Partners in Long Distance Interactions

Sanja Krpic

The studies presented in this thesis manuscript were performed in the Department of Cell Biology of the Erasmus MC in Rotterdam, The Netherlands. The Department is a member of the Medisch Genetisch Centrum Zuid-West Nederland (MGC).

Cover Design: Sanja Krpic and Noa Haim

Inspired by “Starry night” Vincent Van Gogh

Thesis layout: Nir Goldberg

Printed in: Wöhrmann Print Service, Zutphen

Partners in Long Distance Interactions

Partners in lange afstand interacties

Proefschrift

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

Prof.dr. S.W.J. Lamberts

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

Woensdag 25 maart 2009 om 15.45 uur

door

Sanja Krpic

geboren te Belgrado
(Servië)

Promotiecommissie

Promotor:	Prof.dr. F.G. Grosveld
Overige leden:	Prof.dr. J.N.J. Philipsen Dr.ir. N. Galjart Dr.ir. D.N. Meijer

.....for my parents and Nir

CONTENTS

List of abbreviations

3C	Chromosome Conformation Capture
4C	Chromosome Conformation Capture on Chip
ACH	Active Chromatin Hub
APC/C	Anaphase Promoting Complex or Cyclosome
ATP	Adenosine triphosphate
Bp	Base pairs
CAR	Cohesin-associated regions
CdLS	Cornelia de Lange Sydrome
CH	Chromatin Hub
Chip	Chromatin immuno-precipitation
Chip-seq	Chromatin immuno-precipitation and direct sequencing
CTCF	CCCCTC-binding factor
DNA	Deoxyribonucleic acid
DNAaseI	Deoxyribonuclease I
DSB	DNA double-strand break
EB	Enhancer blocker
FISH	Fluorescent in situ hybridization
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HP1	Heterochromatin protein 1
HR	Homologous recombination
HS	Hypersensitive site
ICR	Imprinting control region
IL	Interleukin
KO	Knockout
LCR	Locus control region
LPS	Lipopolysaccharides
Mb	Mega base
MEL	Mouse erythroleukemia
MTA1	Metastasis tumor antigen 1
NIPBL	Nipped-B Like Protein
PCR	Polymerase chain reaction
PEV	Position effect variegation
RNA	Ribonucleic acid
SHH	Sonic Hedgehog
SMC	Structural Maintenance of Chromosomes
SIR	Silent information regulator
Th-1/2	T helper type-1/2

The aim of the thesis

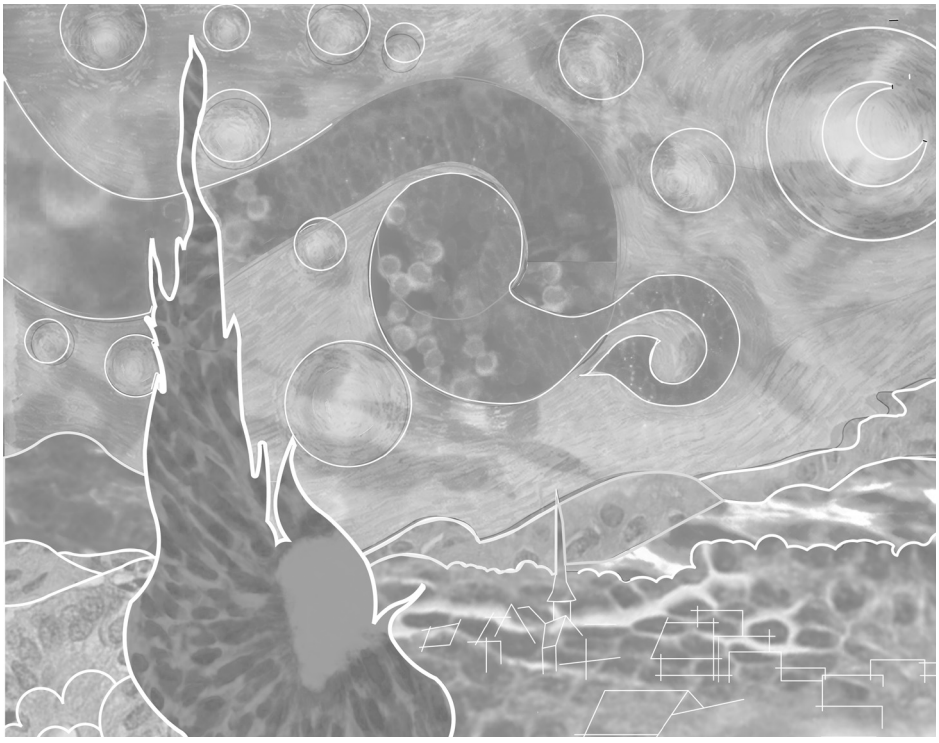
In the DNA of metazoans, long-range interactions between distant regulatory elements are responsible for specific gene expression in many developmentally regulated genes and gene families. Gene-regulatory mechanisms are usually complex and involve the interaction of control elements such as enhancers with the respective promoter of a particular gene. Often these interactions occur over considerable distances as in the some cases control elements are located hundreds or even thousands of kilobases away from the transcriptional initiation site. The main focus of research described in this thesis are several factors that have a role in long-range interactions and chromatin organization.

Chapter 1 provides an introduction and gives an overview of information required for understanding the studies presented in subsequent chapters of this Thesis.

A genetic screen, aimed to isolate genes involved in long- range interactions in *Drosophila*, identified a class of proteins with novel function: Chip (mammalian homolog Ldb1) and Nipped-B (mammalian homolog Delangin). Ldb1 is ubiquitously expressed during development and interacts with a great number of transcriptional regulators. This suggests an important function for Ldb1 in different developmental pathways. Chapter 2 describes novel protein complexes of Ldb1, formed during hematopoietic development. Furthermore, morpholino mediated inhibition of Ldb1 expression in zebrafish show that these factors are essential for hematopoiesis and they are coexpressed in prehematopoietic cells of the early mouse embryo.

Delangin (Nipped-B) protein is involved in loading of Cohesin onto chromatin thereby playing an important role in chromatin dynamics. The role of Delangin in mammals is described in Chapter 3 focuses on description of the Delangin heterozygous mutant mouse that shows a complex phenotype similar to that seen in Cornelia de Lange patients that have mutations in the same gene.

Chapter 4 describes the role of CTCF, the prototype vertebrate insulator implicated in chromatin structure with Cohesins. In the same chapter, we are co-relating functions of Cohesins, better known for their role in mediating sister chromatid cohesion, with CTCF protein. Our study focuses on co-operation of Cohesin and CTCF in dictating chromosome conformation at several loci. The studies presented in the Chapter 5 address the role of CTCF in T-cell differentiation, showing that CTCF is required for T-helper-2 cytokine expression. Finally, in the concluding Chapter 6, an overall discussion of this thesis is presented with general implications of this work and future directions.



Chapter 1 - Introduction

General Introduction

The genome of higher eukaryotes consists of DNA, which in case of the human genome measures 2m in length and is divided over 46 chromosomes. These long DNA molecules are packed in a nucleus that measures about 10μm in diameter. In order to fit the complete DNA into such a small volume, DNA is folded and compacted by proteins in a structure called chromatin. During mitosis, is even further compacted into condensed chromosomes (Kornberg, 1974). All the information needed for the formation and proper function of an organism is stored in these structures and it is reasonable to expect that this overcrowded situation is organized in a very specific manner, with controlled three-dimensional contacts within the nucleus. The need for controlled chromatin contacts is also suggested by the fact that gene regulation is a tightly regulated process. Different levels of control must be involved in regulating proper spatio-temporal expression of genes throughout the process of cellular differentiation. These processes are coordinated by interactions of an "army" of general, cell-type and stage specific proteins that bind to chromatin and DNA.

Several techniques allow the identification and study of chromatin regions that interact with each other. These include functional genetic analysis, microscopic analysis after DNA or RNA fluorescent *in situ* hybridization (FISH) in combination with 3D microscopy, as well as biochemical methods, such as chromosome conformation capture (3C) and the more sophisticated variation thereof (4C). The combination of these methods reveals a network of contacts in the nucleus. These interactions are mediated by insulators and other regulatory sequences, including enhancers and promoters, which mediate/promote certain functional three-dimensional interactions while preventing other enhancer-promoter contacts. In this chapter, I will introduce several factors: Ldb1, Delangin, Cohesin and CTCF which have important role long-range interactions.

1. *Cis*-regulatory elements and gene regulation

Regulatory sequences that are encoded in the primary DNA sequence provide the first level of transcriptional control. Several distinct *cis*-regulatory elements are identified including promoters, enhancers, silencers, Locus Control region (LCR) and insulators. These so-called *cis* regulatory elements, attract general and specific transcription factors that are responsible for the functional and structural organization of chromatin. They are relatively small DNA fragments (200-300bp) that contain information "when, where and how much" should be made of a gene product encoded in the primary DNA sequence. This information is used by the RNA polymerase II machinery to transcribe the gene into RNA. Studies over a number of years have revealed that enhancers, LCR and promoters are in contact with one another when genes are transcribed.

1.1 Long-range enhancer-promoter interaction

In a structural sense, enhancers and promoters are quite similar. The main differences between their sequences and their location are that the sequence of enhancers is usually longer than that of promoters and that enhancers can be located tens or hundreds of kilobases away from the genes whose activity they control. Furthermore, enhancer activity is not dependent on orientation.

One of the most spectacular examples of transcriptional control over long distance is the regulation of SHH (Sonic Hedgehog) expression in the developing limb. The enhancer is positioned 1Mb away from the *Shh* gene itself (Lettice et al., 2003). An important aim of current biology is to understand how remote *cis*-acting elements and the promoters control gene transcription levels at the appropriate times in differentiation and development.

Enhancers are capable of enhancing the basal transcription levels of the linked promoter. They were originally identified in transient transfection assays as sequences capable of transcriptional activation of a linked promoter over large distances in an orientation independent manner (Banerji et al., 1981; Moreau et al., 1981).

According to the current view, the factors participating in transcription are recruited to the activated promoter via protein-protein interactions (Ptashne and Gann, 1997). The enhancer-bound activator proteins increase the local concentration of the transcriptional machinery near the promoter. In higher organisms, long-distance chromatin interactions have been demonstrated for enhancer elements which are separated from a gene by hundreds up to millions of base pairs (Kleinjan and van Heyningen, 2005). However, binding of an activator protein to remote elements, would not increase the concentration of the transcriptional machinery at the promoter in a strictly linear manner. Hence, there must be mechanisms that allow long-distance enhancer-promoter interactions. There are several different models proposed for enhancer action over distance which are not mutually exclusive. The combination of different models is used to explain data generated at different loci in different organisms.

The looping model:

For many years after the discovery that enhancers are separated from the genes, it was believed that enhancers and promoters interact irrespective of the DNA between them. A loop of DNA would be formed, which otherwise would separate promoter and enhancer on a one dimensional level (Ptashne, 1986). In higher organisms, where gene clusters have complex regulation, a number of observations could only be explained by looping. In the case of *trans*-activation, an enhancer situated on one chromosome can establish contact with the promoter on the paired homologous allele (Bickel and Pirrotta, 1990). This phenomenon called transvection was observed in *Drosophila* where genetic results argued for an inter-chromosomal interaction between promoter and enhancer situated on different chromosomes (Wu

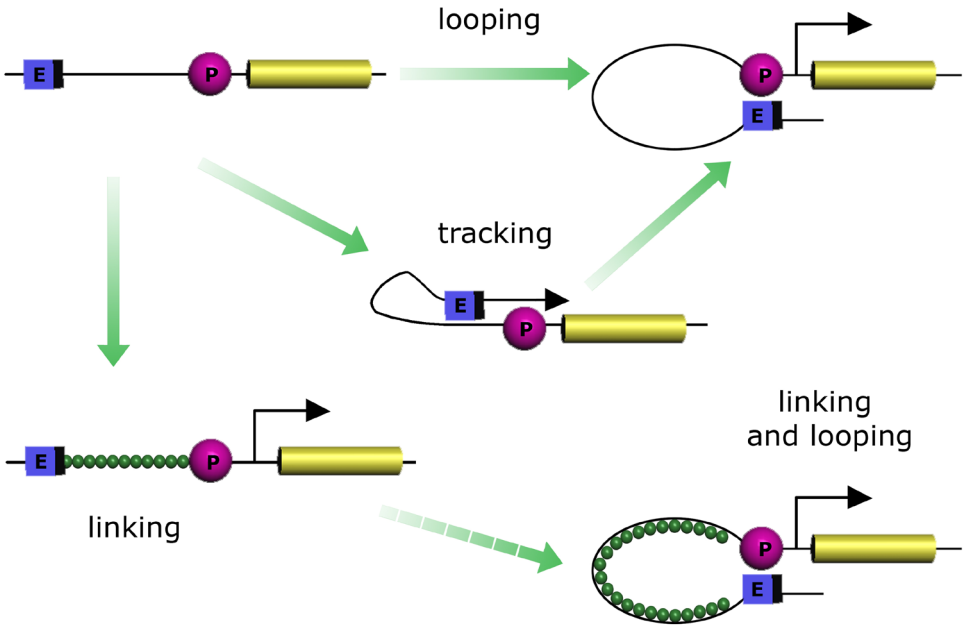


Figure 1

Possible models for long-range enhancer-promoter communication

In the simplest form of "looping" (top), enhancer (E blue) and promoter (P purple) establish contact by random collision, at which point proteins bound to either site contact each other directly to activate transcription of a linked gene (yellow cylinder). As a consequence, the intervening chromatin template is looped out. Note that the looping model, unlike other models does not depend on the presence of an intervening chromatin template for transcription. Other models differ from this largely in the mechanism by which enhancer-promoter interaction occurs. For example, in the tracking model, the enhancer-bound protein complex actively scans the DNA in search of the promoter. The facilitated tracking model unites looping and tracking (center). This model states that complexes nucleated at the enhancer track along intervening chromatin templates towards a potentially active gene. In the linking model, a chain of the protein complexes (green balls) extends along the chromatin fiber to the promoter where gene activation starts.

and Morris, 1999). However in mammalian cells, data that would support functional interchromosomal promoter-enhancer interactions are still missing.

The looping model provides the most plausible explanation for gene competition for a single regulator. For example, in transfection assays where plasmids contain different genes and enhancers, it was shown that genes compete for regulatory sequences (Wasylyk et al., 1983). Such competition also operates in endogenous chromosomal loci. An explanation how competition may work was provided by

fluorescent in situ hybridization (FISH) data on ongoing transcription of the fetal γ- or β-gene in a single cell where it was shown that one or the other gene is active at a given time (Wijgerde et al., 1995). The studies that followed confirmed these results (Gribnau et al., 1998) indicating that alternate transcription that takes place is and is caused by a stochastic 'flip-flop' mechanism of LCR action. Although such studies are most easily explained by a looping model, they do not show directly an interaction between distal and proximal elements. Strong evidence for such interactions was obtained from the mouse β-globin locus using two different biochemical approaches, the recovery of associated protein (RNA TRAP) assay (Carter et al., 2002) and chromosome conformation capture (3C) technology (Tolhuis et al., 2002). Both studies showed that the endogenous mouse β-globin LCR (Locus Control Region) is in close spatial proximity to the active β-globin gene promoter located 50 kbp away. This type of interaction has now been shown for a number of loci, for example in the Th2 locus, it was shown by 3C technology that a lineage-restricted chromatin loop is formed between the Th2 LCR and genes encoding for the interleukins IL-4, IL-5 and IL-13 (Spilianakis and Flavell, 2004). These studies show that long-range interactions are established by contacts of interacting regions with intervening DNA looping out.

The tracking model

In the tracking model, the enhancer acts as a loading platform for a DNA-tracking protein which travels along the chromatin fiber and ends up in the vicinity of the promoter (Herendeen et al., 1992). The best experimental example for the tracking model is the enhancer action of the late genes in the bactoriophage T4 (reviewed in (Kolesky et al., 2002). However, there is no evidence for tracking in eukaryotes and no known example of an activator leaving the enhancer to activate transcription. On the other hand, this model explains well the enhancer-blocking activity of boundary or insulator elements in which an insulator bound protein blocks the tracking protein complex.

A 'facilitated tracking' model is a combination of the looping and the tracking models. It suggests that enhancer bound complexes travel in small steps along the intervening chromatin fiber until it reaches the promoter after which a stable loop is formed. In Chip (Chromatin immunoprecipitation) analysis of a distribution of C/EBPα and HNF-3β enhancer binding proteins at the HNF-4α gene locus, these proteins could be detected at the spacer DNA as well as at the promoter. The proteins were bound after activation of the enhancer, but before actual transcription of the gene (Hatzis and Talianidis, 2002). Since the enhancer-promoter communication is a slow process and transcription starts only 80 hours after induction of the enhancer, it seems that other mechanisms of communication are involved for fast transcription or at larger loci. In addition, it is not possible to explain activation in trans or alternative transcription by this model.

The linking model

The linking model proposes that an enhancer acts as a loading platform for a DNA binding protein that facilitate polymerization of proteins in the direction of the promoter by coating the chromatin fiber (Bulger and Groudine, 1999; Dorsett, 1999). This model was proposed to explain the action of *Drosophila* Chip (Ldb1) protein, that doesn't bind to DNA directly but interacts with several transcription factors and facilitate their action *in vivo* (Morcillo et al., 1997; Torigoi et al., 2000). It is suggested that Chip is recruited by an activator bound at an enhancer where it functions as a protein 'bridge' between the activator bound at the enhancer and proteins having multiple binding sites between an enhancer and a promoter (Dorsett, 1999). Still, there is no functional data to support this model. Recent experiments have shown that Ldb1, the mouse homologue of Chip, forms large complexes with transcription factors (Meier et al., 2006) that bind specific sites in the genome, indicating that Ldb1 complexes more likely play a role in looping (Song et al., 2007).

1.2 LCR (Locus Control Region)

The presence of an enhancer as a part of a transgenic construct is usually not sufficient to ensure (optimal levels of) expression. The expression of the transgene is often low when compared to endogenous levels, and the spatio-temporal expression is usually altered. This is caused by the inability of the transgenic construct to overcome the local effects of chromatin structure at the site of the integration (position effects). In the 1980s, transgenic mouse studies using human **β-globin gene** construct led to the discovery of the locus control region (LCR). It is a dominant regulatory region that confers position-independent and copy-number dependent expression to linked transgenes (Grosveld et al., 1987). Importantly, an LCR is able to overcome heterochromatin-mediated position effect variegation (PEV) (Festenstein et al., 1996; Milot et al., 1996). Transgenic mice carrying a human CD2 or **β-globin gene** linked with LCR show normal levels of expression irrespective of their chromosomal integration site while partial deletion of the LCR resulted in variegated expression. These experiments have shown that both human CD2 and β-globin LCR are essential for establishing an open chromatin configuration. Locus control regions are structurally composed of varying numbers of tissue specific DNAseI hypersensitive sites each consisting of multiple binding sites for transcription factors and they typically include enhancer and insulator elements (Bonifer, 2000; Dillon and Sabbattini, 2000). The most prominent characteristic of LCR is its ability to overcome repressive chromatin structures and provide strong transcriptional enhancer activity. The absence of an LCR in humans leads to an absence of transcription (Kioussis et al., 1983) or in mice to a severely reduced transcription of the human β-globin gene (Epner et al., 1998; Magram et al., 1985; Townes et al., 1985). The mouse β-globin locus is situated on chromosome 7 and contains four functional genes: εγ, βh1, β_{maj} and β_{min} (from 5'to 3'). The major determinant for activation of a gene is the relative distance from the

LCR for example, introduction of a β-globin gene between the LCR and the γ-gene leads to premature β-globin gene activation and reduction of γ-gene expression (Hanscombe et al., 1991).

The competition between genes in the β-globin locus suggests that the LCR functions by physical interaction with the genes. After the discovery of the β-globin LCR many combinations of *cis*-regulatory elements were characterized in various vertebrate species which show the functional characteristics of a LCR. Other domains with LCR-like activity have been characterized in several loci in humans, including the human growth hormone (GH) locus and in the *Rad50* gene in Th2 cytokine locus (Ho et al., 2006; Lee et al., 2003) (reviewed in (Li et al., 2002). The variety and complexity of LCR action points to the importance of studying transcriptional regulation of intact genetic loci ***in vivo***.

1.3 The Active Chromatin Hub (ACH) of the β-globin locus

Although several mechanisms have been proposed involving complex interactions between genes and regulatory elements in the β-globin locus, only the DNA looping model is compatible with all available experimental data. As described above, in the looping model, the physical interactions of an enhancer bound activator and a protein at the promoter result in looping out of the intervening DNA. This model assumes that the chromatin fiber is flexible enough to enable distant DNA fragments to interact with each other. Such chromatin interactions were postulated on the based of genetic experiments (Dillon et al., 1997; Gribnau et al., 1998; Hanscombe et al., 1991; Wijgerde et al., 1995) and later biochemicaly by 3C, a technique used to identifies chromatin regions that are in close spatial proximity. Surprisingly, interactions were much more complex than just a simple LCR-gene contact. First, a number of interactions were identified in cells before differentiation and expression of the globin genes, forming a so called chromatin hub (CH). Once the genes are expressed, an additional set of interactions form an Active Chromatin Hub (ACH) (Palstra et al., 2003; Tolhuis et al., 2002). Apart from these interactions, contacts with distant regulatory elements were also observed (Tolhuis et al., 2002) suggesting that several CHs are generated or that one CH contains additional sequences from other loci. The dynamics of active and inactive genes is different; active genes loop inwards, while inactive genes loop out. CTCF has an important role in forming this structure since interactions are dependent on CTCF in erythroid progenitor cells but there is no impact on globin expression (Splinter et al., 2006). The role of CTCF will be described later in this Chapter.

