/HRV cells and qPCR with primers specific for cohesin binding sites shows a reduced signal in the SMC3 ChIP after RAD21 cleavage. **(f-g)** Analysis of long-range chromosome interactions for six viewpoints at the human chromosome 11 for in RAD21cv cells transfected with control protease (TEV) and cleaving protease (HRV). **f**. Normalized Hi-C interaction frequencies displayed as a two-dimensional heat map previously observed in human IMR90 cells showing the domains and domain boundaries (DB, indicated as rectangles) in this area. Genes annotated by the Reference sequence collection for this region (Refseq genes) are also indicated. **g**. Position of primer pairs used to test cohesin depletion in Fig. 1e and binding sites for cohesin (SMC3) detected by ChIP-sequencing in RAD21cv/TEV cells are indicated. **h**. Chromosomal interactions detected by 3C-seq for six different viewpoints (Vp1-6, marked with green bars) in RAD21cv/TEV cells (blue graph line) and RAD21cv/HRV cells (red graph line). Interactions are presented as line graph with the observed interaction frequency (Reads per million sequenced reads) assigned to the center of each BglII fragment.

To investigate whether cohesin plays a general role in topological domain organization, we performed Hi-C experiments with control RAD21cv/TEV cells and RAD21cv cells after RAD21 cleavage (RAD21cv/HRV). We obtained greater than 370 million non-redundant uniquely mapping read pairs for both control and RAD21 cleaved cells, split between two replicates for each condition. We normalized the Hi-C interaction frequencies according to the iterative correction method (Imakaev, Fudenberg et al., 2012). For each replicate both before and after RAD21 cleavage, we identified the location of topological domains using a previously described algorithm (Dixon, Selvaraj et al., 2012).

We also performed ChIP-seq for the cohesin subunit SMC3 in control (RAD21cv/TEV) cells to determine the cohesin binding sites in the genome in these cells. Similarly to what had previously been observed for CTCF, cohesin appears to be enriched at the borders or boundaries between domains (**Figure 2a**). Notably, only SMC3 sites that co-localize with CTCF show enrichment at boundaries, while CTCF-independent SMC3 sites show no enrichment at boundary regions (**Supplementary Figure 6a**).

To compare Hi-C interaction frequencies with SMC3 binding, we separated the genome into 40kb interacting bin-pairs and stratified them according to if each bin in the pair is bound by SMC3 ("SMC3 2x"), or if only one bin ("SMC3 1x") or no bins ("None") were bound by SMC3 (**Figure 2b**). We observed a higher interaction frequency in control Hi-C experiments between bin-pairs containing SMC3 sites on both ends than when only one or no SMC3 site is present (**Figure 2c**), consistent with the notion that cohesin binding could mediate long-range chromatin interaction frequencies genome-wide. Upon cleavage of RAD21, we observed an overall loss in local chromatin interaction frequency primarily occurring at distances up to 2 Mb, with a maximum in the range between 100-200kb (**Figure 2d**, **Supplementary Figure 6b**). The loss in interaction frequency is highest after RAD21 cleavage when both interacting loci are bound by SMC3 (**Figure 2d inset**).

Using DNA-FISH (van de Corput, de Boer et al., 2012) we observed a spatial separation of cosmid probes placed in the H19/IGF2 domain (**Figures 2e,f**) and in the HOXD domain (**Figures 2g,h**) after RAD21 cleavage, consistent with the aforementioned loss of interactions.

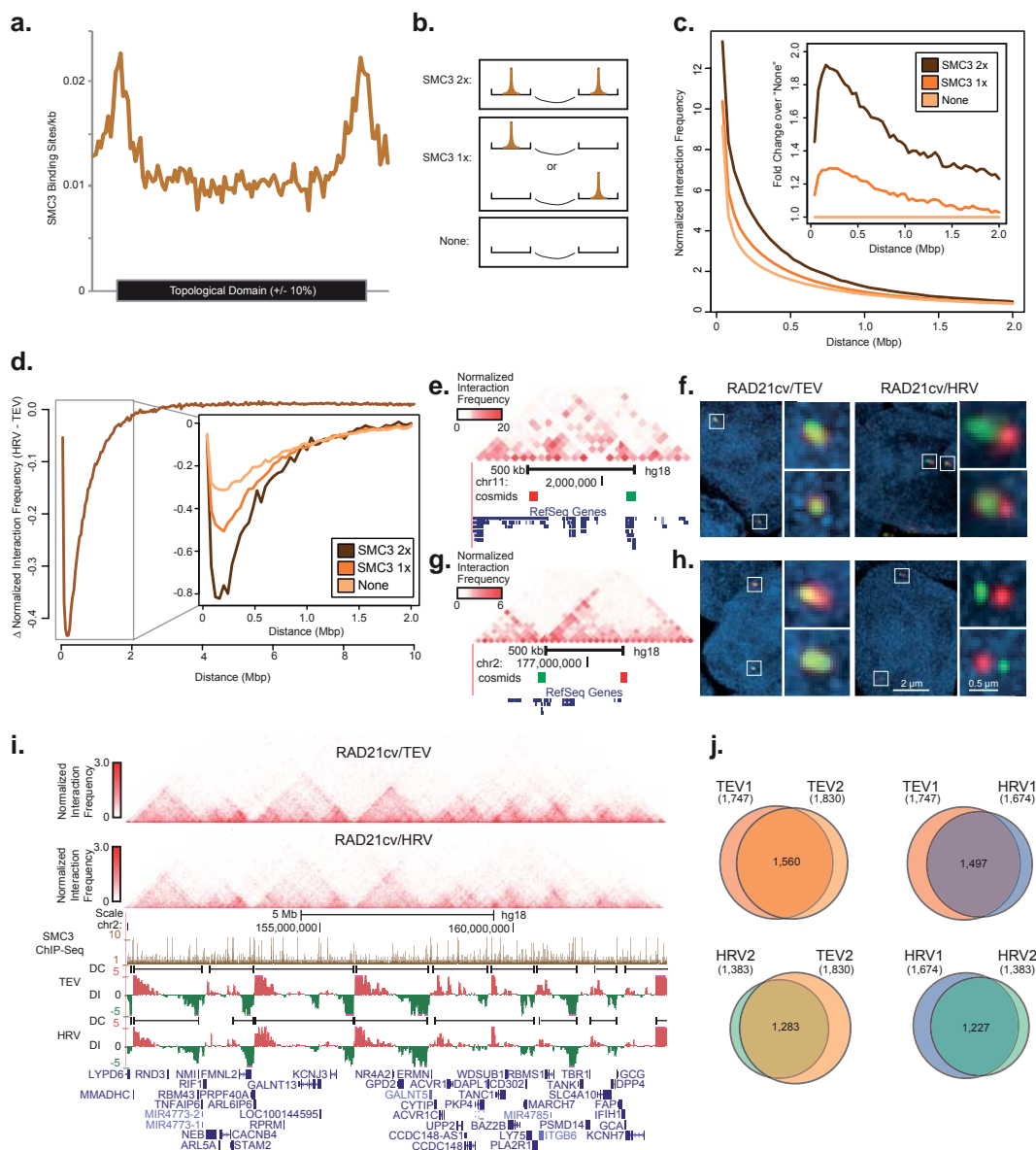


Figure 2: Cohesin cleavage reduces interactions within topological domains genome-wide

a. SMC3 binding frequency across topological domains. SMC3 is enriched at the borders of domains. Each domain was split into 100 bins +/- 10 bins upstream and downstream of the domain boundaries. The frequency of SMC3 binding sites per kb was calculated and averaged over all domains. **b.** Schematic representation of the stratification method of interacting loci based on SMC3 binding. Interacting loci are broken into 3 classes, regions that have at least one SMC3 binding site at each interacting locus (2x), regions that have at least one SMC3 binding site at either interacting locus (1x), and regions that have no SMC3 binding (None). **c.** Hi-C interaction frequency correlated with cohesin. Comparison of the average normalized interaction frequency between SMC3 2x, 1x and none interacting loci at distances from 40kb to 2Mb. The inset is the fold change of the SMC3 2x and SMC3 1x categories relative to the "None" category. The largest fold change in interaction frequency appears between 100-200kb. **d.** Cohesin depletion leads to a loss of local interaction frequency. The y-axis shows that average loss of interaction frequency in the RAD21cv/HRV (HRV) cells compared to RAD21cv/TEV (TEV) cells for distances ranging from 40kb to 10Mb. The largest losses occur between interacting loci that are less than 2Mb apart. The inset shows the degree of depletion

for the SMC3 2x, 1x and none categories. The SMC3 2x category is most affected by RAD21 depletion, and the maximal degree of depletion appears to occur in the 100-200kb range. **e**, Position of the cosmid-based FISH probes in the H19/IGF2 domain relative to the topological domain as shown by Hi-C interaction data. A subset of genes in the region is shown. The color (red, green) of the cosmids corresponds to the FISH images. **f**, DNA-FISH using two cosmid-based probes located in the H19/IGF2 domain in control cells (RAD21cv/TEV, left panel) and after RAD21 cleavage (RAD21cv/HRV, right panel). **g**, Cosmid-based FISH probes at the topological domain of the HOXD gene locus. Only a subset of genes in the region is shown. The color of the cosmid probes (red, green) corresponds to the FISH images. **h**, DNA-FISH using the cosmid probes shown in (g) in control cells (RAD21cv/TEV, left panel) and after RAD21 cleavage (RAD21cv/HRV, right panel). The marked FISH signals (white boxes) are shown enlarged at the right side of each panel. Consistent with the loss of interactions observed in the chromatin conformation capturing experiments we observed a separation of the FISH signals after cohesin cleavage. **i**, The positions of topological domains does not markedly change with RAD21 depletion. Browser shot showing heat maps of interaction frequency in the Control and RAD21 depleted cells. Also shown are the domain calls (DC) and directionality index (DI) over this region. **j**, Comparison of the topological domain boundary calls between replicates (TEV1/TEV2; HRV1/HRV2) and control (TEV) and RAD21 depleted cells (HRV). The differences between replicates and between control and knockdown experiments are comparable and largely unchanging.

We next investigated the effects of cohesin complex destruction on topological domain organization. Surprisingly, the positions of most topological domains do not markedly change upon cleavage of RAD21 (**Figures 2i,j**). The “triangle” pattern of topological domains is still readily apparent in the interaction heat maps, and, though we consistently call fewer domains in the RAD21 depleted cells (**Figure 2j**), there is a strong overlap in domain boundaries called between control and RAD21 depleted cells. The preservation of topological domains after RAD21 cleavage is consistent with live cell imaging observations of histone H2A-RFP in RAD21cv/TEV and RAD21cv/HRV cells showing no general changes of chromatin morphology after RAD21 cleavage (**Supplementary Figure 7**). However, consistent with the previously described general loss in interaction frequency, we also observed a clear reduction in interaction frequency both within and between domains after RAD21 depletion (**Supplementary Figure 8**). Interestingly, the degree of depletion in interaction frequency within domains is most marked when one or both interacting bins is associated with a boundary region (**Supplementary Figure 8**). Taken together, these results suggest that cohesin contributes to the self-association within topological domains by promoting interactions between regions near the boundaries. However, cohesin depletion does not appear to contribute to the positioning and segregation of neighbouring domains from each other.

To determine CTCF's role in mediating chromatin interactions and to compare it to the effects of RAD21 cleavage, we performed two replicates of Hi-C experiments for CTCF and control siRNA knockdowns in HEK293T cells (**Supplementary Figures 9a,b**). We obtained between 95 and 288 million unique reads for each replicate. Similar to SMC3, CTCF is enriched at the boundaries of topological domains in control cells (**Figure 3a**).

Likewise, CTCF binding correlates with the strength of Hi-C interaction frequency, where interacting bin-pairs bound by CTCF on each side form stronger interactions compared to regions with only one or no CTCF sites (**Figure 3b**). Upon knockdown of CTCF, we observed a loss of interactions within topological domains, but with a different pattern with respect to the distance between interacting loci compared to RAD21 cleavage (**Figures 3c,d**).

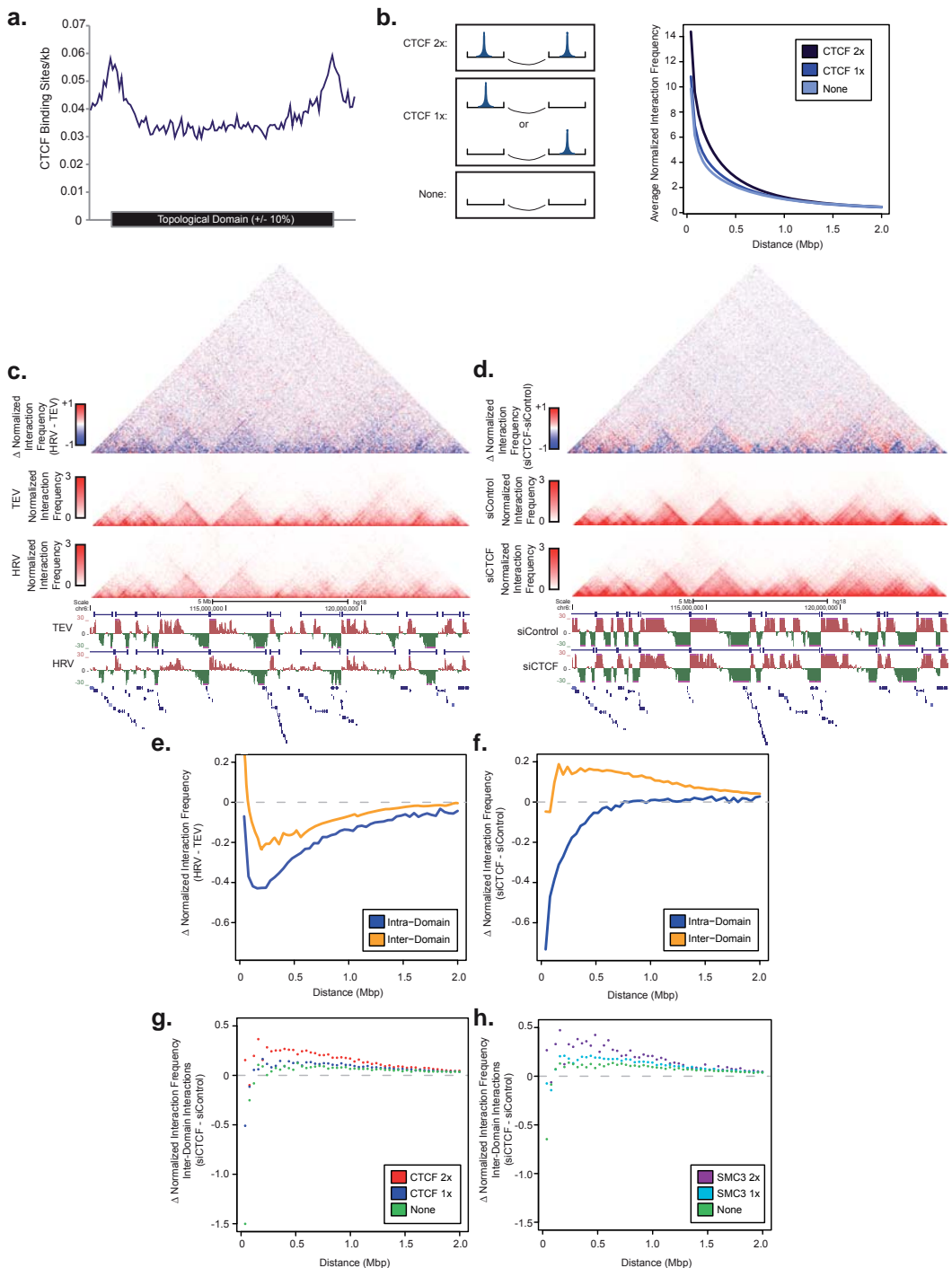


Figure 3: CTF depletion reduces the function of domain boundaries

a, CTF binding frequency across topological domains. Similar to Fig. 2a, CTF is enriched at the borders of topological domains. **b**, Using a stratification scheme similar to what was described in Fig. 2b, CTF binding correlates with interaction frequency. The average interaction frequency at each distance was calculated for each category of interaction (CTCF 2x, CTCF 1x, and none). CTCF 2x shows the strongest interactions. **c**, Heat maps showing changes in interaction frequency between control and RAD21 depleted samples, as well as the actual

interaction frequencies and domain calls. In the top heat map, blue color indicates a loss of interaction frequency, and red indicates a gain in interaction frequency. Upon RAD21 depletion there is predominantly a loss of intra-domain interaction frequency. **d**, Similar to **c**, but showing the changes in interaction frequency over the same locus after CTCF siRNA. A similar pattern of loss of intra-domain interaction frequency is observed (see blue triangles). However, unlike after RAD21 depletion, CTCF siRNA leads to an increase in inter-domain interaction frequency (see red signal in between blue triangles). **e**, Quantification of average change in interaction frequency after RAD21 depletion for intra-domain interaction (blue) and inter-domain interactions (yellow). In both cases, RAD21 leads to a loss of interaction frequency. **f**, Similar to **e**, but showing the change in interaction frequency after CTCF siRNA for intra- and inter-domain interactions. CTCF siRNA leads to a loss of intra-domain interaction frequency, but unlike RAD21 depletion, CTCF siRNA leads to an increase in inter-domain interaction frequency. **g-h**, Changes in inter-domain interaction frequency after CTCF siRNA depletion at sites stratified for either CTCF binding (**g**) or SMC3 binding (**h**). The loci that show the greatest increase in inter-domain interactions after CTCF siRNA tend to have higher frequencies of CTCF or SMC3 binding.

RAD21 depletion appears to most markedly affect interacting loci separated by 100 to 200kb (**Figure 2d**), while CTCF knockdown appears to most prominently affect interacting loci separated by less than 100kb (**Figure 3b**). This implies that CTCF and cohesin may affect intra-topological domain interaction frequency on different spatial scales (**Figures 3e,f blue line**).

The more remarkable difference between CTCF and cohesin depletion concerns the interactions between topological domains. RAD21 depletion leads to a loss in interactions within and between domains but primarily in intra-domain interactions (**Figure 3e yellow line**). CTCF depletion, on the other hand, leads to a significant gain of interactions between neighbouring domains (**Figure 3f yellow line**). This increase in interaction frequency is seen at nearly all distance scales between interacting loci, suggesting that CTCF is necessary to maintain topological domain boundaries throughout the genome. The interactions gained by CTCF depletion could involve delocalised cohesin which now forms “non-specific” interactions, as we have previously shown that cohesin is delocalised but still present on chromatin after CTCF knockdown (Wendt, Yoshida et al., 2008) (**Supplementary Figure 9c**).

Indeed, we observed that the largest gains of inter-domain interaction frequency after CTCF knockdown occur between bins containing CTCF or cohesin sites (**Figures 3g,h**). Altogether our observations suggest that cohesin and CTCF are both important in shaping genomic structure on the level of topological domains in a non-redundant manner.

The above observations suggest that loss of cohesin or CTCF affects chromatin structure in different ways. Given the intimate relationships between chromatin structure and gene regulation, we would predict that loss of these two factors would differentially affect gene expression. To test this prediction, we performed RNA-sequencing (RNA-seq) in the control (RAD21cv/TEV) and RAD21 depleted (RAD21cv/HRV) cells as well as CTCF RNAi and mock treated cells. In both cases we observed only modest changes in gene expression (Supplementary tables 1-4), consistent with earlier observations (Wendt, Yoshida et al., 2008). We observed 48 and 161 differentially expressed genes (FDR <5%) for RAD21 and CTCF depletion, respectively but very little overlap between these sets (**Supplementary table 1, Figure 4a**).

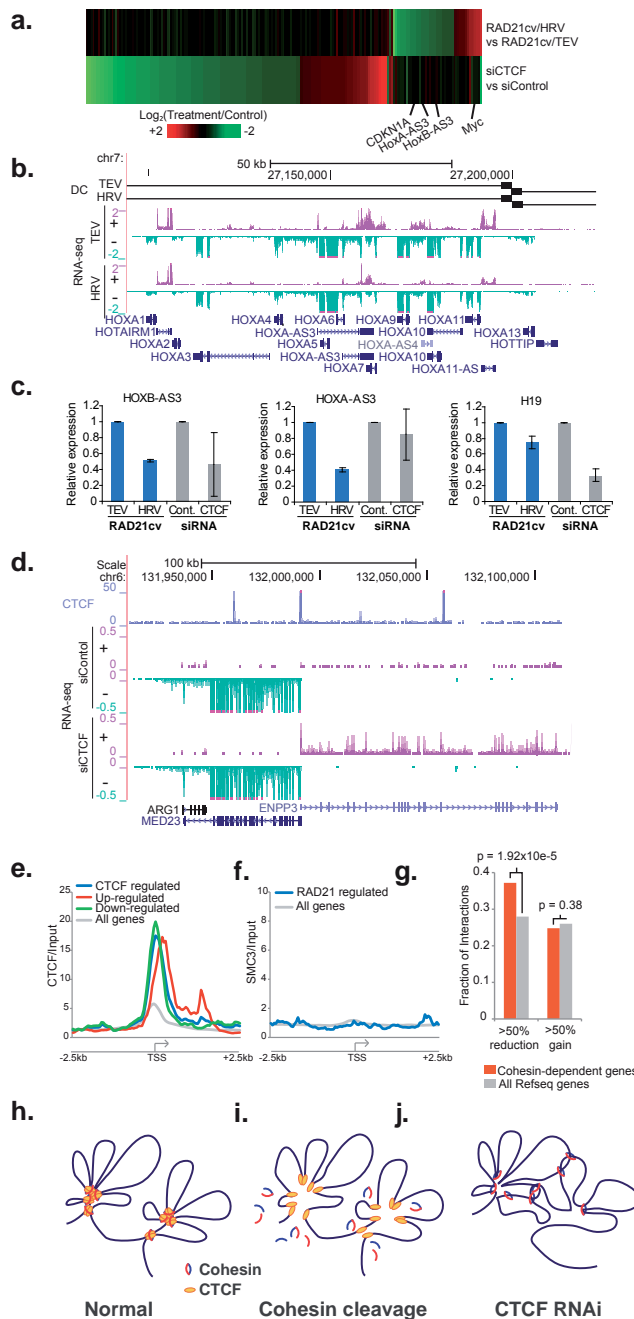


Figure 4: Transcriptional changes after cohesin cleavage and CTCF depletion

a. Changes in expression levels for differentially expressed genes (False Discovery Rate-FDR <5%) are compared between RAD21 cleavage and CTCF RNAi treatment by ranking first the fold changes from highest to lowest for RAD21 cleavage (upper part) and for CTCF depletion (the lower part). Only very few genes behave similar in both experiments. **b.** Expression of HOXA genes is altered after RAD21 cleavage. RNA-seq read coverage normalized to total sequenced reads for both strands are shown for RAD21cv/TEV and RAD21cv/HRV cells (+ strand shown in purple, - strand shown in turquoise). Please note the strong reduction of HOXA-AS3 and HOXA11-AS and HOXA7. **c.** Reduced expression of HOXB-AS3 and HOXA-AS3 after RAD21 cleavage was confirmed by qPCR. CTCF depletion

did not lead to a consistent reduction in expression, as also seen in the analysis of the RNAseq data (Supplementary table 1). Transcription of the H19 noncoding RNA was reduced after CTCF depletion but to a smaller extent by RAD21 cleavage. **d**, Transcription of the ENPP3 gene is increased after CTCF knockdown. RNA-seq read coverage normalized to total sequenced reads for both strands are shown for control siRNA and CTCF siRNA (+ strand shown in purple,- strand shown in turquoise). The gene has CTCF binding sites at the promoter and also intragenic. The upregulation was confirmed by RT-PCR to depend solely on CTCF knockdown (Supplementary fig. 10f). **e**, The position of CTCF sites was analysed relative to transcription start sites of all genes (grey line) and genes with altered expression after CTCF depletion (blue line). Each line represents that average fold-enrichment of CTCF (RPKM) relative to input (RPKM) over a +/- 2.5 kb window surrounding the promoters of CTCF regulated genes. CTCF is clearly enriched at the TSS of differentially expressed genes. **f**, Similar to e, except showing the fold-enrichment of SMC3 (RPKM) over input (RPKM) over the promoter of genes altered after RAD21 depletion. SMC3 does not appear to be enriched at the promoter of the genes regulated by cohesin depletion. **g**, Analysis of change in interaction frequency between restriction fragments containing a promoter and restriction fragments containing a distal DNaseI hypersensitive site (DHS). Shown is the fraction of genes that display a 50% reduction or 50% increase in interaction frequency after RAD21 depletion for either cohesin regulated genes (orange) or all Refseq genes (grey). Cohesin regulated genes are enriched for a loss of interactions with restriction fragments containing distal DHS sites relative to all Refseq genes (Fisher's exact test). **(h-j)** Models describing the different changes in chromosomal interactions after cohesin cleavage (i) and CTCF depletion (j). **h**, Cohesin and CTCF colocalize and are both necessary to shape long-range interactions. **i**, RAD21 cleavage releases cohesin from chromatin but does not change CTCF binding. Loss of cohesin leads to reduced interactions within domains. CTCF binding might still influence the topology of the chromatin fibre and maintain domain identity. **j**, CTCF depletion leads to non-specific cohesin localization which could lead to interactions across domain boundaries normally prevented by CTCF's insulation function.

Among the genes with reduced expression after RAD21 depletion are several Hox genes (HOXA11AS, HOXA-AS3, HOXB-AS3, HOXB5, HOXC9) (**Figure 4b**). We validated the reduced expression of HOXB-AS3, HOXA-AS3 and H19 by RT-PCR and qPCR (**Figure 4c**). Hox genes have been shown to be regulated by antisense transcription as well as the topological organization of the locus (Alexander, Nolte et al., 2009; Noordermeer, Leleu et al., 2011), but have never been reported to depend on cohesin.

Among genes which are differentially expressed after CTCF depletion, we observed a clear enrichment of CTCF binding at their promoters (**Figure 4e**), with a median distance from the TSS to the nearest CTCF binding site being only 191bp (**Supplementary Figure 11**). In contrast, genes that are differentially expressed after cohesin depletion are not directly bound at their promoter by SMC3 (**Figure 4f**), though they are located closer to SMC3 binding sites than would be expected at random (median distance ~4kb, **Supplementary Figure 11**). This indicates that altered expression of genes after RAD21 cleavage may be a product of higher-order chromatin structural changes, as suggested from the observed changes of long-range chromosomal interactions and gene expression of the HOXA and the HOXB locus (**Supplementary Figure 10**). To validate this, we analysed interactions of cohesin-regulated genes with DNaseI hypersensitive sites as markers for potential distal gene regulatory regions at a restriction fragment level resolution. We observed that cohesin regulated genes lose more interactions with distal DNaseI hypersensitive sites than with non-cohesin regulated Refseq genes (**Figure 4g**). These results suggest that cohesin may regulate gene expression by affecting the interaction frequency of genes with distal regulatory elements, while CTCF may directly regulate genes by binding at their promoters.

In summary we showed for the first time how cohesin and CTCF contribute to the topological domain architecture of the human genome. We observed a loss of interactions within and also between domains after cohesin cleavage. On the contrary, CTCF depletion reduces the intra-domain interactions at a somewhat shorter distance while leading to a gain of interactions across domain boundaries. This suggests that cohesin is mainly involved in the self-association property of domains while CTCF is important for their spatial segregation (**Figures 4h-j**). We hypothesize that CTCF maintains boundaries by determining cohesin localization and, in the absence of CTCF, cohesin might form "non-specific" interactions reaching beyond boundaries. Consistent with these differential contributions to the overall architecture, we observed different sets of genes changing after cohesin removal or CTCF depletion. For CTCF, a direct role for transcription is emerging since promoters of CTCF-dependent genes are often bound by CTCF. Genes differentially expressed after RAD21 cleavage have no cohesin bound to their promoters and might be primarily regulated by distal regulatory sequences whose interaction with their target promoters is impaired after chromatin structural changes. This is in good agreement with previous observations suggesting a role for cohesin in enhancer function (Rollins, Korom et al., 2004; Wendt, Yoshida et al., 2008; Kagey, Newman et al., 2010; Seitan, Hao et al., 2011). Taken together, these results provide an initial model for understanding the mechanisms of higher-order chromatin organization and its relationship to gene regulation.

Methods summary

RAD21 cleavage experiments:

HEK293T stable cell lines containing episomes coding for RAD21cv or RAD21wt were grown for 3 days in presence of doxycycline until the endogenous RAD21 was replaced by the engineered RAD21 versions, transfected with either control protease (TEV) or cleavage protease (HRV) according to the manufacturer's instructions and harvested after 24 hours. Cells were then prepared according to the experimental protocols. A detailed description of all methods can be found in the Supplementary Methods.

Acknowledgements:

We acknowledge Laura Schöckel and Olaf Stemmann for providing us the RAD21cv/wt episomes and protease expression constructs. To Ralph Stadhouders, Eric Soler and Robert-Jan Palstra we are very thankful for their advice on 3C-seq. We thank Reinier van der Linden for help with FACS cell cycle analysis. Supat Tongjuea and Boris Lenard we acknowledge for providing the 3C-seq analysis pipeline and help with the analysis of initial HB2 cell data. Lee Edsall and Samantha Kuan we thank for their assistance in NGS DNA sequencing. Niels Galjart and Jan-Michael Peters we thank for the gift of antibodies and Daniele Amadio for the generation and purification of the rabbit anti-EGFP antibodies. Annelies de Klein and Bert Eussen we acknowledge for providing the cosmids and Maiwen Caudron-Herger for providing the H2A-RFP constructs. J.Z. is supported by a NWO grant to K.S.W. K.S.W is supported by the Erasmus MC and the TRR81. Work in the lab of B.R. is supported by the Ludwig Institute for Cancer Research, the California Institute of Regenerative Medicine (RN2-00905-1) and NIH. P.K. was supported by grants from ERASysBio + /FP7.

Author Contributions:

J.Z. performed 3C-seq and ChIP experiments and analysis. J.D. performed HiC and analysis of HiC, RNA-seq and ChIP-seq experiments. M. v.d.R. established RAD21cv and wt cell lines, the experimental set-up of the cleavage system and helped with cleavage experiments and performed live cell imaging. Z.Y. performed RNA-seq experiments. R.B and W. v.IJ. performed genomic sequencing of the SMC3 ChIP-seq experiments and mapping of the reads. P. K. helped with analysis of the 3C-seq data. J.Z. and M.v.d. C. performed the DNA-FISH, image acquisition and processing. F.G. contributed with his expertise to experimental design and provided financial support. K.S.W. helped with experiments and designed the study together with B.R., K.S.W., J.D., B.R. and J.Z. wrote the manuscript.

Author Information:

All Hi-C data, ChIP-sequencing data and RNA-sequencing data described in this study have been deposited in the GEO under accession number XXXXXX. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.R. (biren@ucsd.edu) and K.S.W. k.wendt@erasmusmc.nl.

References

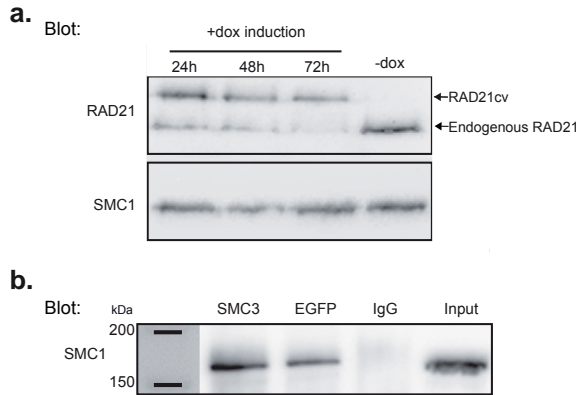
- Alexander T., C. Nolte, et al., 2009. Hox genes and segmentation of the hindbrain and axial skeleton. *Annu Rev Cell Dev Biol* 25, 431-456.
- Bell A.C., A.G. West, et al., 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98, 3, 387-396.
- Dixon J.R., S. Selvaraj, et al., 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 7398, 376-380.
- Hadjur S., L.M. Williams, et al., 2009. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* 460, 7253, 410-413.
- Hou C., R. Dale, et al., 2010. Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc Natl Acad Sci U S A* 107, 8, 3651-3656.
- Imakaev M., G. Fudenberg, et al., 2012. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Methods* 9, 10, 999-1003.
- Kagey M.H., J.J. Newman, et al., 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 7314, 430-435.
- Kurukuti S., V.K. Tiwari, et al., 2006. CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc Natl Acad Sci U S A* 103, 28, 10684-10689.
- Lieberman-Aiden E., N.L. van Berkum, et al., 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 5950, 289-293.
- Mishiro T., K. Ishihara, et al., 2009. Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *Embo J* 28, 9, 1234-1245.
- Nativio R., K.S. Wendt, et al., 2009. Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. *PLoS Genet* 5, 11, e1000739.
- Noordermeer D., M. Leleu, et al., 2011. The dynamic architecture of Hox gene clusters. *Science* 334, 6053, 222-225.
- Nora E.P., B.R. Lajoie, et al., 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 7398, 381-385.
- Parelho V., S. Hadjur, et al., 2008. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 132, 3, 422-433.
- Phillips J.E. and V.G. Corces, 2009. CTCF: master weaver of the genome. *Cell* 137, 7, 1194-1211.
- Rollins R.A., M. Korom, et al., 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. *Mol Cell Biol* 24, 8, 3100-3111.
- Rubio E.D., D.J. Reiss, et al., 2008. CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci U S A* 105, 24, 8309-8314.

- Schockel L., M. Mockel, et al., 2011. Cleavage of cohesin rings coordinates the separation of centrioles and chromatids. *Nat Cell Biol* 13, 8, 966-972.
- Seitan V.C., B. Hao, et al., 2011. A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. *Nature* 476, 7361, 467-471.
- Sexton T., E. Yaffe, et al., 2012. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* 148, 3, 458-472.
- Splinter E., H. Heath, et al., 2006. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev* 20, 17, 2349-2354.
- Stadhouders R., P. Kolovos, et al., 2013. Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nat Protoc* 8, 3, 509-524.
- Uhlmann F., F. Lottspeich, et al., 1999. Sister-chromatid separation at anaphase onset is promoted by cleavage of the cohesin subunit *Sccl*. *Nature* 400, 6739, 37-42.
- van de Corput M.P., E. de Boer, et al., 2012. Super-resolution imaging reveals three-dimensional folding dynamics of the beta-globin locus upon gene activation. *J Cell Sci* 125, Pt 19, 4630-4639.
- Wendt K.S., K. Yoshida, et al., 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451, 7180, 796-801.

Supplementary materials:

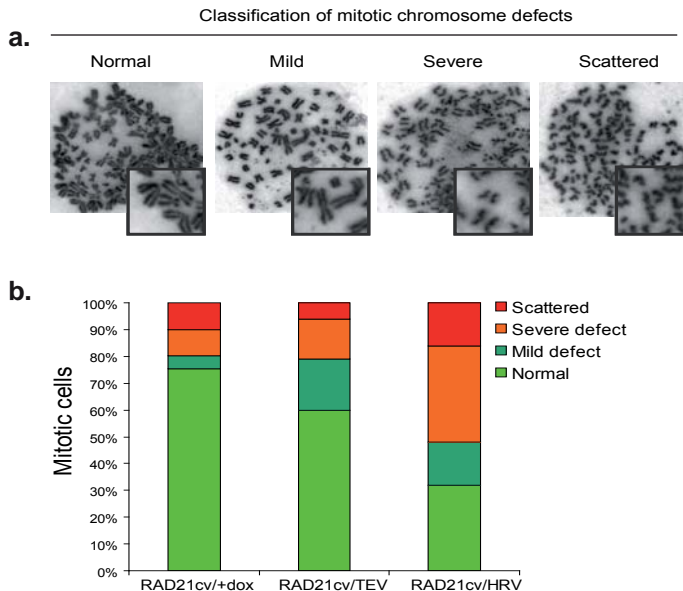
1. Supplementary Figures 1-12 including Legends
2. Supplementary Material and Methods
3. Supplementary tables 1-7

1. Supplementary Figures



Supplementary figure 1: Replacement of the endogenous RAD21 by RAD21cv and incorporation in the cohesin complex.

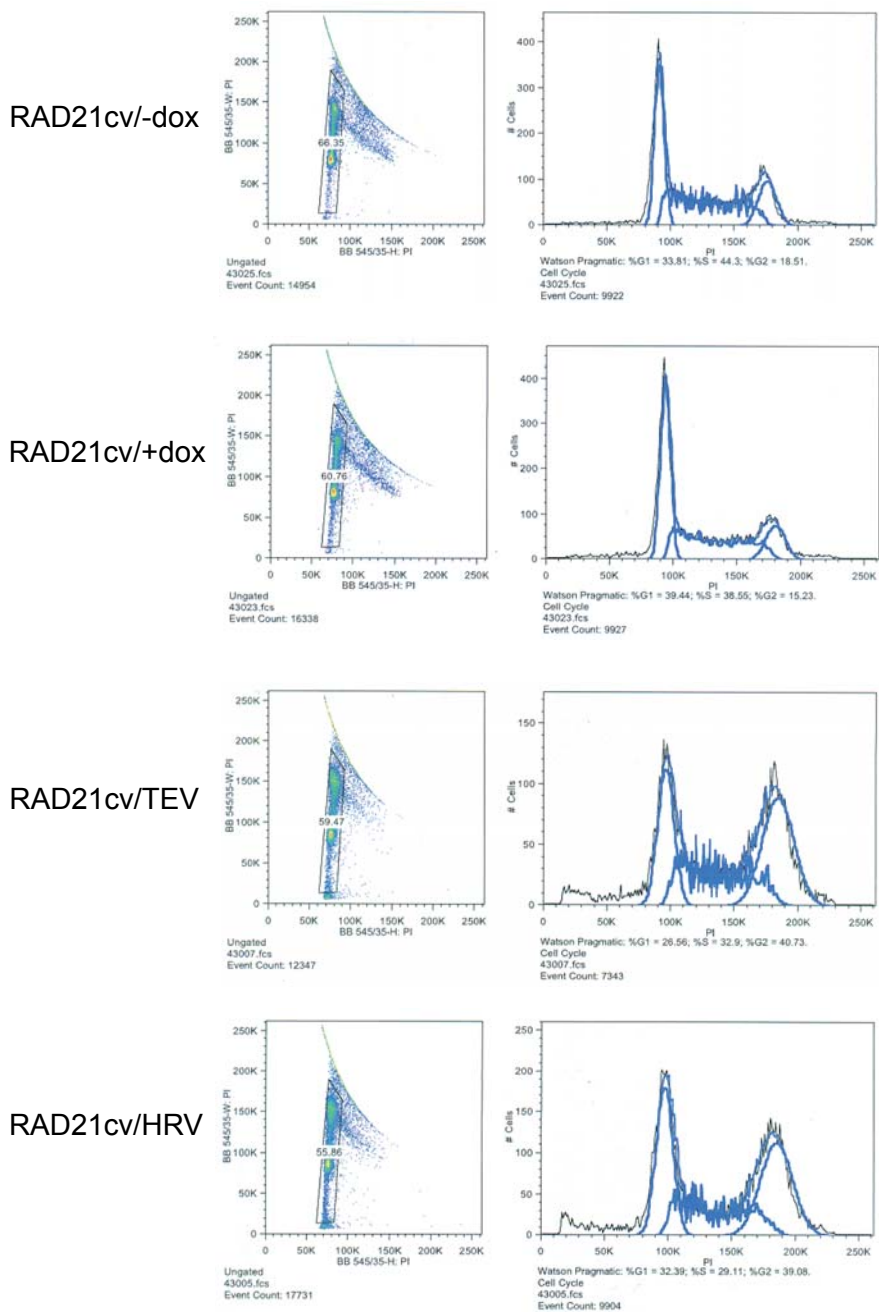
a, Western blot showing the replacement of endogenous RAD21 by RAD21cv in a time-course until 3 days after doxycycline induction. Note that the chromatin-bound level of RAD21cv is similar to the endogenous RAD21 level (Fig. 1d). **b,** Immunoprecipitation with anti-EGFP and anti-SMC3 antibodies was performed from RAD21cv cells and the Western blot probed with anti-SMC1. The co-precipitation of SMC1 with RAD21cv (EGFP-tag) shows that RAD21cv is incorporated in the cohesin complex.



Supplementary figure 2: RAD21 cleavage causes premature loss of sister chromatid cohesion.

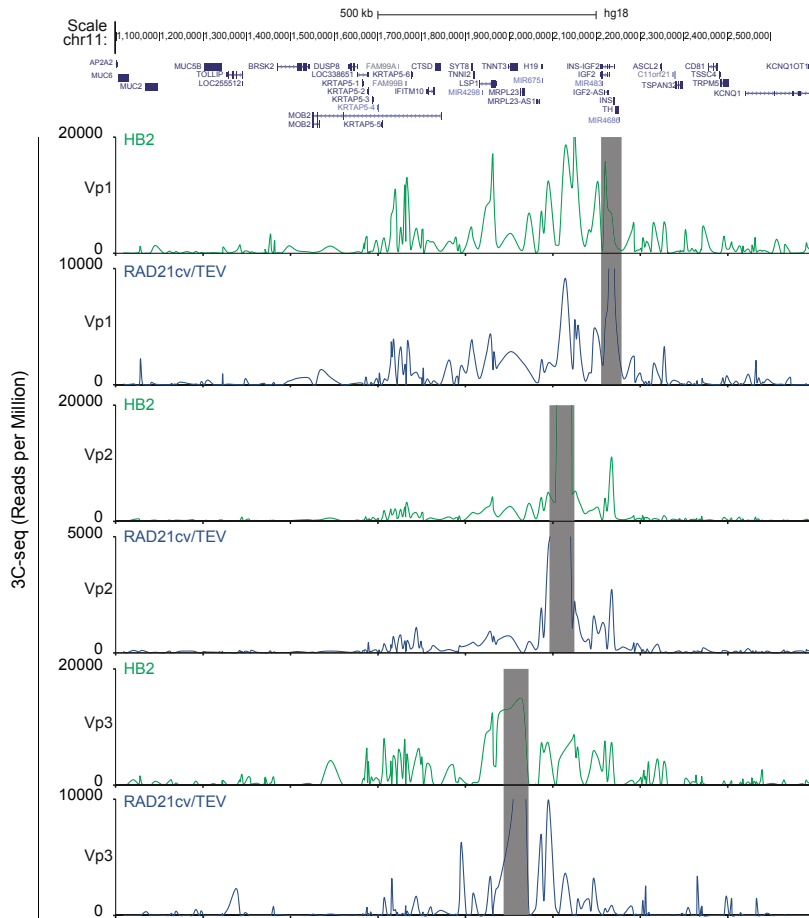
Cells were treated for 2 hours with nocodazole and then spread to analyse the mitotic cells for sister chromatid cohesion defects.

a, Different degrees of sister chromatid cohesion defects used to cluster observed mitotic defects. **b,** Bar chart displaying the percentage of counted mitotic cells displaying different degrees of sister cohesion defects. We analyzed doxycycline induced RAD21cv cells (RAD21cv/+dox), and RAD21cv cells transfected with the different proteases (RAD21cv/TEV and RAD21cv/HRV).



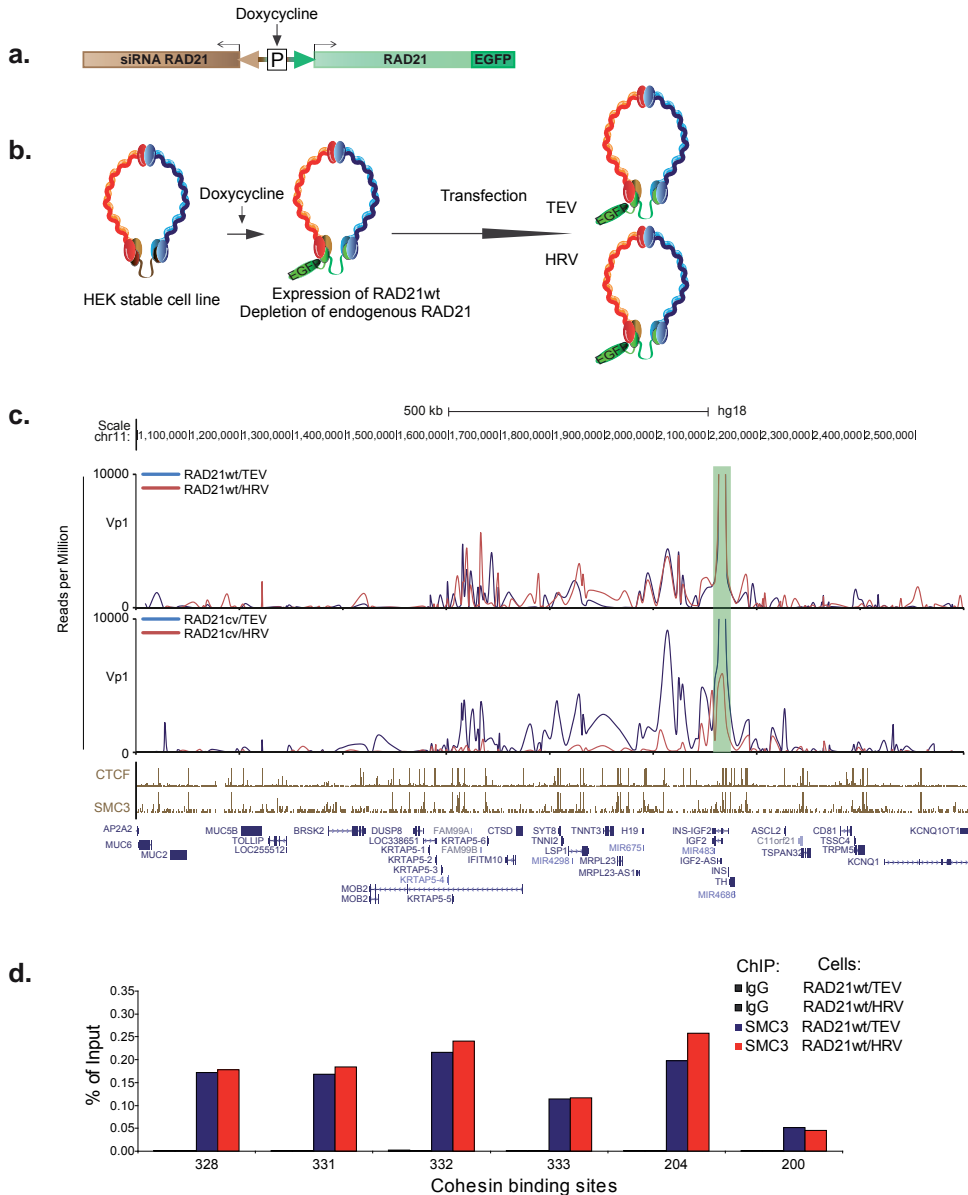
Supplementary figure 3: Cell cycle distribution of treated and untreated RAD21cv cells.

FACS analysis of the cell cycle distribution of the cells used in the RAD21 cleavage experiments: uninduced cells (RAD21cv/-dox), doxycycline induced (RAD21cv/+dox) and cells after transfection with the different proteases (RAD21cv/TEV, RAD21cv/HRV).



Supplementary figure 4: Conservation of long-range chromosomal interactions between different tissues.

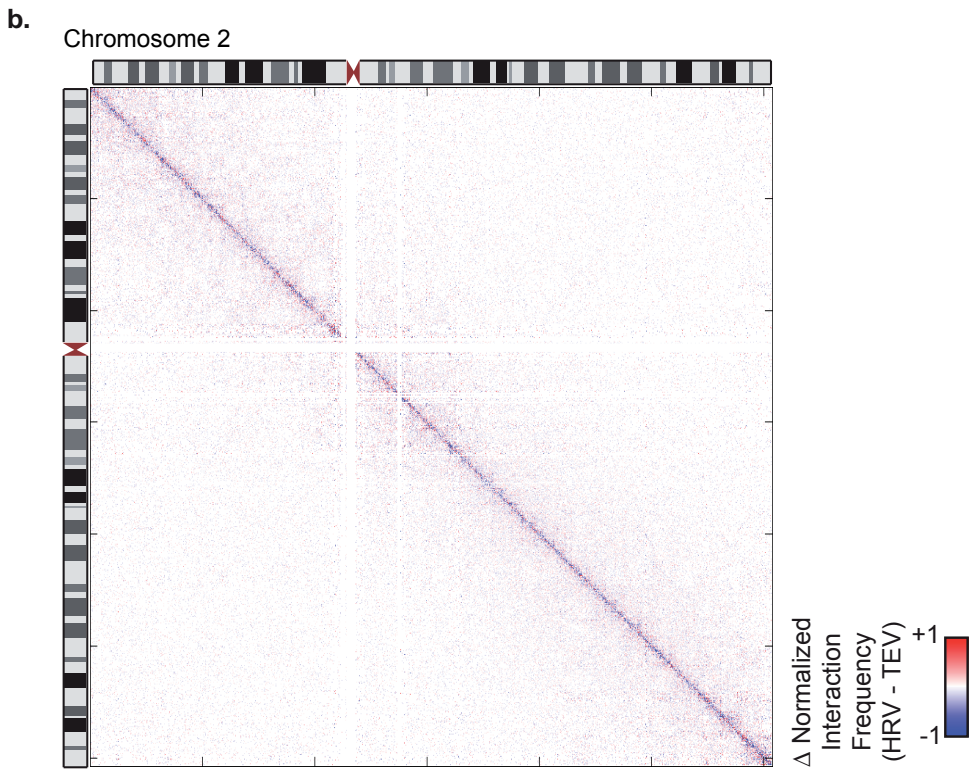
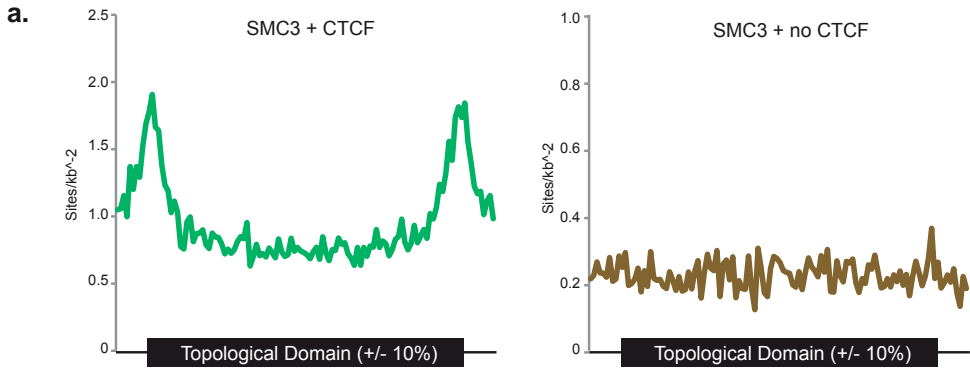
3C-seq was performed for three different viewpoints (Vp1-3) in a breast endothelial cell line with normal karyotype (HB2, green line) and the control cells for the RAD21 cleavage experiment (RAD21cv/TEV, blue line) using the same protocol. The viewpoint positions are marked with grey bars. All three viewpoints in the IGF2-H19 domain show similar interactions, indicating conservation of chromosomal interactions between different cell types.



Supplementary figure 5: Control cells expressing RAD21wt w/o HRV site do not respond to HRV protease transfection.

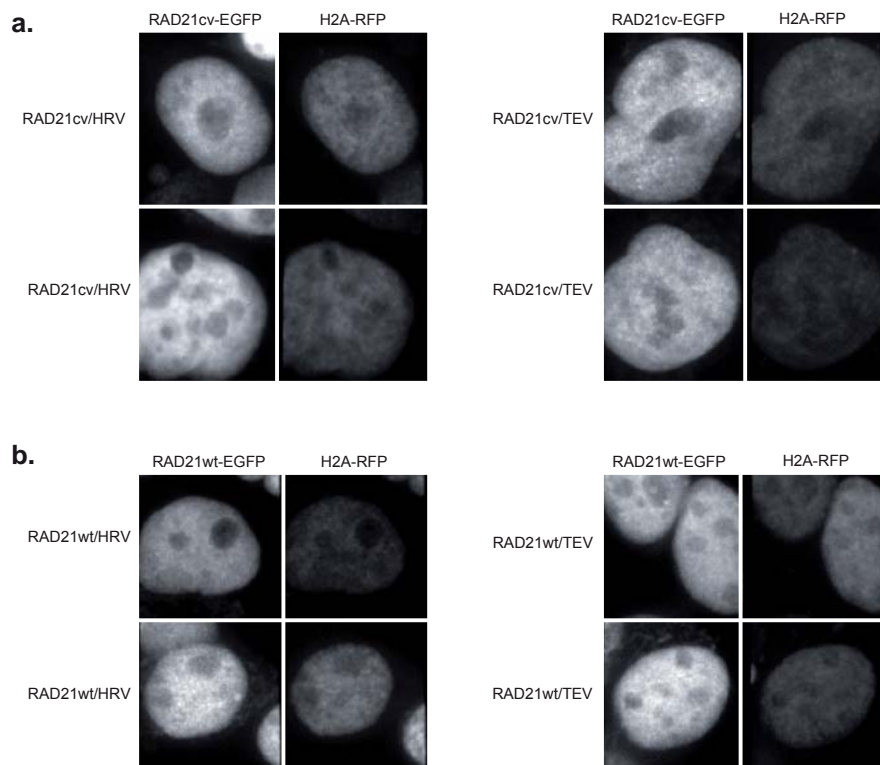
a. Schematic representation of the episomal construct used for the control cell line containing a doxycycline-inducible bidirectional promoter driving the expression of RAD21wt and siRNA targeting endogenous RAD21 simultaneously.

b. Outline of the experiment. **c.** Cells expressing protease-insensitive RAD21-EGFP (RAD21wt) do not show altered long-range interactions for viewpoint 1 (Vp1) after HRV protease transfection (RAD21wt/HRV, red line graph upper panel), in contrast to cells expressing cleavable RAD21 (RAD21cv/HRV, red line graph lower panel). The respective transfections with control protease are shown as blue line graphs. **d.** The ChIP-qPCR assay using different primer sets corresponding to cohesin binding sites for RAD21wt/TEV and RAD21wt/HRV shows no effect on SMC3 binding to chromatin.



Supplementary figure 6: Enrichment of cohesin/CTCF at the boundaries and differential heat map plot after RAD21 cleavage.

a, SMC3 is enriched at the borders of domains when it colocalizes with CTCF but not without. Each domain was split into 100 bins +/- 10 bins upstream and downstream of the domain boundaries. The frequency of SMC3/CTCF or SMC3 only binding sites per kb was calculated and averaged over all domains. **b**, Chromosome 2 - heat map of changes in interaction frequency. The heat map shows the changes in interaction frequency (RAD21cv/HRV – RAD21cv/TEV). The largest differences are very close to the diagonal, indicating that the most prominent changes in interaction frequency are local (<2Mb), and loss of interactions is dominating.

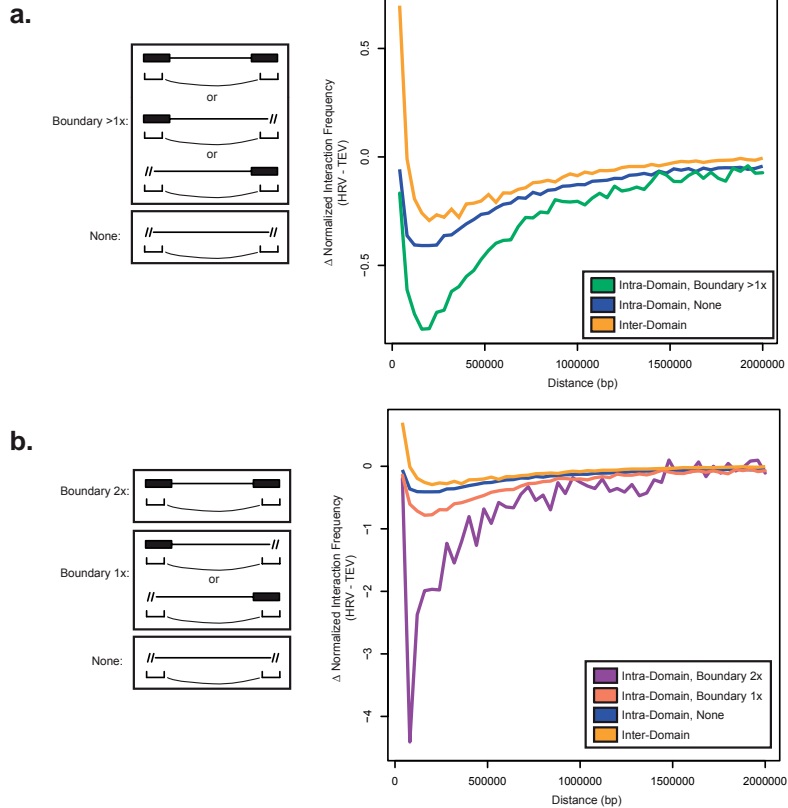


Supplementary figure 7: Live cell imaging shows preservation of chromatin morphology after RAD21 cleavage.

Live cell imaging of RAD21cv and RAD21wt cells transfected with HRV or TEV protease and with a construct encoding histone H2A-RFP. The EGFP signal of RAD21cv/RAD21wt and the RFP signal of the H2A-RFP construct were imaged using a spinning disc microscope. RAD21cv and RAD21wt show a speckled pattern (EGFP) when cells are transfected with TEV.

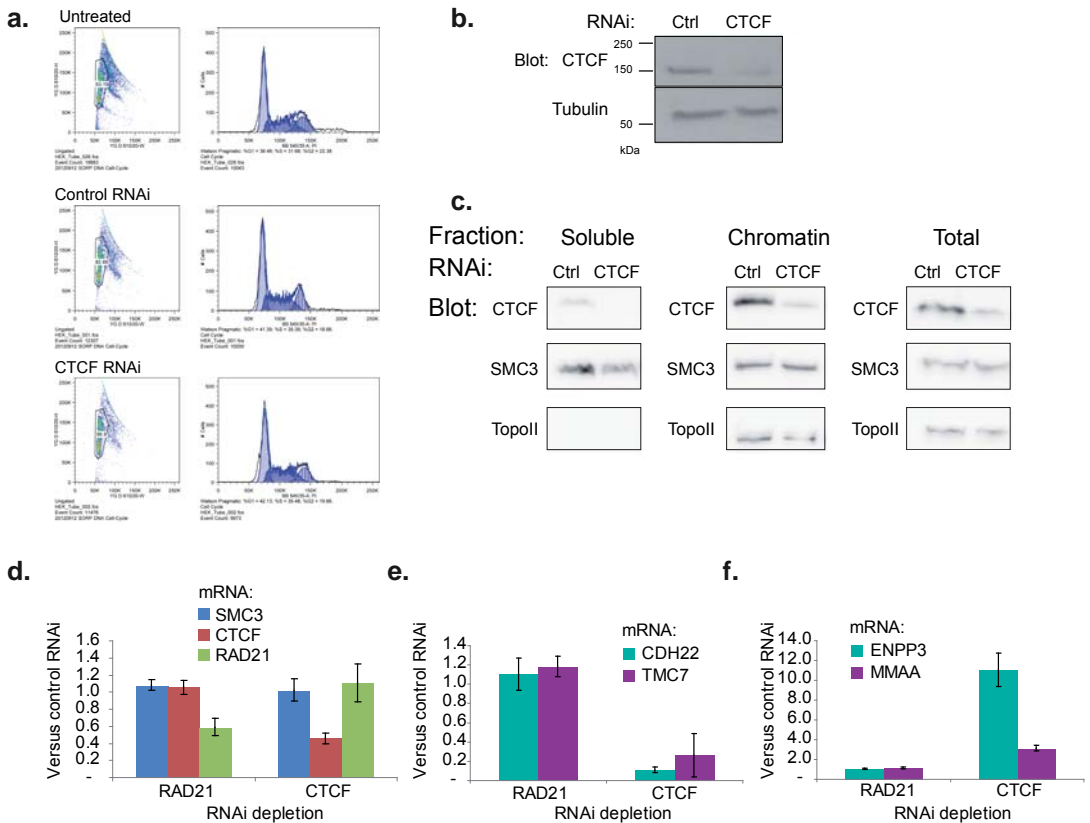
a, The EGFP signal turns into an amorphous pattern when transfected with HRV, consistent with a release of RAD21cv from chromatin. The H2A-RFP patterns do not change visibly between the different protease transfections, indicating that cohesin cleavage does not trigger major changes in chromatin morphology. **b,** The RAD21wt and H2A-RFP patterns do not change when HRV is transfected, indicating that the localization of RAD21wt does not change.

Cohesin Depletion



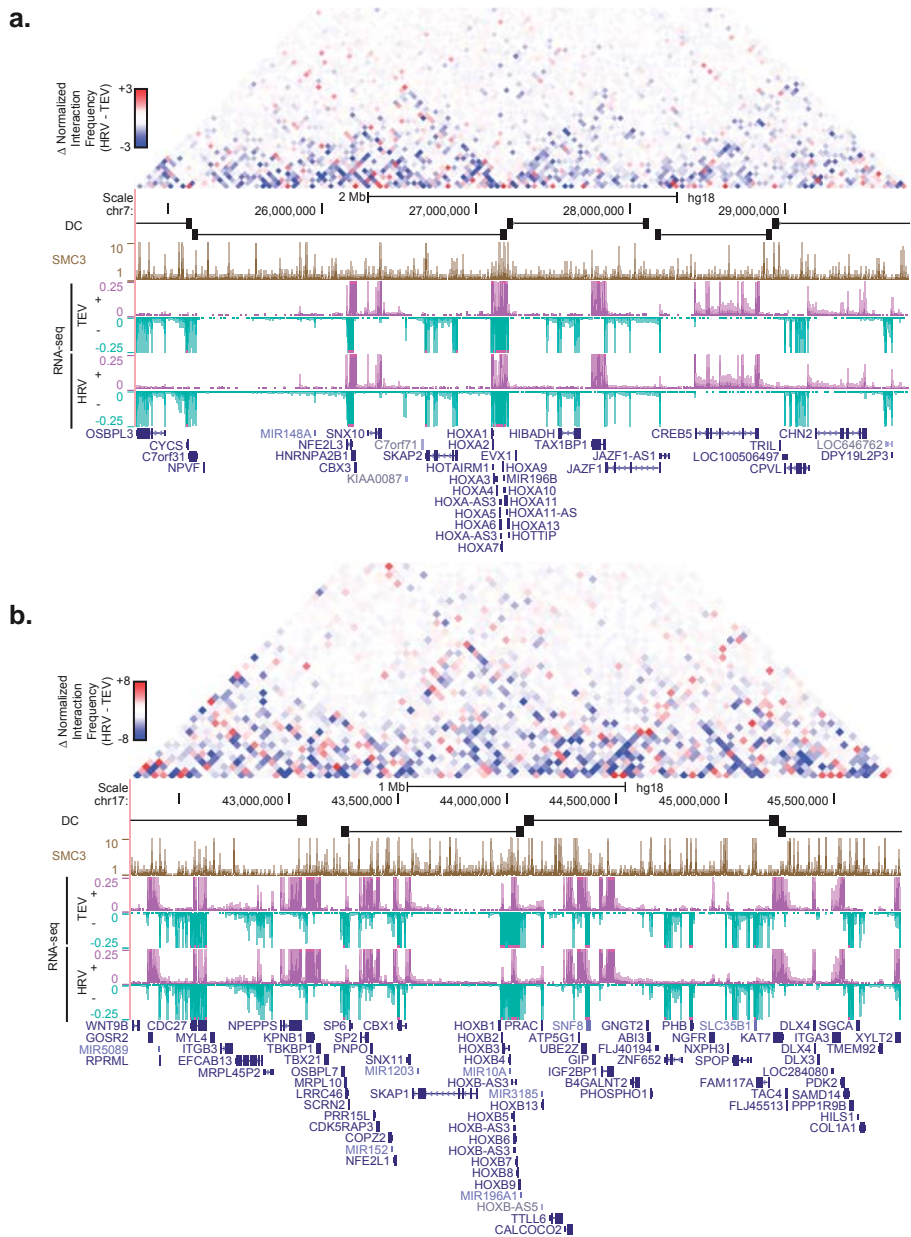
Supplementary figure 8: Cohesin depletion leads to a loss of intra-domain boundary associated interactions.

a. Graphs showing the degree of depletion in interaction frequency between RAD21cv/TEV (TEV) and RAD21cv/HRV (HRV) cells at each distance up to 2Mb after RAD21 depletion. Interacting bin-pairs were stratified either as inter-domain (yellow), non-boundary associated ("None", blue), or where at least one bin was boundary associated ("Boundary >1x", green). The loss of interaction frequency is highest when at least one bin is associated with a boundary. **b.** Similar to panel a, but in this case also showing bin-pairs with exactly one boundary associated bin ("Boundary 1x", orange) or two boundary associated bins ("Boundary 2x", magenta). There are relatively few "Boundary 2x" bin pairs in the genome, which accounts for the jaggedness of the magenta line. In both cases, a schematic explaining the bin-pair segregation scheme is shown on the left.



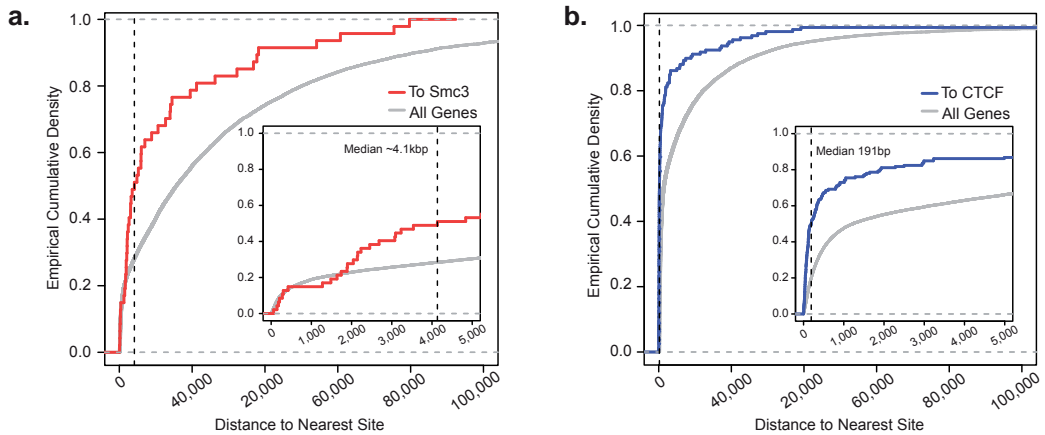
Supplementary figure 9: CTCF RNAi depletes CTCF but does not change the levels of cohesin bound to chromatin.

a, FACS analysis of the cell cycle distribution of untreated HEK293T cells and cells treated with CTCF and control RNAi. **b,** Western blot for CTCF and tubulin as loading control to show the depletion of CTCF by siRNA. **c,** HEK293T cells were transfected with control and CTCF RNAi and the cells fractionated in total cell extract (total), the soluble pool containing cytoplasm and nucleoplasm (soluble) and the chromatin-bound pool (chromatin) of proteins. The western blot for SMC3 shows equal levels of chromatin-bound proteins after CTCF depletion. The chromatin-bound fraction is marked by blotting for Topoisomerase II (TopoII). This confirms observations from our earlier study (Wendt, Yoshida et al., 2008). **d,** Transcripts of SMC3, CTCF and RAD21 were analysed after RNAi depletion of RAD21 and CTCF in HEK293T cells. Consistent with the depletion we observed a reduction of RAD21 and CTCF after the respective siRNA treatments. The fold expression compared to the control RNAi is shown. **e,** Validation of two genes (CDH22 - cadherin 22 precursor, TMC7 - transmembrane channel-like 7) found to be down-regulated after CTCF RNAi by qPCR. The fold expression compared to the control RNAi is shown. **f,** Validation of two genes (ENPP3 - ectonucleotide pyrophosphatase/phosphodiesterase, MMAA - methylmalonic aciduria type A precursor) found to be up-regulated after CTCF RNAi by qPCR. (All SD are from three independent replicates).



Supplementary figure 10: Changes of long-range interactions around the HOXA and the HOXB locus after RAD21 cleavage.

Differences in Hi-C interaction frequency after RAD21 depletion in regions surrounds the HOXA locus (a) and the HOXB locus (b). Heat maps show the difference in interaction frequency between the RAD21cv/HRV and RAD21cv/TEV cells (HRV - TEV). Reduced interactions are shown in blue and increased interactions shown in red. Further we show the location of topological domains using the domain calls (DC), SMC3 ChIP-seq profiles for RAD21cv/TEV cells and RNA-seq experiments with (RAD21cv/HRV) and without (RAD21cv/TEV) RAD21 cleavage. In the RNA-seq tracks, positive stranded reads are shown in purple, while negative stranded reads are shown in turquoise.



Supplementary figure 11

Empirical Cumulative Density plots of the distance from the transcription start site of an RAD21 (a) or CTCF (b) regulated gene to the nearest binding site for either SMC3 (a) or CTCF (b).

a, Shown in red is the empirical cumulative density distribution of the distance between the TSS of RAD21 regulated genes to the nearest SMC3 binding site. The median distance to the nearest site is ~4.1kb (shown with a vertical dashed line). Shown in grey is the distribution for all RefSeq genes, demonstrating that RAD21 regulated genes typically have an SMC3 binding site closer than would be expected at random. The inset shows the same data but zoomed into a 5kb limit. **b**, Similar to a, but showing the distance between the TSS of CTCF regulated genes and the nearest CTCF binding site (blue). The median distance for CTCF regulated genes to the nearest CTCF binding site is 191bp, showing that CTCF regulated genes tend to be directly bound at their promoter by CTCF.

2. Supplementary Material and Methods

Cleavable HRV RAD21-EGFP construct (RAD21cv)

The construct encoding the cleavable RAD21 subunit (RAD21cv) was previously described in (Schockel, Mockel et al., 2011). Briefly, the first RAD21-separase cleavage site was replaced by one for 3C protease of the human rhinovirus (HRV protease) using a PCR-based mutagenesis. The second cleavage site was unchanged to ensure less cell cytotoxicity. RAD21cv was then cloned in front of an EGFP cassette. The tobacco etch virus protease (TEV protease), which does not recognize the HRV cleavage site is used as control.

siRNA cassette for the endogenous RAD21

For the knock-down of the endogenous RAD21 subunit, the following 3'UTR-directed siRNA were used:

5'-ACUCAGACUUCAGUGUAUA-3' (Scc1-1),

5'-AGGACAGACUGAUGGGAAA-3' (Scc1-2).

Episomal system

The vector pRTS-1, described in (Bornkamm, Berens et al., 2005), presents a doxycycline-responsive cassette composed of a bidirectional promoter that drives the expression of two genes in a coordinated fashion. The cleavable RAD21-eGFP and the siRNA for RAD21 cassettes were cloned under the control of the bidirectional promoter; thus after doxycycline treatment both cassettes are expressed simultaneously.

Generation of HEK293T cell stably containing the episomal constructs

HEK293T cell line was cultured in DMEM supplemented with 0.2mM L-glutamine, 100 units/ml penicillin, 100 mg/ml streptomycin and 10% FCS and was grown at 37°C and 5% CO₂. Transfection of the episome was done by Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions and cells carrying the episome vector were selected by growing in a medium containing 150 µg/mL hygromycin. Single clones were picked and analysed for expression of RAD21cv and RAD21wt constructs and depletion of the endogenous RAD21 three days after induction with 2 µg/ml doxycycline.

RAD21 cleavage experiments

To activate transgene expression, cells were cultured for 3 days in the presence of 2 µg/ml of doxycycline. After 3 days, cells were split to 50% confluency and transfected with TEV or HRV construct using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Cells were harvested 24 hours after protease transfection.

Spreading of mitotic chromosomes

Cells were treated with nocodazole (Sigma) for 2 hours and fixed with methanol/acetic acid after hypotonic treatment. After spreading of the cells on cover slips the chromosomes were stained with Giemsa.

Live cell imaging of Histone-RFP after RAD21 cleavage

To activate transgene expression, cells were cultured for 3 days in the presence of 2 µg/ml of doxycycline. After 3 days, cells were split to 50% confluency and transfected with TEV or HRV constructs and H2A-RFP using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Cells were imaged 24 hours after protease transfection using a SpinD1454 Roper/Nikon spinning disk microscope with temperature controller. Cells were imaged using the 60X Objective, 491 nm and 561 nm lasers and 700 ms exposure time. Image stacks were processed with ImageJ and projected in a single plane.

RNAi depletion of CTCF and RAD21

HEK293T cells were seeded in DMEM supplemented with 0.2mM L-glutamine and 10% FCS and transfected with siRNA oligos (Ambion) directed against CTCF and a non-targeting control siRNA using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Cells were harvested 48 hours after transfection according to the respective protocols for Hi-C or western blotting. The following siRNA oligos were used:

CTCF siRNA	sense	GGAGCCUGCCGUAGAAAUUTT
	antisense	AAUUUCUACGGCAGGUCCTC
Control siRNA	sense	CGUACGCGGAAUACUUCGATT
	antisense	UCGAAGUAUCCGCGUACGTT

Depletion of RAD21 in HEK293T by siRNA transfection was only performed for the transcript analysis with the same protocol used for CTCF depletion.

RAD21 siRNA	sense	GGUGAAA AUGGCAUUACGGtt
	antisense	CCGUAAUGCCAUUUUCACctt

Fractionation in soluble and chromatin bound proteins

To prepared soluble and chromatin bound fractions cells were harvested and lysed (20mM Tris-HCl pH 7.5, 100mM NaCl, 5mM MgCl₂, 2mM CaCl₂, 10% glycerol, 0.2% NP-40, 1mM NaF, 0,5mM DTT and protease inhibitors. An aliquot was taken as total lysate and the remaining lysate centrifuged 10 min at 1500 rpm to collect the chromatin pellet. The supernatant was collected as soluble fraction. The pellet was washed 3 times with lysis buffer, resuspended in TBS/T and the chromatin-bound proteins solubilized by sonication and benzonase treatment.

Antibodies

Primary antibodies used were mouse anti-CTCF (BD, for immunoblotting), rabbit anti-CTCF (Millipore, for ChIP), mouse monoclonal anti-EGFP (Sigma, immunoblotting), mouse anti-

tubulin (Sigma, immunoblotting) and rabbit anti-TopoII (Millipore, immunoblotting). Polyclonal rabbit antibodies against RAD21 and STAG1/STAG2 (immunoblotting) were a gift from Jan-Michael Peters and described in (Sumara, Vorlaufer et al., 2000). Rat monoclonal antibodies against SMC1 (immunoblotting) were a gift from Niels Galjart. Polyclonal rabbit antibodies against the human SMC3 (ChIP) were raised against the peptide (C-EMAKDFVEDDTTHG) as described before (Sumara, Vorlaufer et al., 2000) (Absea, China) and purified using the peptide antigen. Polyclonal rabbit antibodies against EGFP were raised against recombinant EGFP produced in *E. coli* (Absea, China) and purified using the protein antigen.

Immunoprecipitation

Immunoprecipitations were performed as described (Kueng, Hegemann et al., 2006) using antibodies against SMC3 and EGFP.

Chromosome conformation capture sequencing (3C-seq) and analysis

Chromosome conformation capture sequencing was performed as previously described in (Simonis, Kooren et al., 2007; Stadhouders, Kolovos et al., 2013). Briefly, cells were crosslinked with 1% (w/v) formaldehyde for 10 minutes and quenched with 120mM glycine. Crosslinked-cells were resuspended in lysis buffer (50mM Tris-HCl pH 8.0, 0.5% NP-40, 50mM NaCl and Complete protease inhibitor (Roche) and subjected to enzymatic digestion using 400 units of BglII (Roche). Digested chromatin was then diluted and ligated using 5 units of T4 DNA ligase (Promega) under conditions favouring intramolecular ligation events. After reversing the crosslink at 65°C over night, the digested and ligated chromatin was subjected to a second enzymatic digest using NlaIII (New England Biolabs) to produce smaller DNA fragment. The resulting digested DNA underwent a second ligation using 10 units of T4 DNA ligase (Promega) under conditions favouring self-ligation events that produce circular DNA molecules. The unknown DNA fragment, ligated to the fragment of interest (called viewpoint), was amplified by inverse-PCR using specific primers designed to anneal to the outer part of the restriction site of the viewpoints, linked with Illumina adapter sequences. The samples were then single-read sequenced using the Illumina Genome Analyzer II generating 76bp reads. The reads were trimmed to remove the illumina adapter sequences and mapped against the human genome (hg18).

The reads were extended to 56bp in the 3' direction using the r3C-Seq pipeline (Thongjuea, Stadhouders et al., in preparation; <http://www.bioconductor.org/packages/2.11/bioc/html/r3Cseq.html>). Interaction frequencies were calculated using the number of reads per million (RPM). The data were visualized using a local UCSC mirror browser.

Cell cycle analysis

Cells were fixed with methanol and after RNase treatment the DNA was stained with propidium iodide. The cells were analyzed with a BD FACS Aria Cell sorter and FlowJo software.

Chromatin immunoprecipitation (ChIP)

Chromatin immunoprecipitation was performed as described before (Wendt, Yoshida et al., 2008). In brief, cells at 70–80% confluence were crosslinked with 1% formaldehyde for 10 minutes and quenched with 125mM glycine. After washing with PBS, cells were resuspended in lysis buffer (50mM Tris-HCl pH 8.0, 1% SDS, 10mM EDTA, 1mM PMSF and Complete protease inhibitor (Roche)) and chromatin was sonicated (Diagenode Bioruptor). After a centrifugation step to remove the debris. The lysate was diluted 1:4 with IP dilution buffer (20mM Tris-HCl pH 8.0, 0.15M NaCl, 2mM EDTA, 1% TX-100, protease inhibitors) and precleared with Affi-Prep Protein A support beads (BioRad). The respective antibodies were incubated over night with the lysate at 4°C, followed by 2 hours incubation at 4°C with blocked protein A Affiprep beads (Bio-Rad) (blocking solution: 0.1 mg/ml BSA or 0.1 mg/ml fish skin gelatine). The beads were washed with washing buffer I (20mM Tris-HCl pH 8.0, 0.15 M NaCl, 2mM EDTA, 1% TX-100, 0.1% SDS, 1mM PMSF), washing buffer II (20mM Tris-HCl pH 8.0, 0.5M NaCl, 2mM EDTA, 1% TX-100, 0.1% SDS, 1mM PMSF), washing buffer III (10mM Tris-HCl pH 8.0, 0.25 M LiCl, 1mM EDTA, 0.5% NP-40, 0.5% sodium deoxycholate) and TE-buffer (10mM Tris-HCl pH 8.0, 1mM EDTA). The beads were eluted twice (25mM Tris-HCl pH 7.5, 5mM EDTA, 0.5% SDS) for 20 minutes at 65°C. The eluates were treated with proteinase K and RNase for 1 hour at 37°C and decrosslinked at 65°C over night. The samples were further purified by phenol-chloroform extraction and ethanol-precipitated. The pellet was dissolved in 50µl TE buffer. CTCF ChIP-seq data in the HEK293T cell line was downloaded from the ENCODE consortium.

ChIP sequencing and peak detection

The ChIP DNA library was prepared according to the Illumina protocol (www.illumina.com). Briefly, 10 ng of ChIPped DNA was end-repaired, ligated to adapters, size selected on gel (200±25 bp range) and PCR amplified using Phusion polymerase as follow: 30sec at 98°C, 18 cycles of (10sec at 98°C, 30sec at 65°C, 30sec at 72°C), 5min at 72°C final extension. Cluster generation was performed using the Illumina Cluster Reagents preparation. The libraries were sequenced with the Illumina HiSeq2000 system. Read lengths of 36 bases were obtained. Images were recorded and analyzed by the Illumina Genome Analyzer Pipeline (GAP 1.6.0. and 1.7.0.). The resulting sequences were mapped against Human_UCSChg18 using the Bowtie (Langmead, Trapnell et al., 2009) alignment software, with the following parameters: -v 3 -m 1 --best --strata -S --time -p 8. Unique reads were selected for further analysis. PCR duplicate reads were removed using Picard MarkDuplicates.

ChIP-qPCR

ChIP samples (2µl) were used for a 25µl PCR reaction. Analyses by qPCR were performed using Platinum Taq and SYBR Green (Invitrogen) on an ABI 9500 cycler. The results were presented as the percentage of input-chromatin that was precipitated.

Transcript analysis by reverse transcription (RT) and qPCR

Total RNA was prepared using TRIzol Reagent (Invitrogen) according to the manufacturer's instructions. After chloroform extraction and isopropanol precipitation, pellets were dissolved in DEPC water. cDNA was generated by reverse transcription using oligo(dT)18 primer (Invitrogen), Superscript II Reverse Transcriptase (RT) (Invitrogen) and RNaseOUT Recombinant Ribonuclease Inhibitor (Invitrogen) according to the manufacturer's instructions. The DNA was then purified using PCR purification Kit (Qiagen) according to the manufacturer's instructions. The amounts of the different transcripts were compared by qPCR using SYBR Green and Platinum Taq Polymerase (Invitrogen) in CFX96 lightcycler (BioRad) and specific primers. The transcripts of the housekeeping gene SNAPIN were used for normalization of the samples. $\Delta\Delta C_t$ method was used to calculate the fold change in gene expression.

RNA-seq experiments and data analysis

Total RNA was isolated using TRIzol reagent (Invitrogen). PolyA RNA was isolated using Dynal beads mRNA purification kit (Invitrogen), and paired-end libraries were prepared as previously described (Parkhomchuk, Borodina et al., 2009).

Reads were aligned to hg18 using Tophat with the following parameters: -g 1 -p 12 --solexa1.3-quals --library-type fr-firststrand --segment-length 25 --bowtie1. A GTF file for UCSC genes was provided for the initial alignment. Wig files were generated using an in house pipeline. We normalized each wig file using trimmed mean of M normalization (Robinson and Oshlack, 2010) using Refseq exons to calculate the scaling factors between experiments.

Read counts for RefSeq genes were calculated using an in house pipeline and differentially expressed genes were called using edgeR. Common and tagwise dispersions were estimated based on all 8 RNA-seq experiments performed (2 replicates for each of 4 conditions: RAD21cv/TEV, RAD21cv/HRV, siRNA CTCF, siRNA Control). Differentially expressed genes were called by an FDR <5%.

Probe labelling for Fluorescence in situ hybridization (FISH)

Cosmids for the human chromosome 11 (G248P89139F12 and G248P85529B4) and for the human chromosome 2 (G248P85616F7 and G248P80003A6) were obtained from the CHORI library. 500ng cosmid DNA was labelled with Alexa 488-5-dUTP or Alexa 594-5-dUTP (Invitrogen) using the BioPrime Random Prime Labeling kit (Invitrogen). After labelling, probes were purified from unincorporated nucleotides using Sephadex G-50 column elution. Fractions containing the labelled cosmids DNA were pooled, ethanol precipitated and dissolved in hybridization mix containing 50% deionized formamide (Sigma), 2XSSC, 100mM phosphate buffer pH 7.5, 5x Denhardt solution, 5% dextran sulphate (Sigma) to a final concentration of 500ng/ml.

3D DNA Fluorescence in situ hybridization (3D DNA-FISH)

3D DNA-FISH was performed as described before (van de Corput, de Boer et al., 2012). In brief, cells were grown on 18mm poly-D-lysine (Sigma) coated coverslips (VWR), fixed with 2% (w/v) formaldehyde/1×PBS for 10 minutes and permeabilized in 1×PBS containing 0.5% (v/v) Triton X-100 (Sigma) and 0.5% Saponin (Sigma) for 10 minutes at room temperature. Next, cells were treated with 0.1N HCl for 5 minutes at room temperature. Cosmid probe mixes with a final concentration of 4-10ng/ml each and 50x excess of human Cot I DNA (Sigma) were added to the slides. Probes and cells were denatured simultaneously at 70°C for 2 minutes on a hot plate and hybridized overnight at 37°C in a humidified chamber. After hybridization, slides were washed with 2xSSC at 37°C for 30 minutes and one time with 2xSSC at RT for 15 minutes. Coverslips were mounted with Prolong Gold contained DAPI (Invitrogen).

Confocal Laser Scanning Microscopy

All cell samples were imaged using a Leica SP5 confocal laser scanning microscope using the LAS software provided with the instrument. The system was equipped with a 63x plan-apochromat oil NA1.4 DIC objective. The pinhole diameter was set to 1 airy unit. DAPI, Alexa 488 and Alexa 594 fluorochromes were excited with a 405nm diode laser, a 488nm Argon laser and a 594nm laser respectively and detected using a multi-track imaging mode of which the band pass filters were 410-450nm (DAPI), 505-585nm (Alexa 488) and 605-700nm (Alexa 594). 8 bit images with a 512 x 512 pixels frame size and 51x51nm pixel size were acquired with 400Hz scan speed, 2-times line averaging and an optical sectioning of 120nm. The point spread function was measured using 100nm red and green beads (Thermo Scientific) and the chromatic shift was measured using 500nm TetraSpeck beads (Invitrogen).

All confocal images were deconvolved using the Huygens Professional software v4.1.0p8 (SVI) using the measured Point Spread Function and the classical maximum likelihood-estimation algorithm. The background, signal to noise ratios and chromatic shift were corrected during the deconvolution process.

Hi-C experiments and data analysis

Hi-C experiments were performed as previously described (Lieberman-Aiden, van Berkum et al., 2009). Reads were aligned as single end reads using bwa with default parameters against hg18 reference genome. Single end reads were filtered for uniquely mapping reads and paired manually using an in house pipeline. Hi-C interaction matrices were generated as previously described (Lieberman-Aiden, van Berkum et al., 2009) and normalized using the iterative correction method either using 40kb bins or at a restriction fragment based level (for Figure 4g) (Imakaev, Fudenberg et al., 2012). To facilitate comparison of Hi-C interaction frequencies between different experiments, interaction matrices were also normalized for “depth,” with the normalized interaction frequency (I_{ij}) between two loci i and j , being normalized by the sum of all I_{ij} in a given chromosome wide normalized interaction matrix. This is analogous to read-depth based normalization schemes (i.e. RPKM) of other high-throughput sequencing

experiments. These normalized interaction matrices serve as the input for generating the directionality index and topological domain calls using previously described methods (Dixon, Selvaraj et al., 2012).

To generate the “delta” interaction matrices (**Figures 3c,d**), we subtracted the normalized interaction frequency I_{ij} at each locus of an experimental treatment (RAD21cv/HRV or siRNA CTCF) from the control treatment (RAD21cv/TEV or siRNA Control) to generate a new ΔI_{ij} for comparison between experiments.

For analysis of interactions between promoters and distal DNaseI Hypersensitive (DHS) sites, we used DHS sites from the ENCODE consortium and used the UCSC liftover tool to convert these coordinates into hg18. We identified restriction fragments containing a DHS site greater than 5kb away from any RefSeq promoter, and considered all possible interactions between these fragments and restriction fragments containing a RefSeq promoter that were within 500kb of each other. We computed the fold-change in interaction frequency between the control (RAD21cv/TEV) and RAD21 depleted (RAD21cv/HRV) samples and calculated the fraction of potential promoter-to-DHS interaction that showed a 50% gain or reduction in interaction frequency. Fisher’s exact test was used to assess the enrichment of cohesin regulated genes versus all genes for a loss or gain of promoter-to-DHS interactions.