1.4 Insulators

A major fraction of vertebrate genomes is composed of repetitive, silenced DNA that exist as large regions of condensed chromatin. Chromatin is compacted by condensation processes that are self-propagating and can spread into neighboring loci affecting their expression (Ghirlando et al., 2004). Therefore, mechanisms must exist that prevent such spreading into neighboring gene loci. Insulators are DNA elements that were first identified based on their ability to protect a gene from the outside influences, which might otherwise lead either to inappropriate activation or silencing of a gene (Kellum and Schedl, 1991). In addition, certain insulator elements, when placed between an enhancer and a promoter, inhibit enhancer-stimulated gene expression. These sequences are necessary to ensure that genes are not activated in the wrong place or at the wrong time by regulatory elements from other genes. Insulator elements have been experimentally identified based on these two defining characteristics; barrier insulators, which block heterochromatization and consequent silencing a gene and enhancer blocking (EB) insulators (Figure 2), which prevent communication between discrete sequence elements (typically enhancers and promoters) when positioned between them (Wallace and Felsenfeld, 2007).

Much of the early work that characterized insulator elements was done in *Drosophila*. The first DNA sequences to be described to have the properties of an insulator were the specialized chromatine structure (scs and scs') elements. These elements were initially identified as DNAse I HSS that mark the chromatin boundaries of a heat shock locus (Udvardy et al., 1985). Moreover, it was shown that these elements could block an enhancer from activating a promoter (Kellum and Schedl, 1992). Several DNA binding proteins have been shown to bind to insulator sequences: Su (Hw), Zw5, BEAF-32, GAGA factor and dCTCF (Gaszner et al., 1999; Geyer and Corces, 1992; Hart et al., 1997; Moon et al., 2005; Schweinsberg et al., 2004). In vertebrates the only known insulator protein is CTCF (Wallace and Felsenfeld, 2007). It is unknown whether the vertebrate CTCF is able to replace the roles of all the insulator proteins in *Drosophila*, or other vertebrate insulators are yet to be identified.

Sequences that are able to act as insulators were identified later in other organisms but no significant similarity is evident among these proteins (Bell et al., 2001). Enhancer-blocking insulators are best exemplified by the *gypsy* element in ***Drosophila*** (Capelson and Corces, 2005) and the CTCF-binding sites identified as insulators initially in vertebrates (Chung et al., 1993) and later in *Drosophila* (Moon et al., 2005).

Although the demand for insulators in genomes with larger distances between genes seems to be less obvious, there are well-established examples of insulator mediated gene regulation in mammals. The first vertebrate insulator described was the HS4 element of the chicken β-globin locus which showed enhancer blocking activity when placed between a reporter gene containing the β-globin promoter and LCR (Chung et al., 1993). The minimal core required for insulation contains binding sites for CTCF and these sites are necessary and sufficient for positional enhancer

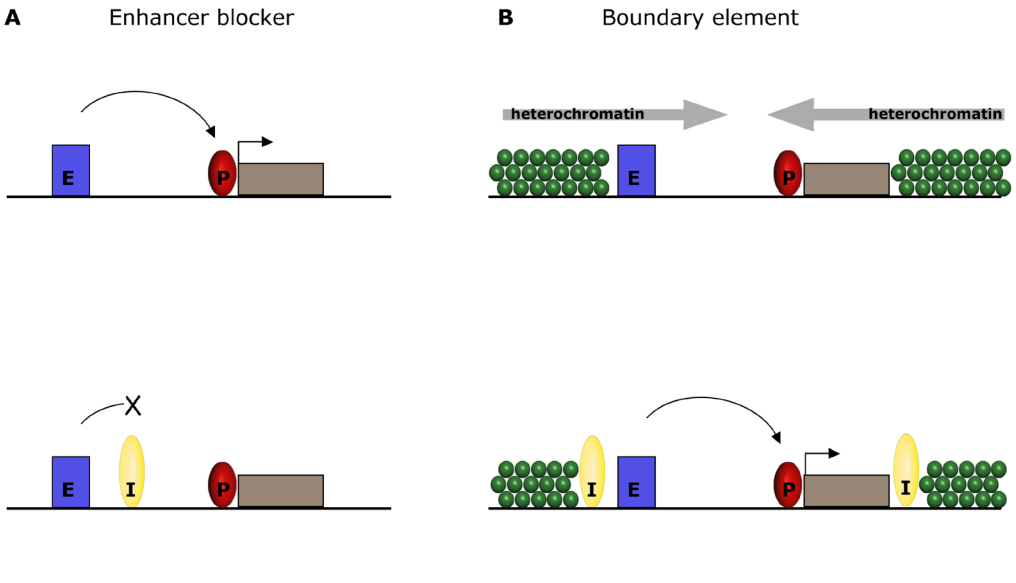


Figure 2
Insulators can function as position dependent enhancer-blockers or boundary elements

Insulators have been experimentally identified based on two defining characteristics: enhancer blocking function and boundary function.

A. Enhancer-blocker: an enhancer (E, blue box) can activate transcription (arrow) from a promoter (P, red ellipse) of a cis-linked gene (grey rectangle) .The enhancer activity is blocked when an insulator element (I, yellow oval) is positioned in between the enhancer and promoter.
B. Boundary element: spreading of nearby heterochromatin into a locus prevents the transcriptional activation by an enhancer of a *cis*-linked gene. Insertion of two insulator elements flanking the locus prevents the spread of condensed chromatin.

blocking activity (Bell et al., 2001). The CTCF site in cHS4 is insufficient to protect against position effects, indicating that the enhancer-blocking activity of CTCF and the boundary function of cHS4 are separable (Recillas-Targa et al., 2002). In additional experiments, the upstream transcription factor 1 and 2 (USF1/USF2) were suggested as candidates that control boundary activity of cHS4.

There are several models explaining how an enhancer-blocking insulators influence communication between enhancers and promoters. The models are not mutually exclusive and potentially any insulator could employ one or a combination of mechanisms. In the promoter decoy model, an enhancer–blocking insulator recruits a component of the transcriptional machinery mimicking a promoter at the molecular level (Geyer, 1997). Consistent with this model, enhancers and insulators have been observed to co-localize (Yoon et al., 2007) though interactions between insulators and promoters have also been observed (Yoon et al., 2007) suggesting an alternative mechanism. The physical barrier model proposes that a molecular

signal coming from the enhancer such as RNA polII complex simply hits the insulator complex and is unable to progress. This model can explain the accumulation of RNA PolII at the cHS4 insulator that is inserted between the human HS2 enhancer and its target globin gene (Zhao and Dean, 2004). The argument for an additional mechanism exists, since insulators located within introns can silence a downstream enhancer without truncating the gene product. The loop domain model proposes that insulator sites interact with each other and/or with other nuclear structures. In the *Igf2*/*H19* locus in mice, functional loop formation has been found at the imprinting control region (ICR), which binds CTCF in an allele-specific manner (Kurukuti et al., 2006; Murrell et al., 2004; Yoon et al., 2007). Although these studies disagree on the specific interactions, they agree that CTCF–bound ICR is involved in loop formation that is necessary for proper enhancer-promoter interactions with the *Igf2* promoter, thereby excluding other interactions.

Models explaining insulator boundary function have in common that boundary activity is linked to a disruption of the reaction responsible for heterochromatin spreading. In these models, the barriers function as chain terminators by either modulating the nucleosomal structure or by formation of looped structures by anchoring the chromatin fiber (Gaszner and Felsenfeld, 2006). For example, the cHS4 vertebrate insulator recruits histone acetyltransferase (HAT) and methytransferase which leads to histone modifications and eventual termination of heterochromatin formation (West et al., 2004). Although histone modification is a necessary step for insulation in cHS4, it is not sufficient to block heterochromatin spreading. Additional protein binding sites in cHS4 without histone modification marks are required for complete boundary activity (West et al., 2004).

2. The Role of CTCF in long-range interactions

The complex three-dimensional organization of gene regulation raises several questions. One of these is which factors are involved in nuclear organization and long-range interactions. CTCF, which is thought to have multiple context dependent functions, is one of the potential candidates to mediate long-range interactions.

CTCF (CCCTC binding factor) has a very important role in genetic and epigenetic maintenance of the genome. It was originally identified as a transcription factor that negatively regulates the chicken *c-myc* gene (Lobanenkov et al., 1990). Other studies showed that CTCF is not only involved in transcriptional silencing but also in the activation of *c-myc* and other genes (Klenova et al., 1993). CTCF can also act as an enhancer-blocking protein (Prioleau et al., 1999; Recillas-Targa et al., 1999) and it can bind to boundary elements to prevent spreading of heterochromatin (Defossez and Gilson, 2002). Moreover, CTCF binding sites have been found at the imprinting center of the X chromosome (Chao et al., 2002) and at the boundaries of domains that are escaping X-inactivation (Filippova et al., 2005). How CTCF carries out all these diverse functions is still not clear. Some investigators have suggested that CTCF might function through topological organization of the genome by folding the

genome into loop domains, in that way isolating the genes from the influence of neighboring genes or regulatory sequences. Additionally, CTCF could be important for folding chromatin to bring distal regulatory elements closer together to form a chromatin hub.

An important study on the function of CTCF in long-range interactions was done in the mouse and human *Igf2*/*H19* locus where it was shown that this factor regulates the imprinting of the both genes in a methylation sensitive manner. On the maternal chromosom, the *Igf2* gene is turned off and *H19* is transcribed while on the paternal chromosome the situation is the opposite.

One of the regions responsible for this expression pattern is the imprinting control region (ICR) that lies between the *Igf2* gene and the downstream enhancers. The sites in the ICR are occupied by CTCF on the maternally inherited allele while the paternal copy is methylated preventing CTCF binding (Bell and Felsenfeld, 2000; Hark et al., 2000; Kanduri et al., 2000). Mice in which the two alleles can be distinguished have been used in 3C experiments to study the insulator effect on higher order structure (Kurukuti et al., 2006; Murrell et al., 2004). Contacts are detected between the ICR and one of the two upstream imprinted DNA methylated regions, DMR1 and DMR2. These contacts differ on the two alleles suggesting allele-specific interactions between the enhancer and the *Igf2* promoter (Lopes et al., 2003; Murrell et al., 2004). A different model of the *Igf2*/*H19* system was proposed based on a similar 3C approach, but with emphasis on the interaction between promoters and downstream enhancers. In the paternal allele, where CTCF does not bind to the methylated ICR, the *H19* downstream enhancer interacts with the *Igf2* promoter consequently expressing Igf2. On the maternal allele, CTCF bound to ICR prevents the *H19* downstream enhancer to activate *Igf2*, which results in activation of *H19*. Furthermore, the chromatin loop between the CTCF-bound ICR and DMR1 are maintained during mitosis, whereas enhancer-promoter loops are absent suggesting a possible role of CTCF in epigenetic memory during cell division (Burke et al., 2005).

How does CTCF mediate long-range interactions? Since it forms dimers and even oligomers, it is possible that CTCF molecules bound to distal elements could interact with each other, thereby driving loop formation (Yusufzai et al., 2004). One interesting study, which shows that CTCF is not absolutely required on both interacting loci, suggested that additional factors are involved in these interactions (Ling et al., 2006). Indeed, in the case of two divergently transcribed MHC class II genes, *HLA-DRB1* and *HLA-DQA1*, CTCF requires at least two additional factors to mediate long-range interactions (Majumder et al., 2008).

Another interesting study (Lefevre et al., 2008) supports a model in which transcription-dependent chromatin remodeling leads to physical dislodgement of CTCF prior to gene activation. CTCF form an insulator complex bound to the upstream *cis* element of the chicken lysozyme gene. Bacterial lypopolysaccharides (LPS) induce expression via this element through stepwise recruitment of transcription factors followed by concomitant alterations in chromatin structure within the upstream cis elements (Lefevre et al., 2005). LPS stimulation triggers destabilization of nucleosomes and exposure of the flanking enhancer elements, which in turn allows

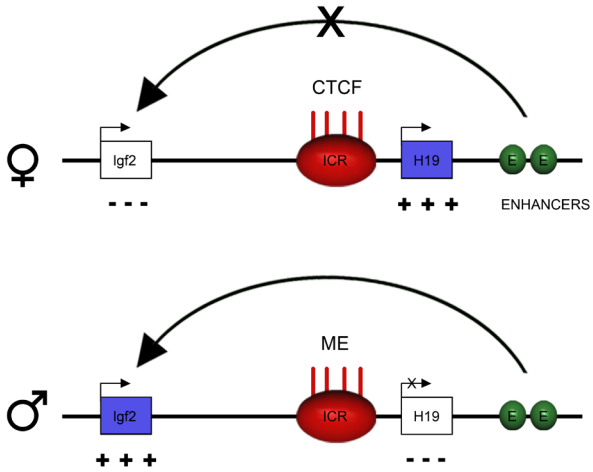


Figure 3

Action of CTCF in the mouse Igf2/H19 locus

On the maternally transmitted allele, CTCF binds to the sites in the imprinted control region (ICR) and prevents downstream enhancers from activating Igf2 expression. The ICR of the paternal allele is methylated, CTCF doesn't bind and the enhancer is no longer blocked.

recruitment of additional factors and the initiation of non-coding RNA synthesis. Transient transcription and the concomitant passage of the RNA PolII complex through this element are correlated with H3 phosphorylation and repositioning of a nucleosome over the CTCF occupancy site (Ong and Corces, 2008).

2.1 The Role of CTCF in the β-globin locus

The first direct evidence for CTCF-mediated chromatin looping was provided by an analysis of the β-globin locus (Splinter et al., 2006). The entire mouse β-globin locus is embedded in a cluster of olfactory receptor (OR) genes that are not expressed in erythroid cells (Bulger et al., 2000). The globin genes are positioned in the order of their developmental expression and all encode for a β-globin like proteins. Together with α-globin- like proteins and heme groups they form a functional hemoglobin molecule. Both α- and β-globin genes duplicated several times, subsequently evolved independently and ended up on different chromosomes in birds and mammals (Hardison, 1998). In the mouse β-globin locus is situated on chromosome 7 and contains four functional genes: εγ, $\beta_{h1,}$ β_{maj} and β_{min} (from 5'to 3'). During primitive eryhropoiesis, yolk sac derived erythroid cells express mostly εy and β_{h1}, while the expression of β_{maj} and β_{min} is low. The generation of definitive erythroid cells starts in fetal liver around day 11 of embryonic development and is characterized by high levels of transcription of adult β-globin genes (β_{maj} and β_{min}) and inactive, or minimal, expression of embryonic (εγ, and β_{h1}) genes (Wawrzyniak and Popp, 1987). Upstream of the genes lies a cluster of erythroid specific regulatory sequences known as the LCR, a dominant regulatory element that confers position-independent and copy number dependent expression of linked transgenes (Grosveld et al., 1987). Additional

erythroid- specific regulatory sequences are located upstream of the LCR (5'HS-84/-85 and 5'HS-60/-62) and one downstream from the genes (3'HS1) (Bulger et al., 2003; Farrell et al., 2000; Tuan et al., 1985). The role of different elements involved in the regulation of β-globin gene expression and chromatin organization was defined by deleting the different hypersensitive (HS) sites in the locus. For example, deletion of the whole β-globin LCR in the mouse resulted in a large decrease in β-globin gene expression but did not alter chromatin structure (Epner et al., 1998). Importantly, none of the individual HS sites deletions affected the formation of the remaining sites, implicating that there is no dominant initiating site whose formation is required for formation of the other HS sites (Bender et al., 2000).

CTCF binding sites and their relative positions are conserved between chicken, mouse and humans. In the mouse β-globin locus the CTCF binding sites are present at the 3'HS1, in the LCR (5'HS5/cHS4) and upstream of the locus (HS-62.5 and HS-85) (Bulger et al., 2003; Farrell et al., 2002). Binding of CTCF to both ends of the locus led to the idea that CTCF may serve as an enhancer blocker that prevents the inappropriate activation of surrounding olfactory receptor genes by the β-globin LCR in erythroid cells (Farrell et al., 2002). Alternetively, it may be required to prevent spreading of heterochromatin into the β-globin locus. Indeed, it was shown that CTCF is required for the long-range interactions within the locus by experiments in which CTCF was depleted (Splinter et al., 2006). However, disruption of a CTCF-binding site had no effect on the expression of the β-globin genes or an effect on surrounding mouse olfactory receptor genes. The balance between active and repressive histone modification only changed locally at the binding sites and not elsewhere in the locus, demonstrating that CTCF does not serve to insulate the β-globin locus in erythroid cells (Splinter et al., 2006). Expression of the β-globin genes was also not changed in the case of deletions of other HS sites (Bender et al., 2006). This raises the question how those evolutionary conserved sites participate in the tissue-specific interactions with the LCR and the active β-globin genes. One of the possible explanations is that β-globin gene expression benefits from the CTCF-mediated loops to an extent that is sufficient for evolutionary selection but too limited to be detected by current technology (de Laat et al., 2008). Another explanation comes from the fact that CTCF is a structural factor that binds to many loci all over the human (Kim et al., 2007) or mouse (Chen et al., 2008) genome to provide proper folding of the genome. It would have specialized functions only in imprinted regions, where it has been shown to act as an enhancer blocker (Bell and Felsenfeld, 2000; Hark et al., 2000).

.

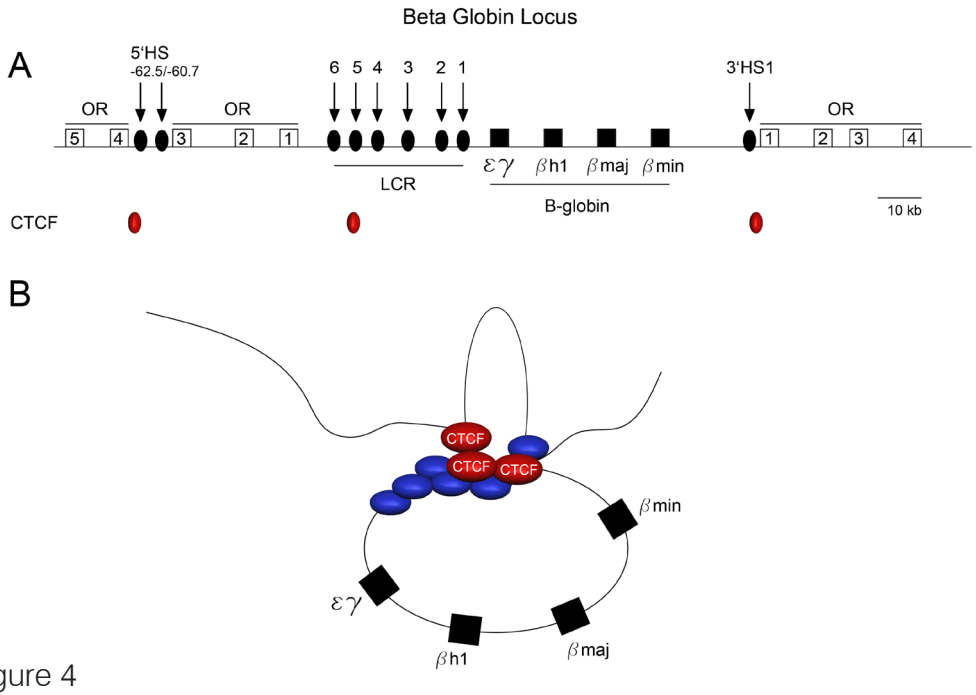


Figure 4

CTCF binding in the mouse β-globin locus

A. Schematic representation of the mouse β-globin locus with the position of CTCF binding sites, which are indicated with red ovals. The β-globin genes, aligned in the order of their developmental expression, are indicated by black boxes. Olfactory receptor genes (OR) are represented by white boxes and flank the locus. DNAse I hypersensitive sites (HS) are indicated with arrows.
B. The presentation of distant CTCF binding sites (red ovals) that occur between regulatory DNA elements of the β-globin locus in erythroud progenitor cells (HS-62, 5'HS5 and 3'HS1). Not all sites are shown (HS-85). Several proteins (blue ovals) that bind to the regulatory elements are indicated but not mentioned for simplicity. The globin genes are expressed at basic levels in these cells (Tolhuis, Palstra et al. 2002; Palstra, Tolhuis et al. 2003; Splinter, Heath et al. 2006). Mutating the CTCF binding site at 3'HS1 destroys the interactions between it and other CTCF binding sites. At later stage of erythroid development, additional contacts are made between LCR and other regulatory sites to form the ACH; this structure becomes independent on CTCF binding (Splinter, Heath et al. 2006)

Palstra, R. J., B. Tolhuis, et al. (2003). "The beta-globin nuclear compartment in development and erythroid differentiation." Nat Genet 35(2): 190-4.
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): 2349-54.
Tolhuis, B., R. J. Palstra, et al. (2002). "Looping and interaction between hypersensitive sites in the active beta-globin locus." Mol Cell 10(6): 1453-65.

2.2 Identification of CTCF-binding sites in the genome

The CTCF protein, originally identified as a transcriptional repressor, has been associated with different functions such as transcriptional activator, enhancer-blocker, boundary definer and genome organizer. It binds to DNA in a sequence-specific manner, with 11-zinc-finger domains.

A first attempt to map CTCF binding sites in a genome-wide fashion was performed using ChIP-on-Chip in fetal liver cells (Mukhopadhyay et al., 2004). In this procedure, chromatin is cross-linked, sonicated and protein-DNA complexes are purified by immunoprecipitation. Following de-crosslinking, the associated proteins are degraded, and the isolated DNA is labelled and hybridized to microarrays that cover mouse genome. The analysis of microarrays can determine the genomic binding location of protein of interest. However, only a subset of CTCF binding sites was identified in this study due to technical limitations.

More recent genome-wide approaches allowed much better characterization of the distribution of CTCF binding sites (Barski et al., 2007; Kim et al., 2007; Xie et al., 2007). It was found that CTCF binds throughout the genome at conserved sites, implying that these sites are functional (Wallace and Felsenfeld, 2007). Identified sites correlate strongly with regions containing genes, suggesting that CTCF's primary role in the genome is to regulate gene expression. The domains depleted of CTCF sites tend to include clusters of related gene families and genes that are transcriptionally co-regulated; CTCF sites flank many of these regions. Domains that are enriched in CTCF binding sites often have multiple alternative promoters (Kim et al., 2007). These observations are consistent with a role of CTCF as an enhancer blocker.