3. Supplementary tables

Supplementary tables 1-4 are presented as excel files

FC: Fold Change

CPM: Counts per millions

FDR: False Discovery Rate

Supplementary table 1:

Significantly changed genes after RAD21 cleavage (FDR < 0.05)

chr.	start	end	strand	Gene Name	RAD21 Cleavage				CTCF depleted			
					FC	CPM	p-val	FDR	FC	CPM	p-val	FDR
chr7	142053693	142055088	+	<i>MTRNR2L6</i>	-3.001885189	3.596146636	3.17E-11	3.63E-07	-0.260664767	-0.240044182	0.629315402	1
chr17	44927653	44947381	+	<i>NGFR</i>	-1.332460534	1.374728621	7.90E-06	0.005264291	0.50693152	0.192570949	0.18376594	1
chr10	90684810	90741127	-	<i>ACTA2</i>	-1.105976433	3.095741633	1.37E-05	0.008599514	-0.222359022	1.858920898	0.438756766	1
chr17	37132416	37144424	-	<i>HAP1</i>	-1.066912668	3.592623019	4.55E-08	9.79E-05	-0.161700289	2.625497341	0.461794637	1
chr6	36752214	36763095	+	<i>CDKN1A</i>	-1.057734202	6.267310189	4.82E-11	3.63E-07	-0.223180959	4.816216738	0.168299898	1
chr13	102179717	102209423	-	<i>CCDC168</i>	-1.053282934	2.162082792	2.72E-05	0.013229621	-0.209998463	0.475209698	0.549919805	1
chr17	44022821	44038773	-	<i>HOXB-AS3</i>	-1.005391316	4.756720225	1.91E-08	6.78E-05	0.042223115	3.946689375	0.830960127	1
chr8	96007375	96030791	-	<i>TP53NP1</i>	-0.953642714	4.772480412	3.71E-06	0.002941424	-0.667379855	3.660971884	0.00079237	0.068202657
chr7	27191551	27195437	+	<i>HOXA-AS5</i>	-0.939639625	4.244340846	6.60E-07	0.000903949	-0.119891338	3.975213846	0.545692795	1
chr19	6482009	6486939	+	<i>TNFSF9</i>	-0.933731443	3.200416601	1.96E-05	0.010529801	-0.078380556	2.891999944	0.741192048	1
chr1	148088382	148088964	-	<i>HIST2H2BC</i>	-0.930196359	3.492540973	1.67E-05	0.009698014	0.004143119	2.324376129	1	1
chr19	45620248	45623772	-	<i>SERTAD1</i>	-0.929249497	3.920717993	3.71E-07	0.000579414	0.155732397	2.115802626	0.507112545	1
chr10	76831291	76838746	+	<i>ZNF503-AS2</i>	-0.866424503	5.188134577	8.12E-07	0.000940965	0.109578851	4.681740106	0.574711018	1
chr8	95961628	95976658	-	<i>CCNE2</i>	-0.851407024	5.626129819	3.21E-08	8.06E-05	-0.048945243	5.214461233	0.755163054	1
chr8	23049048	23077485	-	<i>TNFRSF10D</i>	-0.822805488	4.338144698	3.51E-05	0.015666412	-0.341632678	3.75881083	0.08547648	1
chr22	36531059	36533389	+	<i>H1FO</i>	-0.800324133	8.077945126	1.55E-06	0.001554275	0.159659919	7.689811987	0.297182829	1
chr6	27968455	27968942	-	<i>HIST1H2AM</i>	-0.771660988	3.412599467	0.000100577	0.03523192	0.138836619	3.234626437	0.587823783	1
chr1	191044791	191048030	+	<i>RGS2</i>	-0.762081879	3.844464673	5.35E-05	0.021789455	0.379712207	3.020658809	0.071932119	0.985012275
chr7	27146507	27162072	+	<i>HOXA-AS3</i>	-0.751555564	5.448153081	1.85E-07	0.000347451	0.070078895	4.956775892	0.66782584	1
chr1	148122633	148124856	-	<i>HIST2H2BE</i>	-0.728039139	5.262171119	0.000112056	0.038361325	0.325029909	4.892244097	0.16483912	1
chr16	88515917	88530006	+	<i>TUBB3</i>	-0.721007063	4.339014951	0.000133044	0.042639142	0.003316151	3.627694282	1	1
chr12	52680143	52683387	+	<i>HOXC9</i>	-0.70981464	5.430457659	8.06E-07	0.000940965	0.136377683	5.130894665	0.385238708	1
chr13	65774966	66702469	+	<i>PCDH9</i>	-0.696278647	5.3036176	5.09E-05	0.021789455	0.24375815	5.405859014	0.128919633	1
chr17	44023617	44026102	-	<i>HOXB5</i>	-0.678066117	5.10411087	2.63E-06	0.002358182	0.300926316	4.764859655	0.063056047	0.961349424
chrX	53128266	53134453	-	<i>TSPLY2</i>	-0.669300692	5.535778165	1.49E-05	0.0089777	0.13423463	5.898150297	0.390823806	1
chr6	94006458	94186021	-	<i>EPHAT7</i>	-0.652555553	4.928555035	0.000157854	0.047555016	-0.210801747	5.219287865	0.174737382	1
chr6	41148684	41178124	+	<i>NFYA</i>	-0.623657463	6.120510265	6.80E-05	0.026952827	-0.0571224	6.22124452	0.704533703	1
chr6	43651855	43696238	+	<i>POLH</i>	-0.617921211	6.266450682	9.12E-05	0.033509095	-0.098701118	5.626965954	0.524597131	1
chr1	144149818	144153985	+	<i>TXNIP</i>	-0.615854233	5.317906683	0.000121555	0.04067075	0.953629933	5.490576351	2.01E-09	1.08E-06
chr11	65021808	65030515	+	<i>MALAT1</i>	-0.610835914	9.999367932	0.000143965	0.044974258	0.154358336	9.887225409	0.339596544	1
chr10	63331018	63526713	+	<i>ARID5B</i>	-0.58896163	6.049687687	0.000146301	0.044974258	0.092820129	5.315387819	0.565461753	1
chr17	40580466	40585251	+	<i>HEXIM1</i>	-0.587442969	7.204683941	5.35E-05	0.021789455	0.115721479	6.369671754	0.463157231	1
chr17	3512935	3518722	-	<i>TAX1BP3</i>	-0.584012474	5.029029003	7.69E-05	0.029076327	-0.146654285	4.761383642	0.377025665	1
chr5	172674331	172689112	-	<i>STC2</i>	0.566974935	6.913223998	3.49E-05	0.015666412	0.264153378	6.563617654	0.087735981	1
chr17	3710365	3743086	-	<i>CAMKK1</i>	0.58037408	4.798503337	9.97E-05	0.03523192	-0.217165798	5.038283113	0.180580406	1
chr3	137538688	137953935	-	<i>STAG1</i>	0.592268644	6.42773176	0.000124202	0.04067075	1.160542278	7.649867354	1.33E-13	2.87E-10
chr5	137829079	137832903	+	<i>EGR1</i>	0.623573154	7.178224997	7.25E-06	0.005216845	-0.769177542	2.635577886	6.52E-05	0.009180417
chr8	134536272	134653365	-	<i>ST3GAL1</i>	0.665784759	4.16233909	3.29E-05	0.015497782	-0.133845916	4.422849107	0.417695867	1
chrX	107862382	107866263	-	<i>IRS4</i>	0.726757038	9.525407099	1.30E-06	0.001396489	0.039983274	9.70013011	0.800981158	1
chr8	126511744	12659826	-	<i>TRIB1</i>	0.731163548	5.124976165	1.77E-05	0.009895423	0.112639189	5.209839811	0.470004025	1
chr20	36482652	36497432	-	<i>LOC388796</i>	0.734979802	6.117374784	2.05E-05	0.010651409	0.159391522	5.242480908	0.309527315	1
chr8	128817496	128822862	+	<i>MYC</i>	0.868857148	6.888740089	2.25E-08	6.78E-05	-0.089924204	7.127191259	0.565343548	1
chr11	124010894	124051409	-	<i>SIAE</i>	1.014433973	6.919696157	3.05E-06	0.002551014	0.162075929	5.982727425	0.560600802	1
chr8	67568044	67593313	+	<i>Cbor146</i>	1.219609841	0.796674988	7.72E-05	0.029076327	0.84639459	-1.614571388	0.147179916	1
chr19	18228905	18246319	+	<i>KIAA1683</i>	1.243388536	1.989556024	2.66E-06	0.002358182	-0.381367368	-0.999944872	0.435832629	1
chr7	141342147	141453016	+	<i>MGAM</i>	1.302158211	1.092436362	2.30E-05	0.011556936	-0.15048453	-1.43575098	0.793222467	1
chr14	19881069	19881410	-	<i>RPPH1</i>	1.37098788	5.799066017	2.23E-08	6.78E-05	-0.127201192	4.822174662	0.569473772	1
chr9	138227916	138236776	-	<i>LHX3</i>	1.389571111	1.139565454	8.04E-06	0.005264291	-0.147057004	-0.334166382	0.812908108	1
chr1	109057078	109086890	+	<i>FNDCT</i>	1.430697071	1.487779142	3.85E-07	0.000579414	-0.084310221	-1.158715598	0.907554233	1
chr6	31647854	31650077	+	<i>LTA</i>	1.565977491	0.70410787	7.27E-06	0.005216845	-1.237924802	-3.533576071	0.365768361	1

Supplementary table 2:

Significantly changed genes after CTCF depletion (FDR < 0.05)

chr.	start	end	strand	Gene Name	RAD21 Cleavage				CTCF depleted			
					FC	CPM	p-val	FDR	FC	CPM	p-val	FDR
chr20	44235782	44313741	-	CDH22	0.287052228	2.139319467	0.25878055	1	-2.104646401	-0.040501259	1.45E-07	5.32E-05
chr17	43239731	43254146	-	OSBP1L	0.019973674	2.620075753	0.851724191	1	-1.916062813	0.965818629	8.26E-09	4.01E-06
chr2	70742723	70848883	+	ADD2	0.10696415	5.717907945	0.539120212	1	-1.867310472	5.283533892	3.72E-30	2.81E-26
chr11	63864265	63881582	+	CCDC88B	-0.115969195	1.573594497	0.942063359	1	-1.711427495	0.862346784	7.78E-07	0.000229766
chr1	16256850	16272714	+	FAM131C	0.438309366	2.067954813	0.066870812	1	-1.661430088	1.562561838	3.25E-08	1.40E-05
chr19	51059357	51068895	+	FOXA3	0.010408809	0.290351106	0.843946989	1	-1.633309178	-0.54407043	0.000329562	0.033769972
chr16	18902756	18982763	+	TMC7	0.015303495	3.588202402	0.934655531	1	-1.556498489	2.782003459	1.56E-12	1.56E-09
chr2	29057667	29128600	+	FAM179A	0.548611012	0.759603845	0.057779221	1	-1.545373503	-0.476365303	0.000362358	0.036632217
chr19	45664965	45774205	+	SPTBN4	0.097851786	2.430478667	0.593834884	1	-1.485107849	1.266450549	1.19E-06	0.000325765
chr12	56305817	56313252	-	B4GALNT1	-0.092936186	3.895398765	0.866674831	1	-1.458843301	3.458807614	2.44E-12	2.30E-09
chr12	54361596	54364661	+	METTL7B	-0.392425761	2.599378846	0.225696823	1	-1.420280169	2.259017043	2.52E-07	8.64E-05
chr12	56406289	56408406	+	LOC100130776	0.073189581	3.02736681	0.742070506	1	-1.385444511	2.376124893	3.48E-08	1.46E-05
chrX	76986517	77037721	-	MAGT1	-0.16086069	6.984624242	0.334732982	1	-1.361929317	7.226839394	3.89E-18	1.47E-14
chr22	18835993	18841786	-	RBMBP3	0.013446728	2.508685216	0.656708488	1	-1.349317634	1.77857075	2.41E-06	0.000625215
chr1	27591738	27594904	+	GPR3	0.546440381	1.825786873	0.043537373	1	-1.342543649	0.236843125	0.00020317	0.023010172
chr4	15080586	15212278	+	CC2D2A	-0.1007299	3.376078661	0.559661652	1	-1.298050902	3.343647967	4.64E-10	3.18E-07
chr22	34107059	34120207	+	HMOX1	-0.456991024	4.268754105	0.030993495	0.894213246	-1.292483511	3.841521043	1.66E-10	1.19E-07
chr15	71996861	72007642	-	LOC100287616	0.320017855	2.145175987	0.133387357	1	-1.249295464	2.320733319	4.95E-06	0.001064447
chr11	118544649	118559936	+	NLRX1	-0.042872259	4.627369647	0.80075582	1	-1.232813517	3.993944453	5.74E-13	9.60E-10
chr11	298106	299410	+	IFITM2	-0.104907496	2.742277429	0.707574723	1	-1.205662748	2.457818021	1.12E-06	0.000312481
chr12	55678883	55686564	-	ZBTB39	-0.024469128	5.516374041	0.79653164	1	-1.20474427	4.853465067	1.69E-13	3.17E-10
chr12	56404342	56422211	-	AGAP2	-0.113638894	3.788678709	0.902685942	1	-1.181078406	3.272281681	4.17E-07	0.000133359
chr7	37926687	37958067	+	EPDR1	-0.001154875	3.857316371	0.873572466	1	-1.152130503	3.377704801	2.41E-08	1.07E-05
chr10	123706592	123724733	+	NSMCEA2	0.158712677	5.946472665	0.350812881	1	-1.14295931	5.282412063	1.05E-12	1.42E-09
chr19	11392271	11406980	-	CCDC151	0.021069757	1.285355455	0.915521235	1	-1.13286405	0.897933739	0.000315339	0.032533895
chr1	152168600	152185778	-	DENN4D4	0.194384223	6.515528444	0.251920232	1	-1.126007147	5.930148611	1.18E-12	1.42E-09
chr20	61589809	61600949	+	EEF1A2	-0.050951731	7.083428838	0.963167024	1	-1.12607215	5.986388939	1.32E-12	1.42E-09
chr11	125638023	125644087	+	SRPR	-0.092455657	7.027447611	0.480514641	1	-1.118771497	6.285435775	1.23E-12	1.42E-09
chr9	129968164	130006483	+	CIZ1	-0.026122642	7.952762208	0.770980864	1	-1.108638487	7.382652073	6.14E-12	1.42E-09
chr16	29789980	29818086	-	SEZBL2	0.040319609	3.971662593	0.939845807	1	-1.10370632	3.572079999	6.14E-08	2.37E-05
chr16	66153810	66230589	+	DTKN	0.175387153	6.938425397	0.318715866	1	-1.090678587	6.50276881	3.77E-12	3.34E-09
chr19	40679958	40696400	-	CMFC	-0.101954575	6.309359926	0.453433863	1	-1.073689118	5.939081895	1.25E-11	1.04E-08
chr17	35770430	35774471	+	GJD3	0.181528972	2.228673511	0.326657924	1	-1.067431989	1.790465425	0.000156713	0.019191649
chr7	73848419	73905777	-	GTFRIRD2	0.12850774	3.205586377	0.599100052	1	-1.064871324	1.981721786	3.28E-05	0.005149687
chr22	34952200	34965946	-	APOL2	-0.382766009	3.524574622	0.407664823	1	-1.044895361	3.085940175	5.81E-07	0.000178587
chr1	149235924	149247478	-	FAM63A	-0.252391733	4.589517175	0.136140462	1	-1.022211305	3.982886476	1.56E-09	9.04E-07
chr5	35653745	35850470	+	SPEF2	0.118545814	2.540699334	0.609592228	1	-1.014157365	1.799499414	0.00011072	0.014502337
chr11	61323672	61341105	+	FADS1	-0.017071153	6.362064775	0.909747275	1	-1.014149879	6.114814041	1.23E-10	9.26E-08
chr14	93663870	93665710	-	IF2L2	-0.1049038	2.197466761	0.762487186	1	-1.012954763	1.881072312	0.000332442	0.03834989
chr2	74578151	74583951	-	LBX2	0.004435649	2.681038448	0.938774727	1	-0.993172387	2.102129426	9.57E-05	0.01267486
chr16	54070581	54098087	+	MM2P	-0.310890855	1.973875964	0.334797105	1	-0.988756022	2.546439083	0.00020002	0.02299927
chr11	780474	788269	+	SLC25A22	0.248224584	5.416236545	0.098588604	1	-0.981132989	5.211525474	1.03E-09	6.48E-07
chr7	904062	960815	-	ADAP1	0.139932639	3.677413001	0.371184575	1	-0.975255104	3.168220761	2.67E-06	0.000660551
chr7	36859035	37455420	-	ELMO1	-0.348196835	2.44034968	0.194032123	1	-0.971633544	2.15463096	0.000106886	0.014248018
chr14	75406357	75685553	-	TCFV1	-0.42266517	5.549284904	0.007357548	0.518114086	-0.966071589	5.074474828	1.38E-09	8.32E-07
chr4	39734918	39836267	+	M4BP2	0.029824131	5.635937267	0.58674151	1	-0.96478601	5.086306546	1.89E-09	1.05E-06
chr6	100075590	100087780	+	LOC100130890	-0.080190581	3.050672191	0.729349379	1	-0.954419658	2.545286266	8.02E-05	0.01098033
chr1	23627642	23683337	+	ASAP3	-0.170350422	3.620009276	0.644742927	1	-0.942397163	3.266910921	4.85E-05	0.0071622
chr5	101597590	101660152	-	SLC04C1	-0.281167311	2.675858828	0.306789008	1	-0.931335053	2.19847122	0.000197538	0.022888557
chr2	68364806	68400687	+	CNRP1	0.068265327	3.995133804	0.767333065	1	-0.927559892	3.584428053	5.24E-06	0.00195454
chr11	61276696	61312565	+	C11orf9	-0.191822478	3.672574115	0.409885445	1	-0.912407923	2.797719842	2.08E-05	0.003482653
chr2	241156676	241166822	+	RNPEFL1	0.061558633	4.301421716	0.68841807	1	-0.904571671	4.324625123	5.66E-08	2.24E-05
chr19	46528651	46551671	-	TGFB1	-0.113806109	5.960018035	0.326556818	1	-0.90154991	5.331776429	1.78E-08	8.29E-06
chr5	179961111	180009230	+	FLT4	-0.116266138	5.893422227	0.605097621	1	-0.901186097	5.311181808	1.82E-08	8.29E-06
chr18	50139168	50162240	+	C18orf54	0.025159156	4.458142095	0.847162179	1	-0.88878786	4.093066341	1.04E-07	3.92E-05
chr3	27732889	27738789	-	EOMES	-0.040469904	3.122600739	0.856314147	1	-0.877197748	2.435237048	0.000390965	0.038744071
chr19	18560494	18564147	+	C19orf60	0.044615866	4.128275182	0.946756918	1	-0.864935663	4.015185656	1.31E-05	0.002325127
chr10	69712422	69762690	-	PBLD	-0.027481038	3.757081795	0.780614866	1	-0.858657034	2.851203368	5.20E-05	0.007608371
chr12	129840099	129889772	-	STX2	-0.056365427	5.008409195	0.819755266	1	-0.838449802	4.965578852	1.88E-07	6.75E-05
chr3	101694152	101778975	+	TMEM45A	-0.020724893	3.198927561	0.868790958	1	-0.835119739	2.888905475	0.00045206	0.043649863
chr12	3594754	3732627	+	EFCA84B	0.228882187	2.581056792	0.242180891	1	-0.829464319	2.751035146	0.000515474	0.047929523
chr4	12978444	13095087	-	RAB28	0.181150315	4.526748741	0.216936806	1	-0.823844733	4.074245014	8.50E-07	0.000246107
chrX	151746526	151749957	+	CNTN2	-0.062714149	5.025510307	0.81958119	1	-0.809344763	4.69704168	6.50E-07	0.000195738
chr16	88542639	88561969	+	DEF8	0.010574611	6.361494058	0.925900097	1	-0.805061229	5.549160997	3.83E-07	0.000125323
chr10	69773280	69837057	+	RNFY1	0.085606948	5.464483745	0.532665825	1	-0.8037979781	4.944250343	5.45E-07	0.000171051
chr7	26298039	26380474</										

chr9	130504622	130523020 +	PKN3	-0.050145543	6.104787723	0.78594556	1	-0.744689036	5.631911081	2.69E-06	0.000660551
chr6	127801243	127822228 -	KIAA0408	-0.045990323	5.414041402	0.69316684	1	-0.73929192	4.9320291	4.11E-06	0.00091094
chr15	41597097	41611110 +	MAP1A	-0.414412036	4.773985267	0.033382953	0.9226555828	-0.713686004	5.545881269	3.48E-06	0.000805979
chr10	115428924	115480654 +	CASP7	-0.251623972	4.371058179	0.15445097	1	-0.726067334	4.35788515	1.01E-05	0.001860686
chr3	16273571	16281500 -	DPH3	-0.099268373	5.909386544	0.655303463	1	-0.724737218	5.508463663	4.71E-06	0.001027375
chrX	128501932	128554211 +	OCRL	0.018153527	5.898316979	0.854443216	1	-0.721524272	6.217581592	3.74E-06	0.00085448
chr17	24424653	24531533 -	MYO1B	0.069604245	5.248247823	0.559145927	1	-0.71691285	4.931792546	8.63E-06	0.001631656
chr16	19777793	19803652 -	GPRC5B	-0.326034933	3.826073177	0.103903133	1	-0.712786833	3.795946559	0.000365975	0.036751225
chr7	47581276	47588267 -	TMS3	-0.142722062	3.232344983	0.574054579	1	-0.708623298	3.464902687	0.000483508	0.045796664
chr6	26646550	26655143 +	HMGN4	0.131473179	5.595611497	0.259570653	1	-0.705375218	5.451325597	8.67E-06	0.001631866
chr11	117212900	117253411 +	FXYD6	-0.261395254	4.190138959	0.265171854	1	-0.697457581	3.847359942	0.00051292	0.047929523
chr3	47819402	47866690 +	DHX30	0.149240647	7.981281775	0.528415504	1	-0.693024955	7.764988378	7.58E-06	0.001482345
chr19	11170968	11234168 -	DOCK6	0.182845683	5.177773801	0.242309016	1	-0.689407874	4.592228507	2.07E-05	0.003482653
chr19	54160377	54161948 +	FTL	-0.268331279	8.560057468	0.123367601	1	-0.683759845	8.175721272	9.69E-06	0.001802449
chr2	121271336	121466699 +	GLI2	0.056780575	4.685124862	0.259543637	1	-0.679605688	4.637123913	3.09E-05	0.004893743
chr12	51577237	51629917 -	KRTL8	-0.040923052	3.500015693	0.905681424	1	-0.678351493	4.707195929	0.000443595	0.004110874
chr1	149298774	149307597 +	MLL11	-0.329158908	4.759894898	0.062359053	1	-0.67467216	4.463364608	3.80E-05	0.005841918
chr15	38973919	38983465 +	VPS18	-0.263498895	5.011571031	0.101192209	1	-0.668680271	4.609134071	3.94E-05	0.005999338
chr9	103277428	103289296 -	C9orf125	0.013927128	4.075866974	0.754991629	1	-0.668155657	4.535141187	4.84E-05	0.0071622
chr14	64523259	64599123 +	FNTB	-0.181547066	4.663800152	0.185990919	1	-0.662597897	4.314203818	6.40E-05	0.00909512
chr21	37045058	37284415 -	HLCS	0.169569654	5.286538739	0.275566206	1	-0.654792451	5.286197949	4.06E-05	0.006109363
chr1	152564659	152590404 +	RAB22	-0.038064937	6.785081278	0.983161448	1	-0.623381054	6.847038194	6.06E-05	0.00688732
chr15	61268780	61347026 +	ATBBB	-0.121496759	4.939662237	0.559328432	1	-0.620010877	4.226287745	0.000175739	0.021009163
chr1	40278841	40310908 +	CAP1	-0.203911175	7.826450298	0.179839122	1	-0.613025455	7.29464485	7.45E-05	0.010395458
chr1	167603817	167632404 +	BZLF1	-0.140111404	5.007835446	0.399790981	1	-0.61287194	4.359459667	0.000175047	0.021009163
chr2	74216035	74228547 -	BOLA3	0.027445642	5.891300044	0.98640676	1	-0.607408867	5.345072197	0.000153288	0.018926041
chr3	49032911	49035930 +	NDUFAF3	-0.106066931	6.152654659	0.460567832	1	-0.602518115	6.164675645	0.000113683	0.014762147
chr22	37590193	37598204 -	CBX6	0.149506273	7.678037356	0.485150977	1	-0.600886751	7.051827726	0.000110084	0.014502337
chr11	44073674	44223556 -	EXT2	-0.182271369	5.998134574	0.285211538	1	-0.599991528	5.709041255	0.000148354	0.018520341
chr17	148759386	148789896 -	ZNF777	0.297961145	5.56310396	0.107119837	1	-0.596888223	4.896989269	0.000202185	0.002301072
chr1	170017383	170033479 +	METTL13	0.131212672	5.575550804	0.472744273	1	-0.592348784	5.47886018	0.000194404	0.022700056
chr1	199755654	199756343 +	RPS107	-0.136863287	4.749446947	0.529930401	1	-0.589034604	4.542592278	0.000304512	0.031853233
chr20	32896183	32924232 -	GGT7	-0.221444106	5.489040162	0.204606667	1	-0.585046962	4.80346689	0.00029501	0.031293872
chr14	59020913	59041834 +	KIAMP	-0.04686861	5.568725806	0.868682986	1	-0.582543127	5.061122844	0.000265982	0.028823649
chr5	61744329	61960172 +	IPK1	0.050214029	6.635933029	0.676655688	1	-0.581247299	6.299007702	0.000190595	0.02260578
chr5	133765654	133775497 -	CDKN2AIPNL	0.050623445	5.551046787	0.829911247	1	-0.578981051	5.297820246	0.000278774	0.020978133
chr19	6445329	6453330 +	TUBBA4	-0.209114217	5.162882945	0.198649909	1	-0.578410548	5.221897683	0.000302894	0.031853233
chr5	76047623	76067351 +	F2R	0.067895148	4.512794413	0.425663441	1	-0.574676295	5.008410346	0.000311933	0.032404421
chr6	133132199	133161440 -	C6orf12	0.067163856	5.051658819	0.53739283	1	-0.565394393	5.222083502	0.000414305	0.040523839
chr18	42007985	42100953 +	UHRF125	0.012040549	6.150258783	0.986750513	1	-0.551402828	5.465046779	0.000486479	0.045796884
chr1	160733587	160766043 +	CHMK1	0.042917194	7.779650008	0.752058466	1	-0.544396575	7.420805847	0.000491959	0.040282105
chr20	18516555	18692560 +	DTD1	-0.087277202	5.037427559	0.803345681	1	0.551489312	5.55614381	0.000465384	0.043467979
chr14	67125775	67136770 -	PIGH	-0.008477136	5.198221455	0.927156569	1	0.558964053	5.07013721	0.000458689	0.040407866
chr6	21701950	21706828 +	SOX4	-0.035772443	6.6543047	0.624384941	1	0.582829753	7.489964656	0.000167502	0.020347484
chr6	139735089	139737478 -	CITED2	-0.140856125	4.996716338	0.400843343	1	0.590765324	5.245035182	0.000217964	0.02414147
chr12	122311492	122322640 -	CDK2AP1	0.048942456	6.001220668	0.826706852	1	0.595064922	7.1990635	0.000121024	0.055043137
chr6	121798443	121812572 +	GJA1	-0.073876341	5.230736287	0.999014478	1	0.620716419	6.015474938	3.92E-05	0.0109455
chr12	64816983	64850074 -	BTM1M	0.051793189	4.847661229	0.633627734	1	0.662089274	5.134002023	3.54E-05	0.005496704
chr4	55907144	55934023 +	SRD5A3	0.152221627	5.059577753	0.255691014	1	0.667335518	5.692712242	2.38E-05	0.003891701
chr12	43188324	43593978 -	NELL2	-0.047611549	4.706471495	0.925965029	1	0.674605387	4.949613237	2.80E-05	0.004534574
chr16	53510278	53520580 -	CRND	0.270169019	4.413640914	0.07445282	1	0.675480734	4.767376093	2.84E-05	0.004556489
chr16	65195435	65205296 +	CMT3M	-0.078579979	4.396795818	0.69874618	1	0.682390542	5.053811757	1.98E-05	0.003393812
chr5	52380633	52584946 +	FYVE9	0.101968888	5.156004581	0.35388244	1	0.69913071	5.59500728	1.04E-05	0.001888798
chr8	144878090	144887902 -	FAM83H	0.17192707	4.387782648	0.313603017	1	0.700428961	4.16270336	0.000371558	0.03706471
chr14	35077308	35348183 -	RALGAP1	-0.013372653	6.377334451	0.967388355	1	0.700905212	7.080850863	6.35E-06	0.001274489
chr6	126319553	126343082 -	HINT3	0.121662408	5.941462815	0.349463433	1	0.701232894	6.024689631	8.15E-06	0.001573073
chr10	73703682	73705803 +	DDIT4	0.146999442	5.418488527	0.160458614	1	0.702633559	5.702617867	0.000193019	0.022700056
chr3	195336624	195339095 +	HES1	-0.195725839	4.49662823	0.205992814	1	0.716077175	4.545480332	1.20E-05	0.002145416
chr5	121426788	121441954 -	LOX	-0.092263625	4.64577154	0.886132282	1	0.723749475	5.616187488	5.05E-06	0.00172485
chr5	130627600	130758281 +	CDCA2S2E	0.053231563	5.385869669	0.687409398	1	0.725324026	5.256225328	5.65E-06	0.001157096
chr1	210805298	210860742 +	ATF3	0.17191547	6.685314539	0.001271559	0.18289697	0.72748061	4.232722277	0.000215515	0.02404671
chr2	190319630	190336169 -	OSGEPL1	0.143515276	3.474197973	0.356289238	1	0.729906524	4.113717186	0.000206492	0.023211817
chr17	71892284	71895536 +	SPHK1	0.185693805	3.846901191	0.296949341	1	0.733505427	3.662703876	0.00022856	0.024502791
chr3	198146669	198153861 -	NCBP2	0.149463722	5.808439727	0.372105964	1	0.737562698	6.297139049	2.54E-06	0.000648191
chr14	49161641	49171698 -	DNAAF2	0.271940938	4.476029584	0.063718267	1	0.744248597	5.00255538	3.86E-06	0.000867737
chr19	47509316	47521054 +	TMEM145	0.261758204	3.422392336	0.226719949	1	0.75259587	3.274754546	0.000246287	0.026882785
chr2	217245072	217268517 -	IGFBP5	-0.099413094	3.766774211	0.817277777	1	0.816979613	5.559132955	2.67E-07	8.92E-05
chr16	391858	402488 +	DEC2R	0.221888791	3.186913264	0.225199437	1	0.85142751	3.556603288	2.31E-05	0.00381994
chr11	75157425	75190229 -	DGAT2	-0.102752135	2.564964104	0.754238298	1	0.878941775	2.611773383	0.000269717	0.029019654
chr11	106049947	106394381 -	GUCY1A2	0.65494924	-0.24515094	0.018396439	0.735715694	0.882716151	4.293321974	7.27E-06	0.001441471
chr8	19215360	19299809 +	SH2DA	0.130539753	2.188196686	0.501914762	1	0.896244757	3.004849899	0.000142862	0.018083448
chr1	144149818	144153985 +	TNXP1	-0.615854233	5.317906683	0.000121555	0.04067075	0.953629933	5.490576351	2.01E-09	1.08E-06
chr4	970784	988317 +	IDUA	0.004587952	1.952145038	0.939385261	1	0.954863284	2.565381927	8.67E-05	0.011767874
chr5	76408287	76418786 -	ZBED3								

chr17	6858779	6861567 +	C17orf49	0.126880876	4.463302612	0.489014468	1	1.318307948	4.934123173	5.81E-16	1.75E-12
chr19	45045802	45132373 -	FCGBP	-0.270644414	2.040338183	0.285024971	1	1.461022958	2.695268206	2.68E-09	1.34E-06
chr12	119123476	119135014 +	LOC100506649	-0.028581398	1.936031421	0.991248638	1	1.494337232	2.876909718	7.12E-10	4.66E-07
chr11	111249989	111255391 -	FDXACB1	0.330430642	3.066709528	0.102570831	1	1.51015353	4.316248054	4.14E-14	1.04E-10
chr15	39032927	39036009 +	CHAC1	0.37398851	2.99227481	0.047083288	1	1.585617077	2.471695255	5.24E-08	2.13E-05
chr7	141050606	141084499 -	FLJ40852	-0.062937753	0.691413102	0.902507157	1	1.685541742	0.971100482	9.75E-07	0.00027719
chr4	146759989	146800637 +	MMAA	0.120184832	1.19175043	0.629015322	1	1.791246419	2.525937822	8.44E-11	6.69E-08
chr16	29782504	29786875 +	LOC440356	0.37567817	-0.177403112	0.265962513	1	2.113020711	0.970916243	2.67E-09	1.34E-06
chr6	132000134	132110243 +	ENPP3	0.298202083	0.296262085	0.208687767	1	3.860050305	2.324064363	1.03E-28	5.17E-25
chrX	151833641	151892681 +	ZNF185	0.338869437	1.405751976	0.121427222	1	4.023726175	5.031710915	1.29E-72	1.94E-68

Supplementary table 3: Significantly changed genes after RAD21 cleavage (FDR < 0.2)

chr.	start	end	strand	Gene Name	RAD21 Cleavage			CTCF depleted				
					FC	CPM	p-val	FDR	FC	CPM	p-val	FDR
chr7	142053693	142055088	+	MTRNR2L6	-3.001885189	3.596146636	3.17E-11	3.63E-07	-0.260664767	-0.240044182	0.629315402	1
chr17	44927653	44947381	+	NGFR	-1.332460534	1.374728623	7.90E-06	0.005264291	0.50693152	0.192570949	0.18376594	1
chr10	9068410	90741127	-	ACT2	-1.05976433	3.095741332	1.37E-05	0.008599514	-0.222359022	1.858920898	0.438756766	1
chr17	37132416	37144424	-	HAP1	-1.066912668	3.592623019	4.55E-08	9.79E-05	-0.161700289	2.625497341	0.461794637	1
chr6	36752214	36763095	+	CDKN1A	-1.057734202	6.267310189	4.82E-11	3.63E-07	-0.223180959	4.816216738	0.168299898	1
chr19	18357967	18368086	+	GDF15	-1.054100385	1.60491729	0.000952804	0.157715221	-0.449648611	-0.707031304	0.357929089	1
chr13	102179717	102209423	+	CCDC168	-1.053282934	2.162082792	2.72E-05	0.13229621	-0.209998463	0.475209664	0.549919805	1
chr11	64968504	64968804	+	MIR612	-1.007751009	0.482153697	0.001106845	0.171095159	0.258615979	0.300421964	0.493977567	1
chr17	44022821	44038773	+	HOXB-AS3	-1.005391316	4.756720225	1.91E-08	6.78E-05	0.042223115	3.946689375	0.830960127	1
chr6	11291516	11490567	-	NEDD9	-0.978087774	0.829756181	0.001066992	0.16917999	-0.30400911	0.564428127	0.401067497	1
chr8	96007375	96030791	-	TP53NP1	-0.953642714	4.772480412	3.71E-06	0.02941424	-0.667379855	3.660971884	0.00079237	0.068202657
chr7	27191551	27195437	+	HOXA-AS5	-0.939639625	4.244340846	6.60E-07	0.000903949	-0.119891338	3.975213846	0.545692795	1
chr19	6482009	6486939	+	TNFSF9	-0.876016156	3.200416601	1.96E-05	0.10529801	0.58360251	3.307197502	0.011313355	0.426032668
chr1	148088382	148089864	-	HIST2H2BC	-0.930196359	3.492540973	1.67E-05	0.009698014	0.004143119	2.324376129	1	1
chr19	45620248	45623772	-	SERTAD1	-0.929249497	3.920717993	3.71E-07	0.000579414	0.155732397	2.115802626	0.507112504	1
chr6	26342418	26343195	-	HIST1H1D	-0.927140345	0.916444653	0.001311351	0.18289697	0.376958919	2.141143625	0.224068144	1
chr19	4473543	4486208	-	PLIN5	-0.926208741	0.99437813	0.001416334	0.189560422	0.027952879	-0.183465739	0.941607475	1
chr9	106583104	106730348	-	ABC1	-0.933731443	3.200416601	0.0005514	0.118300818	0.58360251	3.307197502	0.011313355	0.426032668
chr10	76831291	76838746	+	ZNFS03-AS2	-0.866424503	5.188134577	8.12E-07	0.000940965	-0.109578851	4.681740106	0.574711018	1
chr8	95961628	95976658	-	CCNE2	-0.851407024	5.626129819	3.21E-08	8.06E-05	-0.048945243	5.214461233	0.755163054	1
chr19	19398924	19420500	+	ACER2	-0.84921886	3.28412685	0.000179877	0.053127155	0.057131788	2.573644382	0.819866354	1
chr8	23049048	23077485	-	TNFRSF10D	-0.822400548	4.338144698	3.51E-05	0.015566412	-0.341632678	3.75881083	0.08547648	1
chr17	30594198	30618874	+	SLFN5	-0.817744305	2.738685214	0.000467227	0.106638374	-0.417200858	2.421517094	0.086066571	1
chr22	36531059	36533389	+	H1FO	-0.800324133	0.877945126	1.55E-06	0.001554275	0.159659919	7.689811987	0.297182829	1
chr6	27968455	27968942	-	HIST1H2AM	-0.771660988	3.412599467	0.000100577	0.035232192	0.138836819	3.234622947	0.587823783	1
chr5	57785566	57791723	-	PLK2	-0.769562606	3.769586071	0.000972475	0.159221711	-0.098506337	2.472782805	0.695631419	1
chr1	191044791	191048030	+	RGS2	-0.762081879	3.844464673	5.35E-05	0.021789455	0.379712207	3.020658905	0.071932119	0.985012275
chr6	26231673	26232111	+	HIST1H2BC	-0.756517247	3.557840475	0.000790516	0.136868249	0.216994658	3.382161045	0.434452952	1
chr3	168936125	169026051	+	SERPIN1	-0.752725743	3.947070414	0.000329119	0.082625371	0.026850303	2.998816236	0.592824606	1
chr7	27146507	27162072	+	HOXA-AS3	-0.751555564	5.448153081	1.85E-07	0.000347451	0.067707895	4.965775892	0.66782584	1
chr1	148122633	148124856	-	HIST2H2BE	-0.728039139	5.26217119	0.000112056	0.038361325	0.325029909	4.892244097	0.16483912	1
chr16	88515917	88530006	+	TUBB3	-0.721070063	4.339014951	0.000133044	0.042639142	0.003316151	3.627694822	1	1
chr12	52680143	52683387	+	HOXC9	-0.70981464	5.430457659	8.06E-07	0.000940965	0.136377683	5.130894865	0.385238708	1
chr7	64088167	64104559	-	ERV3-1	-0.704771337	3.646883724	0.000423597	0.09969744	-0.184264113	3.431764946	0.359854662	1
chr9	98831779	98841746	+	CTSL2	-0.700219138	4.669374314	0.000632118	0.126102227	0.169433535	4.153275707	0.387611164	1
chr17	70370213	70380751	-	FDXR	-0.700126301	4.238070595	0.000557615	0.118300818	-0.345780429	3.35993722	0.097824619	1
chr13	65774966	66020469	-	PCHD9	-0.696276647	5.3036176	5.09E-05	0.021789455	0.24375815	5.405889014	0.128919633	1
chr17	44023617	44026128	+	HOXB5	-0.678066117	5.10411087	2.63E-06	0.002358182	0.300926516	4.764859665	0.063050647	0.961349424
chrX	53128266	53134453	+	TPSYL2	-0.669300692	5.535778165	1.49E-05	0.0089777	0.13423463	5.898150297	0.390823806	1
chr6	94006458	94186021	-	EPHA7	-0.652555553	4.928555035	0.000157854	0.047555016	-0.210801747	5.219287865	0.177437382	1
chr16	30011135	30015038	-	YPEL3	-0.646197016	3.807047026	0.001294092	0.18289697	-0.243468091	2.946030854	0.250174583	1
chr6	41148684	41178124	+	NFYA	-0.623657463	6.120510265	6.80E-05	0.026952827	-0.0571224	6.22124452	0.704533703	1
chr12	64416664	6450104	-	VAMP1	-0.621902088	3.675515031	0.000735753	0.133525928	0.111557116	3.959192273	0.579717313	1
chr6	43651855	43669238	+	POLH	-0.617921211	6.266450682	9.12E-05	0.033509095	-0.098701118	5.626965954	0.524597131	1
chr1	144149818	144153985	+	TNXP1	-0.61584233	5.317906683	0.000121555	0.04067075	0.953629933	5.490576351	2.01E-09	1.08E-06
chr11	65021808	65030515	+	MALAT1	-0.610835914	9.999367932	0.000143965	0.044974258	0.154358336	9.887225409	0.339596544	1
chr4	78297380	78310237	+	CCNG2	-0.606402376	4.480197338	0.000284277	0.073828773	0.194039752	4.815114655	0.235443823	1
chr4	6627802	6670589	+	MAN2B2	-0.604794546	4.233238672	0.001341913	0.185442474	-0.179787525	3.907368016	0.376368336	1
chr3	180217706	180272350	-	ZMAT3	-0.600908384	5.535934954	0.000239171	0.066715519	-0.218803881	5.453358917	0.165120436	1
chr4	30331127	30375521	+	PCDH7	-0.599322969	5.037224437	0.000691478	0.131156512	0.08618576	5.324976876	0.593999774	1
chr17	44007867	44010742	+	HOXB4	-0.594515854	4.818449417	0.000275245	0.07339944	-0.002657855	4.153162418	0.989436697	1
chr10	63331018	63526713	+	ARID5B	-0.589896163	6.049687687	0.000146301	0.044974258	0.092820129	5.315387819	0.565461753	1
chr17	40580466	40585251	+	HEXIM1	-0.587442969	7.204683941	5.35E-05	0.021789455	0.115721479	6.369671754	0.463157231	1
chr17	3512935	3518722	-	TAX1BP3	-0.584012474	5.029029003	7.69E-05	0.029076327	-0.146654285	4.761383642	0.377025665	1
chr12	67488237	67525479	+	MDM2	-0.57485266	8.615474247	0.000419666	0.09969744	-0.003775538	7.962493061	0.971245761	1
chr12	7174233	7202797	+	CLSTN3	-0.5738985	4.816141079	0.000755141	0.134825171	-0.244113188	4.905444138	0.129471449	1
chr9	73488101	73573620	-	TMEM2	-0.567509143	5.097182815	0.000656595	0.12844541	-0.073562437	5.429299752	0.639127155	1
chr20	34953640	35013660	-	SAMHD1	-0.565337929	5.868479236	0.000411097	0.09969744	-0.123901045	6.032346859	0.417399738	1
chr11	64635916	64640283	+	TMSF2	-0.562614058	5.122134783	0.000705283	0.131156512	-0.23619539	4.569705399	0.143643734	1
chr7	64072264	64088849	-	ZNFI17	-0.557314826	4.835671666	0.000697153	0.131156512	-0.090398376	4.553054227	0.570354634	1
chr14	22664343	22722689	+	SLC7A8	-0.554288335	4.427814275	0.000918181	0.153672883	0.164323599	4.469832459	0.31945121	1
chr17	34279637	34331549	+	LASP1	-0.539820267	5.220619164	0.00141535	0.189560422	-0.00580424	4.941934522	0.989920704	1
chr7	27159862	27162821	-	HOXA7	-0.535880797	4.34754186	0.001422049	0.189560422	-0.014521075	4.138872777	0.960450973	1
chr21	45318920	45479096	+	ADARB1	-0.534961678	5.172945442	0.00044616	0.103034714	-0.162251044	5.558483522	0.302868196	1
chr8	22933592	22982645	-	TNFRSF10B	-0.524315062	6.420891016	0.000998824	0.160545397	0.098737617	6.169029038	0.53597467	1
chr20	32755808	32764898	+	TP53NP2	-0.515654426	4.838656089	0.000778174	0.13629816	-0.284662605	4.367534783	0.084818484	1
chr15	61232591	61236794	-	RPS2L	-0.513862768	5.488526829	0.001309032	0.18289697	0.032052703	4.981028732	0.87434548	1
chr7	78243223	78255583	+	DNAI1B4	-0.512665727	5.054246595	0.001385369	0.189560422	-0.149214144	4.126238978	0.370610189	1
chr7	27176734	27186405	-	HOXA10	-0.498943851	7.2268907	0.000690644	0.131156512	0.325538297	6.7275917	0.034679073	0.783753711
chr7	27187300	27191360	-	HOXA11	-0.476878097	5.509762322	0.000909087	0.153672883	0.063050414	5.263723218	0.701998159	1
chr5	13863739	13864672	+									

chr1	96959762	97053193 +	PTBP2	0.540711924	6.073547118	0.000760814	0.134825171	0.221118111	6.024816145	0.180611699	1
chr7	148590194	148613018 +	ZNF783	0.558365923	5.333199372	0.000629247	0.126102227	0.012374602	5.396288559	0.924133146	1
chr13	109204184	109236915 -	IRS2	0.560224613	5.056153134	0.000277748	0.07339844	0.055441272	4.75967174	0.714592094	1
chr5	172674331	172689112 -	STC2	0.566974935	6.913223998	3.49E-05	0.015566412	0.264153378	6.563617654	0.087735981	1
chr6	31910671	31915520 +	C6orf148	0.57095497	5.861490194	0.000581374	0.121628314	0.393971611	5.314542136	0.01307859	0.463535999
chr16	4247187	4263002 -	TFAP4	0.573792799	4.717927843	0.000226483	0.064368141	0.25928592	5.138233993	0.10326271	1
chr17	3710365	3743086 -	CAMKK1	0.58037408	4.798503337	9.97E-05	0.035232192	-0.217165798	5.038283113	0.180580406	1
chr11	5210634	5212434 -	HBD	0.587599692	7.294474252	0.000635853	0.126102227	-2.356581317	-6.858657874	1	1
chr3	137538688	137953935 -	STAG1	0.595262864	6.42773176	0.000124202	0.04067075	1.160542278	6.749867354	1.33E-13	2.87E-10
chr14	20738077	20744699 -	LOC283624	0.607579811	6.917018994	0.000400639	0.098931456	-0.072061431	6.658168752	0.636769289	1
chr5	137829079	137832903 +	EGR1	0.623573154	7.178224997	7.25E-06	0.005216845	-0.769177542	2.635577886	6.52E-05	0.009180417
chr10	103882776	103900080 +	PPRC1	0.633940095	7.363800075	0.000269151	0.07339844	0.023365222	7.11360633	0.874963947	1
chr5	1306286	1348162 -	TERT	0.644131951	2.702413962	0.001521087	0.199236002	0.287989649	3.769062159	0.144844198	1
chr8	134536272	134653365 -	ST3GAL1	0.665784759	4.16233909	3.29E-05	0.015497782	-0.133845916	4.422849107	0.417695867	1
chr22	49354155	49363962 -	CPT1B	0.678157427	3.794755046	0.000303155	0.077396944	0.300172269	3.742108609	0.132772917	1
chr1	210805298	210860742 +	ATF3	0.717191547	6.685314539	0.001271559	0.18289697	0.72748061	4.232722277	0.000215515	0.024046714
chrX	107862382	107866263 -	IRS4	0.726757038	9.525407099	1.30E-06	0.001396489	0.03983274	9.70013011	0.800981158	1
chr8	126511744	126519826 -	TRIB1	0.731163548	5.124976165	1.77E-05	0.009895423	0.112639189	5.209839811	0.470004025	1
chr20	36482652	36497432 -	LOC388796	0.734979802	6.117374784	2.05E-05	0.010651409	0.159391522	5.242480908	0.309527315	1
chr19	54883753	54886059 +	C19orf76	0.748252522	2.091867921	0.00121	0.181953727	0.062127374	2.477033759	0.82336646	1
chr8	143689411	143692835 -	ARC	0.77543248	3.653537603	0.000916325	0.153672883	-0.261552352	1.34107119	0.344012888	1
chr8	128817496	128822862 +	MYC	0.868857148	6.888740089	2.25E-08	6.78E-05	-0.089924204	7.127191259	0.565343548	1
chr9	103162519	103187108 -	BAA1	0.878094618	1.923570513	0.001001877	0.160545397	0.317121388	0.308152641	0.372598666	1
chr22	17385346	17387761 +	DGCR9	0.896566077	0.947758911	0.001227755	0.181953727	0.084493257	0.230420688	0.859113873	1
chr12	55708567	55730816 -	MYO1A	0.918855821	1.093090385	0.001301038	0.18289697	0.607345853	-0.61243799	0.174062402	1
chr1	67545634	67635171 +	IL12RB2	0.932724361	2.508061497	0.000475505	0.10690335	-0.173482045	1.337553823	0.541586572	1
chrX	154880439	154893676 +	IL9R	0.955426399	1.530204359	0.000187434	0.054294477	0.156800617	0.623084223	0.687190459	1
chr12	48784067	48791362 +	GPD1	0.962102991	0.880647776	0.001460901	0.193031208	-0.228082403	-0.534627466	0.663314739	1
chr1	111818126	111822657 +	C1orf162	0.969985911	1.553366224	0.001092581	0.171095159	0.403559102	-0.388283202	0.369548098	1
chr11	124010894	124051409 -	SIAE	1.014433973	6.919696157	3.05E-06	0.002551014	0.162075929	5.982727425	0.560600802	1
chr17	15432701	15463743 -	CDRT1	1.058407239	0.664787324	0.000735173	0.133525928	-0.377954265	0.028995773	0.358821901	1
chr8	67568044	67593313 +	C8orf146	1.219609841	0.796674988	7.72E-05	0.029076327	0.84639459	-1.614571388	0.147179916	1
chr19	18228905	18246319 -	KIAA1683	1.243388536	1.989556024	2.66E-06	0.002358182	-0.381367368	-0.999944872	0.435832629	1
chr7	141342147	141453016 +	MGAM	1.302158211	1.092436362	2.30E-05	0.011556936	-0.15048453	-1.43575098	0.793222467	1
chr14	19881069	19881410 -	RPPH1	1.370968788	5.799066017	2.23E-08	6.78E-05	-0.127201192	4.822174662	0.569473772	1
chr7	102899466	103417199 -	RELN	1.375288522	-1.594085678	0.000543635	0.118300818	0.375665911	2.235245499	0.170443843	1
chr9	138227916	138236776 -	LHX3	1.389571111	1.139565454	8.04E-06	0.005264291	-0.147057004	-0.334166382	0.812908108	1
chr1	109057078	109086890 +	FNDC7	1.430697071	1.487779142	3.85E-07	0.000579414	-0.084310221	-1.158715598	0.907545233	1
chr6	31647854	31650077 +	LTA	1.565977491	0.70410787	7.27E-06	0.005216845	-1.237924802	-3.533576071	0.365768361	1

Supplementary table 4: Significantly changed genes after CTCF depletion (FDR < 0.2)

chr.	start	end	strand	Gene Name	RAD21 Cleavage				CTCF depleted			
					FC	CPM	p-val	FDR	FC	CPM	p-val	FDR
chr20	44235782	44313741	-	CDH22	0.287052228	2.139319467	0.25870855	1	-2.104646401	-0.040501259	1.45E-07	5.32E-05
chr17	43239731	43254146	-	OSBPL7	0.019973674	2.620075753	0.851724191	1	-1.916062813	0.965818529	8.26E-09	4.01E-06
chr2	70742723	70848883	-	ADD2	0.10696415	5.717907945	0.539120212	1	-1.867310472	5.283538992	3.72E-30	2.81E-26
chr11	36546138	36557886	+	RAG1	-0.271278741	0.51495031	0.386203247	1	-1.712938404	-0.803554828	0.00054653	0.050197486
chr11	63864265	63881582	+	CCDC88B	-0.115969195	1.573594497	0.944206359	1	-1.711427495	0.862346784	7.78E-07	0.000229766
chr1	16256850	16272714	-	FAM131C	0.438309366	2.067954813	0.066870812	1	-1.661430088	1.562561838	3.25E-08	1.40E-05
chr19	51059357	51068895	+	FOXA3	0.010408809	0.290351106	0.843946989	1	-1.633309178	-0.54407043	0.000329562	0.033769972
chr16	18902756	18982763	+	TMC7	0.015303495	3.588202402	0.934655531	1	-1.556498489	2.782003459	1.56E-12	1.56E-09
chr2	29057667	29128600	+	FAM179A	0.548611012	0.759603845	0.057779221	1	-1.545373503	-0.476365303	0.000362358	0.036632217
chr19	45664965	45774205	+	SPTBN4	0.097851786	2.430478667	0.593834884	1	-1.485107849	1.266450549	1.19E-06	0.000325765
chr12	56305817	56313252	-	BGALNT1	-0.002936186	3.895398765	0.866674831	1	-1.458843301	3.458807614	2.44E-12	2.30E-09
chr12	54361596	54364661	+	METTL7B	-0.392425761	2.599378846	0.22569623	1	-1.420280169	2.259017043	2.52E-07	8.64E-05
chr12	56406289	56408406	+	LOC100130776	0.073189581	3.02736681	0.742070506	1	-1.385444511	2.376124893	3.48E-08	1.46E-05
chrX	79698517	77037721	-	MAG1	-0.160860609	6.984624242	0.334732982	1	-1.361929317	7.226839394	3.89E-18	1.47E-14
chr22	18835993	18841786	+	RIMBP3	0.013446728	2.508685216	0.656708468	1	-1.349317634	1.77857075	2.41E-06	0.000625215
chr1	27591738	27594904	+	GPR3	0.546440381	1.825786873	0.043537373	1	-1.342543649	0.236843125	0.00020317	0.023010172
chr11	1972981	1975641	-	H19	-0.64539011	1.779503802	0.031753703	0.900764654	-1.335869398	-0.275212609	0.001280392	0.097406825
chr11	120036237	120362179	+	GRK4	0.101574995	0.284767511	0.752895378	1	-1.314253197	-0.297967902	0.001562879	0.114072882
chr4	15080586	15212278	+	CC2D2A	-0.1007299	3.376078661	0.559661652	1	-1.298050902	3.343647967	4.64E-10	3.18E-07
chr22	34107059	34120207	+	HMOX1	-0.456991024	4.268754105	0.030993495	0.894213246	-1.292483511	3.841521043	1.66E-10	1.19E-07
chr17	7939942	7962959	-	ALOXE3	0.314552807	0.819358695	0.24725781	1	-1.28313453	0.299433303	0.000597835	0.059323245
chr15	71996861	72007642	-	LOC100287616	0.320017855	2.145175987	0.133387357	1	-1.249295464	2.320733319	4.95E-06	0.001064447
chr19	54669277	54681300	+	FLT3LG	0.406162719	0.479026791	0.189529971	1	-1.235551979	0.301659176	0.000933018	0.077220069
chr11	118544649	118559936	+	NLRX1	-0.042872259	4.627369647	0.80075582	1	-1.232813517	3.993944543	5.74E-13	9.60E-10
chr11	298106	299410	+	IFITM2	-0.104907496	2.742277429	0.707574723	1	-1.205662748	2.457818021	1.12E-06	0.000312481
chr12	55678883	55686564	-	ZBTB39	-0.024469128	5.516374041	0.79653164	1	-1.20474427	4.853465067	1.69E-13	3.17E-10
chr12	56404342	56422211	-	AGAP2	-0.113638894	3.788678709	0.902685942	1	-1.181078406	3.272281681	4.17E-07	0.000133559
chr7	37926687	37958067	+	EPDR1	-0.001154875	3.857316371	0.873574866	1	-1.152130503	3.377704801	2.41E-08	1.07E-05
chr8	72917911	73131101	+	LOC100132891	-0.253696154	0.586872728	0.458448931	1	-1.150167678	0.455080526	0.001275519	0.097406825
chr10	123706592	123724733	-	NSMCE4A	0.158712677	5.946472665	0.350812881	1	-1.14295931	5.282412063	1.05E-12	1.42E-09
chr19	11392271	11406980	-	CCDC151	0.021069757	1.285354545	0.915521235	1	-1.13286405	0.897933739	0.000615339	0.032533895
chr12	47951275	47978748	+	PRPH	-0.137180479	0.515779324	0.681830478	1	-1.129164293	1.012426276	0.00001912	0.053967823
chr1	152168600	152185778	-	DENND4B	0.194384223	6.515528444	0.251920232	1	-1.126007147	5.930148611	1.18E-12	1.42E-09
chr20	61589809	61600949	-	EEF1A2	-0.019051731	7.083428838	0.963167024	1	-1.12507215	5.986389939	1.32E-12	1.42E-09
chr11	125638023	125644087	-	IRAP2	-0.092455657	7.027447611	0.480514641	1	-1.118771497	6.285435775	1.23E-12	1.42E-09
chr6	24652310	24754362	-	KIAA0319	0.20125513	0.641123962	0.397614298	1	-1.11589709	0.738940526	0.001131586	0.089241217
chr9	129968164	130006483	-	CIZ1	-0.026122642	7.952762208	0.770980864	1	-1.108636487	7.382652073	1.14E-12	1.42E-09
chr16	29789800	29818086	-	SEZL2	0.040319609	3.971662593	0.939845807	1	-1.10370632	3.572079999	6.14E-08	2.37E-05
chr16	66153810	66230589	+	CTCF	0.175387153	6.938425397	0.318715866	1	-1.090678587	6.50276881	3.77E-12	3.34E-09
chr19	40679958	40696400	-	DMKN	-0.101954575	6.309359926	0.453433863	1	-1.073689118	5.939081595	1.25E-11	1.04E-08
chr17	35770430	35774471	-	GJD3	0.181528972	2.228673511	0.326657924	1	-1.067431989	1.790465245	0.000156713	0.019191649
chr7	73848419	73905777	-	GF2IRD2	0.12850774	3.205586377	0.599100052	1	-1.064871324	1.981721786	3.28E-05	0.005149687
chr9	33431151	33437631	-	ATP3	-0.166672206	1.893433873	0.72716338	1	-1.049364032	1.154868748	0.000562471	0.051348499
chr22	34952200	34965946	-	APOL2	-0.382766009	3.524574622	0.047664823	1	-1.044895361	3.085940175	5.81E-07	0.000178587
chr1	149235924	149247478	-	FAM63A	-0.252391733	4.589517175	0.136140462	1	-1.022211305	3.982886476	1.56E-09	9.04E-07
chr5	35653745	35850470	+	SPEF2	0.118545814	2.540699334	0.609592228	1	-1.014157365	1.799499414	0.00011072	0.014502337
chr11	61323672	61341105	-	FADS1	-0.017071153	6.362064775	0.909747275	1	-1.014149879	6.114814041	1.23E-10	9.26E-08
chr14	93663870	93665710	-	IFIT2L2	-0.1049038	2.197466761	0.762487186	1	-1.012954763	1.881072312	0.000332442	0.033634989
chr16	65981212	65984922	-	TPPP3	-0.279130327	1.235359919	0.486205619	1	-1.003693425	0.729235201	0.003344541	0.197564032
chr2	74578151	74583951	-	LBX2	0.004335649	2.681038448	0.938774727	1	-0.993172387	2.102129426	9.57E-05	0.012867486
chr16	54070581	54098087	+	MMP2	-0.310890855	1.973875964	0.334797105	1	-0.988756022	2.546439083	0.00020002	0.02299927
chr11	780474	788269	-	SLC25A22	0.248224584	5.416236545	0.098588604	1	-0.981132989	5.211525474	1.03E-09	6.48E-07
chr7	904062	908815	-	ADAP1	0.139392639	3.677413001	0.371184575	1	-0.975255104	3.168220761	2.67E-06	0.000605511
chr7	36859035	37455028	-	ELMO1	-0.348196835	2.44034968	0.194032123	1	-0.971635344	2.15463096	0.000106886	0.01428018
chr14	75406357	75658553	-	BCYRN1	-0.422665217	5.549284904	0.007357548	0.518114086	-0.966071589	5.074474828	1.38E-09	8.32E-07
chr4	39734918	39836267	+	NBP2	0.029824131	5.635937267	0.58674151	1	-0.96478601	5.086306546	1.89E-09	1.05E-06
chr6	100075590	100087780	+	LOC100130890	-0.080190581	3.050672191	0.729349379	1	-0.954419658	2.545286266	8.02E-05	0.01098033
chr5	36070875	36102780	-	UGT3A2	-0.120136549	0.978983845	0.932046369	1	-0.947937879	1.518895024	0.00248556	0.157193719
chr1	23627642	23683337	-	ASAP3	-0.170350422	3.620009276	0.644742927	1	-0.942397163	3.266910921	4.85E-05	0.0071622
chr5	101597590	101660152	-	SLC04C1	-0.281167311	2.675858828	0.360789008	1	-0.931330503	2.19847122	0.000197538	0.022888557
chr2	68364806	68400687	-	CNRIP1	0.068265327	3.995133804	0.767333065	1	-0.927559892	3.584428053	5.24E-06	0.001095544
chr11	61276696	61312565	+	C11orf19	-0.191822478	3.672574115	0.409885445	1	-0.912407923	2.797719842	2.08E-05	0.003482653
chr19	1456016	1464188	-	ADAMTSL5	-0.013970058	1.482726474	0.957348381	1	-0.912088965	1.03137257	0.0003188076	0.193637079
chr8	144888297	144900495	+	LOC100128338	0.255270099	1.846169271	0.313570921	1	-0.907447474	1.741423343	0.001460865	0.10863652
chr1	46255803	46299548	-	CREB3L1	-0.723622461	1.367238301	0.030693462	0.894213246	-0.90585451	1.585294202	0.003319127	0.197564032

chr3	27732889	27738789	-	EOMES	-0.040469904	3.122600739	0.856314147	1	-0.877197748	2.435237048	0.000390965	0.038744071
chr19	18560494	18564147	+	C19orf60	0.044615866	4.128275182	0.946756918	1	-0.864935663	4.015185656	1.31E-05	0.002325127
chr10	69712422	69762690	-	PBLD	-0.027481038	3.757081795	0.780614866	1	-0.858657034	2.851203368	5.20E-05	0.007608371
chr14	105012175	105017545	+	CRIP2	-0.216302571	3.013280865	0.434943628	1	-0.852568378	1.661211136	0.001425934	0.106859937
chr14	36055352	36059181	-	NKX2-1	-0.1055063	1.478679256	0.851184332	1	-0.851176086	2.001168637	0.002196732	0.143244701
chr7	72294837	72323594	-	GTF2IRD2P1	0.180262082	2.590518314	0.456113591	1	-0.839973471	1.263276761	0.002811866	0.174300993
chr12	129840099	129889772	-	STM2	-0.056365427	5.008409195	0.819755266	1	-0.838449802	4.965578852	1.88E-07	6.75E-05
chr3	101694152	101778975	+	TMEM45A	-0.020724893	3.198927561	0.868790958	1	-0.835119739	2.888905475	0.00045206	0.036489863
chr12	3594754	3732627	-	EFCAB4B	0.228682187	2.581056792	0.242180891	1	-0.829464319	2.751035146	0.000515474	0.047929523
chr4	12978444	13095087	-	RAB28	0.181150315	4.526748741	0.216936806	1	-0.823844733	4.074245014	8.50E-07	0.000240167
chr17	40827050	40866065	-	ARHGAP27	0.033376942	3.476751977	0.733008742	1	-0.823474988	2.791538121	0.000608883	0.054269816
chrX	151746526	151749957	-	CETN2	-0.062714149	5.025510307	0.819581117	1	-0.809344763	4.69704168	6.50E-07	0.000195738
chr16	88542639	88561969	+	DEF8	0.010574611	6.361494058	0.925900097	1	-0.805061229	5.549160997	3.83E-07	0.000125323
chr11	92851286	92916194	-	C11orf75	0.115710218	2.725791562	0.598636406	1	-0.804088061	2.700229191	0.000818249	0.070030039
chr10	6973280	69837057	-	RUFY2	0.085606948	5.464483745	0.532665825	1	-0.803779781	4.944250343	5.45E-07	0.000171051
chr1	221460783	221604167	-	SUSD4	-0.129271266	2.226280228	0.862607076	1	-0.803691437	2.186036767	0.000333038	0.197564602
chr7	26298039	26380474	+	SNX10	-0.012967191	4.354362744	0.731597728	1	-0.797499461	3.961247722	5.55E-05	0.008038554
chr5	1062167	1091925	+	NKD2	0.106285206	3.629962866	0.587753447	1	-0.794691215	3.30106433	0.000121448	0.015503137
chr20	36867761	36985081	+	PPP1R16B	-0.309616735	2.958684812	0.241568204	1	-0.787663859	2.463048972	0.001218603	0.094563381
chr19	18901009	18913041	+	HOMER3	0.047081943	3.01488798	0.765961429	1	-0.78270594	2.501049089	0.001415321	0.106594874
chr5	137829079	137832903	+	EGR1	0.623573154	7.178224997	7.25E-06	0.005216845	-0.769177542	2.635577886	6.52E-05	0.009180417
chr4	17097113	17122955	-	QDPR	0.070836721	4.697354029	0.646494237	1	-0.764651725	4.507843327	3.40E-06	0.000799354
chr3	173831128	173911702	-	NCEH1	-0.245142335	4.028861196	0.255908649	1	-0.762222556	3.574508592	0.000148773	0.018520341
chr9	111442888	111753577	+	PALM2	0.202542569	0.964918886	0.418309353	1	-0.760486881	3.134601465	0.001088345	0.087060057
chr1	150271605	150276135	-	S100A11	-0.164888224	4.75459977	0.39135624	1	-0.760221271	4.901961884	2.72E-06	0.000660551
chr2	62662270	62344050	+	MYT1	0.098822143	2.399057409	0.586127148	1	-0.759742365	2.179864935	0.002730774	0.169973772
chr16	65868913	65880904	+	PLEKHG4	0.248763011	5.309491918	0.062463674	1	-0.757929664	4.139847361	5.68E-06	0.001157096
chr6	127801243	127882193	-	C6orf174	-0.063797574	6.181349741	0.68856521	1	-0.749958969	5.616275496	2.02E-06	0.000532733
chr9	130504622	130523020	+	PKNA	-0.050145543	6.104787723	0.78594556	1	-0.744869036	5.631911081	2.69E-06	0.000660551
chr6	127801243	127822228	-	KIAA0408	0.045990323	5.414041402	0.69316684	1	-0.73929192	4.9320291	4.11E-06	0.00091094
chr15	41597097	41611110	+	MAP1A	-0.414412036	4.773985267	0.033382953	0.922655828	-0.731680604	5.545881269	3.48E-06	0.000805979
chr10	115428924	115480654	+	CASP7	-0.251623972	4.371058179	0.15445097	1	-0.728067334	4.35788515	1.01E-05	0.001860686
chr3	16273571	16281050	+	DPH3	-0.099268373	5.909386544	0.655303663	1	-0.724737218	5.508463663	4.71E-06	0.001027375
chrX	128501932	128554211	+	OCLR	0.018153527	5.898316979	0.854443216	1	-0.72152472	6.217581592	3.74E-06	0.00085448
chr17	24424653	24531533	+	MYO18A	0.069600425	5.248247823	0.559145927	1	-0.71691285	4.931792546	8.63E-06	0.001631656
chr16	19777793	19803652	-	GPCR5B	-0.326034933	3.826073177	0.103903133	1	-0.712786833	3.795946559	0.000365975	0.036751225
chr7	47281276	47588267	-	TNS3	-0.142722062	3.223244983	0.574054579	1	-0.708623298	3.646902687	0.000483508	0.045796884
chr6	26646550	26655143	+	HMGNA4	0.131473179	5.595611497	0.259570653	1	-0.705375218	5.451325597	8.67E-06	0.001631656
chr11	117212900	117253411	-	FXR1	-0.261395254	4.190136959	0.265171854	1	-0.697457581	3.847399492	0.00051292	0.007929523
chr3	47819402	47866690	+	DHX30	0.149240647	9.981281775	0.528415504	1	-0.693024955	7.764988378	7.58E-06	0.001482345
chr19	11170968	11234168	-	DOCK6	0.182845683	5.177773801	0.242309016	1	-0.689407874	4.592228507	2.07E-05	0.003482653
chr19	54160377	54161948	+	FTL	-0.268331279	8.560057468	0.123367601	1	-0.683759845	8.175721272	9.69E-06	0.001802449
chr2	121271336	121466699	+	GLR2	0.056780575	4.685124862	0.558434637	1	-0.679605868	4.637123913	3.09E-05	0.004893743
chr12	51577237	51629917	-	KR7B	-0.040923052	3.500015693	0.050681124	1	-0.678351493	4.707159529	0.000443595	0.043108874
chr3	69107057	69145464	-	C3orf64	-0.072228832	4.320298598	0.706144538	1	-0.675477737	4.229632535	0.000546444	0.050197486
chr1	149298774	149307597	+	MLL11	-0.329158908	4.759894898	0.062359053	1	-0.674672116	4.463364608	3.80E-05	0.005841918
chr15	38973919	38983465	+	VTS18	-0.263498895	5.011571033	0.101192209	1	-0.668680271	4.609134071	3.94E-05	0.005999338
chr9	103277428	103289296	-	C9orf125	0.013927128	4.507866974	0.754991629	1	-0.668155657	4.535141187	4.84E-05	0.0071627
chr8	96007375	96030791	-	TNS3INP1	-0.953642714	4.772480412	3.71E-06	0.002941424	-0.667379855	3.660971884	0.00079237	0.068202657
chr14	64523259	64599123	+	FTPB	-0.181547066	4.663800152	0.185960919	1	-0.662597897	4.314203818	6.40E-05	0.00909512
chr21	37045058	37284415	-	HLCS	0.169569654	5.286538739	0.275566206	1	-0.654792451	5.286179749	4.06E-05	0.006103663
chr3	11806918	11863352	-	TAMM41	0.355953124	3.835527525	0.116785596	1	-0.653724109	4.168690789	0.000865102	0.072799038
chr1	92712905	92725021	-	GFI1	0.225857924	3.867467541	0.196350427	1	-0.648454922	3.05481278	0.001822283	0.124632147
chrX	19288096	19443300	-	LAP3K15	-0.038253786	2.683156043	0.922966722	1	-0.647838807	3.725368097	0.001256559	0.0965691
chr10	81575637	81577338	+	MOC642361	0.13448254	4.432826366	0.587152833	1	-0.646143806	4.017115057	0.001069625	0.08615946
chr8	142207901	142275082	+	DENN3	0.221541981	3.603909976	0.220119274	1	-0.632131016	3.665335962	0.001741264	0.12572229
chr1	39861689	39877935	-	HEYL	-0.035069357	3.690628932	0.820987827	1	-0.630357846	3.076773127	0.002447451	0.155948556
chr11	64838864	64846476	+	CDC42EP2	0.33620249	4.057103945	0.105310011	1	-0.625171703	3.677648568	0.001967449	0.132302375
chr1	152564659	152590404	+	ATP8B2	-0.038064937	6.785081278	0.983161448	1	-0.623381054	6.847038194	6.06E-05	0.008688732
chr15	61268780	61347026	+	RAB8B	-0.121496759	4.939662237	0.559328432	1	-0.620010877	4.226287745	0.000175739	0.021009163
chr20	3091272	3097207	-	PROSAPIP1	-0.183162187	4.465481714	0.418193936	1	-0.616007911	4.16111092	0.001773797	0.124273043
chr1	40278841	40310908	+	CAP1	-0.203911175	7.826450298	0.179839122	1	-0.613025455	7.29464485	7.45E-05	0.010395458
chr1	167603817	167632404	+	BLZF1	-0.140111404	5.007835446	0.399790981	1	-0.61287194	4.359459667	0.000175047	0.021009163
chr12	97430881	97434135	-	LOC100128191	-0.061596828	4.314661496	0.750986793	1	-0.610029311	4.143539648	0.002064628	0.137002147
chr2	74216035	74228547	-	BOLA3	0.027445642	5.891300044	0.98640676	1	-0.607408867	5.345072197	0.000153288	0.018926041
chr3	49032911	49035930	+	NIDUFAP3	-0.106066931	6.152665465	0.460567832	1	-0.602518115	6.1614675645	0.000113683	0.014762147
chr22	37590193	37598204	-	CBU5	0.149506273	7.678307359	0.485179977	1	-0.600886751	7.05182726	0.000110084	0.014502337
chr9	98711177	98712558	+	LOC441454	-0.164699646	4.17607307	0.226816612	1	-0.600244539	4.822192249	0.001741383	0.122572229
chr11	44073674	44223556	+	XNT2	-0.182271369	5.998134574	0.285211538	1	-0.599991528	5.709041255	0.000148354	0.018520341
chr7	148759386	14878986	-	FIN777	0.297961145	5.563103396	0.107119837	1	-0.598688223	4.896989269	0.000202185	0.023010172
chr3	87069812	87122947	-	VGLL3	-0.194753535	3.531474577	0.563197667	1	-0.594028128	4.091791446	0.002453682	0.155948556
chr1	170017383	170033479	+	METTL13	0.131212672							

chr5	61744329	61960172 +	IPO11	0.052014029	6.635933029	0.676655688	1	-0.581247299	6.299007702	0.000190595	0.022605738
chr5	133765654	133775497 -	CDKN2AIPNL	0.050623445	5.551046787	0.829911247	1	-0.578981051	5.297820246	0.000278774	0.029781331
chr19	6445329	6453330 -	TUBB4A	-0.209114217	5.162282945	0.198649909	1	-0.578410548	5.221897683	0.000320894	0.031853233
chr5	76047623	76067351 +	F2R	0.067895148	4.512794413	0.425663441	1	-0.577662925	5.008410346	0.000311933	0.032404421
chr6	133132199	133161440 -	C6orf192	0.067163856	5.051658819	0.537392823	1	-0.563594393	5.222083502	0.000414305	0.040523839
chr18	42007985	42100953 +	C18orf25	0.012040549	6.150258783	0.986750513	1	-0.551402828	5.465046779	0.000486476	0.045798684
chr1	160733587	160766043 +	UHMK1	0.042917194	7.779650008	0.752065846	1	-0.544396575	7.420805847	0.000409159	0.040282105
chr13	102249399	102291889 +	BHVM1	0.042349377	5.05733952	0.645851627	1	-0.543052902	5.232811108	0.000659286	0.058075029
chr15	89275158	89276778 -	HDDC3	0.199669606	4.664078362	0.912142046	1	-0.541310488	4.344558579	0.00102217	0.083226743
chr1	31965304	32002235 -	BAID2	-0.049467113	4.947647652	0.966537146	1	-0.529790562	4.675270996	0.001221425	0.094563381
chr3	49133550	49145603 -	LAMB2	-0.258695577	5.775805904	0.173383174	1	-0.528104839	5.674819248	0.000787511	0.068202657
chr1	153557263	153567533 +	RUSC1	-0.076300393	5.996689453	0.575480304	1	-0.527135635	5.703716858	0.000846366	0.072027192
chr2	235066424	235070432 -	ARL4C	-0.179457711	4.451038538	0.307871046	1	-0.524192195	4.717454994	0.001205396	0.094563381
chr6	44374440	44389041 -	AARS2	-0.092889727	6.3111338257	0.545222585	1	-0.519326973	6.158207928	0.00087848	0.073514166
chr1	183353840	183392853 -	TRMT1L	0.072740468	5.51577735	0.675703823	1	-0.518246406	5.270362308	0.001123624	0.089079074
chr6	31234281	31239971 +	TCF19	-0.247611003	5.435127213	0.133939526	1	-0.516562078	5.608760905	0.001092369	0.087060057
chr9	76787682	76833130 -	C9orf41	0.037973307	5.649063165	0.634115085	1	-0.514609773	5.776823185	0.000998985	0.081855522
chr1	149290070	149298749 -	CDC42SE1	-0.009351616	6.819919408	0.941246981	1	-0.508536756	6.816226201	0.001068107	0.086159146
chr15	70320575	70350682 -	PARP6	0.302681672	5.832271534	0.073393439	1	-0.506397155	5.071434998	0.001567622	0.114072882
chr6	44329815	44333261 -	SLC35B2	0.087121766	5.236863848	0.562061318	1	-0.504727292	4.973504618	0.001806053	0.124632147
chr11	10781367	107969584 -	EXPH5	0.149199972	4.912119238	0.367641375	1	-0.502296966	5.045117292	0.001647706	0.118187563
chr19	877036	923803 +	ARD3A	-0.060973956	5.196045061	0.818369149	1	-0.500687843	4.545979161	0.002373962	0.145303532
chr12	121915827	121946665 -	VPS37B	0.024200076	5.62802167	0.806455218	1	-0.500592149	4.834082068	0.00198657	0.132490569
chr11	124115038	124122312 +	NRGN	-0.396323932	4.429504549	0.006339498	0.475083894	-0.497817177	4.854690554	0.002115922	0.139179589
chr9	70840298	70904914 +	FXN	0.281131819	5.449211928	0.123032378	1	-0.496676159	5.667752922	0.001624708	0.117095615
chr15	50386772	50608539 -	MYO5A	-0.019604784	5.747544742	0.889124652	1	-0.494849822	5.892310393	0.001575602	0.1141024
chr10	88844932	88941202 +	KAL5A	0.153387936	5.116073212	0.310106234	1	-0.494216314	5.475753491	0.001715301	0.121875389
chr2	74279197	74295932 +	MTHFD2	0.221898968	7.793341149	0.071015628	1	-0.493053652	7.258558521	0.001407059	0.106505155
chr10	46317646	46359151 +	FAM35B	0.119766389	4.181670887	0.401573072	1	-0.491646173	4.435267952	0.002673891	0.167123756
chr17	74360653	74433067 -	TIMP2	-0.392327272	4.692033386	0.016733468	0.704068253	-0.490091006	4.590101931	0.00266477	0.167123756
chr6	24603175	24645414 +	ALDH5A1	0.028404726	5.452837143	0.687426328	1	-0.487763099	5.92148887	0.001818578	0.124632147
chr10	70418482	70446745 +	KAL1279	-0.172345112	6.134691317	0.246962094	1	-0.485440529	5.756888965	0.001968605	0.132302149
chr7	134114703	134306020 +	JIAA1	-0.226169889	6.039663333	0.352523122	1	-0.48526617	6.388224332	0.001828567	0.124632147
chr1	221955917	222033043 +	CAPN2	0.072848233	7.162639742	0.538524888	1	-0.479788515	7.112965393	0.001987842	0.132490569
chr10	55366902	55368285 -	MTRNR2L3	-0.268087749	4.707624176	0.111640607	1	-0.47921999	6.633140977	0.001944865	0.131961721
chr13	49384842	49408626 -	SPRYD7	0.098754612	5.215144314	0.589554921	1	-0.478925659	4.951338036	0.002844988	0.175631401
chr7	94374884	94763663 +	PPP1R9A	0.054406651	4.879465988	0.555462694	1	-0.47480886	4.507388815	0.003316593	0.197564032
chr9	12240316	122516586 -	MEGF9	-0.364785307	5.248872895	0.033687637	0.9227243547	-0.471810491	4.983736695	0.003208058	0.194068194
chr3	57236804	57282538 +	APL1	-0.052135321	6.695944194	0.01126982	1	-0.458135279	6.908159315	0.003045108	0.18702255
chr2	181553356	181636395 +	UBE2E3	0.197881987	6.066858652	0.23334446	1	0.456485376	6.424559131	0.00326592	0.197564032
chr5	141360417	141372804 -	GNBD3A	0.094639	5.605410565	0.506185362	1	0.465461947	6.168455493	0.002871717	0.176557835
chr9	129249773	129253532 -	RPL12	-0.042206312	10.06099287	0.835797041	1	0.475656787	9.886860411	0.001820841	0.124632147
chr12	22492784	22588719 -	KALAO528	0.080297313	5.756316225	0.483471483	1	0.47726554	6.38367627	0.002301578	0.148156683
chr3	106568246	106778447 +	ALCAM	-0.086922905	6.11490157	0.808288717	1	0.477926945	6.639045265	0.002143606	0.140387533
chr14	51025604	51267194 +	FRMD6	-0.246351206	4.781400629	0.197294529	1	0.488726909	4.807988813	0.002494145	0.157193719
chr16	67435009	67676586 +	TMCO7	0.17644935	4.685317083	0.295717592	1	0.49204087	4.80840619	0.002296309	0.140387533
chr15	39853923	39907345 +	MAPKBP1	-0.066131911	4.735438056	0.621703222	1	0.509005376	5.068083999	0.001447033	0.107904268
chr3	184836104	184884998 +	KLHL24	-0.327380961	5.11189353	0.100883734	1	0.512756066	4.572265087	0.00171027	0.121875389
chr2	70376611	70382724 -	FAM136A	0.019182542	6.889785206	0.942535983	1	0.513347464	7.110273539	0.000864136	0.072799038
chr1	126520	1274355 -	DVL1	0.127556725	5.78074189	0.366344321	1	0.514150483	6.140020576	0.000948817	0.078098155
chr3	123585712	123611651 +	FAM162A	-0.096075426	6.060183361	0.501056754	1	0.530574091	6.200288825	0.000670728	0.058739406
chr7	89678935	89704928 +	STEAP2	0.024421807	4.468027106	0.597729765	1	0.533250312	5.148328467	0.000888136	0.073911532
chr12	47682321	47698896 -	PRKAG1	0.024197731	5.151274822	0.810306272	1	0.541790332	5.548373974	0.000612684	0.054287404
chr12	75776626	75796930 -	CSRP2	-0.39292072	4.973587798	0.013958896	0.651625917	0.544942371	5.317467389	0.000572721	0.051928427
chr20	18516555	18692560 +	TRD1	-0.087277202	5.03747259	0.603345681	1	0.551489312	5.55614381	0.000465384	0.044367579
chr14	61712575	61736770 -	PIGH	-0.008477136	5.198221455	0.927156569	1	0.558964053	5.07013721	0.000458689	0.044007866
chr6	21701950	21706828 +	SOX4	-0.035772443	6.6543047	0.624384941	1	0.582829753	7.489964656	0.000167502	0.020347484
chr6	139735089	139734778 -	CITED2	-0.140856125	4.986716338	0.400843343	1	0.590765324	5.245035182	0.000217964	0.024141147
chr12	122311492	122322640 -	CDK2AP1	0.048942456	6.901220668	0.826706852	1	0.595064922	7.1990635	0.000121024	0.015503137
chr6	121798443	121812572 +	GJA1	-0.073876341	5.230736287	0.999014478	1	0.620716419	6.015474938	7.92E-05	0.0109455
chr22	43655706	43784473 -	PHF21B	0.460160203	4.173163639	0.01788585	0.728147444	0.629965031	3.922494698	0.001530372	0.112448724
chr12	64816983	64850074 -	TMBIM4	0.051793189	4.847661229	0.633362774	1	0.662089274	5.13400203	3.54E-05	0.005496704
chr4	55907144	55934023 +	SRD5A3	0.152221627	5.059577753	0.256691014	1	0.667355181	5.692712442	2.38E-05	0.003891701
chr12	43188324	43393978 -	NELL2	-0.047611549	4.706471495	0.925965029	1	0.674605387	4.949613237	2.80E-05	0.004534574
chr16	53510278	53520580 -	CRND2	0.270169019	4.413640914	0.077445282	1	0.675480734	4.767376093	2.84E-05	0.004556489
chr16	65195435	65205296 +	CKTM3	-0.078579979	4.396795818	0.69874618	1	0.682390542	5.053811757	1.98E-05	0.003393812
chr15	31425178	31485321 -	NKAIN1	0.449039757	2.142460573	0.06954819	1	0.692444599	2.860479776	0.003330448	0.017564032
chr1	52380633	52584946 +	ZFYVE9	0.101968888	5.156004581	0.35388244	1	0.69913071	5.59500728	1.04E-05	0.00188798
chr8	144878090	144887902 -	FAM83H	0.17192707	4.387782648	0.313603017	1	0.700428961	4.16270336	0.000371558	0.003706471
chr14	35077308	35348183 -	RALGAP1	-0.013372653	6.377334451	0.967388355	1	0.700905212	7.080850863	6.35E-06	0.001274489
chr6	126319553	126343082 +	HINT3	0.121662408	5.941462815	0.349463433	1	0.701232894	6.024689673	8.15E-06	0.001573005
chr10	73703682	73705803 +	DDIT4	0.146999442	5.418488527	0.160458614	1	0.702633559	5.702617867	0.000193019	0.022700076
chr13	195336624	195339095 +	HES1	-0.195725839	4.49662823	0.205992814					

chr5	130627600	130758281 +	<i>CDC42SE2</i>	0.053231563	5.385869669	0.687409398	1	0.725324026	5.256225328	5.65E-06	0.001157096
chr1	210805298	210860742 +	<i>ATF3</i>	0.717191547	6.885314539	0.001271559	0.18289697	0.72748061	4.232722277	0.000215515	0.024046714
chr2	190319630	190336169 -	<i>OSGEPL1</i>	0.143515276	3.474197973	0.356289238	1	0.729906524	4.113717186	0.000206492	0.023211817
chr17	71892284	71895536 +	<i>SPHK1</i>	0.185693805	3.846901191	0.296949341	1	0.733505427	3.662703876	0.000222856	0.024502791
chr3	198146669	198153861 -	<i>NCBP2</i>	0.149463722	5.808439727	0.372105964	1	0.737562698	6.297139049	2.54E-06	0.000648191
chr14	49161641	49171698 -	<i>DNAAF2</i>	0.271940938	4.476029584	0.063718267	1	0.744248597	5.00255538	3.86E-06	0.000867737
chr19	47509316	47521054 +	<i>TMEM145</i>	0.261758204	3.422392336	0.226719949	1	0.75259587	3.274754546	0.000246287	0.026882785
chr8	23755378	23768265 -	<i>STC1</i>	-0.650918459	1.552717657	0.011853193	0.611454283	0.764475469	3.616197879	0.000791979	0.068202657
chr8	48812028	48813279 -	<i>CEBPD</i>	-0.108217658	1.461553379	0.694624572	1	0.815054479	2.029396379	0.002996914	0.18350615
chr2	217245072	217268517 -	<i>IGFBP5</i>	-0.099413094	3.766774211	0.817277777	1	0.816979613	5.559132955	2.67E-07	8.92E-05
chr22	40852444	40856827 +	<i>CYP2D6</i>	0.111652052	1.844852932	0.815606266	1	0.828496087	2.529383758	0.001819234	0.124632147
chr16	391858	402488 +	<i>DEC2R</i>	0.221888791	3.186913264	0.225199437	1	0.855142751	3.556032288	2.31E-05	0.00381994
chr11	75157425	75190229 +	<i>DGAT2</i>	-0.102752135	2.564964104	0.754238298	1	0.874941775	2.611773383	0.000269717	0.029019654
chr11	106049947	106394381 -	<i>GUCY1A2</i>	0.65494924	-0.24515094	0.018396439	0.735715694	0.882716151	4.293321974	7.27E-06	0.001441471
chr8	19215360	19298009 +	<i>SH2D4</i>	0.130539753	2.188196686	0.501914762	1	0.898244757	3.004849899	0.000142862	0.018083448
chr1	144149818	144153985 +	<i>TXNIP</i>	-0.615854233	5.317906683	0.000121555	0.04067075	0.953629933	5.490576351	2.01E-09	1.08E-06
chr4	970784	988317 +	<i>IDUA</i>	0.004587952	1.952145038	0.939385261	1	0.954863284	2.565381927	8.67E-05	0.011767874
chr5	76408287	76418786 -	<i>ZBED3</i>	0.328578444	3.22962119	0.116112157	1	0.96846413	3.55262833	1.96E-06	0.00052728
chr22	40816882	40851298 +	<i>LOC100132273</i>	-0.164131608	2.436510347	0.451226759	1	1.021200782	3.010992793	1.50E-05	0.002593385
chr16	2829574	2832753 -	<i>PRSS30P</i>	0.158293396	1.781977314	0.560757252	1	1.040641275	2.833219499	1.40E-05	0.002445271
chr18	586997	640293 +	<i>CLUL1</i>	-0.609370206	0.553024514	0.032767755	0.91743623	1.053489023	0.999676771	0.001486706	0.10977574
chr12	81276406	81397147 +	<i>C12orf126</i>	0.323932677	2.814695557	0.10943509	1	1.055127028	3.649622185	2.14E-07	7.48E-05
chr10	32775046	32903498 +	<i>CCDC7</i>	-0.110614281	0.524758663	0.856046252	1	1.110347879	0.781125178	0.001224182	0.094563381
chr18	53448531	53487506 +	<i>LOC100505549</i>	0.141507683	1.980934367	0.623541358	1	1.116770385	2.770852913	3.37E-06	0.000799354
chr3	137538688	137953935 -	<i>STAG1</i>	0.592628644	6.42773176	0.000124202	0.04067075	1.160542278	6.749867354	1.33E-13	2.87E-10
chr17	6858779	6861567 +	<i>C17orf49</i>	0.126880876	4.463302612	0.489014468	1	1.318307948	4.934123173	5.81E-16	1.75E-12
chr19	45045802	45132373 -	<i>FCGBP</i>	-0.270644414	2.040338183	0.285024971	1	1.461022958	2.695268206	2.68E-09	1.34E-06
chr12	119123476	119135014 +	<i>LOC100506649</i>	-0.028581398	1.936031421	0.991248638	1	1.494337232	2.876909718	7.12E-10	4.66E-07
chr11	111249989	111255391 -	<i>FDXACB1</i>	0.330430642	3.066709528	0.102570831	1	1.51015353	4.316248054	4.14E-14	1.04E-10
chr15	39032927	39036009 +	<i>CHAC1</i>	0.37398851	2.99227481	0.047083288	1	1.585617077	2.471695255	5.24E-08	2.13E-05
chr7	141050606	141084499 -	<i>FLJ40852</i>	-0.062937753	0.691143102	0.902507157	1	1.685541742	0.971100482	9.75E-07	0.00027719
chr4	146759989	146800637 +	<i>MMAA</i>	0.120184832	1.19175043	0.629015322	1	1.791246419	2.525937822	8.44E-11	6.69E-08
chr16	29782504	29786875 +	<i>LOC440356</i>	0.37567817	-0.177403112	0.265962513	1	2.113020711	0.970916243	2.67E-09	1.34E-06
chr6	132000134	132110243 +	<i>ENPP3</i>	0.298202083	0.296262085	0.208687767	1	3.860050305	2.324064363	1.03E-28	5.17E-25
chrX	151833641	151892681 +	<i>ZNF185</i>	0.338869437	1.405751976	0.121427222	1	4.023726175	5.031710915	1.29E-72	1.94E-68