3. Cohesin:

Another interesting protein complex, that has several diverse function, is Cohesin. It is involved in many cellular processe which will be discussed bellow. The Cohesin is best known for its role in sister chromatid cohesion. The genomes of eukaryotic organisms are organized in chromosomes, which are individual large DNA molecules folded into complex structures. Before cell division, the DNA replication machinery faithfully copies each of these DNA molecules and a physical linkage is established between the two sister DNA molecules (or sister chromatids). This physical DNA connection is known as sister-chromatid cohesion. This pair-wise organization allows the cell to distribute chromatids during cell division so the daughter cells receive a full complement of chromosomes. The communication between two sister chromatids has two components: the DNA catenation arising from the replication process itself and a multiprotein complex holding sister helices together, known as Cohesin. This complex is deposited on chromatin during the G1 phase and establishes cohesion concomitant with DNA replication. The physical linkage between the sister chromatids is maintained in G2 and is dissolved only during mitosis. In yeast, this occurs in a single step at the onset of anaphase. It requires cleavage, by the enzyme separase, of

Scc1/Rad21, one of the subunits of the Cohesin complex. In Metazoa, most Cohesin dissociates from chromatin during prophase, in a process regulated by Aurora B and Polo kinase, whereas a small population, located preferentially at pericentric heterochromatin, remains on the condensed chromosomes until anaphase. This mitotic Cohesin is essential to prevent the precocious separation of the sister chromatids, and is removed from chromatin by separase cleavage upon activation of the APC/C (Anaphase Promoting Complex or Cyclosome) (Losada and Hirano, 2005; Nasmyth and Haering, 2005).

Biochemical studies in yeast and high resolution imaging of purified human Cohesin complexes suggest that Cohesin forms a tripartite ring. Cohesin contains four core subunits: SMC1 and SMC3, members of the SMC (Structural Maintenance of Chromosomes) protein family, and two non-SMC subunits Scc1/Rad21 and Scc3/SA (Nasmyth and Haering, 2005). The SMC1 and SMC3 subunits of Cohesin both form rod-shape molecules that heterodimerize by means of 'hinge' domains situated at the ends of 30-nm-long intramolecular antiparallel coiled-coiles (Haering et al., 2002)(Figure 5). ATPase 'heads', at the other ends, are connected by the Scc1/kleisin subunit of Cohesin thereby forming a tripartite ring with a 35nm diameter (Gruber et al., 2003; Haering et al., 2002). Although it has been proposed that Cohesin forms a ring that embraces the sister DNA molecules, other models are possible (Nasmyth and Haering, 2005). The ring-like structure has led to the proposal that Cohesin holds sister chromatids together by embracing the two DNA duplexes within its coiled-coil arms, trapping sister DNAs inside its ring (Gruber et al., 2003). This topological model for Cohesin interaction with DNA is further supported by experiments with circular DNA molecules in yeast (Ivanov and Nasmyth, 2005) explaining how cleavage of Scc1/Rad21 by separase at the onset of anaphase opens the ring and thereby triggers sister chromatid separation (Uhlmann et al., 1999). An alternative model suggests that two Cohesin complexes interact directly with DNA and with each other to hold two DNA molecules together. It has also been proposed that one Cohesin ring encircling one sister chromatid may interact with proteins (other than Cohesins) present on the other sister chromatid (Chang et al., 2005). In a variation of the latter model, Cohesin rings embracing a single chromatid could interact with each other either directly or with the participation of accessory factors to establish cohesion.

In normal cells, Cohesin can connect sister chromatids only in S phase, even though newly synthesized Cohesin complexes can associate with DNA in the subsequent G2 phase (Uhlmann and Nasmyth, 1998). Until recently, it was generally assumed that sister–chromatid cohesion can only be generated at replication forks where DNA is synthesized. The finding that cohesion can also occur in G2 phase if DNA is damaged challenged this opinion (Strom et al., 2007; Strom and Sjogren, 2005). In this process, if DNA is damaged by double-strand breaks, Cohesin is loaded *de novo* to the damaged DNA forming physical connections between the sister chromatids (Strom et al., 2004). However, repair of damaged DNA also depends on DNA synthesis, so it remains possible that establishment of cohesion after DNA damage is also mechanistically coupled to DNA synthesis, since those two processes are linked during S phase.

3.1 Cohesin in gene regulation

Early evidence that the sister chromatid cohesion apparatus plays a role in gene expression arose from studies on gene silencing in yeast and on long-range gene activation in *Drosophila* (Donze et al., 1999; Rollins et al., 1999). The yeast data indicated that Cohesin helps establish the boundaries that define the silenced chromatin domains, while *Drosophila* studies showed that Nipped-B (Delangin) is required for long-range activation. The function of Delangin will be explained in details further in this chapter.

Cohesin and yeast gene silencing

In *S.cerevisiae*, the silent loci are surrounded by boundary elements that are able to block the spread of the silent chromatin established by chromatin-bound SIR (silent information regulator) protein complexes. Deletion of the boundaries, that surround silent-mating-type loci, leads to spreading of SIR complexes into the neighboring regions (Donze et al., 1999). A genetic screen of the factors that are able to establish this type of boundary revealed that some mutations of Smc1 cause loss of boundary function (Donze et al., 1999). Supporting this idea, the Scc1/Rad21 Cohesin subunit binds to boundary elements of the silent-mating-type locus, as well as to the locus itself. This is consistent with the idea that Cohesin functions in controlling the spread of silencing complexes (Chang et al., 2005; Glynn et al., 2004;

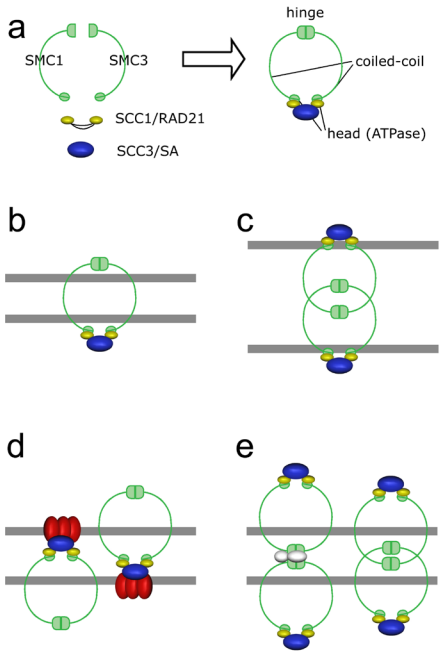


Figure 5

Hypothetical models of Cohesin-mediated cohesion

a Subunit composition of the Cohesin complex and arrangement of the subunits in the Cohesin ring
b – e Models proposed to explain how Cohesin interacts with DNA
b Cohesin molecule entraps the two sister chromatids
c the two Cohesin rings interact with each other and bind each to a chromatid
d Cohesin ring embracing a single chromatid interacts with proteins (red ovals) on the opposite chromatid
e Cohesin ring embracing each a single chromatid interacts with each other. The intermolecular interactions are stabilized by additional factors (light grey ovals).
Notice that only in model c Cohesin binds directly to DNA molecule

Laloraya et al., 2000). Cohesin also binds to non-transcribed region in the ribosomal RNA repeat and to subtelomeric repeats. Binding of Cohesin to these sites requires SIR proteins, indicating that Cohesin binding to this loci dependent on silencing (Chang et al., 2005; Kobayashi et al., 2004). The presence of Cohesin in rDNA and subtelomeric repeats raises the possibility that Cohesin might also be involved in inhibition of spreading of silencing complexes at those locations. Loss of cohesion is essential to establish silencing loci, while loss of Scc1 causes premature silencing (Lau et al., 2002). More recent studies show that silencing has a reciprocal effect on Cohesin binding and cohesion (Chang et al., 2005).

Genome-wide mapping of Cohesin binding sites in yeast revealed a strong correlation of Cohesin binding to non-transcribed spacer regions between convergent transcription units (Glynn et al., 2004; Lengronne et al., 2006). This location of Cohesin might be a consequence of RNA polymerase being involved in "pushing" Cohesin from its initial locations. It should be noted that this idea has not been shown experimentally and hence that it is currently unknown how Cohesin re-localizes and whether it requires factors other than the transcriptional machinery for this "maneuver".

Cohesin in long-range gene activation in *Drosophila*

A genetic screen identified *Nipped-B* gene, the homologue of *Scc2* in yeast and *Delangin* in humans, as a mediator of a long range enhancer-promoter interaction (Rollins et al., 1999). This finding initially suggested that Cohesin itself might facilitate activation by stabilizing interactions between distant enhancers and promoters in a way similar to sister chromatid cohesion (Dorsett, 1999; Hagstrom and Meyer, 2003). However this mechanism appears unlikely after it was shown that reduced levels of Cohesin subunits increased *cut* expression, opposite to the effect of Nipped-B mutations (Dorsett et al., 2005; Rollins et al., 2004). This opposite effect of Cohesin and Nipped-B on *cut* gene expression suggests certain possibilities. It is possible that Cohesin acts as an insulator that blocks enhancer-promoter communication, similar to the manner in which Cohesin blocks spreading of silencing protein complexes in yeast. In this case Nipped-B, in addition to loading Cohesin onto chromosomes, may also remove or help the re-localization of Cohesin to permit gene activation. Reduced levels of Nipped-B would slow the rate of Cohesin mobilization and have an indirect effect on gene transcription through Cohesin (Dorsett, 1999; Dorsett, 2007; Gause et al., 2001).

One could argue that sister chromatid cohesion *per se* inhibits gene activation. In such a model Cohesin would block communication of the enhancers and promoters of the sister chromatids thereby reducing the ability of the enhancer and promoter on each sister chromatid to come together. However it appears unlikely to act in this way since a reduced level of *Pds5*, the factor required for cohesion, but not Cohesin binding, does not change gene expression (Dorsett et al., 2005). Another possibility is that Cohesin inhibits gene activation by interacting with Nipped-B and

preventing it from performing another (unknown) function. Current evidence favors the Cohesin insulator model over the idea of an interaction between Cohesin and Nipped-B (Dorsett, 2007). There is more evidence that the effect of Nipped-B on gene expression involves its role in regulating the binding of Cohesin to chromosomes. The mutation in *pds5* that reduces Cohesin-binding increases *cut* expression. A null *pds5* mutation, which causes loss of sister chromatid cohesion, but not Cohesin binding, does not increase gene activation (Dorsett et al., 2005).

3.2 Genomic mapping of Cohesin

The genomic location of Cohesin has been demonstrated by chromatin-immunoprecipitation followed by hybridization to DNA microarray (Chip-on-chip) studies in different species (Glynn et al., 2004; Misulovin et al., 2008; Parelho et al., 2008; Wendt et al., 2008). In *S.cerevisiae*, Cohesin binds approximately every 11kb correlating with AT content and with a clear preference for intergenic regions between convergent transcription units (Glynn et al., 2004). Centromeric regions are also enriched for Cohesin. Chip-on-chip mapping showed Cohesin subunits to be present at the inactive genes. This suggests that in yeast Cohesin is re-localized in response to transcription: after activation of gene transcription Cohesin is redistributed around the genes (Bausch et al., 2007; Glynn et al., 2004; Wendt et al., 2008). Once a gene becomes inactive, Cohesin returns to its previous location (Bausch et al., 2007). The dynamics nature of Cohesin association with chromatin suggests that Cohesin might be involved in organizing local chromatin structure.

Chip-on-Chip results from *Drosophila* cell lines show a different binding pattern of Cohesin compared to yeast, having a clear preference for active genes. Its binding sites are found mostly in active transcriptional units, but also in untranscribed regions (Misulovin et al., 2008). The binding sites of SA (Scc3) are enriched in 5' untranslated regions (5'UTRs) and in intronic regions. The prevalence of Cohesin subunits at 5'UTRs suggests that there is correlation between Cohesin binding and the presence of a poised PolII at the 5'end of a subset of genes. Thus Cohesin may function in the regulation of gene transcription by aiding the loading of the RNA transcription machinery. In *Drosophila*, there is a negative correlation between Cohesin binding and histone H3K27 trimethylation, a marker of transcriptionally repressed chromatin. This further suggests that heterochromatin either inhibits Cohesin binding or that Cohesin acts as a barrier to heterochromatic spreading (Misulovin et al., 2008). Although yeast and *Drosophila* showed differences in Cohesin binding patterns, they both show that Cohesin is able to relocate in response to a different state of transcription within a region. Additionally, while Cohesin is present in non-coding regions of transcribed genes in *Drosophila*, it has reduced binding over coding regions, similar to the situation in yeast. The mechanism of how Cohesin responds to different states of transcription has not been determined (McNairn and Gerton, 2008).

It is not clear why the distribution of Cohesin binding sites differs so much in yeast and *Drosophila*. It could be that is not particularly important for sister chromatid

cohesion where Cohesin binds, as long as the spacing along the DNA fiber is sufficiently dense to maintain sister chromatid cohesion. Such a loose constraint may have created opportunities during evolution to re-use the unique molecular properties of Cohesin for other regulatory functions. For example insulator activity of Cohesin has not only been observed in mammals but has also been suggested for yeast and *Drosophila* (Dorsett, 2007; Peric-Hupkes and van Steensel, 2008).

3.3 CTCF and Cohesin cooperation

Recently, four groups mapped Cohesin-binding sites in mammalian genome and found a substantial overlap with CTCF binding sites (Parelho et al., 2008; Rubio et al., 2008; Stedman et al., 2008; Wendt et al., 2008). Using Chip-on-Chip, Wendt et al. determined the binding of the Cohesin subunit Rad21 and CTCF through most of the non-repetitive human genome in HeLa cells (Wendt et al., 2008). They found nearly 14,000 sites, similar to the number of CTCF binding sites detected in previous studies (Kim et al., 2007). The high degree of co-localization of CTCF and Cohesin binding sites suggests that CTCF and Cohesin influence each others binding. Indeed, it appears that CTCF influences the positioning of Cohesin, but not overall binding. In CTCF-depleted cells Cohesin was still bound to the chromatin but in most cases not to the shared CTCF sites. Wendt et al. also showed that CTCF-depleted cells do not show an effect on sister chromatid cohesion indicating that the positioning of Cohesin by CTCF is not required for cohesion (Wendt et al., 2008).

The effect of Cohesin on CTCF binding is less clear and may vary from site to site. The knockdown of the Rad21, Cohesin subunit, had no (Parelho et al., 2008) or a moderate (Wendt et al., 2008) effect on CTCF binding. It has also been suggested that Cohesin may increase the accessibility of some sites to CTCF by affecting chromatin structure, or by increasing the binding affinity through CTCF-Cohesin cooperative binding (Gause et al., 2008).

The co-localisation of CTCF and Cohesin raises the question what is the functional significance of shared sites, with respect to gene expression. Several functional analyses have demonstrated that the Cohesin complex plays an important role in gene regulation by CTCF. Using reporter gene assays, Wendt et al show that depletion of the Rad21 Cohesin subunit partially reduces the activity of the chicken β-globin insulator (Wendt et al., 2008). In a further analysis they investigated the effect of depletion either Rad21 or CTCF on *H19* transcript levels at the endogenous *H19*/*Igf2* locus. The transcript levels of *H19* where decreased while *Igf2* transcripts were increased, mimicking what occurs when the insulator is inactivated by DNA methylation on the paternal chromosome during imprinting. The finding that Cohesin accumulates at CTCF-binding sites suggests the exciting possibility that Cohesin stabilizes CTCF-mediated long-range interactions. It is possible that a portion of the available Cohesin is diverted from a role in sister chromatid cohesion to support long-range interactions.

Although the data above strongly suggest a role in long-range interaction, it

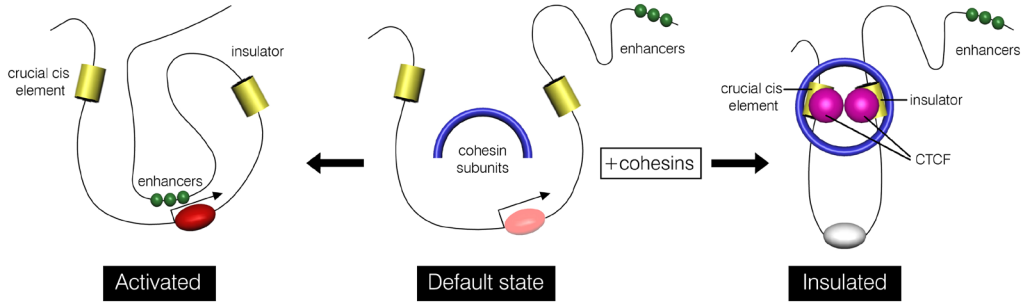


Figure 6
Hypothetical model of how Cohesin complexes cooperate with CTCF
The model is based on data obtained largely with G2/M cells.
The Figure proposes that the CTCF-Cohesin complexes might establish and stabilize long-range interactions. The proposal that these interactions involve CTCF-CTCF interactions is entirely hypothetical.

is important to determine how and where Cohesin subunits organize higher-order chromatin structures. To allow a better description of Cohesin in this process, it is necessary to perform 3C long range interaction experiments on cells that lack CTCF and/or Cohesin.

3.4 The Cohesin accessory factors

The association of Cohesin with chromatin requires the SMC proteins to have a functional hinge domain and ATPase activity, and it requires the activity of the Scc2/Scc4 (Delangin) loading complex (Ciosk et al., 2000; Gruber et al., 2006; Weitzer et al., 2003). Metazoan Delangin is a large chromatin-associated protein that contains multiple HEAT repeats, which are related to Armadillo-like repeats that mediate protein-protein interactions (Neuwald and Hirano, 2000). In addition to the HEAT repeats, the N-terminus of Metazoan Delangin contains a Heterochromatin Protein 1 (HP1) binding domain that was shown to interact with HP1α. This interaction raises the possibility that Delangin is involved in the establishment and maintenance of

heterochromatin domains (Lechner et al., 2005).

The recently identified metazoan Scc4, a binding partner of Delangin, appears to have an important role in neuronal development. Depletion of Scc4 (or Mau-2 in *C. elegans*) causes severe Premature Sister Chromatid Separation and impaired loading of Cohesin in human cells (Seitan et al., 2006; Watrin et al., 2006). The identification of Mau-2 as Scc4 was among the first indications that Cohesin, and Cohesin associated proteins, may have an additional role related to neuronal development.

Chromosome cohesion is established during S-phase and in response to DNA damage in G2-phase. In both cases, Cohesion is dependent on upon the activity of the Eco1 protein (Ball and Yokomori, 2008; Ben-Shahar et al., 2008). *In vitro*, yeast Eco1 and human homologues are capable of autoacetylation. Recently, Eco1 and human paralogues were demonstrated to specifically acetylate the Smc3 protein. This event is essential for the establishment of cohesion (Zhang et al., 2008). Depletion or a mutation in the Eco1 gene can result in the loss of chromosome cohesion and in premature sister chromatid separation (Hou and Zou, 2005).

Once, cohesion has been established in S-phase by the Cohesin complex, it has to be maintained to function properly. Several proteins are important for on the Cohesin complex and Cohesin-mediated cohesion. One of them is *Pds5*, an essential gene in budding yeast (Hartman et al., 2000). The deletion of one of the *Pds5* genes in mouse is embryonic lethal, due to developmental defects, which indicates that Pds5 has an important role in the process of cohesion.

4. Delangin

Drosophila Nipped-B (*Delangin* in mammals) is an essential gene that was identified in a genetic screen for mutants interfering with long-range gene activation of the *cut* and *Ultrabithorax* (Ubx) homeobox genes by enhancers positioned some 80 and 50 kbp away from the gene promoters (Rollins et al., 2004; Rollins et al., 1999). Nipped-B facilitates the activation of the *cut* gene by the distant wing-margin enhancer. Mutations that result in reduced Nipped-B levels result in a lower expression of *cut* gene. In contrast to other *cut* activators, mutations in *Nipped-B* have a much more striking effect when *cut* activation is hindered by a *gypsy* insulator insertion rather than by loss of important components of the enhancer sequence (Rollins et al., 1999). Subsequent work revealed that, as expected, Nipped-B is also required for sister chromatid cohesion (Rollins et al., 2004) and it was shown that Delangin and its orthologues are required for Cohesin to bind to chromosomes (Arumugam et al., 2003; Ciosk et al., 2000; Gillespie and Hirano, 2004; Seitan et al., 2006).

Protein motifs and sequence conservation

Delangin is expressed in two isoforms of 2804 and 2697 amino acids. Delangin isoforms have very few features that provide clues to their function. The amino-terminal region contains several potential nuclear localization sequences.The tandem repeats of an undecapeptide sequence are conserved in the amino-terminal region in mammals, chick and *Xenopus*. The Carboxyl-terminal region has a cluster of HEAT repeats: protein-protein interaction motifs that have been reported in conserved position in Delangin homologues, including *Drosophila Nipped-B*, *Saccharomyces cerevisiae Scc2*, *Schizosaccharomyces pombe Mis4* and *Coprinus cinereus Rad9*.

C. cinereus	*Rad 9*
C.elegans	*Pqn-85*
S.pombe	*Mis 4*
S.cerevisie	*Scc2*
Drosophila	*Nipped-B*
Xenopus	*XScc2*
Mouse	*Delangin*
H.sapiens	*NIPBL*

Delangin orthologues in different species

4.1 Chromosomal and DNA repair functions of Cohesin and Delangin orthologues

DSB (DNA double-strand break) repair is an important component of the machinery that maintains the integrity of the genome after DNA damage. Mistakes in this process lead to genomic instability, for example. chromosomal translocations or aneuploidy. The cell responds to DSBs in several ways, one of which is through the homologous recombination (HR) repair pathway. In this process, an intact sister chromatid serves as a template for the repair that occurs in S/G2 phase, following DNA replication (West, 2003).

Studies in *S.cerevisie* and chicken cells gave the first evidence of the role of Cohesin in postreplicative DNA repair (Sjogren and Nasmyth, 2001; Sonoda et al., 2001). Immunostaining analysis of laser-induced damage in human cells and Chip (Chromatin immunoprecipitation) analysis of the HO-endonuclease-induced DSB site in yeast revealed that Cohesin specifically accumulates at the damage sites, suggesting a specific function of Cohesin in DNA repair (Kim et al., 2002; Strom et al., 2004). Cohesin accumulation at the damage site requires Scc2-Scc4, indicating that the loading mechanism is similar to that for genome-wide Cohesin binding following mitosis (Strom et al., 2004; Unal et al., 2004).