Supplementary table 5: ChIP-qPCR primers

Primer Nr.	Forward	Reverse	Genomic Position
328	CAGAGGGAGTCACACTAACG	GTGCAGGTGTGGAGAATG	chr11:1,493,349-1,493,623
331	AGGTGGCACTGTTGCATAG	CCTGCCAAAGGAGAGCTTTATC	chr11:1,630,677-1,630,849
332	CAACACGCAGGAGAAAGG	AAGCGCAGATGAGCAGTC	chr11:1,671,708-1,671,998
333	GTCTCCGGTGTAAATGGTC	AGTGCCATGCTAACTTG	chr11:1,931,466-1,931,618
204	TTCAGCCGGTTCAAGGGACG	CTAGGGAGGAGGACAGAGGCAAGAG	chr11:2,014,891-2,015,077
200	GAGCTCAACAATAGCCCTTGATC	AGCCTGGGAGAAAGCACATCT	chr11:2,121,433-2,121,703
185	GCTGCTGTTCCGCCATTCAATTC	GCTGATACCAACCACCAATCCATGAG	chr11:59,140,122-59,140,236
338	AGGCATGGCACAAGACAC	GTCTCCCTCTGCCAAGTTAGC	chr11:3,069,776-3,069,983

Supplementary table 6: qPCR primers for transcript analysis

Target gene	Forward	Reverse	Genomic Position
CTCF	CAGCCATTCCACATCATCTACG	CACCATGTCCCACCTACAAATC	chr16:66,157,405-66,157,639
SMC3	TTGACCAGGCTCTGGATG	CTCCCAAACCAGTAGGTAG	chr10:112,352,709-112,354,079
RAD21	GTGATGGAGGCCAGCAGAAC	ACCTGCTGAGGAGGCATCAC	chr8:117,932,176-117,933,456
H19	CTTTACAACCACTGCACCTACCTGAC	GATGGTGTCTTTGATGTTGGGCTGA	chr11:1,973,551-1,973,726
HOXBAS3	AACCCAGCGATATCCGACCAG	AGGGAATGAGTCCGGGAGCAAG	chr17:44,036,386-44,038,655
HOXA-AS3	ATTCCACCCACGCACCTATTC	TGGCTAATTTCTCGTCCTC	chr7:27,157,720-27,158,542
CDH22	TCTTGGTCTGCGTTCTCATCC	CCGCCTTCGTCGTGTAATTTG	chr20:44,237,015-44,240,018
ENPP3	TATCATCCCTCACCGACCTAC	GAGACAGGCTGCACCTTATCC	chr6:132,103,174-132,109,737
MMAA	GCTGTTGCTGACATGGTTGAC	ATCCTTCGAGCTGGCACAATC	chr4:146,791,684-146,794,681
NAD	ACTGGTACTGCGTACATCC	AGATGCGCCTATCTCTTTCC	chrX:118,889,962-118,891,355
SNAPIN	GCTGCATACGTGAGGCTTAC	TGGGTGGACAGAGGTCATTCC	chr1:151,900,673-151,900,856
TMC7	GCCTTTGCAGTTCCCTTTC	CGCTCAGTCTAGTTAGTTCC	chr16:18,975,489-18,980,678

Supplementary table 7: 3C-sequencing primer for amplification of ligation products with the different viewpoint fragments

BglII/NlaII digestion		
View Point	Forward	Reverse
Vp1	CCGAGCCTGACACCAT	CGGGCCTCTCAGTAATAGAT
Vp2	CTGCCTGGGAGATGAAATCA	TGACGGGACAGGCTCC
Vp3	CCGACGACCTCTACAGCA	CGTCTTCTCGGAAACAGAT
Vp4	GGGTCTTCCAATGCACGAAAC	GACTGAGCCAGTCCCAGG
Vp5	CTGGTGTGCTTTAATCCTTG	ACACTGTGGCGGAGGATAC
Vp6	CACCAGAACACCTCCTCCA	CGACGGGACATCATCCACT

Supplementary references

- Bornkamm G.W., C. Berens, et al., 2005. Stringent doxycycline-dependent control of gene activities using an episomal one-vector system. *Nucleic Acids Res* 33, 16, e137.
- Dixon J.R., S. Selvaraj, et al., 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 7398, 376-380.
- Imakaev M., G. Fudenberg, et al., 2012. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Methods* 9, 10, 999-1003.
- Kueng S., B. Hegemann, et al., 2006. Wapl controls the dynamic association of cohesin with chromatin. *Cell* 127, 5, 955-967.
- Langmead B., C. Trapnell, et al., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, 3, R25.
- Lieberman-Aiden E., N.L. van Berkum, et al., 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 5950, 289-293.
- Parkhomchuk D., T. Borodina, et al., 2009. Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res* 37, 18, e123.
- Robinson M.D. and A. Oshlack, 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 11, 3, R25.
- Schockel L., M. Mockel, et al., 2011. Cleavage of cohesin rings coordinates the separation of centrioles and chromatids. *Nat Cell Biol* 13, 8, 966-972.
- Simonis M., J. Kooren, et al., 2007. An evaluation of 3C-based methods to capture DNA interactions. *Nat Methods* 4, 11, 895-901.
- Stadhouders R., P. Kolovos, et al., 2013. Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nat Protoc* 8, 3, 509-524.
- Sumara I., E. Vorlauffer, et al., 2000. Characterization of vertebrate cohesin complexes and their regulation in prophase. *J Cell Biol* 151, 4, 749-762.
- van de Corput M.P., E. de Boer, et al., 2012. Super-resolution imaging reveals three-dimensional folding dynamics of the beta-globin locus upon gene activation. *J Cell Sci* 125, Pt 19, 4630-4639.
- Wendt K.S., K. Yoshida, et al., 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451, 7180, 796-801.

Chapter 4



A cohesin-independent role for NIPBL at promoters provides insights in Cornelia de Lange Syndrome

Jessica Zuin, Vedran Franke, Wilfred F.J. van IJcken, Antoine van der Sloot, Ian D. Krantz, Michael I.J.A. van der Reijden, Ryuichiro Nakato, Boris Lenhard & Kerstin S. Wendt

Manuscript submitted

A cohesin-independent role for NIPBL at promoters provides insights in Cornelia de Lange Syndrome

Short title:

NIPBL as novel transcription co-factor

Authors:

Jessica Zuin^{1*}, Vedran Franke^{2,8*}, Wilfred F.J. van IJcken³, Antoine van der Sloot³, Ian D. Krantz⁷, Michael I.J.A. van der Reijden¹, Ryuichiro Nakato^{4#}, Boris Lenhard^{2,5,6#} and Kerstin S. Wendt¹⁺

¹ Department of Cell Biology, Erasmus MC, 3015GE Rotterdam, Netherlands

² Computational Biology Unit-Bergen Center for Computational Science and Sars Centre for Marine Molecular Biology, University of Bergen, N-5008 Bergen, Norway

³ Center for Biomics, Erasmus MC, 3015GE Rotterdam, Netherlands

⁴ Laboratory of Genome Structure and Function, Institute of Molecular and Cellular Biosciences, The University of Tokyo, Tokyo 113-0032, Japan

⁵ Department of Biology, University of Bergen, N-5008 Bergen, Norway

⁶ Present address: Institute of Clinical Sciences, Imperial College London and MRC Clinical Sciences Centre, Hammersmith Hospital Campus, Du Cane Road, London W12 0NN, UK

⁷ The Children's Hospital of Philadelphia, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA

⁸ Present address: The faculty of Science, University of Zagreb, Croatia

* These authors contributes equally

Corresponding authors for computational biology

+ Corresponding author

Corresponding author contact information:

Kerstin S. Wendt

Dept. of Cell Biology, Erasmus MC, Faculty building

PO Box 2040, 3000 CA Rotterdam, Netherlands

Phone: +31-10-7044007, Fax : +31-10-7044743

E-mail: k.wendt@erasmusmc.nl

Keywords: cohesin, NIPBL, transcription, CHIP-sequencing, Cornelia de Lange Syndrome, developmental defects

Abstract

The cohesin complex is crucial for chromosome segregation during mitosis and has recently also been implicated in transcriptional regulation and chromatin architecture. The NIPBL protein is required for the loading of cohesin onto chromatin, but how and where cohesin is loaded in vertebrate cells is unclear. Heterozygous mutations of NIPBL were found in 50% of the cases of Cornelia de Lange Syndrome (CdLS), a human developmental syndrome with a complex phenotype. However, no defects in the mitotic function of cohesin have been observed so far and the links between NIPBL mutations and the observed developmental defects are unclear.

We show that NIPBL binds in somatic cells independently from cohesin to chromatin and almost exclusively to the promoters of active genes. NIPBL or cohesin knockdown reduce transcription of these genes differently, suggesting a cohesin-independent role of NIPBL for transcription. Motif analysis and comparison to published data show that NIPBL co-localizes with a specific set of other transcription factors. In cells derived from CdLS patients NIPBL binding levels are reduced and several of the NIPBL-bound genes have previously been observed to be mis-expressed in CdLS.

In summary, our observations indicate that NIPBL mutations might cause developmental defects in different ways. First, defects of NIPBL might lead to cohesin-loading defects and thereby alter gene expression and second, NIPBL deficiency might affect genes directly via its role at the respective promoters.

Author summary

The multi-protein complex cohesin is crucial for chromosome segregation during cell divisions but has recently also been implicated in transcriptional regulation and chromatin architecture. Cohesin's binding to chromatin depends on NIPBL, a factor that was found to be mutated in 50% of the cases of the human multisystem developmental disorder Cornelia de Lange Syndrome (CdLS). To understand the molecular mechanism of cohesin, we need to know when and where the complex is loaded onto DNA. Therefore we compared the genomic binding sites of NIPBL and cohesin in different cell lines. To our surprise, NIPBL does not bind to the same sites in the genome as cohesin but instead resides at active promoters and the activity of these genes depends on NIPBL. These observations point to a dual role for NIPBL in cohesin-loading but also as potential transcription co-factor. This has important implications for the research on Cornelia de Lange Syndrome, since NIPBL mutations might directly influence expression of developmentally important genes. Consistent with this, we observed that a number of genes misregulated in CdLS patient samples have NIPBL bound to the promoters.

Introduction

Genomes need to be stably inherited over numerous cell generations. For each cell division the genetic information has to be replicated, the copies identified and then equally distributed between daughter cells. This process crucially depends on the cohesin complex, consisting of the core subunits SMC3, SMC1A, RAD21, SA1/STAG1 or SA2/STAG2 and several transiently associated regulatory proteins (reviewed in Peters, Tedeschi et al., 2008). Cohesin tethers both sister chromatid together from S-phase on, allowing for their proper segregation in mitosis. Furthermore, cohesin is important for DNA damage repair (for review see Sjogren and Strom, 2010), for chromatin insulation in cooperation with the chromatin insulator protein CTCF (Parelho, Hadjir et al., 2008; Stedman, Kang et al., 2008; Wendt, Yoshida et al., 2008), for chromosomal long-range interactions (Hadjir, Williams et al., 2009; Mishiro, Ishihara et al., 2009; Nativio, Wendt et al., 2009) and for development (Horsfield, Anagnostou et al., 2007; Pauli, Althoff et al., 2008; Schuldiner, Berdnik et al., 2008; Pauli, van Bommel et al., 2010). The latter functions implicate cohesin in regulating gene expression; indeed, a large number of genes are misregulated after cohesin depletion (Wendt, Yoshida et al., 2008; Kagey, Newman et al., 2010). How exactly cohesin associates with DNA is not understood since none of the subunits binds directly to DNA. Rather, cohesin is hypothesized to bind to DNA by embracing the DNA strands with a “protein ring” formed by the core subunits (Haering, Lowe et al., 2002; Haering, Farcas et al., 2008).

Cohesin’s binding to chromatin is tightly regulated throughout the cell cycle. To enable chromosome segregation it is removed from chromosomes during mitosis. A prophase pathway depending on WAPL and specific phosphorylation of cohesin subunits dissociates cohesin from chromosome arms. The remaining cohesin is removed by proteolytic cleavage of the RAD21 subunit at anaphase onset (reviewed in Peters, Tedeschi et al., 2008). Cohesin re-associates with chromatin at the G1-S-phase transition in yeast but in vertebrates already earlier during G1 phase.

The chromosomal localization of cohesin is determined by several factors. First, the cohesin loading factors NIPBL (also known as IDN3 or Delangin or Nipped-B, in *D. melanogaster*; Scc2, in *S. cerevisiae*) and MAU2 (also KIAA0892; Scc4 in *S. cerevisiae*) which are crucial for the re-loading of cohesin in G1-phase after its complete dissociation from chromatin during mitosis (reviewed in Peters, Tedeschi et al., 2008). In yeast, it has been shown that cohesin associates first with Scc2 binding sites and then relocates to different positions (Lengronne, Katou et al., 2004; Hu, Itoh et al., 2011). In *D. melanogaster* cohesin colocalizes with NIPBL to actively transcribed genes (Misulovin, Schwartz et al., 2008) and in mouse ES cells a subset of cohesin binding sites was described to colocalize with NIPBL and the mediator complex (Kagey, Newman et al.). Second, factors co-localizing with cohesin on chromatin such as CTCF (Wendt, Yoshida et al., 2008) and Estrogen receptor (Schmidt, Schwalie et al., 2010) determine where cohesin is positioned.

Mutations in NIPBL and cohesin subunits, have been linked to the “Cohesinopathy” Cornelia de Lange syndrome (CdLS, OMIM #122470, #300590 and #610759). This dominant,

genetically heterogeneous developmental disorder has a high degree of variability in its clinical presentation with multiple organ systems affected. It is estimated to occur in 1:60.000 to 1:45.000 live births. Characteristic features include craniofacial anomalies, growth retardation, intellectual disability, upper limb defects, hirsutism, and involvement of the gastrointestinal and other visceral organ systems (Liu and Krantz, 2009). Clinically, CdLS phenotypes can range from very mildly affected (no structural abnormalities, minor intellectual disability) to severely affected (upper limb defects, severe intellectual disability). Heterozygous mutations of NIPBL, ranging from nonsense and frame shift mutations to truncation mutations, have been found in 50% of CdLS patients and mutations of the cohesin subunits SMC1A and SMC3 were found in another 5% (reviewed in Liu and Baynam, 2010).

Observations in patients and mouse models show that in cells with heterozygous NIPBL mutations the NIPBL transcript levels are only reduced by ~30% due to an increased expression from the intact allele (Kawauchi, Calof et al., 2009; Liu, Zhang et al., 2009). A clinical phenotype is observed with a modest 15% reduction in expression (Borck, Zarhrate et al., 2006). This indicates that NIPBL expression levels are tightly regulated and are critical for cells. Defects in cohesin-dependent chromosome cohesion were not observed at this level of NIPBL reduction in CdLS patients or any model systems (Castronovo, Gervasini et al., 2009; Kawauchi, Calof et al., 2009). However, a reduction of cohesin binding sites was observed in cells derived from CdLS patients, which was most obvious in close proximity to genes (Liu, Zhang et al., 2009). This suggested that the clinical features of CdLS are the collective outcomes of changes in the expression level of multiple genes during development. NIPBL has already been linked to gene regulation. In *Drosophila*, NIPBL was found to facilitate the activation of the *cut* and *Ultrabithorax* genes by remote enhancers. In the case of the *cut* gene, NIPBL facilitates its long-range activation while cohesin has an inhibitory effect on *cut* expression (Rollins, Korom et al., 2004). Further, human NIPBL was already shown to bind histone deacetylases (HDAC1, 3) (Jahnke, Xu et al., 2008) and heterochromatin protein 1 (HP1) (Lechner, Schultz et al., 2005). These observations implied a “dual role” for NIPBL, in loading cohesin and in gene regulation. It is not known, whether these two functions are independent of each other, or if NIPBL mediates gene regulation via loading of cohesin onto DNA.

In this study we have aimed to determine when and where NIPBL binds to chromatin to determine where cohesin is initially loaded. Furthermore we wanted to elucidate whether the position of NIPBL binding in the genome accounts for the altered gene expression patterns observed in CdLS patients carrying NIPBL mutations (Liu, Zhang et al., 2009).

Results

Consecutive loading of NIPBL, CTCF and cohesin

To gain insight into the cohesin loading mechanism it is crucial to understand when cohesin interacts with these factors during the loading process. We therefore have compared the timing of the chromatin-localization of cohesin with that of NIPBL and CTCF.

Mitotic HeLa cells were fixed with paraformaldehyde (PFA) and immunostained with antibodies

specific for CTCF, NIPBL and the cohesin subunits RAD21 and SA2/STAG2. It was then determined at which stage the signals of these proteins appeared on chromatin during the exit from mitosis (**Figure 1**).

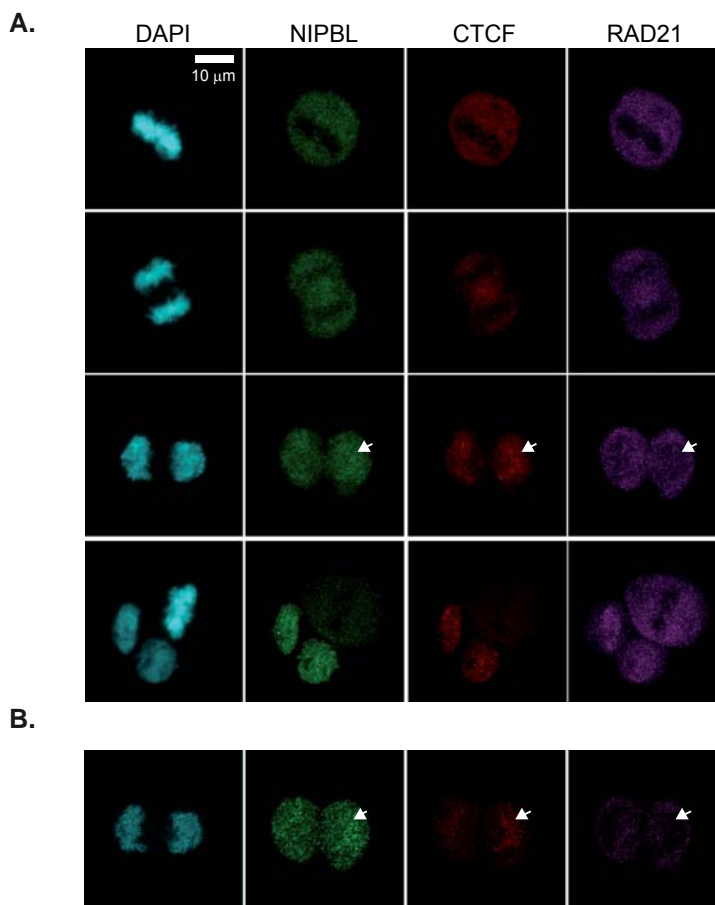


Figure 1: Chromatin association of NIPBL, cohesin and CTCF during exit from mitosis.

(A) To address the association of cohesin, CTCF and NIPBL with chromatin during end of mitosis HeLa cells were fixed with PFA and stained with antibodies against CTCF (CTCF#1), the cohesin subunit RAD21 and NIPBL (NIPBL#2). Image stacks were taken with a confocal microscope and a Z-projection generated with Image J. Different states of mitosis are shown, from top to bottom: metaphase, late anaphase, telophase, complete cytokinesis and metaphase. **(B)** One image slice (100 µm) of the telophase images in (A) is shown to highlight the lack of cohesin signal on chromatin while NIPBL and CTCF are already present.

These results were also correlated with the reassembly of the nuclear envelope in HeLa cells expressing Lamin B-EGFP. Similar to cohesin, we found the signals of NIPBL and CTCF to be (largely) excluded from metaphase chromosomes. However to our surprise NIPBL and also CTCF signals appear on chromatin at an earlier stage of mitotic exit than cohesin (**Figure 1**) and before the nuclear envelope is reassembled as shown by comparison to Lamin B signals (**Figure S1**). Therefore NIPBL and CTCF are already present on chromatin, before the cohesin complex is loaded and the productive interactions that determine the chromatin localization of cohesin might occur on chromatin.

NIPBL localizes in somatic cells independently of cohesin

To analyze the genomic localization of NIPBL-binding sites relative to cohesin and CTCF, we selected the NIPBL antibody (referred as NIPBL#1) that performs best in human cells (**Figure S2**) and performed ChIP-sequencing for NIPBL, cohesin and CTCF using HB2 cells (1-7HB2) (Bartek, Bartkova et al., 1991) enriched in G1 phase (**Figure S3A**) and for NIPBL in lymphoblastoid cells (LCL; B-cell population immortalised by EBV-transformation) derived from a normal control (N5) and CdLS patients (PT1, PT9). Furthermore, we have determined the transcriptional activity by RNA-sequencing, and identified active transcription start sites in HB2 cells by ChIP-sequencing of RNA Pol II. ChIP for NIPBL, SMC3, CTCF and RNA Polymerase II (RNA Pol II) was performed as described in Wendt, Yoshida et al., 2008, but for SMC1A ChIP a SDS-free protocol was used to maximize the ChIP-efficiency (Schmidt, Wilson et al., 2009). RNAi depletion of NIPBL greatly reduced NIPBL ChIP signals when analyzed by qPCR, confirming that NIPBL peaks are specific (**Figure S4**).

Using the criteria described in the materials and methods section, we identified 1138 NIPBL sites, 35668 CTCF sites, 22572 SMC3 sites and 29441 SMC1A sites in HB2 cells and between 1600 and 2000 NIPBL sites in lymphoblastoid cells (LCL). The data from the different LCL's and the conclusions for CdLS are discussed in detail in a later section.

Surprisingly, in HB2 cells the NIPBL binding sites do not overlap with cohesin or CTCF binding sites (**Figure 2A**). Heatmaps centred on NIPBL (**Figure 2B**), or CTCF or cohesin binding sites (**Figures 2C,D**), show no overlap of cohesin or CTCF signals with NIPBL signals. As expected, there was a high correlation between cohesin and CTCF signals. The absence of overlapping NIPBL and cohesin sites was confirmed by qPCR analysis of several NIPBL and cohesin binding sites in SMC3 and NIPBL ChIP experiments, where we observed only background levels of NIPBL binding on cohesin sites and vice versa (**Figure 2E**). The missing colocalization between NIPBL and cohesin is in contrast with observations in mouse embryonic stem cells (mouse ES) that report colocalization of cohesin with NIPBL (Kagey, Newman et al., 2010). To address this, we compared our NIPBL antibody with the one used for this study using the ChIP protocol described (Kagey, Newman et al., 2010) to exclude that different results are due to the use of different antibodies or protocols. Both antibodies perform in mouse ES cells equally well on three NIPBL binding sites at promoters identified in the study by Kagey et al (Nanog, Lefty, Oct4) (Kagey, Newman et al., 2010). We observed that several NIPBL binding sites identified in human cells are conserved in mouse ES cells (Tiam1, Ankhd1, Sp1) and also found that our antibody detects these sites more efficiently (**Figure S7**). A very notable difference between the antibodies is also observed when carrying out immunoprecipitation experiments from HeLa cells (**Figure S2B**). We are therefore confident that we do not miss any binding events compared to the other NIPBL antibody and the differences in the NIPBL binding patterns with respect to overlap with cohesin and enhancers might be explained by variations of the chromatin structure between stem cells and somatic cells. Cohesin may need to localize only transiently at NIPBL binding sites during loading and later re-localize to more permanent positions as suggested by studies in yeast (Lengronne, Katou et al., 2004).

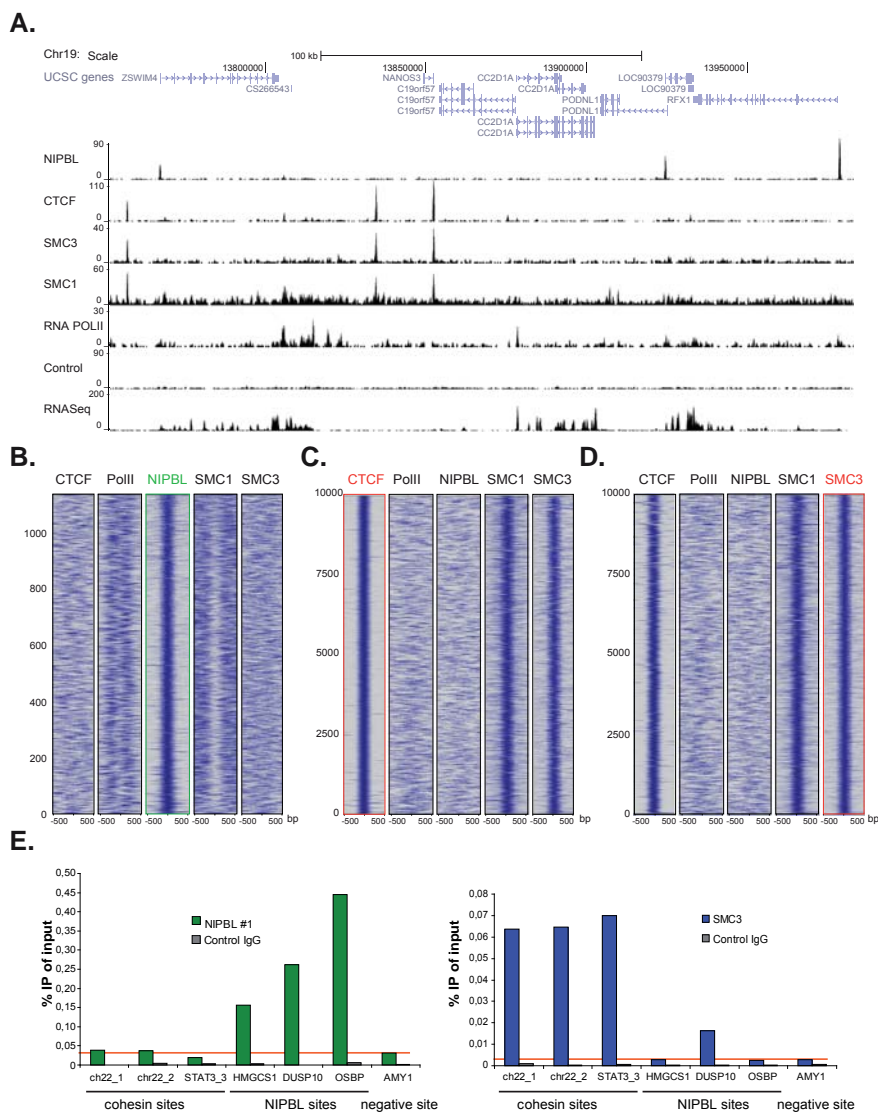


Figure 2: Binding of NIPBL, cohesin and CTCF in the human genome.

(A) Genomic binding of NIPBL, CTCF and the cohesin subunits SMC3 and SMC1A in the breast endothelial cell line HB2 at a selected region of chromosome 19 as determined by ChIP-sequencing. The RNA Pol II binding profile, the control ChIP and the RNA-sequencing data from these cells are also shown. **(B-D)** Heatmaps showing the ChIP signal intensity of the indicated ChIP-sequencing experiments in a window of +/- 500bp around all NIPBL peaks **(B)** as well as the top 10000 CTCF **(C)** and SMC3 **(D)** peaks. Number of binding sites is indicated on the y-axis. Cohesin (SMC3, SMC1A) and CTCF binding does not correlate with NIPBL binding events. RNA Pol II signals are found near NIPBL, consistent with the localization of NIPBL at promoters. Cohesin binding events correlate well between SMC3 and SMC1A and with CTCF. Peaks are ranked by size with the strongest peaks at the bottom of the graph. **(E)** ChIP was performed with NIPBL#1, SMC3 and control antibodies from HB2 cells and analyzed by qPCR with primers specific for cohesin, NIPBL and a negative (AMY) sites. NIPBL ChIP signals on cohesin sites are at background level (red horizontal line). Only the DUSP10 site is higher than the background in the SMC3 ChIP, very likely due to a CTCF/cohesin site close to the NIPBL site. All experiments were at least performed three times and one representative example is shown.

To test this we generated a HEK293T cell line expressing a doxycycline-inducible ATP-hydrolysis-deficient mutant of cohesin's SMC3 subunit SMC3E1144Q fused to a C-terminal EGFP-tag (**Figures S5A,B**). This mutant was suggested to represent a transition state of cohesin which can be loaded onto DNA but does not relocalize from the loading site (Hu, Itoh et al., 2011). Immunoprecipitation with anti-EGFP antibodies showed that SMC3E1144Q-EGFP is incorporated into the cohesin complex (**Figure S5C**). Chromatin immunoprecipitation with anti-EGFP, anti-SMC3 and anti-NIPBL antibodies and qPCR analysis showed that NIPBL binding at the tested sites is conserved in HEK293T cells. However SMC3E1144Q-EGFP localizes to known cohesin binding sites, not to NIPBL sites (**Figure S5D**). Assuming the same general properties of SMC3 mutants between yeast and human cohesin this finding indicates that cohesin is very likely not initially loaded at NIPBL sites. We have also considered that NIPBL might localize on repetitive regions since cohesin was found on centromeric repeats and Alu elements (Tanaka, Cosma et al., 1999; Waizenegger, Hauf et al., 2000; Hakimi, Bochar et al., 2002) and have analyzed the enrichment of sequencing reads mapping uniquely to the repeat sequences (**Table S8**). NIPBL is indeed highly enriched at rRNA repeats (13 fold), in particular at the large subunit (LSU, 15 fold enriched) and small subunit (SSU, 14 fold enriched) repeat families. RRNA repeats are pseudogenes of unknown function distributed all over the human genome (Jurka, Kapitonov et al., 2005). With NIPBL ChIP and qPCR primers specific for LSU and SSU repeats we confirmed four of five LSU repeat regions and one of three SSU regions (**Figure S5E**). In total we observed NIPBL only at 20 out of 467 known LSU/SSU regions (Hg19 assembly of the human genome), a small subset which is unlikely to be of importance for cohesin loading also since it does not match with the repeat classes described for cohesin.

NIPBL binds to active promoters, together with a distinct set of transcription factors

NIPBL binding sites are distributed over the entire genome (repetitive sequences were omitted during the mapping of the reads to the genome) and correlate with gene density (**Figure S6**). The binding of NIPBL is very specific, 912 of 1138 (80%) NIPBL sites in HB2 cells localize in the promoter area (\pm 1000 bp from transcription start sites) (**Figure 3A**) while only ~10% of the cohesin and CTCF sites localize to promoters. About 89% of NIPBL-bound promoters are CpG island promoters (**Table S4**). Analysis of RNA-sequencing data from HB2 cells revealed that >98% of these NIPBL-bound genes are actively transcribed (**Figure 3A** and **Table S3**), indicating a preferential binding of NIPBL to active promoters. Comparison with RNA Pol II binding sites showed that NIPBL preferentially binds 100 – 200 nucleotides upstream of RNA Pol II (**Figure 3B**). This correlation is also visible as bimodal distribution of the RNA Pol II signal (**Figure 2B**).

To analyse the properties of NIPBL-binding sites further, we used the NIPBL-binding sites observed in the control LCL's (N5) since a large number of data for histone modifications and transcription factors is available for lymphoblastoid cells like GM12878 from earlier publications (Ernst, Kheradpour et al., 2011) and ENCODE (Consortium, 2011).

Comparing the pattern of different histone modifications around NIPBL sites we observed that the sites are flanked by histone marks linked to active promoters and enhancers (H3K4me3, H3K27ac, H3K9ac) (**Figure 3C**). However, the H3K4me1 mark, characteristic for enhancers does not show an enrichment (**Figure 3C**). NIPBL itself resides apparently in nucleosome-free areas.

The missing enhancer-specific histone mark is in contrast with observations in mouse ES cells showing a colocalization of NIPBL with enhancers and cohesin (Kagey, Newman et al., 2010). Therefore we also compared the NIPBL binding with the enhancer marker p300 and the cohesin subunit RAD21 (**Figure S8B**) and observed again no correlation, indicating a difference between mouse ES cells and human somatic cells.

Motif analysis of NIPBL binding sites in HB2 cells and LCL's using MEME (Bailey and Elkan, 1994) reveals that the motifs for the transcription factor NFYA (subunit of the NF-Y complex) are present at 80% of NIPBL sites and for SP1 at 50% of the sites (**Figure 3D**). NF-Y binds the CCAAT box which correlates well with the presence of CpG islands at promoters, also a connection between NF-Y and SP1 has often been reported with presence of both motifs at the same promoter. To test whether the presence of the NFYA motif is correlated to the CpG-island promoter or a genuine property of the NIPBL-bound promoters we analyzed NIPBL-bound CpG island promoters versus randomly selected CpG island promoters and observed a statistical significant preference (Fisher test, $p < 0,001$) of NFYA for NIPBL-bound CpG island promoters. ChIP with anti NFYB antibodies from HeLa cells confirms binding of the NF-Y complex to NIPBL binding sites determined above (**Figure S8A**). To investigate whether other transcription factors colocalize specifically with NIPBL we compared the NIPBL sites in LCL's with available ChIP-sequencing data for transcription factors for GM12878 cells collected by ENCODE (Consortium, 2011).

Specifically, we analyzed in total 66 binding profiles and generated heat maps covering +/- 500 bp around NIPBL binding sites conserved in lymphoblastoid cells. By visual inspection of the maps we identified five transcription factors present on NIPBL sites: NFYA/NFYB and SP1, which is consistent with the presence of the motif, as well as PBX3, C-FOS and IRF3 (**Figure 3E**). The heatmaps displaying the signals of the other transcription factors on NIPBL binding sites show a very good correlation between all five factors. When the signals are plotted respective to NFYB sites which are sorted according to peak intensity, it shows that NIPBL and several other factors overlap only with the strongest NFY peaks (**Figure 3F**).

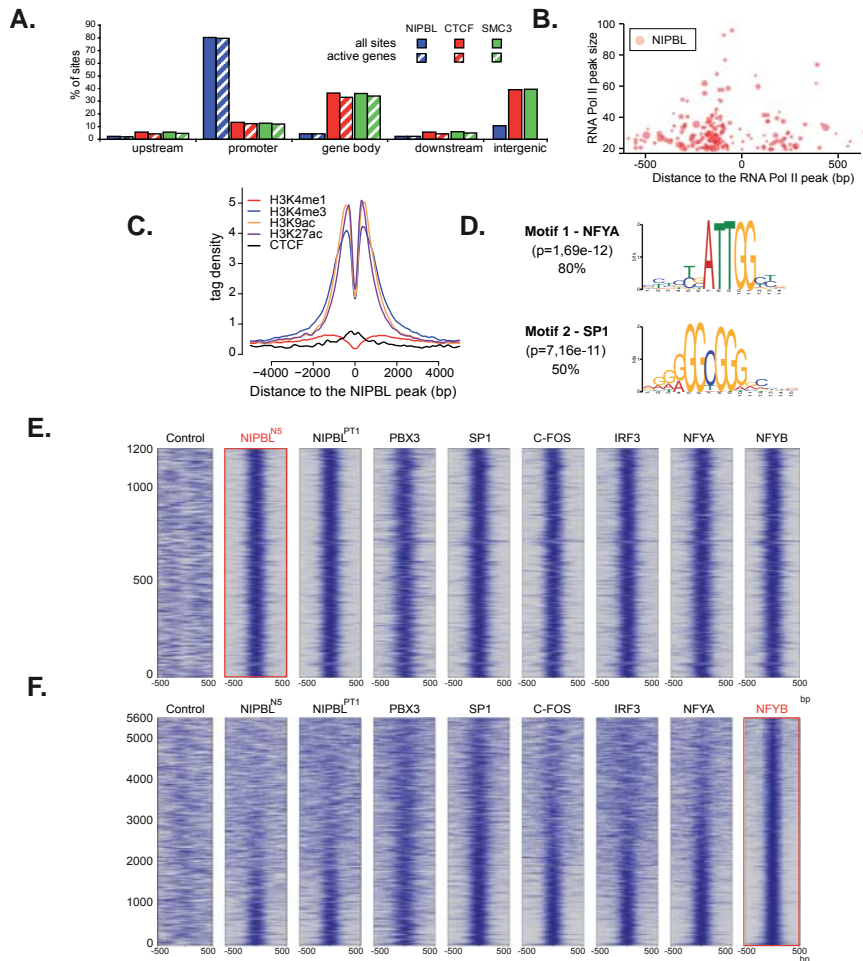


Figure 3: NIPBL binds to active promoters together with other transcription factors.

(A) Binding of NIPBL, CTCF and cohesin (SMC3) relative to active genes in HB2 cells. The different regions were defined as follows; upstream: -5 kbp to -1 kbp from transcription start sites; promoter: 1 kbp upstream and downstream from TSS; gene body: +1 kbp from TSS until end of the coding sequence; downstream: end of the coding sequence - +5 kbp (See also Table S2). **(B)** Bubble plot representation of NIPBL binding around RNA Pol II peaks in HB2 cells. The x-axis denotes the position of NIPBL relative to the closest RNA Pol II peak and the y-axis the strength of the RNA Pol II peak. Bubble size indicates the strength of the NIPBL peak. NIPBL binds 100 - 250 bp around RNA Pol II peaks, preferentially upstream, which is consistent with binding to active promoters. **(C)** NIPBL binding in the control LCL's (N5) was compared with localization of histone modifications and CTCF in the lymphoblastoid cell line GM12878 (Ernst, Kheradpour et al., 2011). The plot is centred on the NIPBL peaks and the y-axis displays the signal intensity of the respective histone modification and CTCF in GM12878 cells. **(D)** Consensus motif derived de-novo from NIPBL binding sites in HB2 cells. The region ± 50 bp around the peak maximum was used to determine motifs with MEME (Bailey and Elkan, 1994). These motifs are nearly identical to the respective motifs of the transcription factors NFYA and SP1, indicating that one or more transcription factors might colocalize with NIPBL. **(E)** Comparison of NIPBL sites in LCL's (N5) with ChIP-sequencing data of various transcription factors revealed a subset of transcription factors colocalizing with NIPBL. The heatmaps reveal a strong correlation of PBX3, SP1, C-FOS, IRF3 and NFYA/B with NIPBL sites. **(F)** Heat maps showing the correlation of the factors in (E) to NFYB sites at GpG island promoters (sites ranked according to strength with the strongest signals at the bottom). The strongest correlation with the other factors is visible for the strongest NFYB peaks. Number of binding sites is indicated on the y-axis.

NIPBL is important for gene activity

NIPBL-bound genes in HB2 cells were analyzed using IPA (Ingenuity® Systems, www.ingenuity.com) and found to be linked to very different cellular functions, such as cell cycle control, gene expression, cell death, RNA post-translational modification and control of cellular growth and proliferation (**Table S5**). Remarkably, 100 of the 912 NIPBL-bound genes (11%) might be transcription factors since they appear on the list of 1391 potential transcription factors identified by Vaquerizas et al. 2009 (Vaquerizas, Kummerfeld et al., 2009). Examples are SP1, SP2, SP3, BBX and STAT3, all are important transcription factors for development and NIPBL binding at their promoters could be important for their appropriate expression. To address whether NIPBL is important for the active transcription of the associated genes we selected functionally different genes with conserved NIPBL binding in the promoter, but no cohesin binding site close to the gene, and asked whether their transcription changes in HB2 cells after knockdown of NIPBL, MAU2 or SMC3. To avoid problems in cell division due to impaired sister chromatid cohesion we synchronized cells in G2 phase during the siRNA treatment (**Figure S3B**).

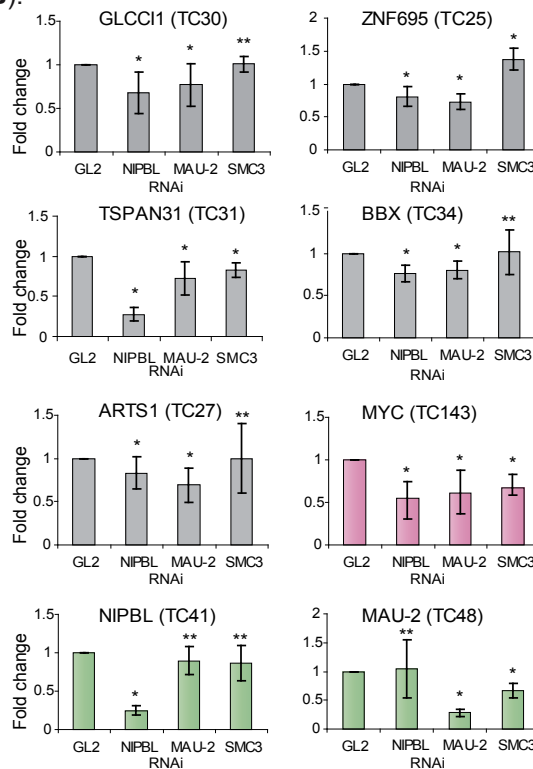


Figure 4: NIPBL is important to maintain gene activity.

Transcript levels of genes with NIPBL-bound promoters and no cohesin sites close to the gene (GLCCI1, BBX, TSPAN31, ARTS-1 and ZNF695) and the cohesin-regulated MYC gene were analyzed by RT-PCR/qPCR after RNAi depletion of NIPBL, MAU2 or SMC3 in HB2 cells. The cells were synchronized in G2 phase and the transcript levels are normalized against the housekeeping gene NAD. Transcripts of NIPBL and MAU2 were also analyzed to exclude that NIPBL affects transcription of MAU2 and vice versa. Standard deviation from three biological replicates is indicated. (P-values were determined using Students test: * significantly different, P-value < 0,05; ** not significantly different, P-value > 0,05).

The following genes were tested: *GLCC1*, a glucocorticoid inducible transcript; *TSPAN31*, encoding a transmembrane protein involved in signal transduction and growth-regulation; *BBX*, encoding a HMG-BOX transcription factor; *ZNF695*, an uncharacterized zinc-finger protein and *ARTS-1/ERAP1*, an endoplasmic reticulum aminopeptidase. Transcript levels were analysed by RT-PCR and qPCR and normalized against the housekeeping gene *NAD*. Depletion of NIPBL and also of MAU2 leads to a statistically significant (*t*-test, *P*-values <0,05) decrease of gene expression levels of the candidate genes (**Figure 4**), indicating that NIPBL and MAU2 dosage are important for maintaining expression levels. The depletion of SMC3 did not significantly reduce the expression of these transcripts, although the expression of the known cohesin-regulated *MYC* gene (Rhodes, Bentley et al., 2010) was reduced. This indicates that the changes in expression as a result of NIPBL depletion are not the indirect result of reduced cohesin binding and cohesin's role for transcription.