In both types of cells, human and yeast in in vitro studies, Cohesin recruitment to the damage site is dependent on the Mre11-Rad50 complex. This complex is one of the first factors to recognize and to accumulate at DNA ends (Kim et al., 2002; Strom et al., 2004). Cohesin clustering in the area surrounding the damage site coincides with and requires H2A phosphorylation (γH2A) in yeast. It is still unclear whether Cohesin binds directly or indirectly to phosphorylated H2A (Ball and Yokomori, 2008).

Since Cohesin loading at the damaged sites appears to correlate with the presence of sister chromatids, it is speculated that Cohesin establishes local sister chromatid cohesion at the damaged sites to facilitate HR repair by promoting homologues pairing of damaged and intact sister chromatids. One elegant experiment in *S.cerevisiae* showed that Cohesin is indeed needed at the damaged site in G2 to promote sister chromatid cohesion. In this experiment it was shown that Cohesin is newly recruited in the G2 phase by inactivating a temperature sensitive mutant Cohesin involved in genome-wide sister chromatid cohesion and complementing it with an inducible temperature-resistant Cohesin (Strom et al., 2007).

In yeast, in the absence of any damage, newly expressed Cohesin accumulates at its normal binding sites in G2 phase, but is unable to establish sister chromatid cohesion (Lengronne et al., 2006; Strom et al., 2004). In contrast, Cohesin at damaged sites can establish sister chromatid cohesion. This fact suggests that function of Cohesin, of keeping sister chromatids together, is reactivated in response to DNA damage (Strom et al., 2004).

How DSB in G2/M induces sister chromatid cohesion at the lesions and on the Cohesin-associated regions (CARs) is still not understood. In the absence of DNA damage, Cohesin is loaded at CARs but cannot mediate cohesion (Lengronne et al., 2006). Upon DSBs, the DNA damage response pathway induces DSB-proximal Cohesin loading (Unal et al., 2004). It has been suggested that DSBs are cohesive through the action of factors that have different action in other phases of the cell cycle. They are activated by DNA damage only in G2/M. In undamaged cells, the same factor or factors are active only during S phase therefore Cohesin loaded at CARs in G2/M generates cohesion (Unal et al., 2007).

The reactivation of Cohesin is not limited to damaged sites but occurs genome-wide when there is Cohesin that is synthesized but not incorporated. A limited number of DSBs on one chromosome is sufficient to induce *de novo* cohesion on other chromosomes i.e. Cohesin complexes could establish cohesion on both damaged and undamaged chromosomes during G2 phase. This implies that the presence of DNA double-strand breaks somehow reactivates the molecular machinery that is normally responsible for cohesion only during S phase (Strom et al., 2007; Unal et al., 2007).

Inactivation of Cohesin subunits or Scc2/Scc4 results in a reduced efficiency of post-replicative DSB in G2 phase of the cell cycle. The local enrichment of Cohesin at a damaged site depends on the Scc2/Scc4 complex (Strom et al., 2007). Additionally, a homologue of *Delangin*, *Pqn-85*, was identified in genome wide screens for genes required for resistance to ionizing radiation in *C.elegans* (van Haaften et al., 2006).

Other interesting evidence of the importance of Delangin for DNA repair comes

from humans, where cells derived from Cornelia de Lange patients, who have a mutation in Delangin gene, have a reduced capacity to tolerate DNA damage. Exposure of CdLS fibroblasts and B-lymphoblastoid cells to the damaging agent mitomycin C (MMC) leads to increased number of chromosomal aberrations. After X-ray exposure, the cells with increased chromosomal aberrations were detected only if they were irradiated in G2 phase when the repair of the double strand breaks is dependent on sister chromatid cohesion (Vrouwe et al., 2007).

4.2 Cornelia de Lange syndrome

Cornelia de Lange syndrome (CdLS; OMIM 122470) is a rare multisystem developmental disorder with characteristic facial dysmorphia, growth and cognitive retardation, malformations of the upper limbs and a variety of abnormalities affecting a wide range of tissues and organs. It is estimated to occur at a frequency of 1 per 10,000 to 30,000 births (Jackson et al., 1993). The deficits begin prenatally and continue after birth. A remarkable feature is the diversity of developmental defects, and thus, CdLS may provide clues regarding the mechanisms of other developmental disorders that display subsets of these features. There is marked heterogeneity in the phenotype. Individuals who show profound growth and neuro-developmental delay, sometimes accompanied by severe limb defect, have been described as exhibiting classical CdLS; in mildly affected individuals, the growth retardation and developmental delay are less severe. Affected individuals are typically identified as sporadic cases, but several reports have documented families with multiple affected members (Russell et al., 2001). In some multicase pedigrees there is strong evidence for autosomal dominant inheritance: in others, germinal mosaicism (mosaicsm in which an individual has a subset of germline cells carrying a mutation that is not found in other germline cells) (Strachan, 2005).

Although genetic data and samples were accumulated over several years, mapping of the gene(s) responsible for CdLS proved challenging because it is genetically dominant, and the vast majority of cases are spontaneous. There are rare examples of a parent having more than one CdLS child that is most likely caused by germline mosaicism, and a few cases in which a mildly affected patient passed on the disorder to their offspring. The first CdLS gene was mapped by linkage exclusion analysis and a *de novo* balanced translocation affecting the 5p13.1 region (Krantz et al., 2004; Tonkin et al., 2004). Subsequent analysis revealed mutations in the *Delangin* (*NIPBL*) gene in 15p13.1 in many of the patients, and then several studies have identified variety a of *Delangin* mutations in CdLS patients (Bhuiyan et al., 2006; Borck et al., 2004),(Gillis et al., 2004; Krantz et al., 2004),(Tonkin et al., 2004).

Delangin mutations have been identified in about half of the CdLS patients. Recently, two mutations in the *Smc1L1* gene (Xp11.21) were identified in certain relatively mild cases of CdLS that lacked a *Delangin* mutation (Musio et al., 2006). *Smc1L1*, which encodes the Smc1 Cohesin subunit, is X-linked, and the affected individuals are male. This indicates that the mutations are unlikely to be strong loss-

of-function alleles, which would likely to be lethal. The *Smc1L1* gene escapes X-inactivation and of the 3 female carriers that passed on *Smc1L1*, one displayed very mild characteristic consistent with CdLS. It is currently unknown how many cases are caused by *Smc1L1* mutations (Dorsett, 2007).

It remains possible that mutations in other genes encoding components of the sister chromatid cohesion apparatus could cause CdLS. However, none of the other potential candidate genes, such as the human *Mau-2* (KIAA0892) or *Pds5* genes, are located in the five candidate regions identified by linkage exclusion (Dorsett, 2007). It is also likely that some CdLS-causing mutations occur in non-transcribed regulatory regions of *Delangin* gene, or have been missed because the large gene size and complex splicing patterns currently preclude sequencing of the entire *Delangin* transcription unit. Improvements in screening for *Delangin* mutations, and large scale screening of patients for *Smc1L1* mutations will resolve some of these questions in the near future (Dorsett, 2007).

The known *Delangin* mutations indicate that reduced levels or activity of the encoded Delangin protein cause CdLS. Many *Delangin* mutations predict a protein truncation, although it is currently unknown whether or not the encoded mutant proteins are expressed. A deletion removing the *Delangin* region was seen in a severe case of CdLS, supporting the idea that loss-of–function mutations cause CdLS. Many of the less severe cases of CdLS are associated with mutations that cause amino acids substitutions, although there are exceptions to the general correlation of milder CdLS cases with missense mutations and more severe forms with truncations (Bhuiyan et al., 2006; Gillis et al., 2004). Many polymorphisms are found in unaffected parents, and in addition to numerous other genetic differences between individuals, these polymorphisms could contribute to the phenotypic variability of the syndrome (Dorsett, 2007).

An intriguing finding from molecular analysis of *Delangin* mutations in CdLS patients is that a central conserved arginine residue (R2298) in one of the HEAT repeats has been altered by missense mutations in multiple independent cases, suggesting that it is a critical residue. Another important finding is that with only a few exceptions, the *Delangin* missense mutations associated with CdLS affect residues conserved in *Drosophila Nipped-B*. Thus, it appears likely that determination of the molecular function of Nipped-B in *Drosophila* and other model organisms will illuminate the molecular etiology of CdLS (Dorsett, 2007).

Evidence from *Drosophila* favors the idea that the developmental disorders in CdLS are caused by altered expression of genes that control development. Although the effects on sister chromatid cohesion are seen in homozygous *Nipped-B* mutants, no effects on cohesion are observed in heterozygous mutants, which shows measurable effects on gene expression. Even partial reduction of Nipped-B by RNAi sufficient to cause lethality had no detectable effect on cohesion (Rollins et al., 2004).

If the cohesion insulator mechanism, proposed to explain the opposite effects on Nipped-B and Cohesin on gene expression in *Drosophila* is correct, one would predict that several genes that control development will be affected in CdLS. Cohesin

likely binds every 20 kb or so, and many mammalian genes rely on distant control elements for appropriate temporal and spatial expression. Many of those could explain some of the developmental deficits in CdLS (Dorsett, 2007).

The *Drosophila* data also suggest that the *Smc1L1* is associated with milder forms of CdLS likely cause alterations in gene expression. A null allele of *Smc1L1* in *Drosophila* dominantly increases *cut* gene expression (Dorsett et al., 2005), as does RNAi knockdown of the Stromalin and Rad21 Cohesin subunit (Rollins et al., 2004). These effects are opposite to those of *Nipped-B* mutations, and thus it might be surprising at first that human *Smc1L1* mutations cause similar developmental problems as *Delangin* mutations. Multiple explanations resolve this apparent paradox. The CdLS patients with *Smc1L1* mutations display relatively mild symptoms, and it is possible that over-expression or inappropriate expression of certain genes could have similar, but not identical, effects on development as their under expression (Dorsett, 2007).

An intriguing possibility is that the identified *Smc1L1* mutations are not null alleles, but slow down Cohesin binding dynamics as is proposed above in the case of reduced Delangin levels. One of the *Smc1L1* mutations alters a conserved residue at the junction of the N-terminal helix and the hinge domain that dimerizes with the Smc3 hinge. It has been shown for a prokaryotic SMC protein that the hinge is critical for DNA binding dynamics, and that hinge-DNA interactions stimulate the ATPase activity of the head domain (Hirano and Hirano, 2002; Hirano and Hirano, 2006). The ATPase activity, along with Scc2/Nipped-B/Delangin, is essential for the binding of Cohesin to chromosomes (Arumugam et al., 2003). The stimulation of the head ATPase activity by the hinge domain may involve direct interactions, as atomic force microscopy has revealed hinge-head contacts in fission yeast Smc1-Smc3 heterodimers (Sakai et al., 2003). These contacts could be affected by the changes in the flexibility of the coiled-coil domain. Thus, it remains possible that *Smc1L1* mutations could slow the kinetics of Cohesin binding and unloading, and thereby alter gene expression in manner similar to that proposed for *Delangin* loss-of–function mutations. Another possibility is that a subset of the CdLS deficits, including those found in patients with *Smc1L1* mutations, reflect changes in sister chromatid cohesion, as opposed to changes in gene expression. Using immortalized cell lines, one study revealed mild cohesion defects in 40 % of CdLS patients (Kaur et al., 2005). The procedures used to prepare metaphase spreads enhance chromatid separation, so it is unclear if such cohesion defects occur *in vivo*, or the analysis reveals a subtle defect that has little role in the etiology of CdLS.

The current challenge is to determine the relative contribution of changes in gene expression and sister chromatid cohesion in CdLS, and to identify the critical target genes responsible for the known developmental effects. This will be aided by studies on the molecular mechanisms by which *Nipped-B* and *Cohesin* subunit mutations affect gene expression, cohesion and development in *Drosophila*, but a mammalian model is essential to identify target genes relevant to specific CdLS birth defects (Dorsett, 2007).

5. Role of Ldb1 in long-range gene activation

Existing evidence indicates that most transcriptional activator proteins recruit the basal transcriptional machinery and increases its binding to a promoter (Ptashne and Gann, 1997). Often, an activator binds to a sequence located many kilobases away from the promoters it activates. In these cases, the local concentration of the activator relative to the promoter it activates is not higher than its concentration relative to many other promoters (Rippe, 2001). This suggests that mechanisms other than diffusion-driven chromatin looping support long-range activation. Indeed, several sequences that facilitate specific long-range interactions have been identified in *Drosophila*, mostly in the *Antennapedia* and *Bithorax* homeotic gene complexes (Calhoun and Levine, 2003; Hopmann et al., 1995; Qian et al., 1992; Ronshaugen and Levine, 2004; Zhou and Levine, 1999). Their function is to support long-range activation in many genes (Dorsett, 1999). One reason for this idea is that insulator sequences, such as the one in the *Drosophila gypsy* transposon, block diverse enhancers in many genes. The location of insulators is important since they are able to block enhancers only when positioned between an enhancer and a promoter. It has been postulated for the *Drosophila gypsy* insulator that they interfere with general factors that function between many enhancers and promoters to facilitate enhancer-promoter communication (Dorsett, 1999).

For the identification of general facilitators of enhancer-promoter communication, genetic screens were performed to isolate of factors that support activation of the *cut* gene by a wing margin-specific enhancer located 85 kbp upstream of the promoter (Morcillo et al., 1997; Morcillo et al., 1996) . The region between this enhancer and the promoter contains many enhancers that activate *cut* in specific tissues during embryogenesis and larval development (Jack and DeLotto, 1995). In addition to tissue-specific activators that bind to the wing margin enhancer, these screens identified two proteins, Chip (Ldb1 in mammals) and Nipped-B (Delangin in Mammals), that are expressed in virtually all cells, and facilitate the expression of diverse genes. Chip (Ldb1) interacts with many DNA-binding proteins, and was suggested to support the cooperative binding of proteins to enhancers and to the sites between enhancers and promoters (Gause et al., 2001; Morcillo et al., 1997; Torigoi et al., 2000). The linking model for long-range enhancer-promoter interactions was proposed to explain the properties of *Drosophila* Chip protein (Dorsett, 1999; Gause et al., 2001). Recent experiments have shown that Ldb1, the mouse homologue of Chip, forms large complexes with transcriptional factors (Meier et al., 2006) that bind specific sites in the genome indicating that Ldb1 complexes more likely play role in looping.

Many proteins, both ubiquitously expressed and erythroid specific are known to be involved in β-globin regulation and chromatin looping. Most of these factors, often present in protein complexes, bind to the promoters of the globin genes and to the HS sites in the β-globin locus. It was shown that *cis*-regulatory elements and their distant target genes come in close proximity during gene activation with the exclusion of intervening chromatin (Palstra et al., 2003; Splinter et al., 2006; Tolhuis

et al., 2002). The erythroid activators EKLF, GATA-1 and FOG-1 are required for close interactions (Drissen et al., 2004; Vakoc et al., 2005).

The widely expressed nuclear protein Ldb1 (CLIM2, NLI) is the human homologue of *Drosophila* Chip protein which was identified in a genetic screen for factors involved in long-range activation of the *cut* gene (Morcillo et al., 1997; Morcillo et al., 1996). The null phenotype of the murine *Chip* homologue *Ldb1* exhibits pleiotropic effects during mouse development, including failure of erythroid differentiation (Mukhopadhyay et al., 2003).

Earlier work showed that Ldb1 participates in a DNA-binding complex formed in murine erythroid cell extracts with the hematopoietic factors GATA-1, LMO-2, TAL-1 and E2A (Wadman et al., 1997). Ldb1 functions locally as a positive regulator of late erythroid gene expression through the promoters of the *Eklf, c-kit, P4.2* and *glycophorin A* genes (Anderson et al., 1998; Lahlil et al., 2004; Lecuyer et al., 2002). Ldb1, together with Rb, Eto2 and BRG1 proteins can also play a repressive role in gene expression (Goardon et al., 2006; Vitelli et al., 2000). Ldb1 is also a member of a complex that assembles at the sites upstream of the α-globin gene at a particular stage of development (Anguita et al., 2004). The same complex binds to the LCR and β-globin promoter in murine erythroleukemic (MEL) cells (Brand et al., 2004).

In recent studies, it was shown that the Ldb1/GATA-1/TAL-1/LMO2 complex binds in vivo to the human β-globin LCR which is located upstream of the globin genes and is required for their high-level expression in erythroid cells (Stamatoyannopoulos, 2005). In these cells, both Ldb-1 and GATA-1 were reported to be present at the promoter of the β_{maj} gene and HS1-HS4 of the LCR in mouse fetal liver and MEL cells and at HS1-HS4 in human K562 cells.

The truncation of Ldb1 reduces the expression of the ε-gene on a minichromosome containing HS2 in K562 cells, while a knockdown of Ldb1 leads to a decreased β_{maj} expression in differentiated MEL cells. Furthermore, if there is a reduced level of Ldb1, the interaction between HS2 and the promoter of the β_{maj} is reduced to background levels. This suggest that Ldb1 is either directly or indirectly involved in long-range chromatin interactions at the β-globin locus (Song et al., 2007). Since Ldb1 is not a DNA-binding protein, but interacts with many factors that are involved in the regulation of erythroid genes, it is more likely that this is an indirect effect where an assembly of complexes on the locus is not appropriate. It would be interesting to understand how the depletion of Ldb1 affects binding of individual partners from the complex.

References:

Anderson, K. P., Crable, S. C. and Lingrel, J. B. (1998). Multiple proteins binding to a GATA-E box-GATA motif regulate the erythroid Kruppel-like factor (EKLF) gene. J Biol Chem 273, 14347-54.

Anguita, E., Hughes, J., Heyworth, C., Blobel, G. A., Wood, W. G. and Higgs, D. R. (2004). Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. Embo J 23, 2841-52.

Arumugam, P., Gruber, S., Tanaka, K., Haering, C. H., Mechtler, K. and Nasmyth, K. (2003). ATP hydrolysis is required for cohesin's association with chromosomes. Curr Biol 13, 1941-53.

Ball, A. R., Jr. and Yokomori, K. (2008). Damage-induced reactivation of cohesin in postreplicative DNA repair. Bioessays 30, 5-9.

Banerji, J., Rusconi, S. and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell 27, 299-308.

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell 129, 823-37.

Bausch, C., Noone, S., Henry, J. M., Gaudenz, K., Sanderson, B., Seidel, C. and Gerton, J. L. (2007). Transcription alters chromosomal locations of cohesin in Saccharomyces cerevisiae. Mol Cell Biol 27, 8522-32.

Bell, A. C. and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature 405, 482-5.

Bell, A. C., West, A. G. and Felsenfeld, G. (2001). Insulators and boundaries: versatile regulatory elements in the eukaryotic. Science 291, 447-50.

Ben-Shahar, T. R., Heeger, S., Lehane, C., East, P., Flynn, H., Skehel, M. and Uhlmann, F. (2008). Eco1-dependent cohesin acetylation during establishment of sister chromatid cohesion. Science 321, 563-6.

Bender, M. A., Byron, R., Ragoczy, T., Telling, A., Bulger, M. and Groudine, M. (2006). Flanking HS-62.5 and 3' HS1, and regions upstream of the LCR, are not required for beta-globin transcription. Blood 108, 1395-401.

Bender, M. A., Mehaffey, M. G., Telling, A., Hug, B., Ley, T. J., Groudine, M. and Fiering, S. (2000). Independent formation of DnaseI hypersensitive sites in the murine beta-globin locus control region. Blood 95, 3600-4.

Bhuiyan, Z. A., Klein, M., Hammond, P., van Haeringen, A., Mannens, M. M., Van Berckelaer-Onnes, I. and Hennekam, R. C. (2006). Genotype-phenotype correlations of 39 patients with Cornelia De Lange syndrome: the Dutch experience. J Med Genet 43, 568-75.

Bickel, S. and Pirrotta, V. (1990). Self-association of the Drosophila zeste protein is responsible for transvection effects. Embo J 9, 2959-67.

Bonifer, C. (2000). Developmental regulation of eukaryotic gene loci: which cis-regulatory information is required? Trends Genet 16, 310-5.

Borck, G., Redon, R., Sanlaville, D., Rio, M., Prieur, M., Lyonnet, S., Vekemans, M., Carter, N. P., Munnich, A., Colleaux, L. et al. (2004). NIPBL mutations and genetic heterogeneity in Cornelia de Lange syndrome. J Med Genet 41, e128.

Brand, M., Ranish, J. A., Kummer, N. T., Hamilton, J., Igarashi, K., Francastel, C., Chi, T. H., Crabtree, G. R., Aebersold, R. and Groudine, M. (2004). Dynamic changes in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics. Nat Struct Mol Biol 11, 73-80.

Bulger, M., Bender, M. A., van Doorninck, J. H., Wertman, B., Farrell, C. M., Felsenfeld, G., Groudine, M. and Hardison, R. (2000). Comparative structural and functional analysis of the olfactory receptor genes flanking the human and mouse beta-globin gene clusters. Proc Natl Acad Sci U S A 97, 14560-5.

Bulger, M. and Groudine, M. (1999). Looping versus linking: toward a model for long-distance gene activation. Genes Dev 13, 2465-77.

Bulger, M., Schubeler, D., Bender, M. A., Hamilton, J., Farrell, C. M., Hardison, R. C. and Groudine,

M. (2003). A complex chromatin landscape revealed by patterns of nuclease sensitivity and histone modification within the mouse beta-globin locus. Mol Cell Biol 23, 5234-44.

Burke, L. J., Zhang, R., Bartkuhn, M., Tiwari, V. K., Tavoosidana, G., Kurukuti, S., Weth, C., Leers, J., Galjart, N., Ohlsson, R. et al. (2005). CTCF binding and higher order chromatin structure of the H19 locus are maintained in mitotic chromatin. Embo J 24, 3291-300.

Calhoun, V. C. and Levine, M. (2003). Long-range enhancer-promoter interactions in the Scr-Antp interval of the Drosophila Antennapedia complex. Proc Natl Acad Sci U S A 100, 9878-83.

Capelson, M. and Corces, V. G. (2005). The ubiquitin ligase dTopors directs the nuclear organization of a chromatin insulator. Mol Cell 20, 105-16.

Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. F. and Fraser, P. (2002). Long-range chromatin regulatory interactions in vivo. Nat Genet 32, 623-6.

Chang, C. R., Wu, C. S., Hom, Y. and Gartenberg, M. R. (2005). Targeting of cohesin by transcriptionally silent chromatin. Genes Dev 19, 3031-42.

Chao, W., Huynh, K. D., Spencer, R. J., Davidow, L. S. and Lee, J. T. (2002). CTCF, a candidate trans-acting factor for X-inactivation choice. Science 295, 345-7.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., Wong, E., Orlov, Y. L., Zhang, W., Jiang, J. et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell 133, 1106-17.

Chung, J. H., Whiteley, M. and Felsenfeld, G. (1993). A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. Cell 74, 505-14.

Ciosk, R., Shirayama, M., Shevchenko, A., Tanaka, T., Toth, A., Shevchenko, A. and Nasmyth, K. (2000). Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. Mol Cell 5, 243-54.

de Laat, W., Klous, P., Kooren, J., Noordermeer, D., Palstra, R. J., Simonis, M., Splinter, E. and Grosveld, F. (2008). Three-dimensional organization of gene expression in erythroid cells. Curr Top Dev Biol 82, 117-39.

Defossez, P. A. and Gilson, E. (2002). The vertebrate protein CTCF functions as an insulator in Saccharomyces cerevisiae. Nucleic Acids Res 30, 5136-41.

Dillon, N. and Sabbattini, P. (2000). Functional gene expression domains: defining the functional unit of eukaryotic gene regulation. Bioessays 22, 657-65.

Dillon, N., Trimborn, T., Strouboulis, J., Fraser, P. and Grosveld, F. (1997). The effect of distance on long-range chromatin interactions. Mol Cell 1, 131-9.

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

Dorsett, D. (1999). Distant liaisons: long-range enhancer-promoter interactions in Drosophila. Curr Opin Genet Dev 9, 505-14.

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

Dorsett, D., Eissenberg, J. C., Misulovin, Z., Martens, A., Redding, B. and McKim, K. (2005). Effects of sister chromatid cohesion proteins on cut gene expression during wing development in Drosophila. Development 132, 4743-53.

Drissen, R., Palstra, R. J., Gillemans, N., Splinter, E., Grosveld, F., Philipsen, S. and de Laat, W. (2004). The active spatial organization of the beta-globin locus requires the transcription factor EKLF. Genes Dev 18, 2485-90.

Epner, E., Reik, A., Cimbora, D., Telling, A., Bender, M. A., Fiering, S., Enver, T., Martin, D. I., Kennedy, M., Keller, G. et al. (1998). The beta-globin LCR is not necessary for an open chromatin structure or developmentally regulated transcription of the native mouse beta-globin locus. Mol Cell 2, 447-55.

Farrell, C. M., Grinberg, A., Huang, S. P., Chen, D., Pichel, J. G., Westphal, H. and Felsenfeld, G. (2000). A large upstream region is not necessary for gene expression or hypersensitive site formation at the mouse beta -globin locus. Proc Natl Acad Sci U S A 97, 14554-9.

Farrell, C. M., West, A. G. and Felsenfeld, G. (2002). Conserved CTCF insulator elements flank the mouse and human beta-globin loci. Mol Cell Biol 22, 3820-31.

Festenstein, R., Tolaini, M., Corbella, P., Mamalaki, C., Parrington, J., Fox, M., Miliou, A., Jones, M. and Kioussis, D. (1996). Locus control region function and heterochromatin-induced position effect variegation. Science 271, 1123-5.

Filippova, G. N., Cheng, M. K., Moore, J. M., Truong, J. P., Hu, Y. J., Nguyen, D. K., Tsuchiya, K. D. and Disteche, C. M. (2005). Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development. Dev Cell 8, 31-42.

Gaszner, M. and Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. Nat Rev Genet 7, 703-13.

Gaszner, M., Vazquez, J. and Schedl, P. (1999). The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer-promoter interaction. Genes Dev 13, 2098-107.

Gause, M., Morcillo, P. and Dorsett, D. (2001). Insulation of enhancer-promoter communication by a gypsy transposon insert in the Drosophila cut gene: cooperation between suppressor of hairy-wing and modifier of mdg4 proteins. Mol Cell Biol 21, 4807-17.

Gause, M., Schaaf, C. A. and Dorsett, D. (2008). Cohesin and CTCF: cooperating to control chromosome conformation? Bioessays 30, 715-8.

Geyer, P. K. (1997). The role of insulator elements in defining domains of gene expression. Curr Opin Genet Dev 7, 242-8.

Geyer, P. K. and Corces, V. G. (1992). DNA position-specific repression of transcription by a Drosophila zinc finger protein. Genes Dev 6, 1865-73.

Ghirlando, R., Litt, M. D., Prioleau, M. N., Recillas-Targa, F. and Felsenfeld, G. (2004). Physical properties of a genomic condensed chromatin fragment. J Mol Biol 336, 597-605.

Gillespie, P. J. and Hirano, T. (2004). Scc2 couples replication licensing to sister chromatid cohesion in Xenopus egg extracts. Curr Biol 14, 1598-603.

Gillis, L. A., McCallum, J., Kaur, M., DeScipio, C., Yaeger, D., Mariani, A., Kline, A. D., Li, H. H., Devoto, M., Jackson, L. G. et al. (2004). NIPBL mutational analysis in 120 individuals with Cornelia de Lange syndrome and evaluation of genotype-phenotype correlations. Am J Hum Genet 75, 610-23.

Glynn, E. F., Megee, P. C., Yu, H. G., Mistrot, C., Unal, E., Koshland, D. E., DeRisi, J. L. and Gerton, J. L. (2004). Genome-wide mapping of the cohesin complex in the yeast Saccharomyces cerevisiae. PLoS Biol 2, E259.

Goardon, N., Lambert, J. A., Rodriguez, P., Nissaire, P., Herblot, S., Thibault, P., Dumenil, D., Strouboulis, J., Romeo, P. H. and Hoang, T. (2006). ETO2 coordinates cellular proliferation and differentiation during erythropoiesis. Embo J 25, 357-66.

Gribnau, J., de Boer, E., Trimborn, T., Wijgerde, M., Milot, E., Grosveld, F. and Fraser, P. (1998). Chromatin interaction mechanism of transcriptional control in vivo. Embo J 17, 6020-7.

Grosveld, F., van Assendelft, G. B., Greaves, D. R. and Kollias, G. (1987). Position-independent, high-level expression of the human beta-globin gene in transgenic mice. Cell 51, 975-85.

Gruber, S., Arumugam, P., Katou, Y., Kuglitsch, D., Helmhart, W., Shirahige, K. and Nasmyth, K. (2006). Evidence that loading of cohesin onto chromosomes involves opening of its SMC hinge. Cell 127, 523-37.

Gruber, S., Haering, C. H. and Nasmyth, K. (2003). Chromosomal cohesin forms a ring. Cell 112, 765-77.

Haering, C. H., Lowe, J., Hochwagen, A. and Nasmyth, K. (2002). Molecular architecture of SMC proteins and the yeast cohesin complex. Mol Cell 9, 773-88.

Hagstrom, K. A. and Meyer, B. J. (2003). Condensin and cohesin: more than chromosome compactor and glue. Nat Rev Genet 4, 520-34.

Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N. and Grosveld, F. (1991). Importance of globin gene order for correct developmental expression. Genes Dev 5, 1387-94.

Hardison, R. (1998). Hemoglobins from bacteria to man: evolution of different patterns of gene expression. J Exp Biol 201, 1099-117.

Hark, A. T., Schoenherr, C. J., Katz, D. J., Ingram, R. S., Levorse, J. M. and Tilghman, S. M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature 405, 486-9.

Hart, C. M., Zhao, K. and Laemmli, U. K. (1997). The scs' boundary element: characterization of

boundary element-associated factors. Mol Cell Biol 17, 999-1009.

Hartman, T., Stead, K., Koshland, D. and Guacci, V. (2000). Pds5p is an essential chromosomal protein required for both sister chromatid cohesion and condensation in Saccharomyces cerevisiae. J Cell Biol 151, 613-26.

Hatzis, P. and Talianidis, I. (2002). Dynamics of enhancer-promoter communication during differentiation-induced gene activation. Mol Cell 10, 1467-77.

Herendeen, D. R., Kassavetis, G. A. and Geiduschek, E. P. (1992). A transcriptional enhancer whose function imposes a requirement that proteins track along DNA. Science 256, 1298-303.

Hirano, M. and Hirano, T. (2002). Hinge-mediated dimerization of SMC protein is essential for its dynamic interaction with DNA. Embo J 21, 5733-44.

Hirano, M. and Hirano, T. (2006). Opening closed arms: long-distance activation of SMC ATPase by hinge-DNA interactions. Mol Cell 21, 175-86.

Ho, Y., Elefant, F., Liebhaber, S. A. and Cooke, N. E. (2006). Locus control region transcription plays an active role in long-range gene activation. Mol Cell 23, 365-75.

Hopmann, R., Duncan, D. and Duncan, I. (1995). Transvection in the iab-5,6,7 region of the bithorax complex of Drosophila: homology independent interactions in trans. Genetics 139, 815-33.

Hou, F. and Zou, H. (2005). Two human orthologues of Eco1/Ctf7 acetyltransferases are both required for proper sister-chromatid cohesion. Mol Biol Cell 16, 3908-18.

Ivanov, D. and Nasmyth, K. (2005). A topological interaction between cohesin rings and a circular minichromosome. Cell 122, 849-60.

Jack, J. and DeLotto, Y. (1995). Structure and regulation of a complex locus: the cut gene of Drosophila. Genetics 139, 1689-700.

Jackson, L., Kline, A. D., Barr, M. A. and Koch, S. (1993). de Lange syndrome: a clinical review of 310 individuals. Am J Med Genet 47, 940-6.

Kanduri, C., Pant, V., Loukinov, D., Pugacheva, E., Qi, C. F., Wolffe, A., Ohlsson, R. and Lobanenkov, V. V. (2000). Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. Curr Biol 10, 853-6.

Kaur, M., DeScipio, C., McCallum, J., Yaeger, D., Devoto, M., Jackson, L. G., Spinner, N. B. and Krantz, I. D. (2005). Precocious sister chromatid separation (PSCS) in Cornelia de Lange syndrome. Am J Med Genet A 138, 27-31.

Kellum, R. and Schedl, P. (1991). A position-effect assay for boundaries of higher order chromosomal domains. Cell 64, 941-50.

Kellum, R. and Schedl, P. (1992). A group of scs elements function as domain boundaries in an enhancer-blocking assay. Mol Cell Biol 12, 2424-31.

Kim, J. S., Krasieva, T. B., LaMorte, V., Taylor, A. M. and Yokomori, K. (2002). Specific recruitment of human cohesin to laser-induced DNA damage. J Biol Chem 277, 45149-53.

Kim, T. H., Abdullaev, Z. K., Smith, A. D., Ching, K. A., Loukinov, D. I., Green, R. D., Zhang, M. Q., Lobanenkov, V. V. and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell 128, 1231-45.

Kioussis, D., Vanin, E., deLange, T., Flavell, R. A. and Grosveld, F. G. (1983). Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. Nature 306, 662-6.

Kleinjan, D. A. and van Heyningen, V. (2005). Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet 76, 8-32.

Klenova, E. M., Nicolas, R. H., Paterson, H. F., Carne, A. F., Heath, C. M., Goodwin, G. H., Neiman, P. E. and Lobanenkov, V. V. (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, 7612-24.

Kobayashi, T., Horiuchi, T., Tongaonkar, P., Vu, L. and Nomura, M. (2004). SIR2 regulates recombination between different rDNA repeats, but not recombination within individual rRNA genes in yeast. Cell 117, 441-53.

Kolesky, S. E., Ouhammouch, M. and Geiduschek, E. P. (2002). The mechanism of transcriptional activation by the topologically DNA-linked sliding clamp of bacteriophage T4. J Mol Biol 321, 767-84.

Kornberg, R. D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science 184,

868-71.

Krantz, I. D., McCallum, J., DeScipio, C., Kaur, M., Gillis, L. A., Yaeger, D., Jukofsky, L., Wasserman, N., Bottani, A., Morris, C. A. et al. (2004). Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B. Nat Genet 36, 631-5.

Kurukuti, S., Tiwari, V. K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., Lobanenkov, V., Reik, W. and Ohlsson, R. (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, 10684-9.

Lahlil, R., Lecuyer, E., Herblot, S. and Hoang, T. (2004). SCL assembles a multifactorial complex that determines glycophorin A expression. Mol Cell Biol 24, 1439-52.

Laloraya, S., Guacci, V. and Koshland, D. (2000). Chromosomal addresses of the cohesin component Mcd1p. J Cell Biol 151, 1047-56.

Lau, A., Blitzblau, H. and Bell, S. P. (2002). Cell-cycle control of the establishment of mating-type silencing in S. cerevisiae. Genes Dev 16, 2935-45.

Lechner, M. S., Schultz, D. C., Negorev, D., Maul, G. G. and Rauscher, F. J., 3rd. (2005). The mammalian heterochromatin protein 1 binds diverse nuclear proteins through a common motif that targets the chromoshadow domain. Biochem Biophys Res Commun 331, 929-37.

Lecuyer, E., Herblot, S., Saint-Denis, M., Martin, R., Begley, C. G., Porcher, C., Orkin, S. H. and Hoang, T. (2002). The SCL complex regulates c-kit expression in hematopoietic cells through functional interaction with Sp1. Blood 100, 2430-40.

Lee, G. R., Fields, P. E., Griffin, T. J. and Flavell, R. A. (2003). Regulation of the Th2 cytokine locus by a locus control region. Immunity 19, 145-53.

Lefevre, P., Lacroix, C., Tagoh, H., Hoogenkamp, M., Melnik, S., Ingram, R. and Bonifer, C. (2005). Differentiation-dependent alterations in histone methylation and chromatin architecture at the inducible chicken lysozyme gene. J Biol Chem 280, 27552-60.

Lefevre, P., Witham, J., Lacroix, C. E., Cockerill, P. N. and Bonifer, C. (2008). The LPS-induced transcriptional upregulation of the chicken lysozyme locus involves CTCF eviction and noncoding RNA transcription. Mol Cell 32, 129-39.

Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G. P., Itoh, T., Watanabe, Y., Shirahige, K. and Uhlmann, F. (2004). Cohesin relocation from sites of chromosomal loading to places of convergent transcription. Nature 430, 573-8.

Lengronne, A., McIntyre, J., Katou, Y., Kanoh, Y., Hopfner, K. P., Shirahige, K. and Uhlmann, F. (2006). Establishment of sister chromatid cohesion at the S. cerevisiae replication fork. Mol Cell 23, 787-99.

Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., Goode, D., Elgar, G., Hill, R. E. and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet 12, 1725-35.

Li, Q., Peterson, K. R., Fang, X. and Stamatoyannopoulos, G. (2002). Locus control regions. Blood 100, 3077-86.

Ling, J. Q., Li, T., Hu, J. F., Vu, T. H., Chen, H. L., Qiu, X. W., Cherry, A. M. and Hoffman, A. R. (2006). CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. Science 312, 269-72.

Lobanenkov, V. V., Nicolas, R. H., Adler, V. V., Paterson, H., Klenova, E. M., Polotskaja, A. V. and Goodwin, G. H. (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, 1743-53.

Lopes, S., Lewis, A., Hajkova, P., Dean, W., Oswald, J., Forne, T., Murrell, A., Constancia, M., Bartolomei, M., Walter, J. et al. (2003). Epigenetic modifications in an imprinting cluster are controlled by a hierarchy of DMRs suggesting long-range chromatin interactions. Hum Mol Genet 12, 295-305.

Losada, A. and Hirano, T. (2005). Dynamic molecular linkers of the genome: the first decade of SMC proteins. Genes Dev 19, 1269-87.

Magram, J., Chada, K. and Costantini, F. (1985). Developmental regulation of a cloned adult beta-globin gene in transgenic mice. Nature 315, 338-40.

Majumder, P., Gomez, J. A., Chadwick, B. P. and Boss, J. M. (2008). The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. J Exp Med 205, 785-98.

McNairn, A. J. and Gerton, J. L. (2008). Cohesinopathies: One ring, many obligations. Mutat Res.

Meier, N., Krpic, S., Rodriguez, P., Strouboulis, J., Monti, M., Krijgsveld, J., Gering, M., Patient, R., Hostert, A. and Grosveld, F. (2006). Novel binding partners of Ldb1 are required for haematopoietic development. Development 133, 4913-23.

Milot, E., Strouboulis, J., Trimborn, T., Wijgerde, M., de Boer, E., Langeveld, A., Tan-Un, K., Vergeer, W., Yannoutsos, N., Grosveld, F. et al. (1996). Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. Cell 87, 105-14.

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

Moon, H., Filippova, G., Loukinov, D., Pugacheva, E., Chen, Q., Smith, S. T., Munhall, A., Grewe, B., Bartkuhn, M., Arnold, R. et al. (2005). CTCF is conserved from Drosophila to humans and confers enhancer blocking of the Fab-8 insulator. EMBO Rep 6, 165-70.

Morcillo, P., Rosen, C., Baylies, M. K. and Dorsett, D. (1997). Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in Drosophila. Genes Dev 11, 2729-40.

Morcillo, P., Rosen, C. and Dorsett, D. (1996). Genes regulating the remote wing margin enhancer in the Drosophila cut locus. Genetics 144, 1143-54.

Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M. P. and Chambon, P. (1981). The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic Acids Res 9, 6047-68.

Mukhopadhyay, M., Teufel, A., Yamashita, T., Agulnick, A. D., Chen, L., Downs, K. M., Schindler, A., Grinberg, A., Huang, S. P., Dorward, D. et al. (2003). Functional ablation of the mouse Ldb1 gene results in severe patterning defects during gastrulation. Development 130, 495-505.

Mukhopadhyay, R., Yu, W., Whitehead, J., Xu, J., Lezcano, M., Pack, S., Kanduri, C., Kanduri, M., Ginjala, V., Vostrov, A. et al. (2004). The binding sites for the chromatin insulator protein CTCF map to DNA methylation-free domains genome-wide. Genome Res 14, 1594-602.

Murrell, A., Heeson, S. and Reik, W. (2004). Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. Nat Genet 36, 889-93.

Musio, A., Selicorni, A., Focarelli, M. L., Gervasini, C., Milani, D., Russo, S., Vezzoni, P. and Larizza, L. (2006). X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. Nat Genet 38, 528-30.

Nasmyth, K. and Haering, C. H. (2005). The structure and function of SMC and kleisin complexes. Annu Rev Biochem 74, 595-648.

Neuwald, A. F. and Hirano, T. (2000). HEAT repeats associated with condensins, cohesins, and other complexes involved in chromosome-related functions. Genome Res 10, 1445-52.

Ong, C. T. and Corces, V. G. (2008). Modulation of CTCF insulator function by transcription of a noncoding RNA. Dev Cell 15, 489-90.

Palstra, R. J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F. and de Laat, W. (2003). The beta-globin nuclear compartment in development and erythroid differentiation. Nat Genet 35, 190-4.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H. C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T. et al. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell 132, 422-33.

Peric-Hupkes, D. and van Steensel, B. (2008). Linking cohesin to gene regulation. Cell 132, 925-8.

Prioleau, M. N., Nony, P., Simpson, M. and Felsenfeld, G. (1999). An insulator element and condensed chromatin region separate the chicken beta-globin locus from an independently regulated erythroid-specific folate receptor gene. Embo J 18, 4035-48.

Ptashne, M. (1986). Gene regulation by proteins acting nearby and at a distance. Nature 322, 697-701.

Ptashne, M. and Gann, A. (1997). Transcriptional activation by recruitment. Nature 386, 569-77.

Qian, S., Varjavand, B. and Pirrotta, V. (1992). Molecular analysis of the zeste-white interaction reveals a promoter-proximal element essential for distant enhancer-promoter communication. Genetics 131, 79-90.

Recillas-Targa, F., Bell, A. C. and Felsenfeld, G. (1999). Positional enhancer-blocking activity of the chicken beta-globin insulator in transiently transfected cells. Proc Natl Acad Sci U S A 96, 14354-9.

Recillas-Targa, F., Pikaart, M. J., Burgess-Beusse, B., Bell, A. C., Litt, M. D., West, A. G., Gaszner, M. and Felsenfeld, G. (2002). Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. Proc Natl Acad Sci U S A 99, 6883-8.

Rippe, K. (2001). Making contacts on a nucleic acid polymer. Trends Biochem Sci 26, 733-40.

Rollins, R. A., Korom, M., Aulner, N., Martens, A. and Dorsett, D. (2004). Drosophila nipped-B protein supports sister chromatid cohesion and opposes the stromalin/Scc3 cohesion factor to facilitate long-range activation of the cut gene. Mol Cell Biol 24, 3100-11.

Rollins, R. A., Morcillo, P. and Dorsett, D. (1999). Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. Genetics 152, 577-93.

Ronshaugen, M. and Levine, M. (2004). Visualization of trans-homolog enhancer-promoter interactions at the Abd-B Hox locus in the Drosophila embryo. Dev Cell 7, 925-32.

Rubio, E. D., Reiss, D. J., Welcsh, P. L., Disteche, C. M., Filippova, G. N., Baliga, N. S., Aebersold, R., Ranish, J. A. and Krumm, A. (2008). CTCF physically links cohesin to chromatin. Proc Natl Acad Sci U S A 105, 8309-14.

Russell, K. L., Ming, J. E., Patel, K., Jukofsky, L., Magnusson, M. and Krantz, I. D. (2001). Dominant paternal transmission of Cornelia de Lange syndrome: a new case and review of 25 previously reported familial recurrences. Am J Med Genet 104, 267-76.

Sakai, A., Hizume, K., Sutani, T., Takeyasu, K. and Yanagida, M. (2003). Condensin but not cohesin SMC heterodimer induces DNA reannealing through protein-protein assembly. Embo J 22, 2764-75.