Insights into Cornelia de Lange Syndrome (CdLS)

Mutations in the *NIPBL* gene have been identified in approximately 50% of CdLS patients. Our discovery that NIPBL binds to active promoters prompted us to identify NIPBL binding sites in lymphoblastoid cells (LCL's) derived from blood samples of severely affected CdLS patients with NIPBL truncation mutations and normal controls (**Figures 5A,B**).

We detected 1612 NIPBL sites in the control (N5) and 2061/2009 sites in the patient-derived lines (PT1/PT9) with 1295 sites overlapping between N5/PT1 and 1273 sites between N5/PT9. In summary 80% of the sites in the control N5 are also found in PT1 and PT9 (**Figure 5C**). The majority (74%) of these conserved sites were also observed in HB2 cells, indicating conservation between different tissues. Consistent with our observations in HB2 cells, most NIPBL binding sites in the LCL's localize to the 5' ends of genes and are enriched for the motifs of the transcription factors NF-Y and/or SP1. Gene ontology analysis of the LCL NIPBL-bound genes showed similar classes of genes as for HB2 cells, but no cell type-specific functions such as immune response. Although expected from patient-derived cell lines with NIPBL haploinsufficiency, we did not observe significant differences in peak number or peak intensity between controls and patient-derived LCL's. This is explained by the rather small differences of NIPBL protein levels between CdLS patients and controls (Liu, Zhang et al., 2009) due to increased transcription from the intact allele. The ChIP-sequencing method is not quantitative and therefore small changes of NIPBL levels might not be reflected by peak intensity. To address this we performed NIPBL ChIP-qPCR from four control cell lines and four CdLS patient cell lines with primers for seven NIPBL binding sites and one cohesin binding site (negative control). QPCR revealed a reduction of the NIPBL signal between the control and patient-derived cell lines (**Figures 5D** and **S9**), but also variations among individual control and patient-derived cell lines. In general, strong NIPBL binding sites (*OSBP*, *GPR108*) seem to be more reduced than weaker binding sites. The position of NIPBL at promoters could be important for the emergence of the developmental defects seen in CdLS cases. Therefore we compared NIPBL binding sites with a list of genes found to be differentially expressed between LCL's from CdLS patients and controls (Liu, Zhang et al., 2009).

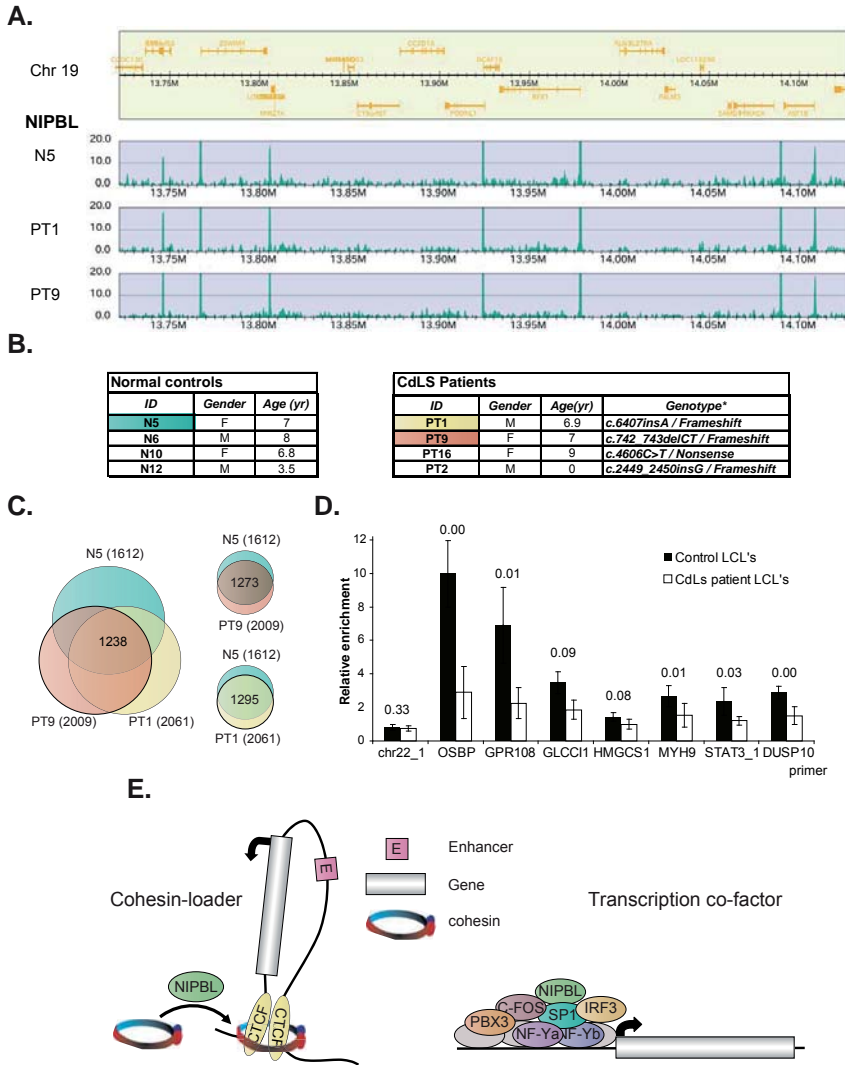


Figure 5: Position of NIPBL sites is conserved but the occupancy is reduced in CdLS.

(A) NIPBL ChIP-seencing data of a region of chromosome 19 showing that NIPBL sites are conserved between CdLS patient cells and the control. **(B)** CdLS patient and control cell lines used in this study. The cell lines highlighted were used for ChIP-seencing. The tables were derived from (Liu, Zhang et al., 2009). Nucleotide numbering refers to the NIPBL B isoform cDNA sequence with GeneBank accession number NM_015384 and starting at the +1 position of the translation initiation codon. **(C)** Venn diagrams indicating the number of NIPBL binding sites observed in the different LCL's and also the sites consistently called in all three lines. The majority of binding sites is conserved, although each cell line displays cell-line specific sites. **(D)** NIPBL binding is reduced in LCL's derived from CdLS patients. NIPBL ChIP was performed for four patient-derived cell lines and four age and gender-matched controls and qPCR analysis was performed for seven NIPBL binding sites and one cohesin site. The enrichment compared to control IgG ChIP was calculated. The data for the individual cell lines are displayed in Fig. S9. Here we present the average relative enrichment for all control and patient-derived lines, p-values derived with a Student test are indicated above the respective columns. **(E)** Model illustrating the dual role of NIPBL: NIPBL can influence gene activity by acting as “cohesin-loader” for the cohesin complexes promoting together with CTCF chromatin insulation and chromosomal long-range interactions. But NIPBL could also act directly on promoters as potential transcription co-factor in context with other colocalizing transcription factors and be required to maintain promoter activity.

We compared the list of 1501 unique genes (FDR < 0,05) found to be differentially expressed between controls and CdLS patients (Liu, Zhang et al., 2009) with our list of 1671 genes neighbouring a NIPBL site (+/- 2 kb) in the patient-derived LCL's (PT1) and found that 155 (10%) of these genes are differentially expressed (**Table S7**), a statistical significant number when compared to a random list of genes (Fisher test, $p < 0,001$).

Discussion

In its classical function, NIPBL promotes the chromatin association of the cohesin complex, but is dispensable for maintenance of the chromatin association of cohesin. Rules that regulate the place and time of cohesin loading and targeting to its various functions (sister chromatid cohesion, transcriptional regulation, mediating long-range chromatin interactions and DNA damage repair) are only partly understood. Factors interacting with chromatin-bound cohesin such as the chromatin insulator CTCF (Parelho, Hadjur et al., 2008; Rubio, Reiss et al., 2008; Wendt, Yoshida et al., 2008) and to a smaller extend estrogen receptor alpha (ERa) (Schmidt, Schwalie et al., 2010) determine the localization of cohesin, but not its general chromatin binding (Wendt, Yoshida et al., 2008). They might either direct NIPBL-dependent cohesin loading to their binding sites or capture cohesin complexes that slide along the DNA fibre.

First, we have addressed when cohesin, CTCF and NIPBL associate with chromatin. So far, only very weak and probably transient interactions have been reported between cohesin and NIPBL in the non-chromatin-bound pool of nuclear proteins (Watrin, Schleiffer et al., 2006). If these transient interactions are sufficient for NIPBL and CTCF to bind cohesin and recruit it onto chromatin, we would expect the proteins to appear on chromatin at the same time after mitosis. Analysis of cells exiting mitosis by immunofluorescence staining showed that NIPBL, CTCF and cohesin are largely excluded from metaphase chromosomes, as seen before (Wendt, Yoshida et al., 2008). The signals of NIPBL and CTCF reappear on DNA before the nuclear envelope reassembles but cohesin overlaps with chromatin only during or after the nuclear envelope reformation, reinforcing what was previously described by Gerlich et al. (Gerlich, Koch et al., 2006). NIPBL and CTCF are therefore already present when cohesin is loaded; supporting the idea that a functional interaction between NIPBL and cohesin occurs on chromatin.

Second, we determined the genomic localization of NIPBL by ChIP-sequencing in HB2 cells enriched in G1 phase and observed about 1100 NIPBL sites, mostly at active CpG-island promoters but without any overlap with cohesin or CTCF sites. Missing colocalization between NIPBL and cohesin was observed before. In yeast, non-overlapping foci were observed for Scc2 (NIPBL ortholog in *S. cerevisiae*) and Scc1 (RAD21 ortholog in *S. cerevisiae*) by immunofluorescence microscopy on spread chromatin (Ciosk, Shirayama et al., 2000). Further, a ChIP-microarray study in budding and fission yeast observed a transient overlap between cohesin and Scc2 in G1 phase cells and a subsequent relocalization of cohesin to more permanent positions between convergently transcribed genes (Lengronne, Katou et al., 2004). Another study in yeast confirmed this property of cohesin (Hu, Itoh et al., 2011) while a third study observed that colocalization of Scc2 with cohesin persists also after cohesin

loading (Kogut, Wang et al., 2009). In *D. melanogaster* the NIPBL ortholog, Nipped-b, was found to colocalize with cohesin and often overlap with RNA polymerase II, decorating entire active transcriptional units (Misulovin, Schwartz et al., 2008). Remarkably, cohesin does not colocalize with CTCF in the fruit fly. A study in mouse embryonic stem cells reported that NIPBL occupies enhancers and core promoter regions of transcriptionally active genes which are also bound by cohesin and by Mediator, a huge transcriptional co-activator complex (Kagey, Newman et al., 2010) (for review see Malik and Roeder, 2010). Although we observe a similar localization of NIPBL we did not detect cohesin at NIPBL sites, even with relaxed parameters for peak calling and different ChIP protocols. Possible explanations for the difference might be a different occupancy of chromatin by cohesin and NIPBL in embryonic stem cells than in differentiated cells. This is actually supported by the same study demonstrating that the binding of cohesin and mediator is very different between mouse ES cells and mouse embryonic fibroblasts (Kagey, Newman et al., 2010). The NIPBL antibodies used in our study recognized more NIPBL forms in HeLa cells than antibodies used by Kagey et al. (Kagey, Newman et al., 2010) (**Figure S2B**). In mouse ES cells our antibody performs similarly and better than the commercial NIPBL antibody (**Figure S7**). It is therefore unlikely that we missed NIPBL binding events due to the performance of the antibody.

Could the NIPBL-binding sites be cohesin loading sites, despite the lack of colocalization? Promoters are open chromatin regions which would allow NIPBL direct access to the DNA strand to load cohesin. In yeast it was found that the cohesin-loading site at the centromere affects cohesin-loading up to 50 kb away (Hu, Itoh et al., 2011). The same study suggested that cohesin containing ATP-hydrolysis deficient SMC1 or SMC3 mutants represents a transition state in cohesin loading that does not permit cohesin to relocalize from its loading sites (Hu, Itoh et al., 2011). The Walker B/DA box motif of SMC3 is highly conserved between yeast and human. Therefore we expressed the respective human SMC3 mutant (SMC3E1155Q) in HEK293T cells and observed binding to classical cohesin but not NIPBL sites. Thus cohesin might be loaded to its binding sites very rapidly and a very transient interaction with NIPBL during loading may not be captured by ChIP.

We have also considered repetitive DNA elements as potential loading sites for mammalian cohesin since it has been found before on centromeric repeats (Losada, Yokochi et al., 2000; Waizenegger, Hauf et al., 2000), Alu repeats (Hakimi et al. 2002) and D4Z4 repeats (Zeng, de Greef et al., 2009) and NIPBL was found to associate with D4Z4 repeats (Zeng, de Greef et al., 2009). We therefore analysed our NIPBL ChIP-sequencing data for binding to repetitive regions and found enrichment at several LSU and SSU rRNA repeat regions, but not at the repeats previously described for cohesin.

The striking localization of NIPBL to promoter of active genes suggested that NIPBL may have a direct role for the transcription of the associated genes. We observe that the transcript levels of several NIPBL-bound genes decrease after RNAi depletion of NIPBL and MAU2. An effect on the transcripts by impaired cohesin loading cannot be entirely excluded but we observe that depletion of SMC3 does not have the same effect on the transcripts. Therefore we hypothesize that NIPBL could have a role as transcription factor, independent from its function

for cohesin. A differential effect of NIPBL and cohesin has already been observed in the fly. Nipped-b facilitates activation of the cut gene, but stromalin/Scc3, the fly orthologs of the SA1/SA2 cohesin subunit, inhibits its activation. A recent study in zebrafish using morpholino knockdown observed a reduced transcription of several genes, including the transcription factors *sox17*, *foxa2* and *sox32*, after NIPBL knockdown but not in *smc3* and *rad21* morphants (Muto, Calof et al., 2011).

We also found that a large number of NIPBL-bound genes are transcription factors, 100 of 912 NIPBL-bound genes in HB2 cells localize at promoters of transcription factors. A number of them are very important during development and can also be found on the list of genes differentially expressed in CdLS, for example *STAT3* and *YBX1* (Table S8). Studies using mouse models show that the absence of these factors (*STAT3*, *YBX1*) leads to severe developmental defects and embryonic lethality (Takeda, Noguchi et al., 1997; Lu, Books et al., 2005; Lu, Books et al., 2006). NIPBL deficiency could therefore interfere with the proper timing and expression of transcription factors during development.

The observation that NIPBL might be important for gene expression lead us to ask whether NIPBL haploinsufficiency in CdLS can be linked to transcriptional changes observed in these patients. We have determined NIPBL sites in unsynchronized LCL's derived from CdLS patients with NIPBL haploinsufficiency and normal controls. These binding sites are again mostly located at CpG-island promoters, not overlapping cohesin or CTCF. The sites are in part conserved between different tissues, 85% of NIPBL sites conserved between different LCL's are also found in HB2 cells, indicating that there are constitutive and cell-type specific sites. The positions of the NIPBL binding sites are conserved between the LCL's from patients and controls, but the actual levels of NIPBL binding are reduced in patients with a hypomorphic NIPBL truncation. To link NIPBL sites to differential gene expression we compared NIPBL-bound genes identified in a patient cell line (PT1) with candidate CdLS target genes identified by Liu et al. (Liu, Zhang et al., 2009) and observed that a significant percentage (10%, Fischer test $P < 0,001$) of these genes have a NIPBL binding site. When we asked whether NIPBL RNAi affects gene expression (**Figure 4**) a subset of these genes was tested and found to be sensitive to NIPBL knockdown. This leads us to the conclusion that a part of the differentially expressed genes in CdLS could be direct targets of NIPBL, and the observed CdLS phenotype could be a cumulative effect of small changes in the transcriptional program of a larger number of genes.

Comparison of NIPBL sites in LCL's with published binding profiles of transcription factors in the lymphoblastoid cell line GM12878 revealed that NIPBL colocalizes with several transcription factors (*SP1*, *NFY*, *PBX3*, *c-FOS*, *IRF3*). *Pbx3* belongs to the Pbx family of TALE (three amino acid loop extension)-class of homeodomain transcription factors which are implicated in developmental and transcriptional gene regulation in numerous cell types. *Pbx3*-deficient mice die after birth due to neuronal malfunctions (Rhee, Arata et al., 2004). The factor is important for facial development in mice (Di Giacomo, Koss et al., 2006) together with *Pbx1* and *Pbx2* and a human *Pbx3* mutation was linked to heart defects (Arrington, Dowse et al., 2012). *IRF3* (interferon regulatory factor 3) is an IRF family transcription factor which

translocates from the cytoplasm to the nucleus upon activation, where it acts together with CBP/p300 to activate transcription of interferons alpha and beta, as well as other interferon-induced genes (for review see Yoneyama, Suhara et al., 2002). C-FOS is part of the AP-1 (activator protein 1) transcription factor complex, which also contains the JUN, ATF and MAF proteins. The complex regulates genes involved in cell proliferation, differentiation, apoptosis, angiogenesis and tumour invasion and can have oncogenic but also anti-oncogenic properties depending on cell type or differentiation state (Eferl and Wagner, 2003). How these factors functionally interact with NIPBL remains to be investigated.

In summary, we show in this study when and where NIPBL binds to the human genome and discovered that NIPBL preferentially localizes to active promoters, together with a specific set of other transcription factors. NIPBL is important for the activity of the bound genes, suggesting that NIPBL influences transcription in two ways; directly due to its binding to the promoters and indirectly by loading of cohesin complexes which then regulate genes by chromatin insulation and chromosomal long-range interactions (**Figure 5E**). The possibility that NIPBL directly affects expression suggests that NIPBL-deficiency also directly contributes to the complex CdLS phenotype by altering the transcriptional program of developmentally important genes.

Materials and methods

Antibodies

If different antibodies for the same protein were used the antibodies were numbered to clearly identify them in the different experiments.

NIPBL#1 - polyclonal rabbit anti-NIPBL antibody raised against residues 2598-2825 of the *X. laevis* Scc2-1B, purified using the epitope used for immunization (133M).

NIPBL#2 - polyclonal rabbit anti-NIPBL antibody raised against residues 787-1164 of *X. laevis* Scc-1B, purified using the epitope used for immunization (114M). Generation and characterisation of the NIPBL #1 and NIPBL #2 antibodies have been published already (Watrin, Schleiffer et al., 2006).

NIPPBL#3 - monoclonal rat anti-NIPBL, isoform A (long isoform) NP_597677 (Absea, China, 010702F01 clone KT54)

NIPPBL#4 - monoclonal rat anti-NIPBL, isoform B (short isoform) NP_056199 (Absea, China, 010516H10 clone KT55)

NIPPBL#5 - polyclonal rabbit anti-NIPBL antibody raised against a region between amino acid residues 550 and 600 of human NIPBL (Bethyl Laboratories A301-778A)

NIPPBL#6 - polyclonal rabbit anti-NIPBL antibody raised against a region between amino acid residues 1025 and 1075 of human NIPBL (Bethyl Laboratories A301-779A)

CTCF#1 - monoclonal mouse anti-CTCF (BD 612149)

CTCF#2 - polyclonal rabbit anti-CTCF antiserum (Millipore 07-729)

SA2 - monoclonal rat anti-SA2 (STAG2) antibody (Frank Sleutels and Niels Galjart)

SMC1A#1 - polyclonal rabbit anti-SMC3 antibodies (Bethyl Laboratories A300-055A)

SMC3 - polyclonal rabbit anti-SMC3 antibodies obtained from Jan-Michael Peters, described for immunoprecipitation and ChIP in (Sumara, Vorlaufer et al., 2000) and (Wendt, Yoshida et al., 2008).

MAU2 - polyclonal rabbit anti-MAU2(Scc4), described in (Watrin, Schleiffer et al., 2006).

RNA Pol II - polyclonal rabbit antibody (N-20) (Santa Cruz sc-899)

Tubulin - mouse anti-tubulin (Sigma)

Control IgG - rabbit whole serum

Rad21 - polyclonal rabbit anti-RAD21 (Jan-Michael Peters), described in (Waizenegger, Hauf et al., 2000)

Cell culture

HeLa cells were cultured in DMEM supplemented with 0.2mM L-glutamine, 100 units/ml penicillin, 100 mg/ml streptomycin and 10% FCS.

HB2 cells (1-7HB2, a clonal derivative of the human mammary luminal epithelial cell line MTSV1-7, (Bartek, Bartkova et al., 1991) were cultured in DMEM supplemented with 0.2mM L-glutamine, 100 units/ml penicillin, 100 mg/ml streptomycin, 10% FCS, 5 µg/ml hydroxycortisone and 10 µg/ml human insulin.

Lymphoblastoid cell lines derived from controls and Cornelia de Lange syndrome patients (**Figure 5B**) were obtained from Ian Krantz (The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, United States of America) and cultured in RPMI medium supplemented with 0.2mM L-glutamine, 100 units per ml penicillin, 100 mg per ml streptomycin, 20% FCS. SMC-LAP and Lamin-LAP HeLa cells were cultured in DMEM supplemented with 0.2mM L-glutamine, 100 units/ml penicillin, 100 mg/ml streptomycin and 10% FCS and 0.2 mg/ml G418.

RNAi depletion

The following siRNA oligos purchased from AMBION were used to deplete the respective proteins for ChIP-analysis and analysis of transcription

GL2

sense CGUACGCGGAUACUUCGAtt

antisense UCGAAGUAUCCGCGUACGtt

NIPBL

sense GCAUCGGUAUCAAGUCCCAtt

antisense UGGGACUUGAUACCGAUGCtt

MAU2

sense GCAUCGGUAUCAAGUCCCAtt

antisense UGGGACUUGAUACCGAUGCtt

SMC3

sense AUCGAUAAAGAGGAAGUUUtt

antisense AAACUCCUCUUAUCGAUtg

The following hairpin siRNA constructs in the pLKO.1-puro vector were obtained from the TRC Mission human library (Sigma) and were used to deplete NIPBL demonstrate the specificity of the NIPBL antibodies:

Control (clone SHC002) non-targeting sequence

NIPBL (clone TRCN0000129033) targeting sequence GCAGAGACAGAAGATGATGAA

The transfection of the siRNA oligos was performed with Lipofectamine RNAiMAX (Invitrogen) according to the manufacturer's instructions. The transfection of the hairpin siRNA constructs was performed with Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Cells were harvested 48 hours after transfection.

To perform isoform-specific RNAi hairpin siRNA oligos targeting the different isoforms were cloned in the pSUPERcherry.neo vector. The sequences are available upon request. The constructs express the hairpin constructs but also mcherry and were used to transfect HeLa cells. The cells were fixed for immunofluorescence staining 4 days after transfection.

Immunofluorescence staining

HeLa cells were grown on 18-mm cover slips and fixed with 4% PFA. After permeabilization with TX100 and blocking with 3% BSA the slides were stained with the respective antibodies.

Images were taken on a Leica DMRBE microscope equipped with a Hamatsu CCD (C4880) camera with a 100X objective. Images were processed with Image J, the colouring, overlay of the images was done with Adobe Photoshop.

Cell cycle analysis

Cells were fixed with methanol and after RNase treatment the DNA was stained with propidium iodide. The cells were analysed with a BD FACS Aria Cell sorter and FlowJo software.

Primers

See Table S6.

Immunoprecipitation

To prepare nuclear extracts from HeLa cells the cells were first lysed by gentle resuspension in hypotonic buffer (20mM Hepes-KOH pH 8.0, 5mM KCl, 1.5mM MgCl₂, 0.1mM DTT). Nuclei were collected by centrifugation and extracted for 30 min on ice with extraction buffer (15mM Tris-HCl pH 7.5, 1mM EDTA, 0.4 M NaCl, 10% sucrose, 0,01%TX-100, 1mM DTT and 1 tablet Complete (Roche) per 50 ml buffer). Debris were removed by centrifugation (14.000 rpm, 30 min).

The nuclear extract was diluted to 50% with IP buffer (20mM Tris-HCl pH 7.5, 100mM NaCl, 5mM MgCl₂, 0,2% NP40, 1mM NaF, 0,5mM DTT) and incubated for 1 h at 4°C with the respective antibodies. Affi-Prep Protein A support beads (BioRad) were added and incubated 1 h at 4°C. The beads were washed 3 times with IP buffer and proteins were eluted by boiling with SDS-page loading buffer. Western blots were analyzed with ECL+ reagent and Alliance imaging system.

Chromatin immunoprecipitation

Chromatin immunoprecipitation was performed as described before (Wendt, Yoshida et al., 2008). In brief, cells at 70–80% confluence were crosslinked with 1% formaldehyde for 10 min and quenched with 125mM glycine. After washing with PBS cells were resuspended in lysis buffer (50mM Tris-HCl pH 8.0, 1% SDS, 10mM EDTA, 1mM PMSF and Complete protease inhibitor (Roche)) and chromatin was sonicated (Diagenode Bioruptor) to around 500 bp DNA fragments. Debris were removed by centrifugation, the lysate diluted 1:4 with IP dilution buffer (20mM Tris-HCl pH 8.0, 0.15 M NaCl, 2mM EDTA, 1% TX-100, protease inhibitors) and precleared with Affi-Prep Protein A support beads (BioRad).

The respective antibodies were incubated with the lysate for 14 h at 4°C, followed by 2 h incubation at 4°C with blocked protein A Affiprep beads (Bio-Rad) (blocking solution: 0.1 mg/ml BSA or 0.1 mg/ml fish skin gelatine). The beads were washed with washing buffer I (20mM Tris-HCl pH 8.0, 0.15 M NaCl, 2mM EDTA, 1% TX-100, 0,1% SDS, 1mM PMSF), washing buffer II (20mM Tris-HCl pH 8.0, 0.5 M NaCl, 2mM EDTA, 1% TX-100, 0,1% SDS, 1mM PMSF), washing buffer III (10mM Tris-HCl pH 8.0, 0.25 M LiCl, 1mM EDTA, 0.5% NP-40, 0,5% sodium desoxycholate) and TE-buffer (10mM Tris-HCl pH 8.0, 1mM EDTA). The chromatin

from the beads were eluted twice (25mM Tris-HCl pH 7.5, 5mM EDTA, 0.5% SDS) for 20 min at 65°C. The eluates were treated with proteinase K and RNase for 1 h at 37°C and decrosslinked at 65°C over night. The samples were further purified by phenol-chloroform extraction and ethanol-precipitated. The pellet was dissolved in 50µl TE buffer.

This protocol was used to perform ChIP-qPCR or ChIP-sequencing for CTCF, SMC3, NIPBL and RNA polymerase II. For SMC1A a milder ChIP protocol from Duncan Odom's group was used (Schmidt, Wilson et al., 2009).

For NIPBL ChIP sequencing, HB2 cells were synchronized in G1 phase by a double thymidine block as described (Nativio, Wendt et al., 2009) (**Figure S8**). All other preparations were done from unsynchronized cells. For NIPBL ChIP, the cells were synchronized in G1 phase by double thymidine block, starting 6 hours after transfection of the NIPBL targeting or control siRNA oligos. Details of the thymidine block to obtain HeLa cells in G1 phase have been described (Wendt, Yoshida et al., 2008).

Samples were either submitted for genomic sequencing or analysed by qPCR using Platinum taq (Invitrogen) according to the manufacturers instructions as described (Wendt, Yoshida et al., 2008). ChIP-qPCR experiments were performed at least three times and one representative example is shown (SD was determined from qPCR replicates).

ChIP sequencing and peak detection

The ChIP DNA library was prepared according to the Illumina protocol (www.illumina.com). Briefly, 10 ng of ChIPped DNA was end-repaired, ligated to adapters, size selected on gel (200±25 bp range) and PCR amplified using Phusion polymerase as follow: 30sec at 98°C, 18 cycles of (10sec at 98°C, 30sec at 65°C, 30sec at 72°C), 5min at 72°C final extension. Cluster generation was performed using the Illumina Cluster Reagents preparation. The libraries for NIPBL, CTCF, SMC3, RNA PolII and the respective controls generated from HB2 cells were sequenced on the Illumina Genome Analyzer II, the SMC1A ChIP samples from HB2 cells, the NIPBL ChIP samples from LCLs and the respective controls were sequenced with the Illumina HiSeq2000 system. Read lengths of 36 bases were obtained. Images were recorded and analyzed by the Illumina Genome Analyzer Pipeline (GAP 1.6.0. and 1.7.0.). The resulting sequences were mapped against Human_UCSChg18 using the Bowtie (Langmead, Trapnell et al., 2009) alignment software, with the following parameters: bowtie -m 1 -S -k 1 -n 1. Unique reads were selected for further analysis.

Peak calling for the ChIP sequencing data derived from HB2 cells was performed with SWEMBL (Schmidt, Schwalie et al., 2010; Wilder, Thybert et al., manuscript in preparation) with the respective parameters described in Table S1.

Co-localization read density profiles were done by extending a region around each peak summit by +/- 200 bp. Regions from each data set were chosen in succession as viewpoints, and the position of 5'ends of the reads in corresponding regions in all data sets was plotted. The profiles were ordered by the significance score determined by the Swembl peak caller.

Peak annotation:

Complete Ensembl hg18 gene dataset was downloaded on 13.04.2011. The genome was

separated into 4 regions: promoter (+/- 1kb from the TSS), upstream (-5000 from the TSS), downstream (+5000 from the gene end) and gene body. A region of +/-150 bp was extended around each peak and overlapped with the genomic annotation.

Peaks were designated into one category based on the following order of preference: promoter -> gene body -> upstream -> downstream.

Peak calling for the ChIP sequencing data derived from CdLS patients and controls was performed as follows. We switched at this point to a different peak calling algorithm to ensure optimal comparability of the data with other datasets generated for lymphoblastoid cells in the Shirahige lab.

Sequenced reads of both ChIP and WCE were aligned to the human genome (UCSC hg18) using Bowtie (Langmead, Trapnell et al., 2009) allowing 3 mismatches. We analysed only uniquely aligned reads for calling peaks. Each aligned read was extended to a predicted fragment length of 150bp. Reads were summed in 10bp bins along the chromosome for ChIP and WCE, respectively. The read number of each bin was smoothed with a 500bp width. We scanned the genome with a 300bp sliding window, which consisted of 30 bins. One-sided Wilcoxon rank-sum test was performed to estimate the enrichment p-value for each window. The fold enrichment (ChIP/WCE) for each window was also calculated. We identified windows which satisfied both fold enrichment >3.0 and p-value <1e-4 criteria as candidate binding sites [2]. To eliminate uncertain sites, we filtered the regions with low ChIP reads. The criteria for this were: 1. the average number of ChIP reads in region / the average number of ChIP reads in the genome >3.0 or 2. the maximum read intensity in ChIP bins (less than 30 reads per a bin).

Repeat analyses

To investigate the repeat enrichment pattern, we used both uniquely- and multiply-aligned reads. Multiply-aligned reads were divided equally amongst all locations (N-times matched reads were weighted as 1/N reads). The reads which were aligned to reference genome more than 10 times were discarded. We applied RPKM measure (reads per kilobase per million reads) which was utilized for RNA-seq analyses (Mortazavi, Williams et al., 2008), but we used “per 10 million reads” instead of “per million reads”. We counted the reads which were aligned to each repeat class and normalized the counts against the total number of aligned reads (whole-genome) and the total length of each repeat class.

RNA sequencing

HB2 cells were enriched in G1 phase by double thymidine block as described (Nativio, Wendt et al., 2009). The RNA was isolated using TRI reagent (Sigma) according to the supplier's protocol. Two microgram of total RNA was converted into a library of template molecules suitable for sequencing according to the Illumina mRNA Sequencing sample prep protocol. Briefly, polyA containing mRNA molecules were purified using poly-T oligo attached magnetic beads. Following purification, the mRNA is fragmented into ~200 bp fragments using divalent cations under elevated temperature. The cleaved RNA fragments are copied into first strand

cDNA using reverse transcriptase and random primers. This is followed by second strand synthesis using DNA polymerase I and RNaseH treatment. These cDNA fragments are end repaired, a single A base is added and Illumina adaptors are ligated. The products are purified and size selected on gel and enriched by PCR. The PCR products are purified by Qiaquick PCR purification and used for cluster generation according to the Illumina cluster generation protocols (www.illumina.com). The sample was sequenced for 36bp and raw data was processed using Narwhal (Brouwer, van den Hout et al., 2011).

RNA sequencing analysis

RNA Seq reads were mapped to the Human UCSCChg18 genome with Bowtie using the same parameters as for the ChIP seq analysis. The coverage vector was calculated from unique reads and the expression value was determined for each gene by taking the RPKM (Mortazavi, Williams et al., 2008) of the most highly expressed isoform (the sum of coverage over exons was used as the numerator of the equation). All genes with RPKM > 0.6 were designated as expressed.

Motif analysis

Motif analysis was performed with the tools MEME and MEME-ChIP (Bailey and Elkan, 1994). Residues +/- 50 bp of NIPBL binding site peaks were retrieved and submitted to MEME-ChIP using standard parameters.

To analyse whether the presence of the NF-Ya motif at NIPBL sites is due to the presence of CpG islands or is a genuine property of NIPBL binding we selected NIPBL binding sites close to only one CpG island promoter (692 sites) and selected the same number of CpG island promoters at random. The presence of the NF-Ya motif was detected and the counts statistically analyzed using a Fischer-test.

Identification of colocalizing transcription factors

We obtained from ENCODE (Consortium, 2011) ChIP-sequencing data tracks for transcription factors generated from GM12878 cells and deposited by the Myers lab (HudsonAlpha Institute for Biotechnology) and the Snyder lab (Yale University). When called peaks were available they were used, else replicates were pooled and peak calling performed with MACS (Zhang, Liu et al., 2008). Peaks were sorted for intensity and for the 10000, 5000 and 1000 (in case of NIPBL) strongest peaks heatmaps were generated centred on NIPBL binding sites conserved in the different lymphoblastoid cell lines and also centred on the peaks of the respective transcription factors. Overlapping patterns were selected by visual inspection of the maps.

Myers lab (Haib)

Atf2, Atf3, Batf, Bcl1, Bcl3, Bclaf, Bhlh, Brca1, Cfos, Chd2, Ctcf, Ebf1, Egr1, Elf1, Ets1, Foxm1, Gabp, Gcn5, Irf3, Irf4, Jund, Max, Mef2, Mta3, Mxi1, Nfatc1, Nfe2, Nfic, Nfya, Nfyb, Nrf1, Nrfsf, P300, Pax5, Pbx3, Pml, Pol2, Pol3, Pou2, Pu1, Rad21, Rfx, Runx3, Rxlch, Rxra, Six5, Smc3, Sp1, Spt, Srf, Stat1, Stat3, Stat5, Tbp, Tcf1, Tcf3, Tr4, Usf1, Usf2, Whip, Yy1, Zbtb3, Zeb1, Znf143, Znf274, Zzz3

Snyder lab (SYDH)

Bhlh, Brca1, Cfos, Chd2, Ctf, Ebf1, Gcn5, Irf3, Jund, Max, Mxi1, Nfe2, Nfya, Nfyb, Nrf1, P300, Rad21, Rfx, Smc3, Spt, Stat1, Stat3, Tbp, Tr4, Usf2, Whip, Yy1, Znf143, Znf274, Zzz3

Transcript analysis

HB2 cells were transfected with the respective siRNA oligos using Lipofectamine 2000 and were harvested after 48 hours. The RNA was prepared using TRI reagent (Sigma). Remaining DNA was removed by DNase treatment and cDNA synthesis was performed with Superscript reverse Transcriptase (Invitrogen) using oligo-dT primers. The qPCR analysis was performed as described (Wendt, Yoshida et al., 2008). The transcripts of the housekeeping gene NAD were used for normalization of the samples. $\Delta\Delta\text{Ct}$ method was used to calculate the fold change in gene expression.

Acknowledgements:

We are very thankful to Koichi Tanaka and Kim Nasmyth for providing us with anti-NIPBL antibodies. Further we want to thank Mathias Madalinsky and Venugopal Bhaskara for the antibody purification. Frank Sleutels and Niels Galjart we thank for providing us with the anti-STAG2 antibodies. Jan-Michael Peters we thank for the gift of SMC3, RAD21 and MAU2 antibodies and the input for the project. Jinglan Liu we thank for helpful discussion and sharing unpublished data. Reinier van der Linden we thank for the FACS analysis. To Ina Poser and Antony A. Hyman we are grateful for providing us with the LaminA-LAP cells. This work is supported by a Dutch Royal Academy Professorship Award to Frank Grosveld. V.F. and B.L. were supported by EUTRACC.

Author Contributions:

KSW designed the study. JZ and KSW designed, performed and interpreted all experiments. AvdS and WvIJ performed next generation sequencing. MvdR performed confocal microscopy and established the SMC3 mutant cell line. VF, RK and BL analyzed and interpreted the sequencing data. IDK collected and established CdLS patient and control cell lines. KSW, JZ and VF wrote the manuscript with input and comments from BL and IDK.

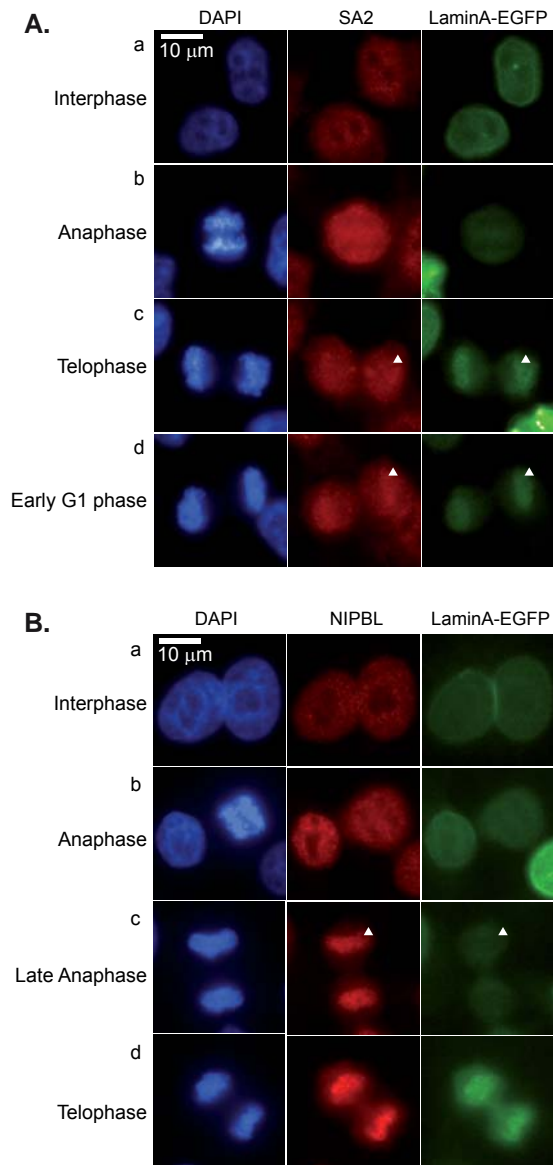
Conflict of Interests:

The authors declare that they have no conflict of interest.

Database accession:

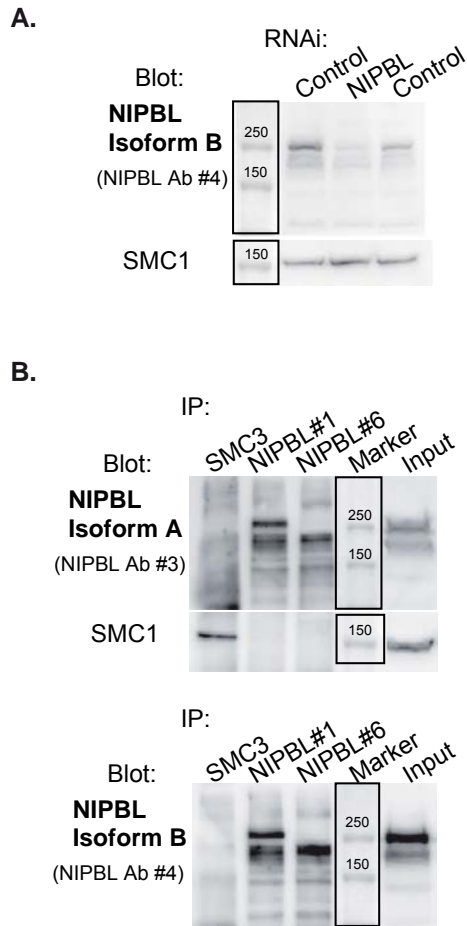
Datasets generated during this study can be obtained from the European Nucleotide Archive with the accession number ERP001494.

Supplementary figures



Supplementary figure 1: Cohesin loading occurs after nuclear envelop reformation.

Lamina-LAP expressing HeLa cells (EGFP, green) were stained with antibodies against (A) SA2/STAG2 (red) and (B) NIPBL (red). Images from different stages during the exit from mitosis were taken. In (A) the cohesin signal can only be observed overlapping with chromatin when a nuclear envelop is visible (white arrows in c and d). In contrast the NIPBL signal in (B) appears on chromatin before a nuclear envelop is visible (white arrows in c).



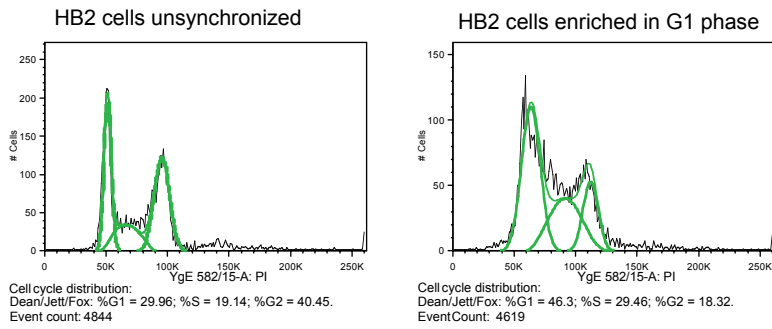
Supplementary figure 2: Characterization of NIPBL antibodies.

We first characterized different antibodies raised against NIPBL, a 320 kDa protein that is difficult to detect by immunoblotting and immunofluorescence staining. For detection by western blotting we used two rat monoclonal antibodies against the two major isoforms of NIPBL, Isoform A (NP_597677, NIPBL#3) and Isoform B (NP_056199, NIPBL#4). The isoforms are splice variants of the last exon, residues 1-2683 are identical but isoform A contains 121 and isoform B 14 unique C-terminal residues.

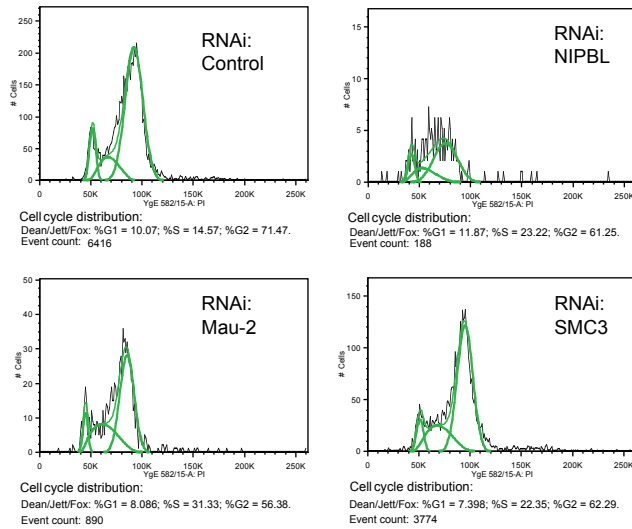
(A) Western blot showing that the band recognized by NIPBL#4 can be depleted by NIPBL-specific siRNA in unsynchronized HeLa cells while it remains well visible in two control siRNA transfections.