Schweinsberg, S., Hagstrom, K., Gohl, D., Schedl, P., Kumar, R. P., Mishra, R. and Karch, F. (2004). The enhancer-blocking activity of the Fab-7 boundary from the Drosophila bithorax complex requires GAGA-factor-binding sites. Genetics 168, 1371-84.

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

Sjogren, C. and Nasmyth, K. (2001). Sister chromatid cohesion is required for postreplicative double-strand break repair in Saccharomyces cerevisiae. Curr Biol 11, 991-5.

Song, S. H., Hou, C. and Dean, A. (2007). A positive role for NLI/Ldb1 in long-range beta-globin locus control region function. Mol Cell 28, 810-22.

Sonoda, E., Matsusaka, T., Morrison, C., Vagnarelli, P., Hoshi, O., Ushiki, T., Nojima, K., Fukagawa, T., Waizenegger, I. C., Peters, J. M. et al. (2001). Scc1/Rad21/Mcd1 is required for sister chromatid cohesion and kinetochore function in vertebrate cells. Dev Cell 1, 759-70.

Spilianakis, C. G. and Flavell, R. A. (2004). Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. Nat Immunol 5, 1017-27.

Splinter, E., Heath, H., Kooren, J., Palstra, R. J., Klous, P., Grosveld, F., Galjart, N. and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev 20, 2349-54.

Stamatoyannopoulos, G. (2005). Control of globin gene expression during development and erythroid differentiation. Exp Hematol 33, 259-71.

Stedman, W., Kang, H., Lin, S., Kissil, J. L., Bartolomei, M. S. and Lieberman, P. M. (2008). Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. Embo J 27, 654-66.

Strachan, T. (2005). Cornelia de Lange Syndrome and the link between chromosomal function, DNA repair and developmental gene regulation. Curr Opin Genet Dev 15, 258-64.

Strom, L., Karlsson, C., Lindroos, H. B., Wedahl, S., Katou, Y., Shirahige, K. and Sjogren, C. (2007). Postreplicative formation of cohesion is required for repair and induced by a single DNA break. Science

317, 242-5.

Strom, L., Lindroos, H. B., Shirahige, K. and Sjogren, C. (2004). Postreplicative recruitment of cohesin to double-strand breaks is required for DNA repair. Mol Cell 16, 1003-15.

Strom, L. and Sjogren, C. (2005). DNA damage-induced cohesion. Cell Cycle 4, 536-9.

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell 10, 1453-65.

Tonkin, E. T., Wang, T. J., Lisgo, S., Bamshad, M. J. and Strachan, T. (2004). NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. Nat Genet 36, 636-41.

Torigoi, E., Bennani-Baiti, I. M., Rosen, C., Gonzalez, K., Morcillo, P., Ptashne, M. and Dorsett, D. (2000). Chip interacts with diverse homeodomain proteins and potentiates bicoid activity in vivo. Proc Natl Acad Sci U S A 97, 2686-91.

Townes, T. M., Chen, H. Y., Lingrel, J. B., Palmiter, R. D. and Brinster, R. L. (1985). Expression of human beta-globin genes in transgenic mice: effects of a flanking metallothionein-human growth hormone fusion gene. Mol Cell Biol 5, 1977-83.

Tuan, D., Solomon, W., Li, Q. and London, I. M. (1985). The “beta-like-globin” gene domain in human erythroid cells. Proc Natl Acad Sci U S A 82, 6384-8.

Udvardy, A., Maine, E. and Schedl, P. (1985). The 87A7 chromomere. Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. J Mol Biol 185, 341-58.

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

Uhlmann, F. and Nasmyth, K. (1998). Cohesion between sister chromatids must be established during DNA replication. Curr Biol 8, 1095-101.

Unal, E., Arbel-Eden, A., Sattler, U., Shroff, R., Lichten, M., Haber, J. E. and Koshland, D. (2004). DNA damage response pathway uses histone modification to assemble a double-strand break-specific cohesin domain. Mol Cell 16, 991-1002.

Unal, E., Heidinger-Pauli, J. M. and Koshland, D. (2007). DNA double-strand breaks trigger genome-wide sister-chromatid cohesion through Eco1 (Ctf7). Science 317, 245-8.

Vakoc, C. R., Letting, D. L., Gheldof, N., Sawado, T., Bender, M. A., Groudine, M., Weiss, M. J., Dekker, J. and Blobel, G. A. (2005). Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. Mol Cell 17, 453-62.

van Haaften, G., Romeijn, R., Pothof, J., Koole, W., Mullenders, L. H., Pastink, A., Plasterk, R. H. and Tijsterman, M. (2006). Identification of conserved pathways of DNA-damage response and radiation protection by genome-wide RNAi. Curr Biol 16, 1344-50.

Vitelli, L., Condorelli, G., Lulli, V., Hoang, T., Luchetti, L., Croce, C. M. and Peschle, C. (2000). A pentamer transcriptional complex including tal-1 and retinoblastoma protein downmodulates c-kit expression in normal erythroblasts. Mol Cell Biol 20, 5330-42.

Vrouwe, M. G., Elghalbzouri-Maghrani, E., Meijers, M., Schouten, P., Godthelp, B. C., Bhuiyan, Z. A., Redeker, E. J., Mannens, M. M., Mullenders, L. H., Pastink, A. et al. (2007). Increased DNA damage sensitivity of Cornelia de Lange syndrome cells: evidence for impaired recombinational repair. Hum Mol Genet 16, 1478-87.

Wadman, I. A., Osada, H., Grutz, G. G., Agulnick, A. D., Westphal, H., Forster, A. and Rabbitts, T. H. (1997). The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. Embo J 16, 3145-57.

Wallace, J. A. and Felsenfeld, G. (2007). We gather together: insulators and genome organization. Curr Opin Genet Dev 17, 400-7.

Wasylyk, B., Wasylyk, C., Augereau, P. and Chambon, P. (1983). The SV40 72 bp repeat preferentially potentiates transcription starting from proximal natural or substitute promoter elements. Cell 32, 503-14.

Watrin, E., Schleiffer, A., Tanaka, K., Eisenhaber, F., Nasmyth, K. and Peters, J. M. (2006). Human Scc4 is required for cohesin binding to chromatin, sister-chromatid cohesion, and mitotic progression. Curr Biol 16, 863-74.

Wawrzyniak, C. J. and Popp, R. A. (1987). Expression of the two adult beta-globin genes in mouse yolk sac and fetal liver erythrocytes. Dev Biol 119, 299-301.

Weitzer, S., Lehane, C. and Uhlmann, F. (2003). A model for ATP hydrolysis-dependent binding of cohesin to DNA. Curr Biol 13, 1930-40.

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

West, A. G., Huang, S., Gaszner, M., Litt, M. D. and Felsenfeld, G. (2004). Recruitment of histone modifications by USF proteins at a vertebrate barrier element. Mol Cell 16, 453-63.

West, S. C. (2003). Molecular views of recombination proteins and their control. Nat Rev Mol Cell Biol 4, 435-45.

Wijgerde, M., Grosveld, F. and Fraser, P. (1995). Transcription complex stability and chromatin dynamics in vivo. Nature 377, 209-13.

Wu, C. T. and Morris, J. R. (1999). Transvection and other homology effects. Curr Opin Genet Dev 9, 237-46.

Xie, X., Mikkelsen, T. S., Gnirke, A., Lindblad-Toh, K., Kellis, M. and Lander, E. S. (2007). Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. Proc Natl Acad Sci U S A 104, 7145-50.

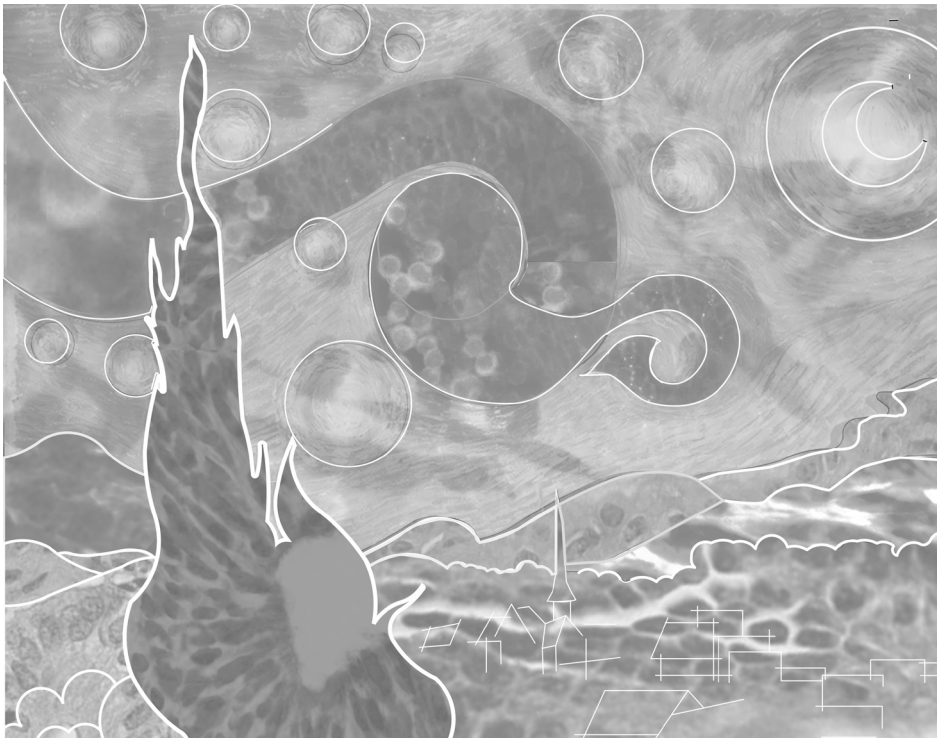
Yoon, Y. S., Jeong, S., Rong, Q., Park, K. Y., Chung, J. H. and Pfeifer, K. (2007). Analysis of the H19ICR insulator. Mol Cell Biol 27, 3499-510.

Yusufzai, T. M., Tagami, H., Nakatani, Y. and Felsenfeld, G. (2004). CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. Mol Cell 13, 291-8.

Zhang, J., Shi, X., Li, Y., Kim, B. J., Jia, J., Huang, Z., Yang, T., Fu, X., Jung, S. Y., Wang, Y. et al. (2008). Acetylation of Smc3 by Eco1 is required for S phase sister chromatid cohesion in both human and yeast. Mol Cell 31, 143-51.

Zhao, H. and Dean, A. (2004). An insulator blocks spreading of histone acetylation and interferes with RNA polymerase II transfer between an enhancer and gene. Nucleic Acids Res 32, 4903-19.

Zhou, J. and Levine, M. (1999). A novel cis-regulatory element, the PTS, mediates an anti-insulator activity in the Drosophila embryo. Cell 99, 567-75.



Chapter 2 Novel binding partners of Ldb1 are required for haematopoietic development.

Meier N, **Krpic S**, Rodriguez P, Strouboulis J, Monti M, Krijgsveld J, Gering M, Patient R, Hostert A, Grosveld F.

Development. 2006 Dec; 133 (24):4913-23.

Development 133, 4913-4924 (2006) doi:10.1242/dev.02656

Novel binding partners of Ldb1 are required for haematopoietic development

Natalia Meier[1], Sanja Krpic[1], Patrick Rodriguez[1], John Strouboulis[1], Maria Monti[2], Jeroen Krijgsveld[2], Martin Gering[3], Roger Patient[3], Arnd Hostert[1] and Frank Grosveld[1,*]

Ldb1, a ubiquitously expressed LIM domain binding protein, is essential in a number of tissues during development. It interacts with Gata1, Tal1, E2A and Lmo2 to form a transcription factor complex regulating late erythroid genes. We identify a number of novel Ldb1 interacting proteins in erythroleukaemic cells, in particular the repressor protein Eto-2 (and its family member Mtgr1), the cyclin-dependent kinase Cdk9, and the bridging factor Lmo4. MO-mediated knockdowns in zebrafish show these factors to be essential for definitive haematopoiesis. In accordance with the zebrafish results these factors are coexpressed in prehaematopoietic cells of the early mouse embryo, although we originally identified the complex in late erythroid cells. Based on the change in subcellullar localisation of Eto-2 we postulate that it plays a central role in the transition from the migration and expansion phase of the prehaematopoietic cells to the establishment of definitive haematopoietic stem cells.

KEY WORDS: Ldb1, Transcription factor complexes, Haematopoietic stem cells, Haematopoiesis

INTRODUCTION

Erythrocytes are derived from the haematopoietic stem cell (HSC) (Durand and Dzierzak, 2005). From approximately E8.5 to E10.5 in the mouse, definitive HSCs are derived from the aorta–gonado–mesonephros region (AGM) (Medvinsky and Dzierzak, 1996; Yoder et al., 1997; de Bruijn et al., 2000; Cumano et al., 2001). Recently the placenta has been identified as a further source of adult HSCs (Gekas et al., 2005; Ottersbach and Dzierzak, 2005). At E11 the mouse fetal liver becomes the main organ of haematopoiesis, later replaced by the bone marrow.

HSC differentiation involves coordinated changes in transcription, often by functionally conserved genes such as *Gata2*, *Tal1*, *Lmo2*, *Gata1* and *Runx1/Aml1* (Cantor and Orkin, 2001).

Often the binding sites for transcription factor complexes are located at great distance from the genes that they control. In the human and mouse β-globin locus, the interacting binding sites and genes are spread over a distance of 100 kb. A three-dimensional structure resulting from long-range interactions, the Active Chromatin Hub (ACH), has recently been demonstrated (Tolhuis et al., 2002; Palstra et al., 2003; Patrinos et al., 2004). We anticipate that (novel) classes of proteins will mediate such 3D interactions of distal regulatory elements. An example of such a protein is Chip/Ldb1 (Morcillo et al., 1997), a protein that can interact with the insulator protein Su(Hw) (Torigoi et al., 2000). Initially isolated in a screen for proteins that bind LIM domains (Agulnick et al., 1996), orthologues of Ldb1 have now been identified in a range of other species. Ldb1 is a ubiquitously expressed nuclear protein that does not bind DNA but appears to participate in transcriptional control by acting as a co-factor for other proteins. It is part of a protein complex in murine erythroid cells composed of the haematopoietic transcription factors Lmo2, Tal1, Gata1 and E2A (Tcfe2a – Mouse Genome Informatics) (Wadman et al., 1997) binding to a GATA-E box motif. This complex binds to the locus control region and β-globin promoter of murine erythroleukemic (MEL, C88) cells (Brand et al., 2004), to the erythroid specific glycophorin A (*Gypa*) promoter (Lahlil et al., 2004), and to multiple sites in the α-globin locus during erythroid differentiation (Anguita et al., 2004).

Consistent with its interaction with a broad range of transcription factors involved in development, the *Ldb1* knockout mouse dies between E9.5 and E10.5 from a series of developmental defects, including absence of haematopoieisis (Mukhopadhyay et al., 2003) (A.H., unpublished). The latter partly resembles the knockout phenotypes of the haematopoietic transcription factors Lmo2 and Tal1 (Warren et al., 1994; Robb et al., 1995; Shivdasani et al., 1995).

In order to understand the role of Ldb1 in erythroid transcriptional interactions, we performed a biochemical screen to identify its binding partners. Using in vivo biotinylation (de Boer et al., 2003; Rodriguez et al., 2005) we describe a number of novel partners. We show that Ldb1 forms complexes that change composition during C88 cell differentiation, and that these complexes are bound to the target genes in vivo. We show that Ldb1 and its binding partners are co-expressed at early stages of development of the murine haematopoietic system, and that the novel erythroid binding partners are required for development of the definitive but not the primitive haematopoetic system of zebrafish embryos.

MATERIALS AND METHODS

Ldb1 cDNA and bio-Ldb1 construct cloning

Ldb1 cDNA was cloned from D14.5 fetal liver RNA (Trizol, Life Technologies) by RT-PCR (SuperScriptII Reverse transcriptase, Invitrogen; Pfu Polymerase, Promega) into the EcoRI site of pBluescript (pBS) (construct AH-3).

ATG-less Ldb1 cDNA was amplified and cloned between the BamHI and NotI sites of pBS. The bio-tag was cloned directly in front of the ATG-less cDNA. Tagged cDNA was isolated as a XhoI–NotI fragment and cloned into the SalI and NotI sites of pEV-Neo (de Boer et al., 2003) to give construct pEV-Neo-bio-Ldb1.

[1]Department of Cell Biology and Genetics, Erasmus MC, PO Box 2040, 3000 CA, Rotterdam, The Netherlands. [2]Department of Biomolecular Mass Spectrometry, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Sorbonnelaan 16, 3584CA, Utrecht, The Netherlands. [3]MRC Molecular Haematology Unit, The Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, UK.

*Author for correspondence (e-mail: f.grosveld@erasmusmc.nl)

Accepted 22 September 2006

MEL cell transfection and culture

C88 cells were cultured and induced for differentiation with DMSO as previously described (Antoniou, 1991). C88BirA cells (Rodriguez et al., 2005) were transfected by electroporation with ScaI-linearised pEV-Neo-bioLdb1, cultured in 96-well plates containing medium with 1 μg/ml puromycin and 0.8 mg/ml neomycin to select single clones (de Boer et al., 2003; Rodriguez et al., 2005).

Nuclear extract preparation

Small-scale nuclear extract preparation of C88 cell cultures (30-50 ml) and nuclear extracts of larger cultures (5.5-9 L) were prepared as described by de Boer et al. (de Boer et al., 2003).

Strepatavidin pulldown and mass spectrometry

Strepatavidin-coated Dynabeads M-280 (Dynal) were blocked for 1 hour with chicken serum albumin/PBS (200 ng/μl). The salt and detergent concentrations of nuclear extract samples of 5-6 mg from induced C88BirA/bio-Ldb1 cells or 15 mg from noninduced C88BirA/bio-Ldb1 cells were adjusted to 200 mM KCl and 0.3% NP40 with 10 mM KCl buffer (10 mM KCl, 10 mM HEPES-KOH, 1.5 mM $MgCl_2$, 25% glycerol, 0.75% NP40, 2 mM PMSF) prior to overnight incubation with blocked beads at 4°C. The beads were washed for 5 minutes six times in washing solution (150/200 mM KCl, 10 mM HEPES-KOH pH 7.9, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.3% NP40, 0.2 mM PMSF) at room temperature. Pulled down proteins were processed and analysed by mass spectrometry (LC-MSMS, MS) as described (Rodriguez et al., 2005).

Immunoprecipitations and western blot analysis

Immunoprecipitations and western blot analysis were performed as previously described (Rodriguez et al., 2005). Bio-Ldb1 was detected using a 1/10,000 dilution of Streptavidin-Horseraddish Peroxidase (HRP) conjugate (NEN).

Antibodies

Monoclonal rat antibodies against Mtgr1 and Lyl1 were produced by Absea. The antibody against E2-2 was obtained from Abcam (ab2233), and all others were from Santa Cruz Biotechnology: α-Ldb1 (sc-11198), α-Lmo2 (sc-10499), α-Lmo4 (sc-11121, sc-22833), α-Eto-2 (sc-9741), α-Runx1 (sc-8563), α-Cdk9 (sc-484), α-HEB (sc-357), α-E2A (sc-349), α-Gata1-N6 (sc-265), α-HDAC1 (sc-7872), α-NMP 238 (sc-15259). The antibody against NMP 238 was used as a loading control for the IP experiments, because it showed no change after induction.

Chromatin immunoprecipitations

Fixation, lysis of cells and sonication of chromatin were performed as previously described (Rodriguez et al., 2005). Primers for real-time PCR were as described (Rodriguez et al., 2005). Primers for the Gata1 hypersensitive site (HS) and negative control sequences were:

Gata1 HS-3.5 Reverse primer: 5′-CCGGGTTGAAGCGTCTTCT-3′
Gata1 HS-3.5 Forward primer: 5′-TCAGGGAAGGATCCAAGGAA-3′
Gata1 Negrev: 5′-TGCCGCTTGCCTTTGTAAG-3′
Gata1 Negfor: 5′-CACTAGCAGCTGGGTGGGTTA-3′

Zebrafish maintenance and morpholino injections

Wild-type zebrafish were kept and staged according to Westerfield (Westerfield, 1993). ATG morpholinos (MO) (Gene-Tools) and corresponding mismatch MOs were derived from the genebank cDNA of ldb1 (NM_131313, 5′-GCCCACGTCTCGGTCCAGCATGGTG-3′), tcf4 (NM_131259, 5′-AGCTGCGGCATTTTTCCCGAGGAGC-3′), cdk9 (BC055634, 5′-CGACGCCATCGTAGTATTTGGACAT-3′, control mismatch MO 5′-CGAgGaCATCGTAcTATTTaGAgAT-3′), lmo4 (NM_177984, 5′-AGCTTTCCACACGACTGTTCACCAT-3′, control mismatch MO 5′-AGgTTTgCACACcACTGTTgACgAT-3′), mtgr1 (XM_695328 5′-CTCTTAAAAGCGTGAAAGACCGCAT-3′, control mismatch MO 5′-CTgTTAAAAcCGTcAAAGAgCGgAT-3′), eto-2 (EST AF164710 5′-AACATGACGGTTGGAACTCTGGTT-3′, control mismatch MO 5′-AAgATcACGGTTcGAACTgTGcTT-3′).

All MOs were dissolved in water to a concentration of 1 mM and injected at three doses (0.1 nl, 0.5 nl and 1.2 nl) into zebrafish embryos at the two- to eight-cell stage. As an injection control, rhodamin-dextrane or phenol red were added to a concentration of 10% vol/vol before use.

Whole mount in situ hybridisation

Digoxigenin-UTP (Roche) labelled antisense and sense RNA probes against the zebrafish orthologues of *hemoglobin beta embryonic-1 (βE1)* (Quinkertz et al., 1999) and *runx1* (M. Gering) were synthesized from linearised plasmids using T3 and T7 RNA polymerases. In situ hybridisation was performed as described previously (Jowett and Yan, 1996). To remove pigmentation, embryos older than 32 hours postfertilisation (hpf) were treated with 5% H_2O_2/PBS for 2-3 hours prior to in situ hybridisation.

Probes were detected by incubation with alkaline phosphatase coupled anti-Digoxigenin antibody (Roche) and colour reaction with Fast Red (Roche) for *βE1* and BM Purple (Roche) for *runx1*.

Immunohistochemistry

E9.5 FVB mouse embryos were fixed in 2% paraformaldehyde/PBS for 2 hours at room temperature. After overnight equilibration in 20% Sucrose/PBS at 4°C, embryos were orientated and quick frozen in Tissue Tek (Sakura Finetek).

Immunohistochemistry was essentially carried out as previously described (de Bruijn et al., 2002). All animal experiments were carried out according to the Dutch Welfare of Animals Act.

RESULTS

Generation of bio-Ldb1 cells

In order to identify Ldb1 interacting partners, C88BirA MEL cells expressing the *Escherichia coli* BirA protein-biotin ligase (de Boer et al., 2003) were stably transfected with pEV-Neo-bio-Ldb1 (Fig. 1A).

A total of 16 C88BIR/bioLdb1 clones were isolated and induced to differentiate with DMSO. Nuclear extracts were tested for presence of the fusion protein using an α-Ldb1 antibody and streptavidin-HRP (Fig. 1B). Clone #3F4 was chosen because it expresses low levels of bio-Ldb1 (Fig. 1C) in the uninduced state. This allowed us to compare Ldb1 complexes before and after terminal differentiation.

Identification of Ldb1 interaction partners

Nuclear extracts prepared from noninduced and induced #3F4 cells and untransfected C88 cells were incubated with streptavidin-coated paramagnetic beads, and separated by polyacrylamide gel electrophoresis (PAGE) (de Boer et al., 2003; Rodriguez et al., 2005) (Fig. 1D). Proteins were trypsin digested, eluted and analysed by mass spectrometry (LC-MSMS). Table 1 shows the pulled down proteins when washed with 150 mM and 200 mM KCl. The pattern of background proteins was similar to that observed for Gata1 (de Boer et al., 2003; Rodriguez et al., 2005) and other transcription factors (H. Braun, J. Demmers and J. Philipsen, personal communication).

The screen was validated by the fact that known Ldb1 partners (Lmo2, Tal1, Gata1 and E2A) (Wadman et al., 1997) were readily identified. We also found three proteins of the Ssdp family, members of which interact with Ldb1 in HeLa cells and in *Drosophila* (Chen et al., 2002; van Meyel et al., 2003; Nishioka et al., 2005). These Ssdp interactions were not analysed further. In addition, a number of novel (potential) interaction partners of different functional classes were found, including transcription (co-)factors, cell-cycle proteins, chromatin remodelling and DNA repair proteins. Some of the previously identified binding partners of Ldb1 (Tal1, E2A), the basic helix-loop-helix (bHLH) proteins HEB (Tcf12 – Mouse Genome Informatics), E2-2 (Tcf4 – Mouse Genome Informatics),

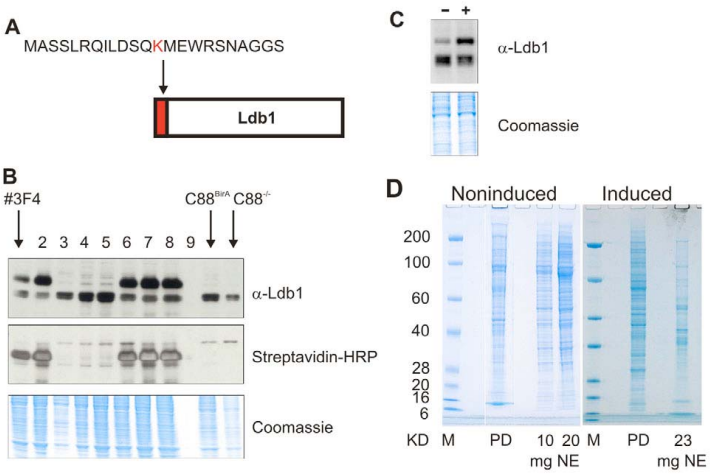


Fig. 1. Ldb1 biotinylation and streptavidin pulldown. (**A**) Schematic representation of bio-Ldb1. The 23 amino acid sequence recognised by BirA and a triple haemagglutinin tag are fused in tandem to the amino-terminal end of the ATG-less Ldb-cDNA. (**B**) Expression of *Ldb1* and *bio-Ldb1* in nuclear extracts of induced C88BirA/bio-Ldb1, C88-/BirA and C88$^{-/-}$ cells. Eight of 16 transfectants are shown. Three clones (lanes 3-5) did not express *bio-Ldb1*, lane 9 not loaded. Clone #3F4 was chosen for further experiments. Lanes on the right are C88-/BirA and C88$^{-/-}$ controls. (**C**) *Ldb1* and *bio-Ldb1* expression in equal amounts (see Coomassie stained gel) of nuclear extracts of noninduced (–) and induced (+) #3F4 cells. Expression of endogenous *Ldb1* is reduced in induced cells. (**D**) PAGE of proteins bound to bio-Ldb1. Pulldowns of noninduced and induced cell extracts washed at lower stringency conditions are shown. PD, pulled down proteins; NE, untreated nuclear extract.

Lyl1 and the novel interacting protein Eto-2 (Cbfa2t3h – Mouse Genome Informatics) were detected in all three analyses. Eto-2, an orthologue of the *D. melanogaster* gene *nervy*, was the most abundant protein in two of the three LC-MSMS outputs. Mtgr1 (Cbfa2t2 – Mouse Genome Informatics), another member of the ETO protein family, was also identified by MS analysis of noninduced and induced cell extracts. The known partner Lmo2 was present in extracts from noninduced and induced cells (Fig. 2), whereas Lmo4, Gata1 and Runx1 were present in the MS analyses of induced cells only, although it should be noted that Gata1 may have been absent due to the fact that only very few Gata1 peptides are detectable in our MS analyses (Rodriguez et al., 2005). In contrast to the proteins mentioned above, the cell-cycle protein Cdk9 was only found in nuclear extracts of noninduced cells. Proteins of different chromatin remodelling complexes were also identified only in induced cell extracts, for example Hdac1 (see Fig. S1B in the supplementary material), but these have as yet not been investigated in detail.

Western blots of equal amounts of nuclear extracts of noninduced and induced C88 cells showed that the levels of all binding partners except for the newly identified interacting proteins Eto-2, Cdk9 and Lmo4, did not change significantly (Fig. 2, input lanes) in normal untransfected cells and bio-Ldb1 cells. Levels of Lmo4 and Cdk9 increased with induction, whereas there was considerably less Eto-2 in induced extracts (Fig. 2A, input Lmo4, Cdk9 and Eto-2 panels). Interestingly, the much less abundant 55 kD isoform of Cdk9 (Shore et al., 2003) is upregulated with induction, while the 41 kD isoform of Cdk9 decreases in untransfected MEL cells, and is present at the same levels in noninduced and induced bio-Ldb1 extracts (Fig. 2, input Cdk9 panels); this phenomenon is also seen in differentiating macrophages (Liu and Herrmann, 2005).

Identification of different complexes

To confirm the interactions found in the MS analysis, immunoprecipitations of equal amounts of nuclear extracts from induced and noninduced C88 cells (Fig. 2) were performed with an α-Ldb1 antibody. Immunoprecipitations carried out on normal C88 cells and bio-Ldb1 transfected cells showed essentially the same results. The α-Ldb1 antibody depleted the extracts of noninduced cells almost completely of Ldb1 (Fig. 2A,B). Lmo2, Tal1, the two E2A isoforms E12 and E47, HEB and Gata1, precipitated equally with Ldb1 before and after induction (not shown). The only difference we observed between the bio-Ldb1 transfected cells and untransfected cells was a small change in the ratio of expression of E12 and E47 (Fig. 2A,B, E2A panels). Consistent with its lower level in induced cell extracts, less Eto-2 precipitated with Ldb1 from induced cell extracts compared with that of noninduced cells (Fig. 2A,B, Eto-2 panels). Mtgr1, the other identified Eto-family member, was also enriched to a lesser extent in induced cells (Fig. 2A,B, Mtgr1 panels). As expected, more Lmo4 precipitated with Ldb1 in induced cells compared with noninduced cells (Fig. 2A,B, Lmo4 panels), whereas the amount of coprecipitated Lmo2 did not change (not shown). These results suggest that the newly identified partner Lmo4 may play a more important role at later stages of erythroid differentiation. Most notably, the two isoforms of Cdk9 immunoprecipitated with Ldb1 only in noninduced cells (Fig. 2A,B, Cdk9 panels). As Cdk9 is involved in cell-cycle progression (Bettencourt-Dias et al., 2004), its interaction with Ldb1 in noninduced, proliferating C88 cells only, may link the complex containing Ldb1 and Cdk9 to the maintenance of the proliferative state. Although their level is unchanged, there was less coprecipitation of E2-2 and Lyl1 with either α-Ldb1 (Fig. 2A,B) or α-Eto-2 (not shown) in induced cells.

Table 1. Proteins identified by LC-MSMS in bio-Ldb1 pulldown experiments

Proteins pulled down	C88/BirA	Non-induced LS	Induced 1 LS	Induced 2 HS
Transcription factors				
LIM only proteins				
Lmo2	–	+	–	+
Lmo4	–	–	+	+
Zinc finger proteins				
Gata1	–	–	+	+
Basic helix-loop-helix				
Tal1	–	+	+	+
E2A	–	+	+	+
Lyl1	–	+	+	+
HEB	–	+	+	+
E2-2	–	+	+	+
ETO-family				
Eto-2	–	+	+	+
Mtgr1	–	+	-	+
Runt domain				
Runx1	–	–	+	+
Ssdp				
Ssdp2	–	+	+	+
Ssdp3	–	+	+	+
RIKENcDNA1210001E11 (Ssdp4)	–	+	+	+
Cell-cycle proteins/kinase				
Cdk9	–	+	–	–

The lower amount of identified proteins and their corresponding peptides in the lower stringency pulldown experiment of induced cells (induced 1) is due to the lower amount of input for the LC-MSMS analysis.
LS, low stringency; HS, high stringency.

In order to gain a better understanding of some of the Ldb1-containing complexes, and to determine whether Ldb1 binding partners bind to Ldb1 in the absence of Eto-2 and vice versa, we performed sequential immunoprecipitation experiments. We first depleted nuclear extracts of either Ldb1 or Eto-2 with their respective antibodies, and then incubated the supernatants with α-Eto-2 and α-Ldb1 respectively (Fig. 3 and see Fig. 1A in the supplementary material). Immunoprecipitations were then analysed for Ldb1 interacting proteins identified and validated in the single immunoprecipitation experiments (Fig. 3). It should be noted that we have not set out to characterise all the possible complexes that may be formed by every single Ldb1 interacting partner identified in the MS analysis, a task which is outside the scope of this paper.

As noted above, the α-Ldb1 antibody (almost) completely depletes Ldb1 protein from the extract (Figs 2, 3, Ldb1 panels). A precipitation with α-Eto-2 antibody brings down high amounts of Ldb1, indicating that a high proportion of complexes contains both Ldb1 and Eto-2. However, α-Eto-2 did not deplete the extract of Ldb1 protein, as more Ldb1 protein is precipitated with α-Ldb1 from the α-Eto-2 treated supernatant (Fig. 3, Ldb1 panels). Conversely, precipitation of Ldb1 with α-Ldb1 antibody brings down large amounts of Eto-2 (Figs 2, 3, Eto-2 panels) but does not deplete Eto-2 completely from the extract, as there is more Eto-2 left in the supernatant (Fig. 3, Eto-2 panel). Thus, there are probably at least three complexes, one containing both Eto-2 and Ldb1, one that contains Ldb1 but not Eto-2, and one that contains Eto-2 but not Ldb1 (Fig. 3B).

To investigate the binding behaviour of Cdk9, the only protein that does not interact with Ldb1 upon induction of differentiation, we tested the sequential immunoprecipitations for its presence. α-Ldb1 antibody precipitated Cdk9 with a substantial amount left in the supernatant, which was not precipitable by α-Eto-2 antibody (Fig. 3, Cdk9 panel). In reverse, incubation of nuclear extracts with α-Eto-2 antibody also precipitated Cdk9 only partially, the remainder also not being bound to Ldb1 (Fig. 3, Cdk9 panel). We suggest that Cdk9 requires both Ldb1 and Eto-2 for its interaction with either of the two proteins. E2-2 and Lyl1 exhibited similar interaction behaviour (Fig. 3), with the exception that Lyl1 is also bound to Ldb1 alone (Fig. 3B).

To test whether the Ldb1-Lmo2 interaction requires Eto-2, we first depleted Eto-2 followed by Ldb1 immunoprecipitation and testing for Lmo2 (Fig. 3, Lmo2 panels). α-Eto-2 brought down only some Lmo2. As all of Lmo2 is in complex with Ldb1 (Fig. 3), we conclude that there is a complex containing Lmo2-Ldb1-Eto-2, and that the remaining Lmo2 left in the supernatant after Eto-2 depletion is bound to Ldb1, but not Eto-2. Lmo4 behaves similarly but shows an important quantitative difference. There is little Lmo4 before differentiation, but this increases several fold after induction of differentiation (Fig. 2).

The two isoforms of E2A show a different binding behaviour. Both the larger E47 and smaller E12 isoforms are precipitated equally by Ldb1, albeit incompletely. The remaining E47, but not E12, is bound by Eto-2 (Fig. 3 and see Fig. S1A in the supplementary material; E2A panels). Conversely, α-Eto-2 completely depletes the extracts of the E47 isoform but binds only some E12 (Fig. 3, E2A panel). Some of this remaining E12 is precipitable by Ldb1. Taking into account the Ldb-1/E2A data from above, this indicates the presence of a complex containing Ldb1/Eto-2/E47/E12, two further complexes composed of at least Eto-2/E47 or Ldb1/E12, and 'free' E12 that can participate in other complex formation.

Mtgr1 was the second member of the Eto family identified in our MS analysis. α-Ldb1 antibody precipitates some Mtgr1, but not all, with a substantial amount left in the supernatant that is all precipitated with an α-Eto-2 antibody (Fig. 3, Mtgr1 panel and see Fig. S1A in the supplementary material). Precipitation with α-Eto-2 antibody, however, completely depletes Mtgr1 from the nuclear extracts, indicating that all Mtgr1 is complexed with Eto-2 (Fig. 3 and see Fig. S1A in the supplementary material; Mtgr1 panels). We conclude that there are at least two complexes containing Mtgr1: one complex containing Mtgr1/Eto-2/Ldb1, and that the Mtgr1/Ldb1 interaction requires Eto-2; a further complex containing Mtgr1/Eto-2, but not Ldb1.

Tal1 also forms several complexes, which contain either Ldb1 and/or Eto-2. α-Ldb1 antibody precipitates most, but not all Tal1 from the nuclear extract, part of the remaining Tal1 is pulled down with Eto-2 (Fig. 3, Tal1 panel). Testing first with α-Eto-2 shows that Eto-2 brings down some but not all Tal1; some is precipitable with an α-Ldb1 antibody (Fig. 3, Tal1 panel). This indicates that there are at least four Tal1-containing complexes: Eto-2/Tal1/Ldb1, Ldb1/Tal1, Eto-2/Tal1 and 'free' Tal1.

HEB also forms complexes that contain either Ldb1 and/or Eto-2. Both α-Ldb1 and α-Eto-2 precipitate large amounts of HEB indicating that HEB forms complexes with Ldb1 or Eto-2, and probably a complex containing all three of these proteins. A second precipitation with either α-Eto-2 or α-Ldb1 precipitates more HEB, although small amounts remain in the supernatant after this second

immunoprecipitation (Fig. 3, HEB panel). Therefore, HEB forms at least three complexes: HEB/Ldb1, HEB/Eto-2 and 'free' HEB. When an α-HEB immunoprecipitation is carried out, it does precipitate all the partners, including Lyl1, Cdk9 and E2-2, suggesting that it is part of a large Ldb1/Eto-2 complex (Fig. 3B), or perhaps a smaller one containing Ldb1 and Eto-2.

Finally, Gata1, a transcription factor known to participate in a multitude of complexes (Rodriguez et al., 2005), also forms separate complexes with either Eto-2 or Ldb1. Treatment of extracts with α-Ldb1 antibody brings down a small part of Gata1. Further precipitation with α-Eto-2 provides evidence of an Eto-2-Gata1 interaction separate from Ldb1 (Fig. 3, Gata1 panel, and see Fig. S1A in the supplementary material). In reverse, α-Eto-2 antibody indeed precipitates part of Gata1. Very little Gata1 is then precipitable with an α-Ldb1 antibody, indicating that possibly most of the Eto-2/Gata1 complex also contains Ldb1 (Fig. 3, Gata1 panel). We conclude that Gata1 forms at least four complexes: Gata1/Ldb1, Gata1/Eto-2, Gata1/Ldb1/Eto-2 and 'free' Gata1, which is known to participate in other complexes.

Single immunoprecipitations using antibodies for the Ldb1 interacting proteins confirmed all the pairwise interactions described above (not shown).

Although it is difficult to distinguish the complexes from each other, the data strongly suggest that there are several subcomplexes formed by Ldb1 and its interaction partners that can form larger, functional complexes (possibly via the homodimerisation of Ldb1). We can distinguish two large subcomplexes with either Ldb1 or Eto-2 and one large complex containing both Ldb1 and Eto-2 (Fig. 3B and Fig. 7). Upon differentiation, association with Cdk9 is lost and the level of Eto-2 is substantially decreased, whereas the amount of Lmo4 is increased several fold. The reduction in association between Ldb1 and Eto-2 is also reflected in the reduced coimmunoprecipitation of E2-2 and Lyl1 with α-Ldb1 and α-Eto-2 in induced cells. In addition, these complexes appear to interact with Runx1, but we have as yet not been able to characterise this interaction due to the poor quality of the available antibody.

Chromatin immunoprecipitations

To confirm that Eto-2 and Ldb1 are bound to chromatin at specific regulatory sites, we carried out chromatin immunoprecipitations using a number of erythroid genes: *Gata1*, *Gypa*, *Myb*, *Myc* and *Eklf* (*Klf1* – Mouse Genome Informatics), for which the binding of Gata1 complexes is known (Rodriguez et al., 2005). Among the *Gata1* gene regulatory regions, the HS-3.5 is known to bind the

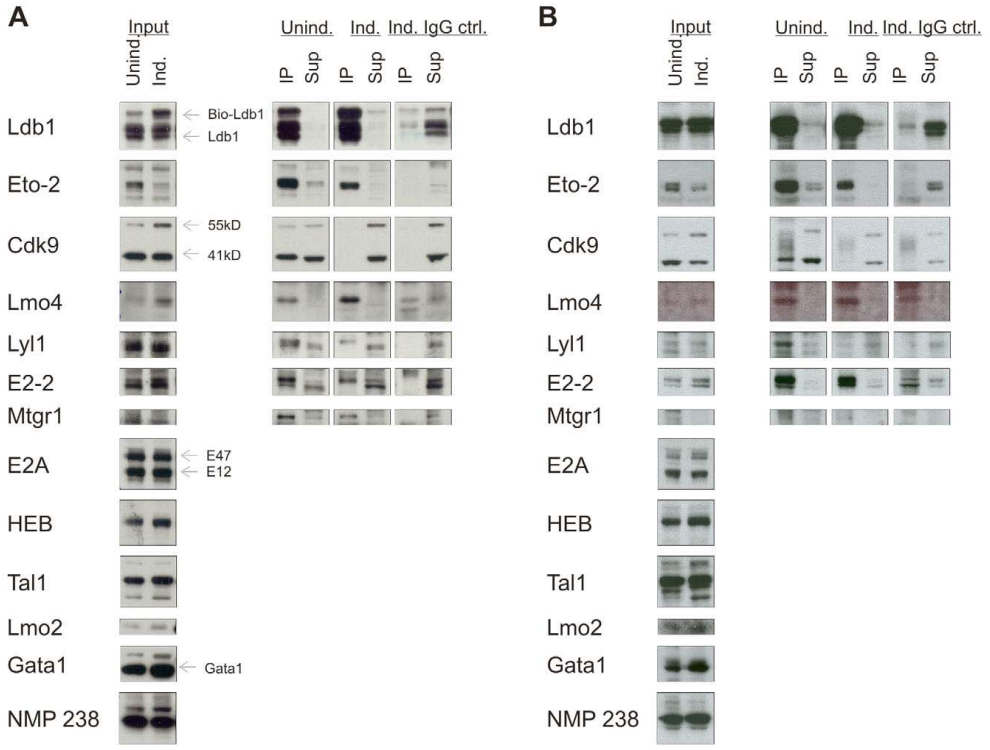


Fig. 2. Analysis of Ldb1 interacting protein complexes. Protein levels of Eto-2, Cdk9 and Lmo4 change with induction of C88 cell differentiation. Western blot analysis of noninduced (Unind.) and induced (Ind.) C88 cells containing bio-Ldb1 (**A**) or untransfected C88 cell nuclear extracts (**B**). Input lanes indicate levels of tested proteins in untreated nuclear extracts diluted to the same concentration as in the immunoprecipitation experiments. Proteins immunoprecipitating with Ldb1 (IP lane) and supernatant (Sup) are shown and labelled accordingly. The isoforms of bio-Ldb1, Cdk9 and E2A, are indicated.