(B) Immunoprecipitations with the rabbit anti-NIPBL antibodies NIPBL#1 and NIPBL#6 antibodies and anti-SMC3 antibodies were performed from nuclear extract of G1-phase enriched HeLa cells. Two identical western blots were generated which were probed with rat monoclonal antibodies against the two isoforms of NIPBL (NIPBL#3 for isoform A and NIPBL#4 for isoform B) and one re-probed with anti-SMC1 (rabbit) after quenching of the rat antibody signal. Both isoform-specific antibodies detected one major (> 250 kDa) and minor NIPBL bands in the G1-phase nuclear extracts (input lane). Multiple bands for NIPBL could occur due to posttranslational modifications of NIPBL. Significant difference between NIPBL#1 and #6 are visible in the immunoprecipitates. NIPBL#1, used by us for ChIP-seq, immunoprecipitates all bands, while NIPBL#6, used by Kagey et al. (Kagey, Newman et al., 2010) for ChIP-seq from mouse ES cells, precipitates only the lower bands. We concluded that the NIPBL#1 antibody recognizes a wider spectrum of NIPBL (posttranslationally modified) forms.

Interestingly, the antibody against the cohesin subunit (SMC3) did not precipitate any of the NIPBL isoforms (Fig. 1C), consistent with previous observations of very weak interactions between NIPBL and cohesin (Watrin, Schleiffer et al., 2006).

A.**B.**

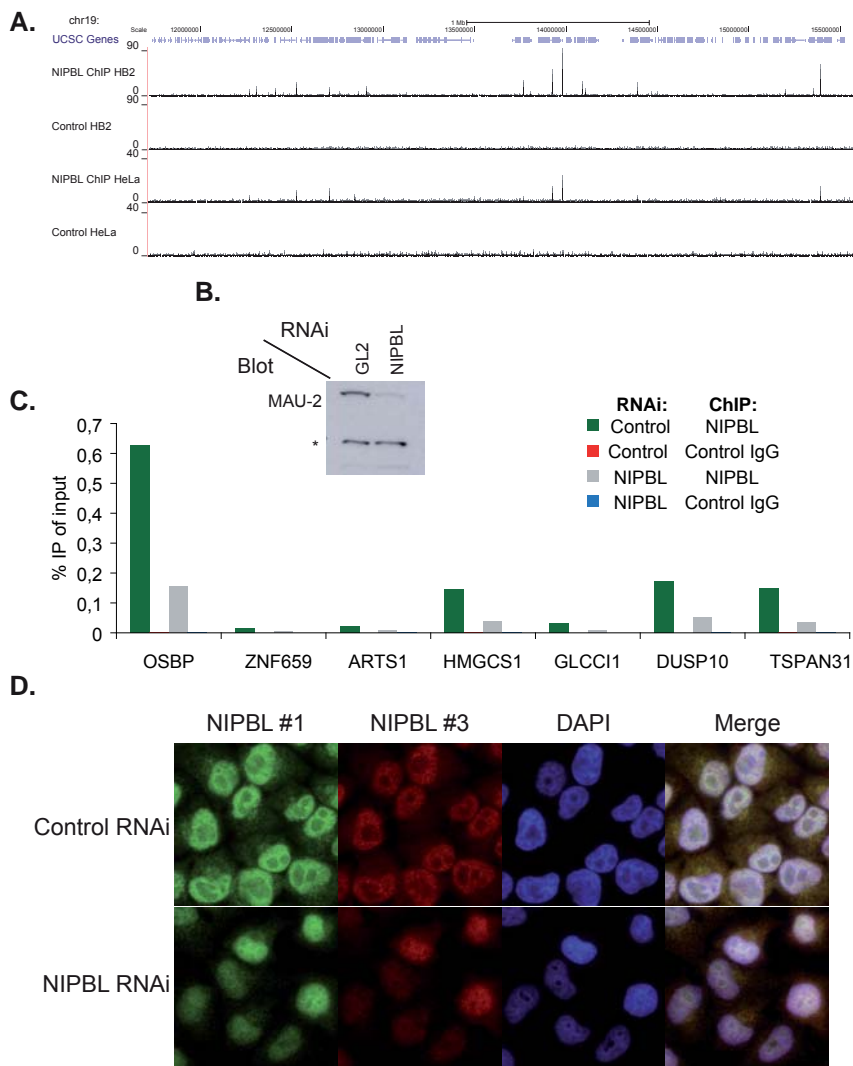
HB2 cells treated with different siRNA and enriched in G2 phase by double thymidine block



Supplementary figure 3: Determination of cell cycle stages by FACS analysis.

(A) HB2 cells growing logarithmically or enriched in G1 phase for NIPBL ChIP were fixed with methanol, stained for the DNA content with propidium iodide and analyzed by FACS.

(B) HB2 cells treated with different siRNA's were enriched in G2 phase. Cells were fixed with methanol, stained for the DNA content with propidium iodide and analyzed by FACS.



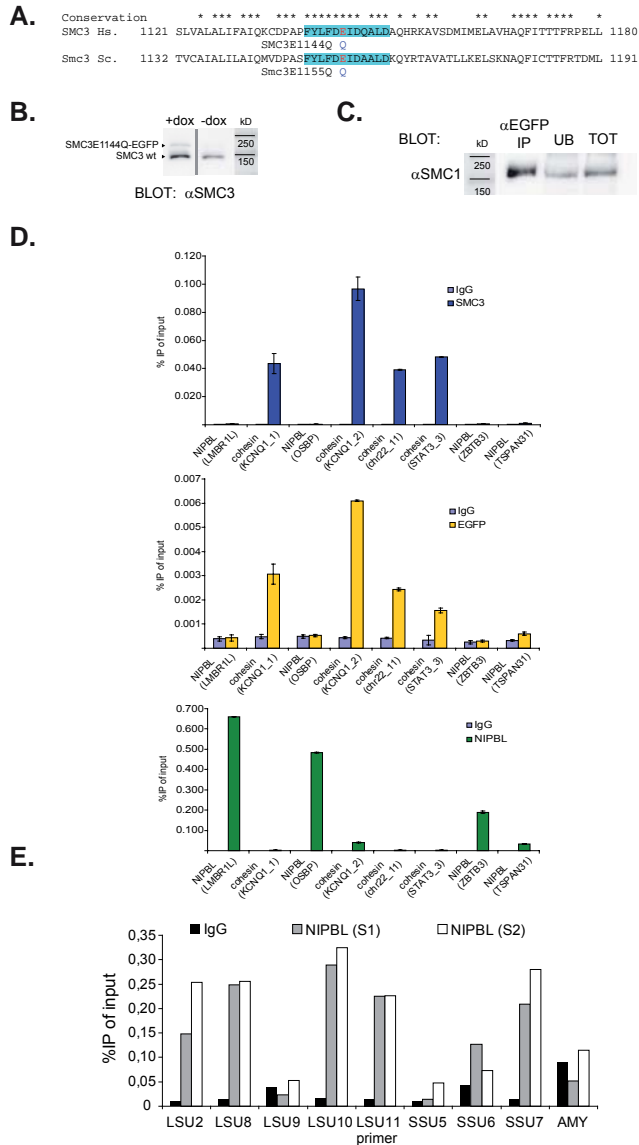
Supplementary figure 4: Specificity of the NIPBL antibody used for ChIP-sequencing.

(A) Genomic binding of NIPBL in a selected region on chromosome 19 in comparison between HB2 cells and HeLa cells. Both cell lines were enriched in G1 phase for the ChIP-sequencing experiment. The position of the peaks is similar between HB2 and HeLa cells, although the enrichment in HeLa was much weaker. As controls the sequencing data from the respective input materials are shown.

(B) Western blot showing the depletion of NIPBL in HeLa cells. Since MAU2 is also destabilized when NIPBL is depleted it can be used as marker for NIPBL depletion (Watrin, Schleiffer et al., 2006), which is rather difficult to blot. The band indicated with * is an unspecific signal of the MAU2 antibodies and can be used as loading control.

(C) NIPBL and control ChIP was performed from HeLa cells treated with NIPBL and control siRNA. QPCR analysis with primers specific for several NIPBL binding sites identified in HB2 cells shows that NIPBL RNAi dramatically reduces the NIPBL ChIP signal. The experiment was performed three times and one representative example is shown.

(D) HeLa cells were treated with control and NIPBL RNAi and stained with different antibodies against NIPBL (green – NIPBL#1, rabbit polyclonal; red - NIPBL#3, rat monoclonal) and with DAPI to visualize DNA. Both antibodies show similar reduction of the signal after NIPBL RNAi, indicating that both recognize the same target protein. The images we selected to show also cells not targeted by the siRNA to visualize the efficiency of the depletion.



Supplementary figure 5: Genomic localization of an ATP hydrolysis deficient SMC3 mutant.

(A) The SMC3E1144Q mutation of the human cohesin subunit is identical to the ATP-hydrolysis-deficient Smc3E1155Q mutation in *S. cerevisiae* described by Hu et al., 2011 as transition state mutant blocking the translocation of cohesin from its loading sites. The Walker B/DA box motifs in human and yeast SMC3 are highlighted in turquoise.

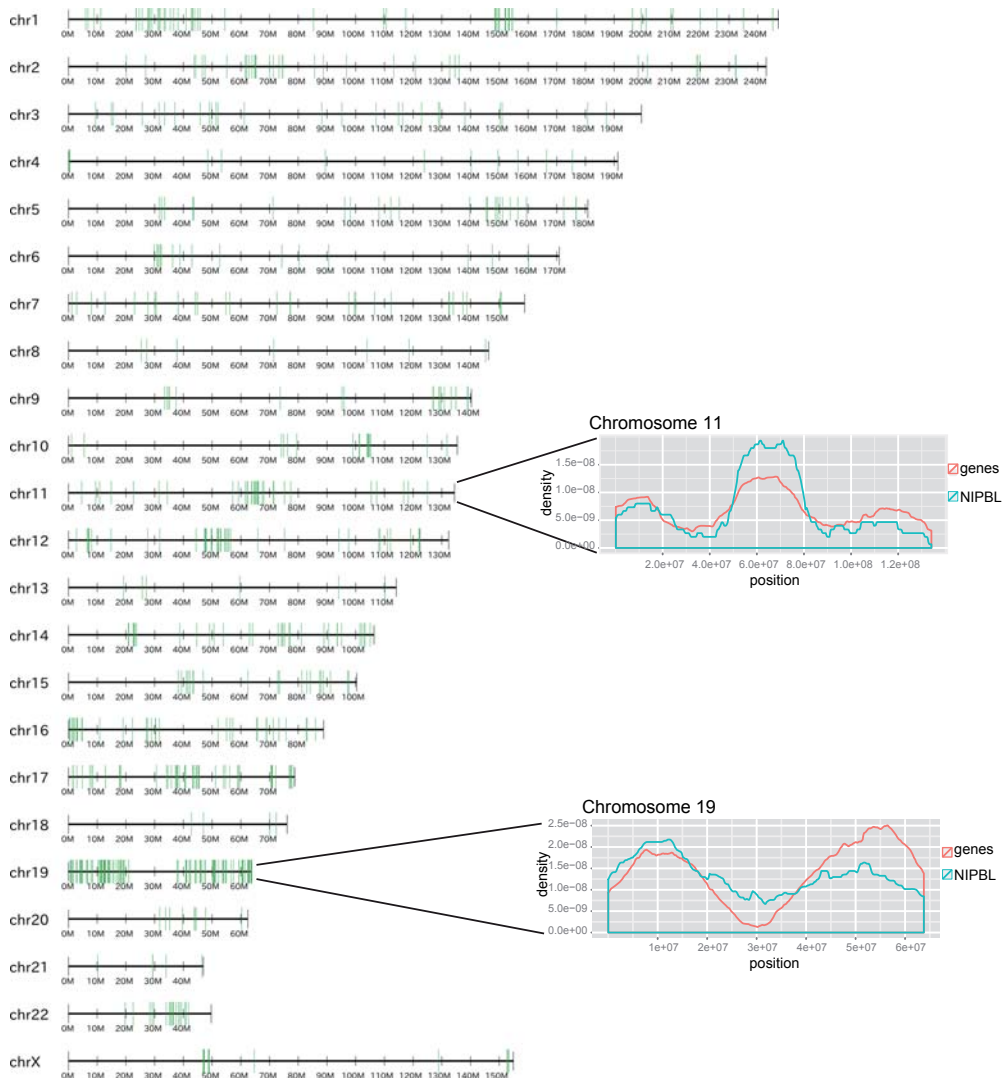
(B) A HEK293T cell line was generated that expresses SMC3E1144Q-EGFP after doxycycline (dox) induction.

(C) Immunoprecipitation of SMC3E1144Q-EGFP with anti-EGFP antibodies and blotting for the other SMC subunit of cohesin, SMC1, shows that the SMC3E1144Q-EGFP mutant is incorporated in the complex.

(D) ChIP with antibodies against SMC3, EGFP (targeting SMC3E1144Q-EGFP) and NIPBL and qPCR analysis of NIPBL and cohesin sites shows that SMC3E1144Q-EGFP binds to cohesin but not NIPBL sites. Therefore the SMC3 hydrolysis mutant might either behave different from the yeast protein or cohesin is directly loaded onto its binding sites and the identified NIPBL sites serve a different purpose.

(E) ChIP/q-PCR validation of NIPBL-binding sites on repetitive regions. The experiment was performed in duplicate and both samples are shown. Five primers for LSU repeats (LSU) and three for SSU repeats (SSU) and one negative control region (AMY) were analysed.

Distribution of NIPBL binding sites over human chromosomes 1 - 22 and X and correlation between density of genes and NIPBL binding sites on chr11 and chr19

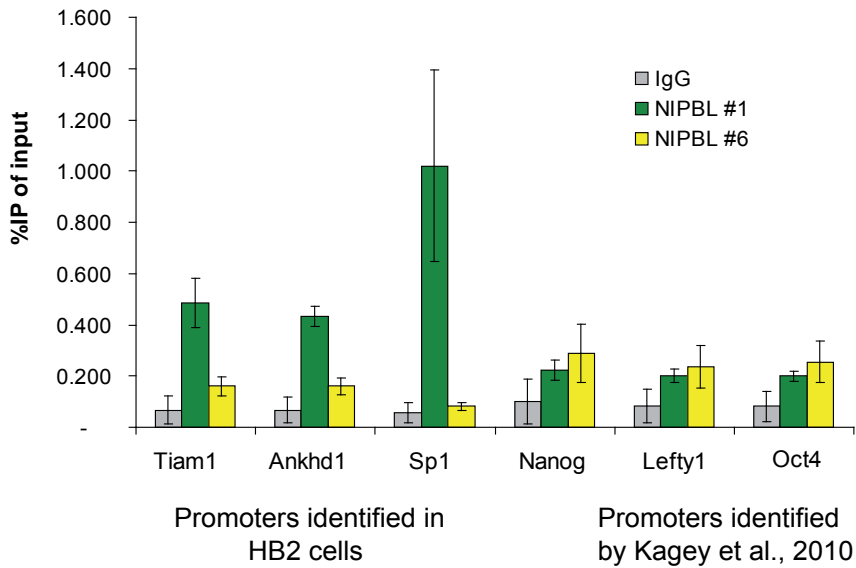


Supplementary figure 6: Clusters of NIPBL binding in the human genome.

The human chromosomes 1-22 are depicted as grey bars and the position of the NIPBL sites determined in HB2 cells are marked with green stripes. NIPBL binding sites occur in clusters and the graphs for chromosomes 11 and 19 showing NIPBL binding events together with gene density visualize that regions rich in NIPBL binding events are also gene-rich regions.

Repetitive regions in the genome were omitted from the analysis; therefore gaps between NIPBL binding events cannot be interpreted as absence of NIPBL binding in these regions.

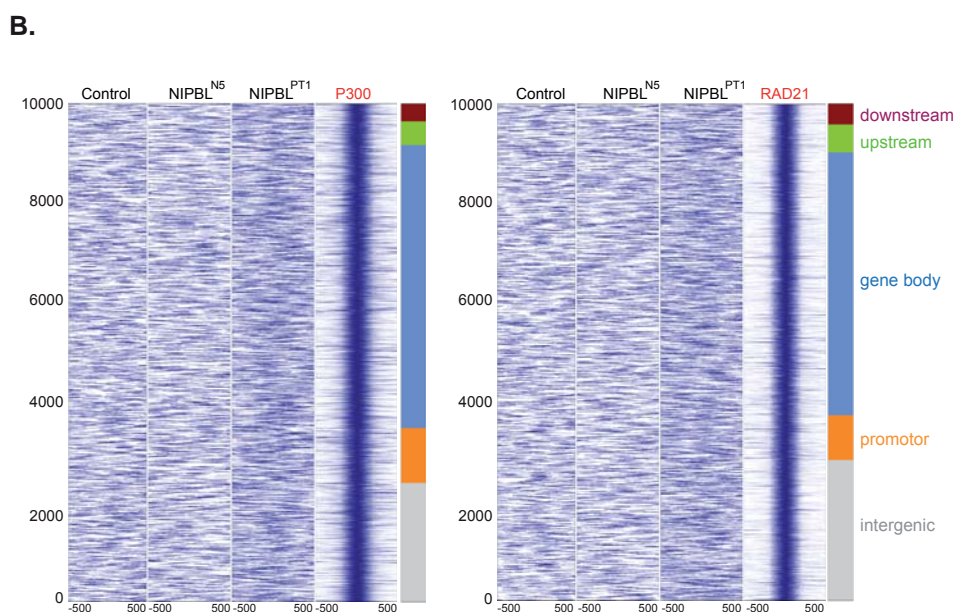
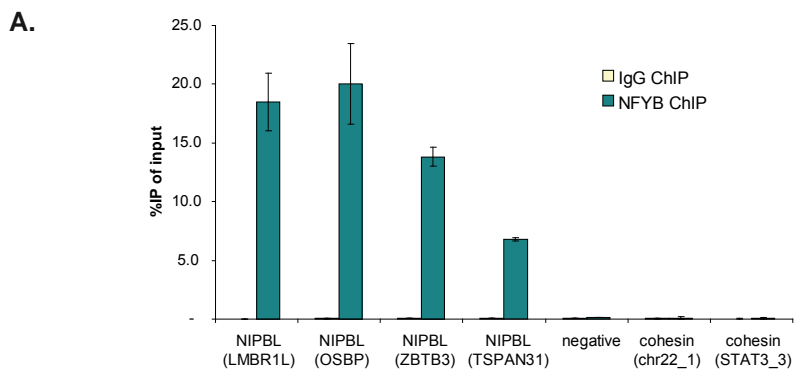
ChIP from mouse ES cells, protocol by Kagey et al., 2010



Supplementary figure 7: Comparison of the NIPBL#1 and NIPBL#6 antibodies in mouse ES cells.

To compare both antibodies under the ChIP conditions used by Kagey et al. we performed ChIP-qPCR from mouse ES cells using the described ChIP protocol. By qPCR we tested several NIPBL sites at promoters identified by Kagey et al. (Nanog, Lefty1, Oct4) and several promoters where we observed initially NIPBL binding in HB2 cells but find it conserved in mouse ES cells (Tiam1, Ankhd1, Sp1). Both antibodies perform similar on the sites found by Kagey et al., but NIPBL#1 performed much better on the sites we identified for this antibody.

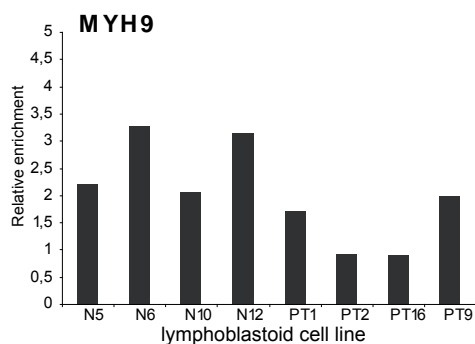
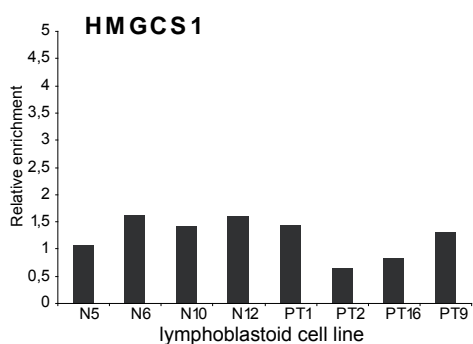
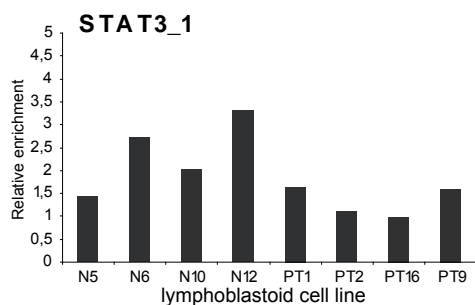
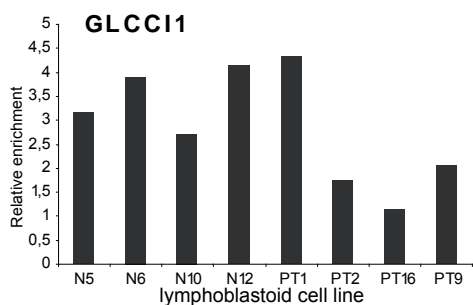
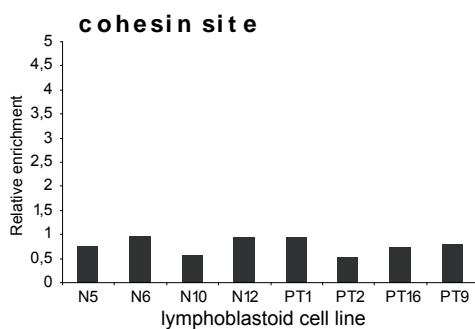
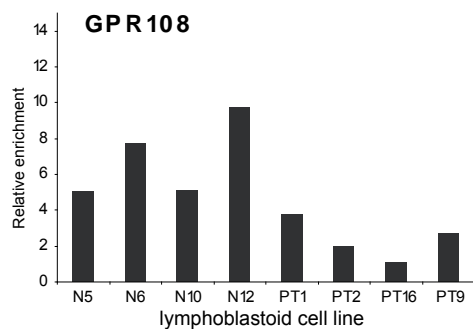
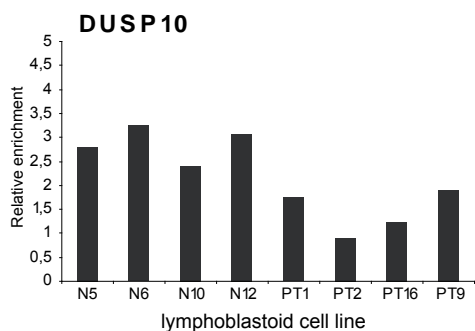
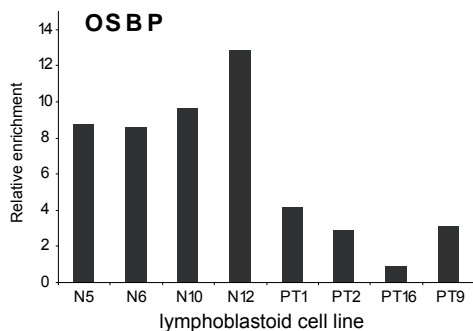
This is consistent with the observation that both antibodies show different efficiencies in immunoprecipitations from human cells (Fig. S2).



Supplementary figure 8: NFYB does localize to NIPBL site but not P300.

(A) Binding of NFYB to NIPBL sites as discovered by the motif analysis and comparison to other transcription factor datasets was confirmed by ChIP-qPCR with NFYB antibodies.

(B) In contrast to observations by Kagey et al. that NIPBL localizes to enhancers in mouse ES cells we do not observe an overlap of the signals when we generate a heatmap of P300 sites found in GM12878 cells in comparison with NIPBL data from LCL's. The same is true for RAD21 sites found in GM12878 cells.



Supplementary figure 9: NIPBL-binding is reduced in LCL cells derived from CdLS patients.

NIPBL (NIPBL#1) and negative control ChIP (IgG) was performed from lymphoblastoid cells derived from CdLS patients and age-matched controls and analyzed by qPCR with primers specific for seven NIPBL binding sites, one cohesin binding site. The sites analysed are indicated above the graph. The enrichment compared to the control IgG ChIP was calculated. The experiment was performed more than three times and a representative example is shown.

Supplementary table 1:

Parameters for peak calling with SWEMBL

	CTCF	RNA Pol II	SMC1A	SMC3	NIPBL
#Sequence length	36	36	36	36	36
#Fragment length	150	150	150	150	150
#Background	0.000000	0.000000	0.000000	0.000000	0.000000
#Position Background	0.010000	0.010000	0.010000	0.010000	0.010000
#Long Background	0.050000	0.050000	0.050000	0.050000	0.050000
#Threshold	5.000000	5.000000	5.000000	5.000000	5.000000
#Minimum count above bg	10	10	15	8	10
#Penalty increase	25	25	25	25	25
#Quality cutoff	0.000000	0.000000	0.000000	0.000000	0.000000
#Result cutoff	4.000000	4.000000	8.000000	4.000000	4.000000
#Penalty factor	1.000000	1.000000	1.000000	1.000000	1.000000

Supplementary table 2:

A. Peaks identified with SWEMBL in the different datasets

	CTCF	SMC1A	SMC3	RNA Pol II	NIPBL
Total number of peaks	35672	29448	22573	9897	1138
Number of peaks associated with genes*	21743	16501	13657	9148	1016
% of peaks associated with genes	61	56	61	92	89
Number of peaks associated with Expressed genes	14501	10775	9801	8802	953
% of peaks associated with Expressed genes	41	37	43	89	84

* as peaks were counted as associated with genes when they matched to one of the following groups:

upstream: -5000 to -1000 bp from transcription start site

promoter: -1000 to +1000 bp around transcription start site

gene body: +1000 bp from transcription start site until end of the coding sequence

downstream: end of the coding sequence + 5000 bp

B. Overlap statistics (%)

Percent overlap in a window of +/- 100 bp:

Total number of sites		CTCF	SMC1A	SMC3	NIPBL	SMC1A+SMC3
CTCF	35672		51.4	38.3	0.3	59.3
SMC1A	29448	62.3		39.2	0.2	100.0
SMC3	22573	60.5	51.1		0.4	100.0
NIPBL	1138	9.8	4.3	8.4		10.4
SMC1+SMC3	40465	52.3	72.8	55.8	0.3	

Supplementary table 4:

Classification of promoters and expression status of genes bound by NIPBL¹, cohesin (SMC1A and SMC3), RNA Pol II and CTCF

(¹ the identical definition of the NIPBL-gene association as in supplementary table 3)

		NIPBL		
		CpG island in promoter		
		no	yes	
Expression	no	27	36	Total: 63
	yes	85	868	953
Total sites:	1138	Sites around genes:		1016

		SMC1A		
		CpG island in promoter		
		no	yes	
Expression	no	3102	2624	Total: 5726
	yes	1816	8959	10775
Total sites:	29441	Sites around genes:		16501

		SMC3		
		CpG island in promoter		
		no	yes	
Expression	no	2132	1724	Total: 3856
	yes	1578	8223	9801
Total sites:	22572	Sites around genes:		13657

		RNA Pol II		
		CpG island in promoter		
		no	yes	
Expression	no	149	197	Total: 346
	yes	710	8092	8802
Total sites:	9879	Sites around genes:		9148

		CTCF		
		CpG island in promoter		
		no	yes	
Expression	no	3894	3348	Total: 7242
	yes	2234	12267	14501
Total sites:	35668	Sites around genes:		21743

Supplementary table 5:

Molecular function of genes with NIPBL binding sites in HB2 cells identified by IPA analysis (IPA version 9.0, IGENUITY SYSTEMS)

Category	Genes	p-value
<i>Cell Cycle</i>	105	2,17E-08-3,56E-02
<i>Gene Expression</i>	147	4,83E-08-3,59E-02
<i>Organismal Development</i>	84	1,68E-07-3,64E-02
<i>RNA Post-Transcriptional Modification</i>	30	7,3E-07-1,43E-02
<i>Cell Death</i>	165	1,6E-06-3,66E-02
<i>Cellular Growth and Proliferation</i>	108	3,14E-05-2,9E-02
<i>Infectious Disease</i>	62	3,42E-05-2,9E-02
<i>Embryonic Development</i>	54	6,9E-05-3,46E-02
<i>Cardiovascular System Development and Function</i>	18	6,93E-05-3,35E-02
<i>Cancer</i>	220	7,32E-05-3,74E-02
<i>Developmental Disorder</i>	39	1,09E-04-3,79E-02
<i>Hematological System Development and Function</i>	9	1,3E-04-1,63E-02
<i>Hematopoiesis</i>	13	1,3E-04-2,07E-02
<i>Renal and Urological Disease</i>	33	1,55E-04-3,74E-02
<i>Dermatological Diseases and Conditions</i>	27	1,96E-04-3,79E-02

Supplementary table 6:

Primers used for ChIP/qPCR and RT-PCR/qPCR

ChIP/qPCR primer

Name	forward	reverse	Position (hg18)
Cohesin binding sites			
chr22_11	GGCTCAGGACAGAAGTGACC	AGGTCGGCAGAGGCTCC	chr22:31255080-31255266
chr22_12	ACATGTGGCCAGCTCAGG	GGCGCTATAAGCCAGAGAAC	chr22:31522929-31523153
STAT3_3	CCTCCCTTACCCTGATGTC	CTAAGCCTCCAGGCACCTTC	chr17:37728538-37728637
KCNQ1_1	AGGACCCGCATGAGGGATTCC	CCGCATATCCCTGGTTCAGC	chr11:2510878-2511089
KCNQ1_2	AAAGGTGTTTGACGCCCTTG	AACATCGTGTCTCGGAATC	chr11:2510713-2510912
NIPBL binding sites			
OSBP	GCTGCTGTTCCGCCATTCAATTC	GCTGATACCAACCACCAATCCATGAG	chr11:59140122-59140236
HMGCS1	AATGACGGAGCTCGGAGATAC	TTGCTAGGCAACCTGACAGAC	chr5:43349567-43349749
DUSP10	TCGGCTTCATTGATCTCCAG	AGAGCAGCTTCGGATAAACC	chr1:219982022-219982252
STAT3_1	CCTCCCTTACCCTGATGTC	CTAAGCCTCCAGGCACCTTC	chr17:37728538-37728637
GPR108	GGAGAGCCGATAACGCTTAAC	GCCGGATATCAGATCCATGAC	chr19:6688766-6688880
GLCCI1	CACCTACCTCAAGGCTGTCAAC	GCGAGCATTTCCATTGGCTTAG	chr7:7974693-7974884
MYH9	GCAGACGTATTGGCCTGTGG	GCCGCCTCCTGATTGGATTCC	chr22:35114119-35114326
STAT3_1	CCGAAACCCTGAATTAC	TAGCTGCTCTCCTCATTGG	chr17:37794031-37794204
ARTS1	TAGCGTTGCGAGGGTTAGG	GCACCTTGCCACGACAGAG	chr5:96169763-96169976
ZNF695	TCTGGCTGCACGCCTGATTG	GGTCGAAGTGCCGTGAGGAATG	chr1:245238016-245238199
TSPAN31	ATGCGGGTTGATTGGCATGCAG	AGGCCAGCCAAACCAATACTC	chr12:56424994-56425106
LMBR1L	CTCTCCAGGAGCCAATGAGTTC	CGCGCTCACGTTTCAATG	chr12:47,790,854-47,790,999
Control			
AMY	TCTGCTGGGCTCAGTATTCTC	TGTTGCCAAGCTTCACGTAG	chr1:104000510-104000730
Repeats			
LSU_2	CAGGTGCAGATGTTGGTGGTAG	GAACCCGACTCCCTTTCGATTG	chr1:107914874-107915060
LSU_8	GCCGCCACAAGCCAGTTATC	AACCTCCCGTGGAGCAGAAG	chr1:91625487-91625643
LSU_9	ACTCAGGATTGGCGTCTTCG	AGCAGTTGAACACGGGTCAG	chr2:229753693-229753871
LSU_10	TGCGGTAACGTGACCGATCC	TTTCAAGGGCCAGCGAGAGC	chr5:71182503-71182673
LSU_11	GCTCAACAGGGTCTTCTTTCC	AGCCAAATGCCTCGTCATC	chr8:70764913-70765053
SSU_5	GCTGGATAGCTGGATAGCTGTC	GAAGGGCAACTGCTTGAATG	chr15:86163096-86163239
SSU_6	GCCGGACTTATCAAGGCAAAC	ACAGGTCTGTGACGCTCATAG	chr5:174474329-174474548
SSU_7	GTAGGGCAAGTCTGGTGCCAAC	CTGCTCCCAGATCCAACACTAG	chr7:80800911-80801011

RT-PCR/qPCR primer

Name	forward	reverse	Gene
TC27	CATGGTGTGAGAGCACTGAAG	GGTGACTATGAGGATGATGG	ARTS1
TC25	AGCGGAGTTTGTATAGGG	GTTTGGCTGTCTTCTCTG	ZNF695
TC30	AGGCGAACCTCCTCTTTGG	GAGGCAGAAGCTGATGTG	GLCCI1
TC31	TGTCCAGCATCCACATCATC	TAGCCAGACATGAGCAAGAG	TSPAN31
TC34	GGAATCCTGCTGCATCAC	TCGTCTCTCTTCTTCC	BBX
TC41	AAGCAGTGGCTGGTATGAAG	TGAACACAAGCGCTAGAGG	NIPBL
TC48	CACCTTGAGCCAGGAGATTG	TGCCTTTCTGCGACCTTG	MAU2
TC35	ACTGGCTACTGCGTACATCC	AGATGCGCCTATCTTCTCC	NADH
TC143	CCAGCAGCGACTCTGAGGAGGAACAAG	GTTTGTGTGGCTCCAGCAGAAGGTG	MYC

Supplementary table 7:

Genes with NIPBL binding sites in LCL's (PT1) found to be differentially expressed in CdLs patient cells by Liu et al. (2009) with FDR < 0.05

ID	Entrez Gene Name	Location	Type(s)
Symbol	Entrez Gene Name	Location	Type(s)
ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1	Plasma Membrane	transporter
ABHD2	abhydrolase domain containing 2	unknown	enzyme
ADSL	adenylosuccinate lyase	Cytoplasm	enzyme
AKAP13	A kinase (PRKA) anchor protein 13	Cytoplasm	other
AMD1	adenosylmethionine decarboxylase 1	Cytoplasm	enzyme
ANAPC5	anaphase promoting complex subunit 5	Nucleus	enzyme
ANKRD11	ankyrin repeat domain 11	Nucleus	other
AP3D1	adaptor-related protein complex 3, delta 1 subunit	Cytoplasm	transporter
ARRB2	arrestin, beta 2	Cytoplasm	other
ATP6V1E1	ATPase, H+ transporting, lysosomal 31kDa, V1 subunit E1	Cytoplasm	transporter
ATPIF1	ATPase inhibitory factor 1	Cytoplasm	other
BBX	bobby sox homolog (Drosophila)	Nucleus	other
BSDC1	BSD domain containing 1	unknown	other
BUB3	budding uninhibited by benzimidazoles 3 homolog (yeast)	Nucleus	other
C16orf46	chromosome 16 open reading frame 46	unknown	other
C2orf44	chromosome 2 open reading frame 44	unknown	other
CAT	catalase	Cytoplasm	enzyme
CCDC12	coiled-coil domain containing 12	unknown	other
CCDC130	coiled-coil domain containing 130	unknown	other
CCR7	chemokine (C-C motif) receptor 7	Plasma Membrane	G-protein coupled receptor
CDKN3	cyclin-dependent kinase inhibitor 3	Nucleus	phosphatase
CDR2	cerebellar degeneration-related protein 2, 62kDa	Cytoplasm	other
CETN2	centrin, EF-hand protein, 2	Nucleus	enzyme
CHAF1A	chromatin assembly factor 1, subunit A (p150)	Nucleus	other
CHCHD7	coiled-coil-helix-coiled-coil-helix domain containing 7	Cytoplasm	other
COPS3	COP9 constitutive photomorphogenic homolog subunit 3	Cytoplasm	other
COPZ1	coatamer protein complex, subunit zeta 1	Cytoplasm	transporter
COTL1	coactosin-like 1 (Dictyostelium)	Cytoplasm	other
COX5B	cytochrome c oxidase subunit Vb	Cytoplasm	enzyme
CREB3L2	cAMP responsive element binding protein 3-like 2	Nucleus	other
CS	citrate synthase	Cytoplasm	enzyme
CYLD	cylindromatosis (turban tumor syndrome)	Nucleus	transcription regulator
DCTN4	dynactin 4 (p62)	Nucleus	other
DDIT3	DNA-damage-inducible transcript 3	Nucleus	transcription regulator
DDX11/DDX	DEAD/H (Asp-Glu-Ala-Asp/His) box helicase 11	Nucleus	enzyme
DMXL1	Dmx-like 1	Extracellular Space	other
DUSP22	dual specificity phosphatase 22	Cytoplasm	phosphatase
DYNLRB1	dynein, light chain, roadblock-type 1	Cytoplasm	other
EBP	emopamil binding protein (sterol isomerase)	Cytoplasm	enzyme
EIF1	eukaryotic translation initiation factor 1	unknown	translation regulator
ELMO2	engulfment and cell motility 2	Cytoplasm	other
EMC2	ER membrane protein complex subunit 2	Nucleus	other
EXOC4	exocyst complex component 4	Cytoplasm	transporter
F11R	F11 receptor	Plasma Membrane	other

FAM102A	family with sequence similarity 102, member A	unknown	other
FAM111B	family with sequence similarity 111, member B	unknown	other
FASN	fatty acid synthase	Cytoplasm	enzyme
FBXO8	F-box protein 8	Cytoplasm	other
FCHO1	FCH domain only 1	Plasma Membrane	other
FDFT1	farnesyl-diphosphate farnesyltransferase 1	Cytoplasm	enzyme
GART	phosphoribosylglycinamide formyltransferase	Cytoplasm	enzyme
GCSH	glycine cleavage system protein H (aminomethyl carrier)	Cytoplasm	enzyme
GINS2	GINS complex subunit 2 (Psf2 homolog)	Nucleus	other
GLCC1	glucocorticoid induced transcript 1	Cytoplasm	other
GLS	glutaminase	Cytoplasm	enzyme
GOLGB1	golgin B1	Cytoplasm	other
GRPEL1	GrpE-like 1, mitochondrial (E. coli)	Cytoplasm	other
GTF2A1	general transcription factor IIA, 1, 19/37kDa	Cytoplasm	transcription regulator
GTPBP3	GTP binding protein 3 (mitochondrial)	Cytoplasm	enzyme
HDAC1	histone deacetylase 1	Nucleus	transcription regulator
HIPK1	homeodomain interacting protein kinase 1	Nucleus	kinase
HIPK1	homeodomain interacting protein kinase 1	Nucleus	kinase
HMGCS1	3-hydroxy-3-methylglutaryl-CoA synthase 1 (soluble)	Cytoplasm	enzyme
INTS3	integrator complex subunit 3	Nucleus	other
ITFG2	integrin alpha FG-GAP repeat containing 2	unknown	other
KDM4B	lysine (K)-specific demethylase 4B	unknown	other
KHSRP	KH-type splicing regulatory protein	Nucleus	enzyme
KIF22	kinesin family member 22	Nucleus	other
LAPTM4A	lysosomal protein transmembrane 4 alpha	Cytoplasm	other
LARP4B	La ribonucleoprotein domain family, member 4B	Cytoplasm	other
LAS1L	LAS1-like (S. cerevisiae)	Nucleus	other
MCM3	minichromosome maintenance complex component 3	Nucleus	enzyme
MCM5	minichromosome maintenance complex component 5	Nucleus	enzyme
MIA3	melanoma inhibitory activity family, member 3	Cytoplasm	other
MIB1	mindbomb E3 ubiquitin protein ligase 1	Cytoplasm	other
MMAB	methylmalonic aciduria (cobalamin deficiency) cblB type	Cytoplasm	enzyme
MSH2	mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli)	Nucleus	enzyme
MTMR4	myotubularin related protein 4	Cytoplasm	phosphatase
NAP1L1	nucleosome assembly protein 1-like 1	Nucleus	other
NCAPD3	non-SMC condensin II complex, subunit D3	Nucleus	other
NFE2L1	nuclear factor (erythroid-derived 2)-like 1	Nucleus	transcription regulator
NGRN	neugrin, neurite outgrowth associated	Nucleus	other
NR1H2	nuclear receptor subfamily 1, group H, member 2	Nucleus	ligand-dependent nuclear receptor
NR2C2	nuclear receptor subfamily 2, group C, member 2	Nucleus	ligand-dependent nuclear receptor
NUP98	nucleoporin 98kDa	Nucleus	transporter
OGDH	oxoglutarate (alpha-ketoglutarate) dehydrogenase (lipoamide)	Cytoplasm	enzyme
ORC2	origin recognition complex, subunit 2	Nucleus	other
OS9	osteosarcoma amplified 9, endoplasmic reticulum lectin	Nucleus	other
OSBPL3	oxysterol binding protein-like 3	Cytoplasm	other
OXR1	oxidation resistance 1	Cytoplasm	other
P2RX1	purinergic receptor P2X, ligand-gated ion channel, 1	Plasma Membrane	ion channel
PBXIP1	pre-B-cell leukemia homeobox interacting protein 1	Nucleus	transcription regulator
PCCB	propionyl CoA carboxylase, beta polypeptide	Cytoplasm	enzyme
PHPT1	phosphohistidine phosphatase 1	Cytoplasm	phosphatase
PIAS3	protein inhibitor of activated STAT, 3	Nucleus	transcription regulator
PJA2	praja ring finger 2, E3 ubiquitin protein ligase	Cytoplasm	enzyme
PLCG1	phospholipase C, gamma 1	Cytoplasm	enzyme
PNN	pinin, desmosome associated protein	Plasma Membrane	other
POLD2	polymerase (DNA directed), delta 2, accessory subunit	Nucleus	enzyme

POLR3K	polymerase (RNA) III (DNA directed) polypeptide K, 12.3 kDa	Nucleus	transcription regulator
PPP3CA	protein phosphatase 3, catalytic subunit, alpha isozyme	Cytoplasm	phosphatase
PSAT1	phosphoserine aminotransferase 1	Cytoplasm	enzyme
PTMA	prothymosin, alpha	Nucleus	other
RAB5B	RAB5B, member RAS oncogene family	Cytoplasm	enzyme
RCL1	RNA terminal phosphate cyclase-like 1	Nucleus	enzyme
RFWD3	ring finger and WD repeat domain 3	Nucleus	enzyme
RHEB	Ras homolog enriched in brain	Cytoplasm	enzyme
RPS6	ribosomal protein S6	Cytoplasm	other
RRAGC	Ras-related GTP binding C	Cytoplasm	enzyme
RUNX3	runt-related transcription factor 3	Nucleus	transcription regulator
RUVBL1	RuvB-like 1 (E. coli)	Nucleus	transcription regulator
SEC22C	SEC22 vesicle trafficking protein homolog C (S. cerevisiae)	Cytoplasm	other
SEC31A	SEC31 homolog A (S. cerevisiae)	Cytoplasm	other
SEC61G	Sec61 gamma subunit	Cytoplasm	transporter
SFXN5	sideroflexin 5	Cytoplasm	transporter
SLC25A37	solute carrier family 25	Cytoplasm	transporter
SLC25A42	solute carrier family 25, member 42	Cytoplasm	transporter
SLC39A3	solute carrier family 39 (zinc transporter), member 3	Plasma Membrane	transporter
SNRNPB	small nuclear ribonucleoprotein polypeptides B and B1	Nucleus	other
SNX11	sorting nexin 11	unknown	transporter
SORT1	sortilin 1	Plasma Membrane	transmembrane receptor
STAT2	signal transducer and activator of transcription 2, 113kDa	Nucleus	transcription regulator
STAT3	signal transducer and activator of transcription 3	Nucleus	transcription regulator
STT3B	STT3, subunit of the oligosaccharyltransferase complex	Cytoplasm	enzyme
SUCO	SUN domain containing ossification factor	unknown	other
TCOF1	Treacher Collins-Franceschetti syndrome 1	Nucleus	transporter
TLE4	transducin-like enhancer of split 4 (E(sp1) homolog, Drosophila)	Nucleus	transcription regulator
TMEM9	transmembrane protein 9	Cytoplasm	other
TMEM97	transmembrane protein 97	Extracellular Space	other
TMTC4	transmembrane and tetratricopeptide repeat containing 4	unknown	other
TNIP1	TNFAIP3 interacting protein 1	Nucleus	other
TP53BP1	tumor protein p53 binding protein 1	Nucleus	transcription regulator
TSPAN31	tetraspanin 31	Plasma Membrane	other
UBR1	ubiquitin protein ligase E3 component n-recogin 1	Cytoplasm	enzyme
UGP2	UDP-glucose pyrophosphorylase 2	Cytoplasm	enzyme
USP22	ubiquitin specific peptidase 22	Nucleus	peptidase
UTP14C	UTP14, U3 small nucleolar ribonucleoprotein, homolog C (yeast)	Nucleus	other
UVSSA	UV-stimulated scaffold protein A	unknown	other
VAMP3	vesicle-associated membrane protein 3 (cellubrevin)	Plasma Membrane	other
WDR74	WD repeat domain 74	Nucleus	other
YBX1	Y box binding protein 1	Nucleus	transcription regulator
YPEL5	yippee-like 5 (Drosophila)	unknown	other
ZFP91	zinc finger protein 91 homolog (mouse)	Nucleus	transcription regulator
ZNF141	zinc finger protein 141	Nucleus	other
ZNF211	zinc finger protein 211	Nucleus	transcription regulator
ZNF599	zinc finger protein 599	unknown	other
ZNF671	zinc finger protein 671	Nucleus	other
ZNF688	zinc finger protein 688	unknown	other
ZNF688	zinc finger protein 688	unknown	other
ZNF688	zinc finger protein 688	unknown	other
ZNF691	zinc finger protein 691	Nucleus	other
ZNF695	zinc finger protein 695	Nucleus	other
ZNF720	zinc finger protein 720	unknown	other
ZNF91	zinc finger protein 91	unknown	other

Supplementary table 8:

NIPBL ChIP signals in HB2 cells on different repeat classes.

RPKM measure (reads per kilobase per 10 million reads) was calculated similar to the RNA-seq analyses (Mortazavi et al. 2008a) and an enrichment compared to the input material (control) was calculated.

class	length (bp)	% of total genomic sequence		ChIP signal (RPKM)		Enrichment
		NIPBL	ChIP control	ChIP	control	
SINE-Alu	306745172	9.91	1.68	1.56	1.08	
SINE-MIR	84349685	2.72	4.75	4.97	0.96	
SINE-Other	571757	0.02	5.94	5.76	1.03	
LINE-L1	512445461	16.55	2.08	1.96	1.07	
LINE-L2	103990071	3.36	5.08	5.18	0.98	
LINE-CR1	10860002	0.35	4.55	4.48	1.02	
LINE-RTE	3643770	0.12	4.08	3.79	1.08	
LINE-Other	201745	0.01	4.89	4.75	1.03	
LTR-ERVL-MaLR	110849894	3.58	4	3.88	1.03	
LTR-ERVL	56136298	1.81	4.86	4.81	1.01	
LTR-ERV1	83494001	2.7	3.2	3.06	1.04	
LTR-ERVK	8842596	0.29	1.94	1.72	1.13	
LTR-Gypsy	2301837	0.07	5.28	5.15	1.03	
LTR-ERV	191609	0.01	5.36	4.77	1.12	
DNA-TcMar-Tigger	33986503	1.1	3.15	3.03	1.04	
DNA-TcMar-Mariner	2830187	0.09	3.29	3.11	1.06	
DNA-TcMar-Tc2	1669939	0.05	3.24	3.08	1.05	
DNA-TcMar	318232	0.01	3.82	3.53	1.08	
DNA-hAT-Charlie	45014464	1.45	4.17	4.03	1.04	
DNA-hAT-Blackjack	3419498	0.11	4.31	4.22	1.02	
DNA-hAT-Tip100	6614899	0.21	4.35	4.32	1.01	
DNA-hAT	1687099	0.05	4.11	4.09	1.01	
DNA-Other	2734962	0.09	4.06	3.79	1.07	
Satellite-acro	31038	0	14.04	11.47	1.22	
Satellite-centr	8243430	0.27	10.45	9.17	1.14	
Satellite-telo	254147	0.01	3.2	4.27	0.75	
Satellite	4025177	0.13	18.3	17.69	1.03	
rRNA	175474	0.01	265.35	19.96	13.3	
<i>LSU-rRNA_Hsa</i>	68641	0.002	442.08	28.92	15.28	
<i>SSU-rRNA_Hsa</i>	16619	0.001	947.81	66.34	14.29	
5S	90214	0.003	5.15	4.59	1.12	
tRNA	102988	0	8.09	13.2	0.61	
scRNA	122042	0	3.5	3.5	1	
snRNA	337551	0.01	5.57	4.13	1.35	
srpRNA	263136	0.01	4.09	4.24	0.96	
RNA	118748	0	4.73	5.74	0.82	
Low_complexity	16922589	0.55	2.79	2.35	1.19	
Simple_repeat	26087578	0.84	2.78	3.15	0.88	
Other_repeat	6109487	0.2	1.85	1.83	1.01	

References

- Arrington C.B., B.R. Dowse, et al., 2012. Non-synonymous variants in pre-B cell leukemia homeobox (PBX) genes are associated with congenital heart defects. *European journal of medical genetics* 55, 4, 235-237.
- Bailey T.L. and C. Elkan, 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2, 28-36.
- Bartek J., J. Bartkova, et al., 1991. Efficient immortalization of luminal epithelial cells from human mammary gland by introduction of simian virus 40 large tumor antigen with a recombinant retrovirus. *Proc Natl Acad Sci U S A* 88, 9, 3520-3524.
- Borck G., M. Zarhrate, et al., 2006. Father-to-daughter transmission of Cornelia de Lange syndrome caused by a mutation in the 5' untranslated region of the NIPBL Gene. *Human mutation* 27, 8, 731-735.
- Brouwer R.W., M.C. van den Hout, et al., 2011. NARWHAL, a primary analysis pipeline for NGS data. *Bioinformatics (Oxford, England)* 28, 2, 284-285.
- Castronovo P., C. Gervasini, et al., 2009. Premature chromatid separation is not a useful diagnostic marker for Cornelia de Lange syndrome. *Chromosome Res* 17, 6, 763-771.
- Ciosk R., M. Shirayama, et al., 2000. Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. *Mol Cell* 5, 2, 243-254.
- Consortium E.P., 2011. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* 9, 4, e1001046.
- Di Giacomo G., M. Koss, et al., 2006. Spatio-temporal expression of Pbx3 during mouse organogenesis. *Gene Expr Patterns* 6, 7, 747-757.
- Eferl R. and E.F. Wagner, 2003. AP-1: a double-edged sword in tumorigenesis. *Nature reviews* 3, 11, 859-868.
- Ernst J., P. Kheradpour, et al., 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473, 7345, 43-49.
- Gerlich D., B. Koch, et al., 2006. Live-cell imaging reveals a stable cohesin-chromatin interaction after but not before DNA replication. *Curr Biol* 16, 15, 1571-1578.
- Hadjur S., L.M. Williams, et al., 2009. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* 460, 7253, 410-413.
- Haering C.H., A.M. Farcas, et al., 2008. The cohesin ring concatenates sister DNA molecules. *Nature*.
- Haering C.H., J. Lowe, et al., 2002. Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol Cell* 9, 4, 773-788.
- Hakimi M.A., D.A. Bochar, et al., 2002. A chromatin remodelling complex that loads cohesin onto human chromosomes. *Nature* 418, 6901, 994-998.
- Horsfield J.A., S.H. Anagnostou, et al., 2007. Cohesin-dependent regulation of Runx genes. *Development* 134, 14, 2639-2649.

- Hu B., T. Itoh, et al., 2011. ATP hydrolysis is required for relocating cohesin from sites occupied by its Scc2/4 loading complex. *Curr Biol* 21, 1, 12-24.
- Jahnke P., W. Xu, et al., 2008. The Cohesin loading factor NIPBL recruits histone deacetylases to mediate local chromatin modifications. *Nucleic Acids Res* 36, 20, 6450-6458.
- Jurka J., V.V. Kapitonov, et al., 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110, 1-4, 462-467.
- Kagey M.H., J.J. Newman, et al., 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 7314, 430-435.
- Kawauchi S., A.L. Calof, et al., 2009. Multiple organ system defects and transcriptional dysregulation in the Nipbl(+/-) mouse, a model of Cornelia de Lange Syndrome. *PLoS Genet* 5, 9, e1000650.
- Kogut I., J. Wang, et al., 2009. The Scc2/Scc4 cohesin loader determines the distribution of cohesin on budding yeast chromosomes. *Genes Dev* 23, 19, 2345-2357.
- Langmead B., C. Trapnell, et al., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, 3, R25.
- Lechner M.S., D.C. Schultz, et al., 2005. The mammalian heterochromatin protein 1 binds diverse nuclear proteins through a common motif that targets the chromoshadow domain. *Biochem Biophys Res Commun* 331, 4, 929-937.
- Lengronne A., Y. Katou, et al., 2004. Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature* 430, 6999, 573-578.
- Liu J. and G. Baynam, 2010. Cornelia de Lange syndrome. *Adv Exp Med Biol* 685, 111-123.
- Liu J. and I.D. Krantz, 2009. Cornelia de Lange syndrome, cohesin, and beyond. *Clin Genet* 76, 4, 303-314.
- Liu J., Z. Zhang, et al., 2009. Transcriptional dysregulation in NIPBL and cohesin mutant human cells. *PLoS Biol* 7, 5, e1000119.
- Losada A., T. Yokochi, et al., 2000. Identification and characterization of SA/Scc3p subunits in the Xenopus and human cohesin complexes. *J Cell Biol* 150, 3, 405-416.
- Lu Z.H., J.T. Books, et al., 2005. YB-1 is important for late-stage embryonic development, optimal cellular stress responses, and the prevention of premature senescence. *Mol Cell Biol* 25, 11, 4625-4637.
- Lu Z.H., J.T. Books, et al., 2006. Cold shock domain family members YB-1 and MSY4 share essential functions during murine embryogenesis. *Mol Cell Biol* 26, 22, 8410-8417.
- Malik S. and R.G. Roeder, 2010. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat Rev Genet* 11, 11, 761-772.
- Mishiro T., K. Ishihara, et al., 2009. Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *Embo J* 28, 9, 1234-1245.
- Misulovin Z., Y.B. Schwartz, et al., 2008. Association of cohesin and Nipped-B with transcriptionally active regions of the Drosophila melanogaster genome. *Chromosoma* 117, 1, 89-102.
- Mortazavi A., B.A. Williams, et al., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5, 7, 621-628.

- Muto A., A.L. Calof, et al., 2011. Multifactorial origins of heart and gut defects in nipbl-deficient zebrafish, a model of Cornelia de Lange Syndrome. *PLoS Biol* 9, 10, e1001181.
- Nativio R., K.S. Wendt, et al., 2009. Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. *PLoS Genet* 5, 11, e1000739.
- Parelho V., S. Hadjur, et al., 2008. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 132, 3, 422-433.
- Pauli A., F. Althoff, et al., 2008. Cell-Type-Specific TEV Protease Cleavage Reveals Cohesin Functions in Drosophila Neurons. *Dev Cell* 14, 2, 239-251.
- Pauli A., J.G. van Bommel, et al., 2010. A direct role for cohesin in gene regulation and ecdysone response in Drosophila salivary glands. *Curr Biol* 20, 20, 1787-1798.
- Peters J.M., A. Tedeschi, et al., 2008. The cohesin complex and its roles in chromosome biology. *Genes Dev* 22, 22, 3089-3114.
- Rhee J.W., A. Arata, et al., 2004. Pbx3 deficiency results in central hypoventilation. *The American journal of pathology* 165, 4, 1343-1350.
- Rhodes J.M., F.K. Bentley, et al., 2010. Positive regulation of c-Myc by cohesin is direct, and evolutionarily conserved. *Developmental biology* 344, 2, 637-649.
- Rollins R.A., M. Korom, et al., 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. *Mol Cell Biol* 24, 8, 3100-3111.
- Rubio E.D., D.J. Reiss, et al., 2008. CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci U S A* 105, 24, 8309-8314.
- Schmidt D., P.C. Schwalie, et al., 2010. A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res* 20, 5, 578-588.
- Schmidt D., M.D. Wilson, et al., 2009. ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods* 48, 3, 240-248.
- Schuldiner O., D. Berdnik, et al., 2008. piggyBac-Based Mosaic Screen Identifies a Postmitotic Function for Cohesin in Regulating Developmental Axon Pruning. *Dev Cell* 14, 2, 227-238.
- Sjogren C. and L. Strom, 2010. S-phase and DNA damage activated establishment of sister chromatid cohesion--importance for DNA repair. *Exp Cell Res* 316, 9, 1445-1453.
- Stedman W., H. Kang, et al., 2008. Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. *Embo J* 27, 4, 654-666.
- Sumara I., E. Vorlaufer, et al., 2000. Characterization of vertebrate cohesin complexes and their regulation in prophase. *J Cell Biol* 151, 4, 749-762.
- Takeda K., K. Noguchi, et al., 1997. Targeted disruption of the mouse Stat3 gene leads to early embryonic lethality. *Proc Natl Acad Sci U S A* 94, 8, 3801-3804.
- Tanaka T., M.P. Cosma, et al., 1999. Identification of cohesin association sites at centromeres and along chromosome arms. *Cell* 98, 6, 847-858.
- Vaquerizas J.M., S.K. Kummerfeld, et al., 2009. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* 10, 4, 252-263.

- Waizenegger I.C., S. Hauf, et al., 2000. Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase. *Cell* 103, 3, 399-410.
- Watrin E., A. Schleiffer, et al., 2006. Human Scc4 is required for cohesin binding to chromatin, sister-chromatid cohesion, and mitotic progression. *Curr Biol* 16, 9, 863-874.
- Wendt K.S., K. Yoshida, et al., 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451, 7180, 796-801.
- Wilder S., D. Thybert, et al., manuscript in preparation.
- Yoneyama M., W. Suhara, et al., 2002. Control of IRF-3 activation by phosphorylation. *J Interferon Cytokine Res* 22, 1, 73-76.
- Zeng W., J.C. de Greef, et al., 2009. Specific loss of histone H3 lysine 9 trimethylation and HP1gamma/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). *PLoS Genet* 5, 7, e1000559.
- Zhang Y., T. Liu, et al., 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9, 9, R137.

Chapter 5

Discussion

Discussion

The chromatin fiber is organized in self-interacting regions, called topological domains. The understanding of how the chromatin fibre is folded within these domains and which proteins are involved in the formation of these contacts will be fundamental to elucidate how these structures influence genes. In this thesis we investigated two candidate factors for the organization of topological domains, the cohesin complex and the chromatin insulator protein CTCF. Both factors are very abundantly bound throughout mammalian genomes and they are linked to the formation of long-range interactions at individual loci. A recent study further showed that cohesin and CTCF are enriched at the boundaries of topological domains.

The aim of this work was to understand the roles of the cohesin complex and CTCF in shaping the topological domains. We used the multiplexing 3C-sequence technique (described in detail in *Chapter 2*) and the genome-wide method, Hi-C, to study the consequences of cohesin or CTCF depletion for higher-order chromatin structure. We showed that both proteins are important for the organization of topological domains but we also observed differences between the contributions of cohesin and CTCF (*Chapter 3*).

Cohesin-regulatory proteins and the cohesin complex are important during development as mutations of these factors can lead to different developmental syndromes. We studied the role of the cohesin loading factor NIPBL in the molecular etiology of the Cornelia de Lange Syndrome (CdLS) and showed a direct role of NIPBL in gene regulation. This suggest that defects in NIPBL might alter expression of developmental genes thereby contributing to the phenotype observed in CdLS patients (*Chapter 4*).

Nuclear architecture at high resolution

Chromosomes are not randomly distributed in the nucleus but they occupy distinct territories (Stack, Brown et al., 1977). The position of the chromosomes within the nucleus is correlated with gene density (Boyle, Gilchrist et al., 2001) and transcriptional activity (Schneider and Grosschedl, 2007). Several microscopy studies showed an arrangement of the chromosome territories in loop domains. These studies have also tried to understand how the fiber folds within the territory and how this influences genome function (Visser, Jaunin et al., 2000; Branco and Pombo, 2006). However due to poor resolution and technical limitations these methods are not suitable to analyze the overall genome organization.

In the past decade, the development of chromosome capturing methods has facilitated the study of the organization of the genome at higher resolution when compared to previous microscopy methods. These technologies allow the analysis of interactions that occur between fragments of interest (3C-PCR/qPCR) or within a specific locus (4C; 5C) or in the whole genome (Hi-C). 3C-PCR/qPCR (Dekker, Rippe et al., 2002; Hagege, Klous et al., 2007) was the first technique to be developed. It is laborious, requires some knowledge about the locus of interest and can investigate only one particular interaction at a time. Thus it is not suitable to

find novel contacts but it can be used to confirm already known interactions. The 4C approach (Simonis, Klous et al., 2006) can analyze all chromatin interactions with a chosen fragment (anchor or viewpoint). The detection of the unknown fragments ligated to the viewpoint is done by microarray; thus the coverage of the interactions is dependent on the array used. The 5C method (Dostie, Richmond et al., 2006) uses pools of oligonucleotides that target a specific part of the genome to detect long-range interactions. It offers a higher coverage compared to 4C but the detection by oligonucleotides does not allow a genome-wide coverage of interactions due to the large number of oligonucleotides that would be needed to detect all possible interactions. Both 4C and 5C methods yield information about a particular region of the genome but they are inappropriate for genome-wide studies. By contrast, the Hi-C technique (Lieberman-Aiden, van Berkum et al., 2009) can identify all possible interactions within the genome. It requires a biotin pull-down of the ligated fragments which are then enriched and sequenced. The data analysis needs bioinformatics expertise and the method requires a massive sequence effort to have high resolution data making it very expensive.

To have a method that combines high resolution and reduction of the sequencing costs, we developed the multiplexed 3C-seq (m3C-Seq) technology, which is described in detail in *Chapter 2*.

This method allows the analysis of up to 12 viewpoints using one sequencing lane and in total 192 viewpoints per Illumina HiSeq2000 platform can be assessed in parallel, giving the possibility to study the region of interest at high resolution. This provides high throughput and the possibility to multiplex the samples has the advantage to reduce the sequencing costs. Novel interactions found in any of the “C” methods should be supported by alternative experiments, such as FISH.

An example of the use of the m3C-Seq is provided in *Chapter 3* where six different viewpoints were pooled and sequenced in one lane. These different viewpoints probed a domain of 1.5Mb on the human chromosome 11 providing a detailed interaction map for this chromosome region.

All the “C” methods are fundamental to understand which factors promote functional contacts by correlating interactions with the genomic binding of particular proteins and analyzing chromosomal interactions after depletion of the candidate protein.

Contribution of cohesin and CTCF to the organization of the human genome

Several chromosome conformation capturing studies have indicated the importance of the cohesin complex and CTCF for long-range chromosomal interactions (Hadjur, Williams et al., 2009; Mishiro, Ishihara et al., 2009; Nativio, Wendt et al., 2009).

Cohesin is a multisubunit protein complex highly conserved in eukaryotes. The complex is necessary for sister chromatid cohesion (Michaelis, Ciosk et al., 1997; Losada, Hirano et al., 1998; Sumara, Vorlaufer et al., 2000) and it is also required for efficient DNA double strand break repair (Sjogren and Nasmyth, 2001; Kim, Krasieva et al., 2002). Several studies

have also demonstrated that the cohesin complex has a function in gene regulation (Donze, Adams et al., 1999; Rollins, Korom et al., 2004; Horsfield, Anagnostou et al., 2007; Schuldiner, Berdnik et al., 2008) and the finding that cohesin highly co-localizes genome-wide with CTCF in mammalian cells gave another important hint for a role of cohesin in regulation (Parelho, Hadjur et al., 2008; Wendt, Yoshida et al., 2008).

CTCF is known to function as repressor (Lobanenkov, Nicolas et al., 1990), activator (Vostrov and Quitschke, 1997) and insulator (Bell, West et al., 1999). Several experiments demonstrated that cohesin is important to maintain CTCF insulation function and CTCF is necessary to bring cohesin to the shared sites (Wendt, Yoshida et al., 2008). How cohesin and CTCF influence gene transcription is not known but different observations by chromosome conformation capturing methods suggested that both factors are involved in promoting long-range chromosomal interactions and their depletion leads to the loss of architecture and impaired gene transcription (Hadjur, Williams et al., 2009; Mishiro, Ishihara et al., 2009; Nativio, Wendt et al., 2009). It was proposed that CTCF can bring cohesin to particular regulatory sites and cohesin might stabilize the architecture of the locus by embracing the DNA with its ring-like structure, as was suggested for the sister chromatid cohesion function (Haering, Lowe et al., 2002; Gruber, Haering et al., 2003). Recently, genome-wide studies using Hi-C to analyse the 3D structure of the mouse and human genome identified large chromatin interaction domains, termed topological domains. Boundaries between the domains were found to be enriched for CTCF and cohesin sites indicating that both proteins might be important to establish the overall structure of topological domains (Dixon, Selvaraj et al., 2012). Cohesin and CTCF have already been involved in mediating long-range chromosomal interactions at individual loci (Hadjur, Williams et al., 2009; Mishiro, Ishihara et al., 2009; Nativio, Wendt et al., 2009) but their role in shaping the topological domains throughout the genome is not known. In *Chapter 3* we elucidated the contribution of both factors to the folding of the chromatin fiber genome-wide.

To establish the role of cohesin, we depleted the complex from interphase chromosomes by proteolytic cleavage of the subunit RAD21. The use of this cleavable cohesin system has some advantages compared to the classical cohesin depletion by RNAi: first the cleavage of cohesin occurs when the complex is functionally bound to the chromatin and second the cleavage occurs very rapidly, taking only 24 hours compared to the 48 hours required for RNAi treatment. This rapid release of cohesin allows studying the immediate effects of cohesin loss on chromatin structure, without interfering with the function of cohesin in cell division.

In order to study local changes in chromatin structure after cohesin depletion we performed m3C-seq to probe the interior and the borders of one topological domain at the human chromosome 11 p15.5 region comprising the Igf2/H19 locus and other imprinted genes. To investigate genome-wide changes after cohesin cleavage we performed Hi-C. The m3C-seq profiles revealed that cohesin cleavage causes a dramatic loss of interaction in the domain analysed, indicating that the complex is important to maintain the organization of this area. The loss of interaction within domains was also observed by Hi-C throughout the entire genome. The Hi-C data also showed a loss of interaction between domains. Interestingly, the domain

boundaries can be still observed after RAD21 cleavage suggesting that the cohesin complex is involved in the organization of the topological domains but not in the segregation of the neighbouring domains.

Since cohesin and CTCF functionally interact and highly colocalize genome-wide, we asked whether depletion of CTCF has the same effect on chromatin structure as cohesin cleavage. The Hi-C data after depletion of CTCF showed loss of interactions within domains, similar to that seen with cohesin cleavage, but the remarkable difference was a gain of interactions between different domains. These observations indicated that CTCF might be important to maintain the domain boundaries to ensure the integrity of the domains.

It was already proposed that CTCF might function as barrier element at the boundaries preventing the formation of contacts across the boundaries (Dixon, Selvaraj et al., 2012).

This would agree with the gain of interactions observed after CTCF depletion, however it could also be due to de-localization of cohesin from CTCF sites. It was already shown that CTCF is important to position cohesin to the shared sites and depletion of CTCF affects cohesin enrichment at the binding sites but not its binding to chromatin (Wendt, Yoshida et al., 2008). The combination of these observation suggests that cohesin could be present at (novel) different locations in absence of CTCF promoting novel interactions.

The Hi-C results indicate that cohesin and CTCF are important for the topological domain organization in a non-redundant manner. This would imply that the simultaneous depletion of both factors would lead to a complete loss of structure, which remains to be investigated.

Long-range chromosomal contacts were already linked to gene regulation since DNA regulatory elements such as enhancers, silencers, locus control regions (LCR) or insulator elements are often found very distant from their target promoters (reviewed in Kadauke and Blobel, 2009; Sexton, Bantignies et al., 2009). Since we observed a different contribution of cohesin and CTCF to the topological domain organization, we hypothesized that both factors might also influence genes in a different way. Consistent with the differential contribution of cohesin and CTCF to the organization of the human genome, we observed a different set of misregulated genes after cohesin cleavage and CTCF depletion. For genes misregulated after CTCF depletion we frequently observed CTCF binding close to the promoter. This is supported also by previous data showing that around 20% of CTCF binding sites localize in proximity of transcription start sites (Kim, Abdullaev et al., 2007; Jothi, Cuddapah et al., 2008). This suggested that CTCF might directly regulate a subset of genes by binding to their promoters. In contrast, genes misregulated after cohesin depletion do not have cohesin enriched at the promoter. Instead, we observed that these genes more likely loose the interactions with distal DNase hypersensitive sites compared to non-cohesin regulated genes. These sites are marks for DNA regulatory regions; thus the altered gene regulation after cohesin removal might occur due to the loss of interactions with regulatory elements.

One example for such a mechanism could be the observation that cohesin cleavage caused a loss of interactions in the HOXA and HOXB gene clusters that coincide also with altered transcription of several HOX genes. The HOX genes encode a family of conserved transcription factors that play important roles in specifying anterior-posterior body patterning

during development. The expression levels of HOX genes are controlled throughout embryonic development and aberrant expression or mutation of HOX genes can lead to body malformations and multiple types of cancer (reviewed in Mallo, Wellik et al., 2010; Shah and Sukumar, 2010).

The finding that cohesin depletion leads to misregulation of gene clusters fundamental during development is very interesting because mutations in cohesin-regulatory proteins or in the cohesin complex are linked to the human developmental diseases *Roberts/SC phocomelia syndrome* (RBS) and *Cornelia de Lange Syndrome* (CdLS). Among other defects, both syndromes show severe limb abnormalities. Alteration in human HOX genes have also been linked to limb abnormalities (Del Campo, Jones et al., 1999) and it will therefore be interesting to study whether defects in cohesin or cohesin-regulators could influence the proper expression of HOX genes during development.

Insight into Cornelia de Lange syndrome (CdLS)

Defects in cohesin regulatory proteins and in the cohesin complex have been linked to several human developmental diseases also referred as “cohesinopathies”. CdLS is a cohesinopathy characterized by growth defects, developmental abnormalities and mental retardation. In 60% of the cases mutations of the cohesin loading factor NIPBL are observed (Krantz, McCallum et al., 2004). In a smaller fraction of patients mutations are found in the cohesin core subunits SMC1A (5%) or SMC3 (<1%) and recently in the SMC3-deacetylase HDAC8 (Musio, Selicorni et al., 2006; Deardorff, Kaur et al., 2007; Deardorff, Bando et al., 2012).

The most severe phenotype of the syndrome is linked to mutations of NIPBL while mutations of SMC1A or SMC3 correspond to a milder phenotype. Patients affected by CdLS do not show cohesion defects, suggesting that the developmental abnormalities are likely due to alteration in cohesin functions that affect gene expression (reviewed in Dorsett and Krantz, 2009). This is supported by a study on *lymphoblastoid cells lines* (LCLs) derived from CdLS patients with NIPBL mutations showing that several misregulated genes in these patients lose cohesin binding next to the transcription start sites. These data suggested that the absence of cohesin, due to NIPBL mutation, might affect transcriptional regulation contributing to the CdLS phenotype (Liu, Zhang et al., 2009). On the other hand, the majority of the cases of CdLS are caused by NIPBL mutations but to date no direct link of NIPBL to the altered transcription found in CdLS patients has been shown. NIPBL has already been linked to gene regulation. In *Drosophila* it was found to facilitate activation of the *cut* and *Ultrabithorax* genes by remote enhancers (Rollins, Korom et al., 2004). A study in mouse embryonic stem cells reported that NIPBL co-localizes with cohesin and the Mediator complex to the promoters of actively transcribed genes (Kagey, Newman et al., 2010).

In *Chapter 4* we analysed the localization of NIPBL in the human genome and elucidated whether NIPBL has a role in the altered gene expression pattern observed in CdLS patients. NIPBL is crucial for the association of cohesin with chromatin. Therefore we were expecting to find a co-localization of NIPBL with cohesin if NIPBL would load cohesin directly to its

binding sites. Surprisingly, this was not observed suggesting that NIPBL and cohesin might only transiently associate during cohesin loading. Instead, we observed that NIPBL binds preferentially to the promoters of active genes. Further, NIPBL depletion altered the transcription of these genes, suggesting a direct role for NIPBL in gene regulation.

This observation prompted us to identify the localization of NIPBL in lymphoblastoid cells lines (LCLs) derived from controls and CdLS patients to understand whether NIPBL could be involved in the differential gene expression pattern observed in LCL's derived from CdLS patients compared to controls. Consistent with our previous results, we observed also in LCL's that NIPBL was preferentially bound to promoters and found reduced binding levels of NIPBL in CdLS patients cells. Therefore we hypothesized that the NIPBL haploinsufficiency observed in CdLS probands could be linked to the altered transcription that contributes to the phenotype of the syndrome. Our comparison of the NIPBL associated genes with a published list of misregulated genes found in cells from CdLS patients (Liu, Zhang et al., 2009) showed that a significant percentage of these genes have NIPBL bound to the promoter, indicating that at least part of the altered genes in CdLS might be a direct target of NIPBL. Nevertheless, NIPBL might also influence transcription indirectly via the loading of cohesin which in turn regulates genes by promoting long-range chromosomal interactions, as we observed in *Chapter 3*. An example of a cohesin-dependent structure sensitive to the partial reduction of NIPBL was shown by Chien et al. for the mouse β -globin locus (Chien, Zeng et al., 2011). During the differentiation of mouse erythroleukemia cells (MEL) into mature erythroid cells, cohesin and NIPBL bind at the LCR and to the β -major and β -minor promoters. This also correlates with the formation of interactions between the LCR and β -major promoter and with an increase in expression of the adult β -globin gene (Chien, Zeng et al., 2011). Furthermore, the analysis of liver tissue from NIPBL heterozygous knockout mice (*Nipbl*^{+/-}) showed a decrease of all cohesin binding sites at the mouse β -globin locus resulting in a reduction of interaction between the LCR and β -globin promoters as well as reduced β -major and β -minor expression. It will be interesting to study whether in CdLS patients with NIPBL mutation the chromatin structure around specific genes is impaired to link the altered expression with changing in the structure.

The data presented in this thesis have improved our understanding of the functional organization of the human genome and have highlighted the roles of cohesin and CTCF as factors important for the organization of the topological domains that constitute a basic structural feature of the genome. Furthermore, we proposed that the cohesin loading factor, NIPBL, might directly influence gene expression by binding to the promoter of genes.

The future prospective will be to study the roles of cohesin and cohesin-regulators in organizing higher-order chromatin structure during development to elucidate the function of the complex in the molecular etiology of CdLS.

References

- Bell A.C., A.G. West, et al., 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98, 3, 387-396.
- Boyle S., S. Gilchrist, et al., 2001. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum Mol Genet* 10, 3, 211-219.
- Branco M.R. and A. Pombo, 2006. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol* 4, 5, e138.
- Chien R., W. Zeng, et al., 2011. Cohesin mediates chromatin interactions that regulate mammalian beta-globin expression. *J Biol Chem* 286, 20, 17870-17878.
- Deardorff M.A., M. Bando, et al., 2012. HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. *Nature* 489, 7415, 313-317.
- Deardorff M.A., M. Kaur, et al., 2007. Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of cornelia de Lange syndrome with predominant mental retardation. *Am J Hum Genet* 80, 3, 485-494.
- Dekker J., K. Rippe, et al., 2002. Capturing chromosome conformation. *Science* 295, 5558, 1306-1311.
- Del Campo M., M.C. Jones, et al., 1999. Monodactylous limbs and abnormal genitalia are associated with hemizygoty for the human 2q31 region that includes the HOXD cluster. *Am J Hum Genet* 65, 1, 104-110.
- Dixon J.R., S. Selvaraj, et al., 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 7398, 376-380.
- Donze D., C.R. Adams, et al., 1999. The boundaries of the silenced HMR domain in *Saccharomyces cerevisiae*. *Genes Dev* 13, 6, 698-708.
- Dorsett D. and I.D. Krantz, 2009. On the molecular etiology of Cornelia de Lange syndrome. *Ann N Y Acad Sci* 1151, 22-37.
- Dostie J., T.A. Richmond, et al., 2006. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res* 16, 10, 1299-1309.
- Gruber S., C.H. Haering, et al., 2003. Chromosomal cohesin forms a ring. *Cell* 112, 6, 765-777.
- Hadjur S., L.M. Williams, et al., 2009. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* 460, 7253, 410-413.
- Haering C.H., J. Lowe, et al., 2002. Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol Cell* 9, 4, 773-788.
- Hagege H., P. Klous, et al., 2007. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* 2, 7, 1722-1733.
- Horsfield J.A., S.H. Anagnostou, et al., 2007. Cohesin-dependent regulation of Runx genes. *Development* 134, 14, 2639-2649.

- Jothi R., S. Cuddapah, et al., 2008. Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res* 36, 16, 5221-5231.
- Kadauke S. and G.A. Blobel, 2009. Chromatin loops in gene regulation. *Biochim Biophys Acta* 1789, 1, 17-25.
- Kagey M.H., J.J. Newman, et al., 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 7314, 430-435.
- Kim J.S., T.B. Krasieva, et al., 2002. Specific recruitment of human cohesin to laser-induced DNA damage. *J Biol Chem* 277, 47, 45149-45153.
- Kim T.H., Z.K. Abdullaev, et al., 2007. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 128, 6, 1231-1245.
- Krantz I.D., J. McCallum, et al., 2004. Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of *Drosophila melanogaster* Nipped-B. *Nat Genet* 36, 6, 631-635.
- Lieberman-Aiden E., N.L. van Berkum, et al., 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 5950, 289-293.
- Liu J., Z. Zhang, et al., 2009. Transcriptional dysregulation in NIPBL and cohesin mutant human cells. *PLoS Biol* 7, 5, e1000119.
- Lobanenkov V.V., R.H. Nicolas, et al., 1990. A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. *Oncogene* 5, 12, 1743-1753.
- Losada A., M. Hirano, et al., 1998. Identification of *Xenopus* SMC protein complexes required for sister chromatid cohesion. *Genes Dev* 12, 13, 1986-1997.
- Mallo M., D.M. Wellik, et al., 2010. Hox genes and regional patterning of the vertebrate body plan. *Developmental biology* 344, 1, 7-15.
- Michaelis C., R. Ciosk, et al., 1997. Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell* 91, 1, 35-45.
- Mishiro T., K. Ishihara, et al., 2009. Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *Embo J* 28, 9, 1234-1245.
- Musio A., A. Selicorni, et al., 2006. X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. *Nat Genet* 38, 5, 528-530.
- Nativio R., K.S. Wendt, et al., 2009. Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. *PLoS Genet* 5, 11, e1000739.
- Parelho V., S. Hadjur, et al., 2008. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 132, 3, 422-433.
- Remeseiro S., A. Cuadrado, et al., 2012. A unique role of cohesin-SA1 in gene regulation and development. *Embo J* 31, 9, 2090-2102.
- Rollins R.A., M. Korom, et al., 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. *Mol Cell Biol* 24, 8, 3100-3111.

- Schneider R. and R. Grosschedl, 2007. Dynamics and interplay of nuclear architecture, genome organization, and gene expression. *Genes Dev* 21, 23, 3027-3043.
- Schuldiner O., D. Berdnik, et al., 2008. piggyBac-Based Mosaic Screen Identifies a Postmitotic Function for Cohesin in Regulating Developmental Axon Pruning. *Dev Cell* 14, 2, 227-238.
- Sexton T., F. Bantignies, et al., 2009. Genomic interactions: chromatin loops and gene meeting points in transcriptional regulation. *Semin Cell Dev Biol* 20, 7, 849-855.
- Shah N. and S. Sukumar, 2010. The Hox genes and their roles in oncogenesis. *Nature reviews* 10, 5, 361-371.
- Simonis M., P. Klous, et al., 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* 38, 11, 1348-1354.
- Sjogren C. and K. Nasmyth, 2001. Sister chromatid cohesion is required for postreplicative double-strand break repair in *Saccharomyces cerevisiae*. *Curr Biol* 11, 12, 991-995.
- Stack S.M., D.B. Brown, et al., 1977. Visualization of interphase chromosomes. *J Cell Sci* 26, 281-299.
- Sumara I., E. Vorlaufer, et al., 2000. Characterization of vertebrate cohesin complexes and their regulation in prophase. *J Cell Biol* 151, 4, 749-762.
- Visser A.E., F. Jaunin, et al., 2000. High resolution analysis of interphase chromosome domains. *J. Cell Sci.* 113, 2585-2593.
- Vostrov A.A. and W.W. Quitschke, 1997. The zinc finger protein CTCF binds to the APBbeta domain of the amyloid beta-protein precursor promoter. Evidence for a role in transcriptional activation. *J Biol Chem* 272, 52, 33353-33359.
- Wendt K.S., K. Yoshida, et al., 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451, 7180, 796-801.

Abbreviations

Summary

Curriculum Vitae

PhD Portfolio

Acknowledgements

Abbreviations

3C	Chromosome Conformation Capture
3C-seq	Chromosome Conformation Capture sequencing
4C	Chromosome Conformation Capture-on-chip
5C	Chromosome Conformation Capture Carbon Copy
ATP	Adenosine Triphosphate
CdLS	Cornelia de Lange Syndrome
CPM	Counts per millions
CT	Chromosome Territory
CTCF	CCCTC-binding factor
cv	Cleaved
DHS	DnaseI Hypersensitive site
dox	Doxycycline
EGFP	Enhance Green Fluorescent protein
ER α	Estrogen Receptor alpha
ES cells	Embryonic Stem cells
FACS	Fluorescent-activated cell sorting
FC	Fold Change
FDR	False Discovery Rate
FISH	Fluorescent in situ hybridization
H2A	Histone 2A
HB2	Breast Endothelial cell line
HDACs	Histones Deacetylase
HEK293T	Hunan Embryonic Kidney
HepG2	Hepatocellular Carcinoma cells
HRV	Human rhinovirus protease
ICR	Imprinting Control Region
Kb	Kilo base pairs
KDa	Kilo Dalton
LCLs	Lymphoblastoid Cell Lines
LCR	Locus Control Region
Mb	Mega base pairs
MCF-7	Breast cancer cell line
MEF	Mouse Embryonic Fibroblast
p-val	p-value
RBS	Roberts/SC phocomelia Syndrome
RFP	Red Fluorescent Protein
RPKM	Reads Per Kilobase Million reads mapped
SA	Stromalin Antigen
SMC	Structural Maintenance of Chromosomes
TEV	Tobacco Etch Virus protease
TSS	Transcription Start Sites
VP	View Point
wt	Wild-type

Curriculum Vitae

Personal Information:

Name: Jessica Zuin
Date of birth: August 2nd, 1983
Nationality: Italian

Education:

Present: **PhD Student**, Erasmus Medical Center, Rotterdam, The Netherlands

2005-2008: **MSc Medical Biotechnology**, University of Padua, Italy
Grade:110/110 *cum laude*

2002-2005: **BSc Biotechnology**, University of Padua, Italy
Grade:102/110

1997-2002: **High School Degree**, Liceo Scientifico Pietro Scalcerle, Padua, Italy
Grade: 97/100

Professional Training:

2009-2013: PhD program at Erasmus Medical Center, Department of Cell biology, Rotterdam, The Netherlands.
“The role of cohesin and CTCF in organizing the human genome.”

2007-2009: Laboratory of Nano Biotechnology Xeptagen SpA, Venice, Italy.
“The diagnostic accuracy in the determination of biomarker-IgM immune complexes for hepatocellular carcinoma diagnosis.”

2004: Laboratory of Molecular Biology, Veterinary Pathology and Hygiene, Department of Public Health, University of Padua, Italy.
“Isolation and characterization of genes coding Toll-like receptor in *Sparus aurata*.”

Publication:

Zuin J, Dixon J.R, van der Reijden M.I.J.A, Kolovos P, Ye Z, Brouwer R.W.W, van de Corput M.P.C, van IJcken W.F.J, Grosveld F.G, Ren, Wendt K.S.

Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells.

Manuscript Submitted

Zuin J, Franke V, van IJcken W.F.J, van der Sloot A, Krantz I.D, van der Reijden M.I.J.A, Nakato R, Lenhard B, Wendt K.S.

A cohesin-independent role for NIPBL at promoters provides insights in CdLS.

Manuscript Submitted

Stadhouders R, Kolovos P, Brouwer R, **Zuin J**, van den Heuvel A, Kockx C, Palstra RJ, Wendt KS, Grosveld F, van IJcken W, Soler E.

Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions.

Nat Protoc. 2013

Turato C, Biasiolo A, Pengo P, Frecer V, Quarta S, Fasolato S, Ruvoletto M, Beneduce L, **Zuin J**, Fassina G, Gatta A, Pontisso P.

Increased antiprotease activity of the SERPINB3 polymorphic variant SCCA-PD.

Exp Biol Med. 2011

Veggiani G, **Zuin J**, Beneduce L, Gallotta A, Pengo P, Fassina G.

Combinatorial semisynthesis of biomarker-IgM complexes.

J Biomol Screen. 2010

Zuin J, Veggiani G, Pengo P, Gallotta A, Biasiolo A, Tono N, Gatta A, Pontisso P, Toth R, Cerin D, Frecer V, Meo S, Gion M, Fassina G, Beneduce L.

Experimental validation of specificity of the squamous cell carcinoma antigen-immunoglobulin M (SCCA-IgM) assay in patients with cirrhosis.

Clin Chem Lab Med. 2010

PhD Portfolio

Name PhD Student Jessica Zuin
 Erasmus MC department Biomedical Science
 PhD period June 2009-June 2013
 Promoter Prof. dr. Frank G. Grosveld
 Co-promoter Dr. Kerstin S. Wendt

Year

General courses

Cell and Developmental Biology (CDB) course	2010
Genetics (G) course	2010
Biochemistry and Biophysics (BB) course	2010
Biomedical English writing course	2012

Specific courses

Next generation sequence (NGS) analysis, Leiden MC, Leiden, The Netherlands	2009
Course Biomedical Research Techniques VIII, Erasmus MC Rotterdam, The Netherlands	2009
In vivo imaging "From Molecule to Organism", Erasmus MC Rotterdam, The Netherlands	2009
Literature course on "Chromatin organization and transcription", Erasmus MC Rotterdam, The Netherlands	2009
"Leica confocal introduction course", Erasmus MC Rotterdam, The Netherlands	2011

Seminars and workshops

The workshop browsing genes and genomes with Ensembl V, Erasmus MC Rotterdam, The Netherlands	2009
8 th Winter School "Transcriptional Control in Developmental Processes", Kleinwalsertal, Austria Oral Presentation	2010
16 th MGC PhD Student workshop, Cologne, Germany	2010
9 th Winter School 2011 "Chromatin Changes in Differentiation and Malignancies", Kleinwalsertal, Austria Oral Presentation	2011
17 th MGC PhD Student workshop, Maastricht, The Netherlands Poster Presentation	2011
"Nuclear Structure and Gene Expression", Hubrecht Institute, Utrecht, The Netherlands	2011
18 th MGC PhD Student workshop, Dusseldorf, Germany Oral Presentation	2012
22 th MGC Symposium, Leiden MC, Leiden, The Netherlands Oral Presentation	2012

National conferences

7 th Dutch Chromatin Meeting, Utrecht, The Netherlands	2009
8 th Dutch Chromatin Meeting, Leiden, The Netherlands	2010
9 th Dutch Chromatin Meeting, Groningen, The Netherlands	2011
10 th Dutch Chromatin Meeting, Swammerdam Institute for Life Sciences, University of Amsterdam, The Netherlands	2012

International conferences

Chromatin Meeting, Zeche Zolleverein, Germany	2010
EuTRACC 2 nd Young Scientist Meeting, Dubrovnik Croatia Oral Presentation	2010
"Cohesin biology and the cohesinopathy", Lucca, Italy	2011
"Regulation from a distance: long-range control of gene expression in development and disease", Royal Society, London, United Kingdom	2012
"Epigenetics & Chromatin: Interactions and processes", Boston, USA Oral and Poster Presentation	2013