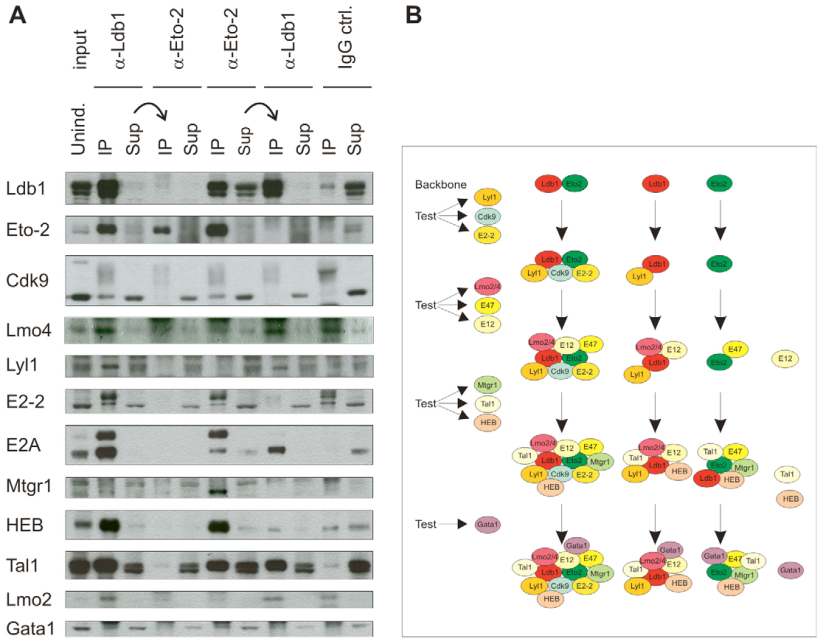


Fig. 3. Sequential immunoprecipitations to investigate Ldb1- and Eto-2-containing complexes. (**A**) Nuclear extracts of noninduced normal C88 cells were first depleted of either Ldb1 or Eto-2 with their respective antibodies. The supernatants were then incubated with α-Eto-2 or α-Ldb1, respectively. The second supernatant was also loaded to determine which proteins do not interact with either Eto-2 or Ldb1. IgG lanes are control immunoprecipitations carried out with a nonspecific, isotype-matched antibody. (**B**) Scheme of interacting factors from IP experiments in A (see also Fig. S1A in the supplementary material). Each new line represents the analysis of an additional three transcription factors.

Gata1/Tal1/Ldb1 complex, whereas the DNaseI hypersensitive site in the gene (Gata1 IE) does not (Guyot et al., 2004). The GATA-E box sites in the *Gypa* and *Eklf* promoters are also known targets of Gata1/Ldb1 (Lahlil et al., 2004; Rodriguez et al., 2005). In contrast, the *Myb* and *Myc* genes bind a Gata1/Gfi1b complex (Rodriguez et al., 2005). In all cases we find an enrichment for *Gata1* (as expected) when compared with non GATA site negative control fragments (not shown) (Rodriguez et al., 2005). In contrast, Eto-2 and Ldb1 were bound to *Gata1*, *Eklf* and *Gypa* (Fig. 4), but not to *Myc* and *Myb* (not shown). The ratio of Eto-2/Ldb1 binding to the three elements, decreases during differentiation in C88 cells (Fig. 4) in accordance with the fact that there is less Ldb1/Eto-2 complex (see Fig. 2). Gata1, glycophorin A and Eklf proteins are expressed late in erythroid differentiation [the –3.5 HS of the *Gata1* gene also regulates *Gata1* expression in megakaryocytic cells (McDevitt et al., 1997; Onodera et al., 1997)], whereas *Myc* and *Myb* are

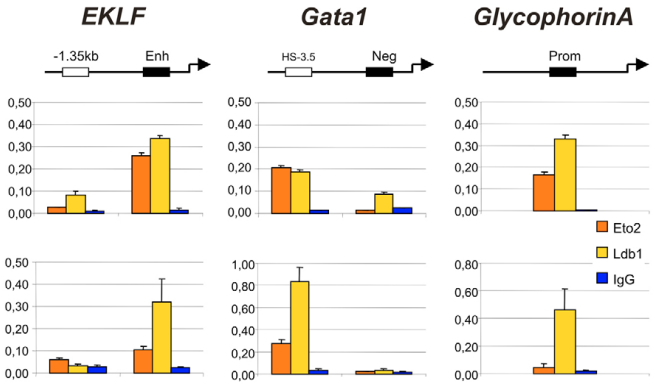


Fig. 4. ChIP of Ldb1 and Eto-2. (Top) The boxes indicate the localisation of the upstream HS in the *Eklf*, *Gata1* and *Gypa* promoters that contain Gata-1/Tal1 binding sites and negative controls not containing such sites. Bar graphs show the relative enrichment of sequences immunoprecipitated by Eto-2 (red), Ldb1 (yellow) and the IgG control (blue) in noninduced (middle row) and induced (bottom row) C88 cells. All values were normalised to a GAPDH control.

downregulated. The decrease of Eto-2 levels in late erythroid cells suggests that Eto-2 adds a repressive function to the Gata1/Tal1/Ldb1 complex to keep late erythroid genes off early during differentiation. In contrast, Eto-2 is not essential for Gfi1b repressed genes such as *Myb* and *Myc* expressed at early stages (Rodriguez et al., 2005).

Eto-2, Cdk9 and Lmo4 are required for definitive haematopoiesis in a zebrafish model system

Ldb1 and its constitutive binding partners Lmo2 and Tal1 are essential for embryonic blood formation (Warren et al., 1994; Robb et al., 1995; Shivdasani, 1995). We next asked whether the newly identified Ldb1 interacting partners E2A, Cdk9, Eto-2, Lmo4 and Mtgr1 are required for haematopoietic development. As the genetic regulation of embryonic and definitive haematopoiesis is highly conserved between zebrafish and mammals, we tested the role of these proteins by MO-mediated inhibition of mRNA translation in zebrafish embryos (Nasevicius and Ekker, 2000). In zebrafish, primitive erythrocytes expressing embryonic haemoglobin derive from the intra-embryonic intermediate cell mass (ICM) and start circulating at 24 hpf. Shortly thereafter, with development of the definitive haematopoietic system, a ventrally located *flk1*-positive subpopulation of the dorsal aorta precursors starts to express *runx1*. *runx1* is also expressed in primitive erythrocytes, the olfactory epithelium, Rohon-Beard neurons (Kalev-Zylinska et al., 2002) and the anterior paraxial mesoderm.

ATG-MOs and mutated MOs targeted against the zebrafish orthologues of *Cdk9*, *Eto-2*, *Lmo4*, *Mtgr1* and *Ldb1* were injected at three increasing doses into one- to eight-cell stage embryos. To test the effects of these MOs on the embryonic and definitive haematopoeitic system, we analysed embryos after onset of blood circulation by in situ hybridisation with probes against embryonic *βE1* and *runx1*, respectively (Fig. 5).

Ldb1-MO-injected embryos displayed variable phenotypes at all doses, including deformation of the body axis, dysmorphic somites, abnormal tail morphology, haematopoietic defects resembling that of

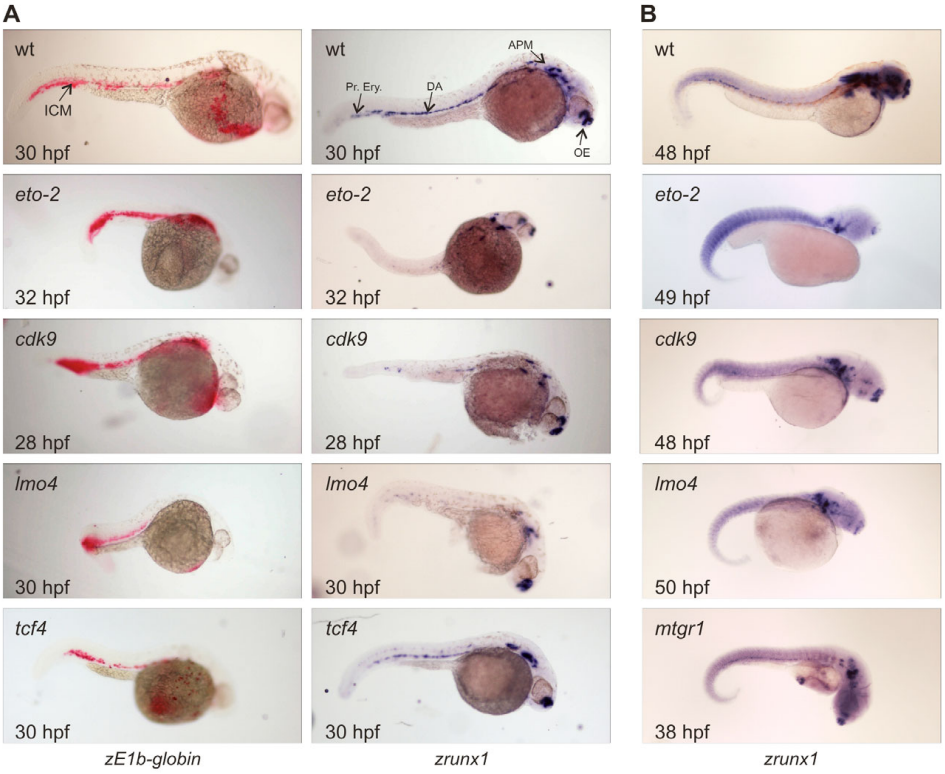


Fig. 5. Analysis of zebrafish embryos after MO injections. The developmental stage of each embryo is in the bottom left corner; the targeted gene is indicated in the top left corner. All pictures were taken at the same magnification. (**A**) Noninjected wild-type (wt) control embryos stained for *βE1*- (*zE1b-globin*, left column) and *runx1*-mRNAs (right column). The ICM, dorsal aorta (DA), primitive erythrocytes (Pr.Ery.), anterior paraxial mesoderm (APM) and olfactory epithelium (OE) are indicated. The *βE1*signal is red, *runx1* signal is blue. Treated embryos were injected with 1 pmol *eto-2*-MO (32 hpf), 1 pmol *cdk9*-MO (28 hpf),1 pmol *lmo4*-MO (30 hpf), or with 0.5 pmol of the control *tcf4*-MO (30 hpf). (**B**) The reduction of *runx1* expression in the dorsal aorta region resulting from injection with the MOs directed against *eto-2* (*n*=23/27), *cdk9* (*n*=33/56) and *lmo4* (*n*=36/54) was still observable after 2 days. Embryos injected with the *mtgr1*-MO were analysed at 38 hpf and showed only a slight effect on the definitive haematopoietic system. All embryos shown were injected with 1 pmol of the corresponding MO.

the mouse (A.H., unpublished) and necrosis in the brain (not shown). The severity of defects increased with increasing dose of injected Ldb1-MO. As a negative control, we used mutated ATG-MOs and an MO targeted against *tcf4*, the zebrafish orthologue of *Tcf7/2*. The latter is not expressed in haematopoietic tissues and, when deleted, has a specific effect in the intestinal epithelium of the mouse (Korinek et al., 1998). As expected, no effects of the mutated ATG-MOs (not shown) or *tcf4*-MO on the haematopoietic system were observed in the injected zebrafish at any dose. Embryos injected with the highest dose of *tcf4*-MO displayed some tail abnormalities.

The *cdk9*-MO had no effect on *βE1* expression or expression of *runx1* in the primitive erythrocytes located in the posterior ICM, the olfactory epithelium and the anterior paraxial mesoderm. However, they had a severe effect on definitive erythropoiesis. A 0.5-pmol dose of *cdk9*-MO showed a clear reduction of *runx1* signal in the dorsal aorta region (*n*=10/18), which decreased further in embryos injected with 1 pmol (*n*=8/10). Surprisingly the effect appears to be specific to the haematopoietic system, although *Cdk9* is expressed in many tissues (Bagella et al., 1998).

Embryos injected with the *eto-2*-MOs had a similar phenotype. The reduction of *runx1* expression in embryos injected with 0.5 pmol (*n*=13/16) and 1 pmol of *eto-2*-MO (*n*=5/7) was more severe compared with the *cdk9*-MO: runx1 in the dorsal aorta was either almost or completely abolished. Embryonic *βE1* expression was normal in all *eto-2*-MO-injected embryos; however, some primitive erythrocytes in the caudal region were located laterally to the midline as opposed to their location in wild-type embryos. This is similar to the effect observed when sonic hedgehog signalling is inhibited (Gering and Patient, 2005), suggesting that *eto-2* may play a role in the response to extracellular signals. Injection of 1-2 pmol of the *mtgr1*-MO did not affect embryonic haematopoiesis. Some reduction of *runx1* expression in dorsal aorta precursor cells was observed at 38 hpf (*n*=5/25; Fig. 5B).

Embryos treated with the *lmo4*-MO were comparable to knockdowns of *cdk9* and *eto-2*. Expression of embryonic *βE1* was normal, whereas reduced levels of *runx1* were observed in the dorsal aorta region at 0.5 pmol MO (*n*=8/13) and decreased further with *lmo4*-MO injected at 1 pmol. In addition, some of the treated embryos appeared to have brain or neural tube abnormalities, which are the cause of perinatal death of the corresponding knockout mouse (Hahm et al., 2004; Tse et al., 2004; Lee et al., 2005). These mice showed no defects in the haematopoietic system, although only half of the homozygous *Lmo4* null mutants were born. The other half died around E9 of gestation, possibly due to a haematopoietic phenotype. The effects of the *eto-2*-, *cdk9*- and *lmo4*-MOs on the definitive haematopoietic system were still observable after 2 days (Fig. 5B) with normally circulating embryonic blood cells. At this stage of development the reduction in body size of the embryos injected with the *eto-2*-, *cdk9*-, *lmo4*- and *mtgr1*-MOs also became more apparent (Fig. 5B).

We conclude that the newly identified Ldb1 interaction partners eto-2, cdk9 and lmo4 are essential for definitive erythropoiesis in zebrafish, whereas mtgr1 plays a less critical role. The fact that embryonic haematopoieis is intact shows that the expression of *tal1* and *gata1* is not affected by knockdown of *eto-2*, *cdk9* and *lmo4*. The dorsal aorta is normally formed in such treated embryos, indicating that the *eto-2*, *cdk9* and *lmo4* are not required for vasculogenesis. It is noteworthy in this context that Eto-2 in zebrafish (this paper) and most of its constitutive binding partners, namely HEB and E47 in mouse (Zhuang et al., 1996) and Mtgr1 in mouse (Amann et al., 2005) and zebrafish (this paper), are not required for embryonic haematopoiesis.

Ldb1 interacting partners are expressed in the same cells in the para-aortic splanchnopleura of the early mouse embryo.

The results obtained for the novel Ldb1 interaction partners in the zebrafish suggest that they would be expressed in the early mouse embryo at stages prior to the 'birth' of the definitive haematopoietic stem cells in the AGM (Durand and Dzierzak, 2005). We therefore performed immunohistochemistry and immunofluorescence on E9.5 embryo sections (Fig. 6). Immunohistochemical analysis with α-Ldb1, α-E2A, α-Lmo2, α-Gata1, α-Eto-2 and α-Cdk9 showed that all are expressed in the para-aortic splanchnopleura (P-Sp), the region destined to contribute to the AGM (Fig. 6A,B).

To further determine the expression pattern of the interacting proteins and to confirm that cells within the P-Sp co-express these proteins, we performed in situ immunofluorescence experiments on the cryosections. We found that cells positive for *Gata1* expression also expressed *Eto-2*, *Ldb1* and *Runx1* (Fig. 6C). Cells expressing *Ldb1* were also positive for Runx1, E2A and Lmo2 (not shown). Although we cannot analyse all of the proteins due to antibody incompatibilities, the observation that *Gata1* and *Eto-2*, *Gata1* and *Ldb1*, and *Ldb1* and *Runx1* are co-expressed in the pre-AGM cells of the P-Sp suggests that they form a complex that may be essential for the induction of the definitive haematopoietic system. The fact that Eto-2 is predominantly cytoplasmic at that time (Fig. 6C, column 3, panels Eto-2 and merge) suggests that it may respond to extracellular signals at the prehaematopoietic stage. At later stages (fetal liver), it is abundant in the nucleus but its cytoplasmic expression is very weak. (see Fig. S2 in the supplementary material).

DISCUSSION

Ldb1 forms dynamic complexes during erythroid differentiation

By using a systems biology approach we identified all known and many new binding partners of Ldb1 and determined their importance for haematopoietic development. The important implication of this work is that the analysis of proteins co-expressed and interacting with each other in a late mature cell type are already implicated at the earliest stages of, in this case, blood development.

The Ldb1 proteome has a number of interesting interactions: (1) with Eto-2 (and Mtgr1); (2) with a large number of different bHLH proteins; (3) with Cdk9; and (4) with the Ssdp proteins, although this latter interaction was not analysed further. Recently we have shown that Gata1 forms at least five clearly identifiable protein complexes (Rodriguez et al., 2005). We did not observe such distinct complexes for Ldb1. Clearly, Ldb1 forms a core complex with the known partners Gata1, Tal1, Lmo2 and E2A, and the newly identified partner Eto-2 (Fig. 6). Eto-2 also forms complexes with E2A and Tal1 without Ldb1. In the sequential immunoprecipitations we can distinguish between the preferential binding partners of Eto-2 and Ldb1, and deduce the existence of a higher order complex the formation of which is favoured in proliferating cells. At the same time it is difficult to separate groups of interacting proteins from each other by immunoprecipitations, especially if a protein interacts with others in different combinations. Hence, we are in the process of purifying the different complexes. With induction, levels of Eto-2 decrease and formation of the large complex is lower (Fig. 7). Cdk9, which has been linked to cell-cycle progression (Bettencourt-Dias et al., 2004), would no longer be part of the complex. This may explain why the cells stop proliferating upon differentiation. At the same time, the levels of Lmo4 increase, possibly replacing Lmo2/Eto-2 and leading to an activation of transcription of genes expressed after terminal differentiation.

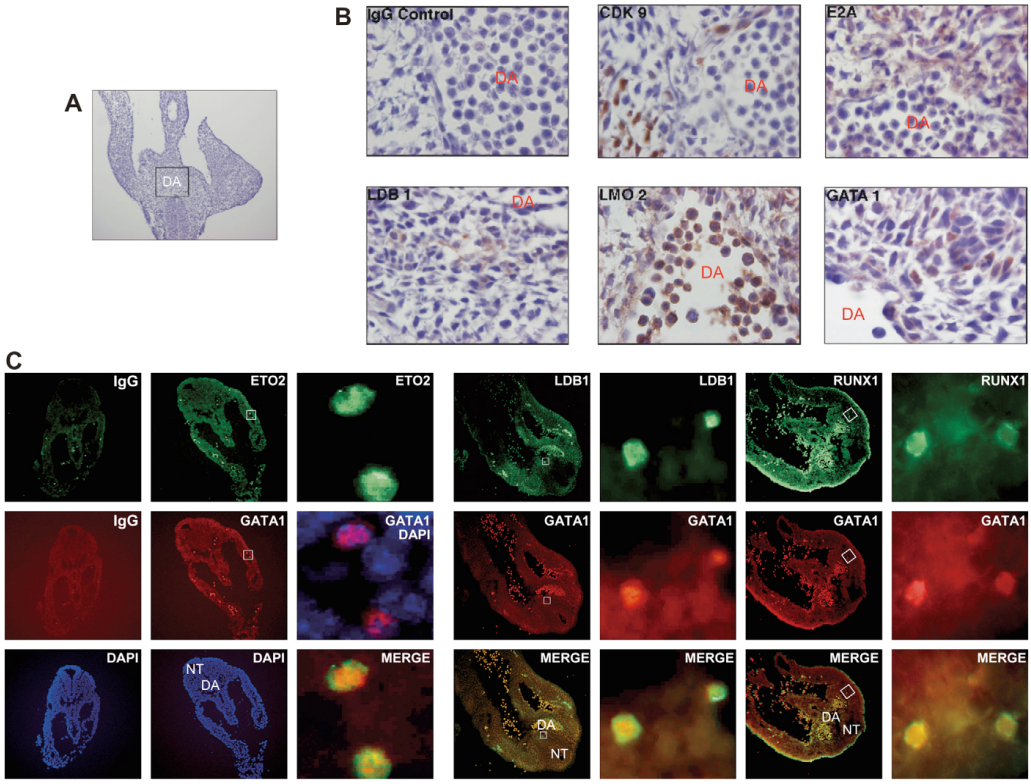


Fig. 6. Ldb1 interacting partners are expressed in the P-Sp. (**A**) Haematoxylin stained 10 μm transversal cryosection of a E9.5 FvB embryo. The area of the P-Sp is indicated by the box. (**B**) High magnification (×1000) of the areas such as shown in (A). Expression of *Cdk9*, *E2A*, *Ldb1*, *Lmo2* and *Gata1* was detected with specific antibodies and visualised with DAB. IgG panel shows the background staining with an unspecific IgG antibody. (**C**) Immunofluorescence analysis of E9.5 embryos with α-Eto-2, α-Ldb1, α-Runx1 and α-Gata1 antibodies. Specific staining is seen for both Eto-2 (green) and Gata1 (red) antibodies (second column), compared with the IgG control (first column). (Third column) Enlargement of the section shown in the second column. Merge is a superposition of images of the Eto-2 and Gata1 detection; cells that express both *Eto-2* and *Gata1* appear as yellow. The overlay of Gata1 and DAPI staining shows exclusive nuclear localisation of Gata1, whereas Eto-2 is seen in the nucleus and cytoplasm. (Fourth column) *Ldb1*- (green) and *Gata1*- (red) expressing cells are located in the dorsal aorta region; most of the cells positive for Ldb1 are also positive for Gata1 (Ldb1 is also expressed in other cell types). A similar pattern is shown for *Gata1*- (red) and *Runx1*- (green) expressing cells in E9.5 embryos (sixth column). (Fifth and seventh columns) High magnification of the same sections reveals more detail. DA, dorsal aorta; NT, neural tube.

Noteworthy is the presence of a large number of DNA binding proteins within the complex, particularly the presence of at least five bHLH proteins and the zinc finger transcription factor Gata1. It is possible that such a complex may very well be involved in the establishment and/or facilitation of long-range interactions, processes in which Ldb1 has been implicated (Morcillo et al., 1997). Specifically, in *Drosophila melanogaster*, Chip was identified in a screen for factors involved in the long-range gene activation of the *cut* gene. Chip was proposed to bridge the Pannier (GATA) and Achaete/Scute (bHLH) complexes causing the intervening DNA to loop out bringing DNA control elements into close proximity (Ramain et al., 2000). The erythroid Ldb1-Lmo2-Tal1-E2A-Gata1 complex may have a similar role. It was recently shown that the complex binds to multiple sites in the α-globin locus (Anguita et al., 2004). We envisage that these complexes also interact and promote long-range interactions in other gene loci, for example in the β-globin locus ACH (Tolhuis et al., 2002).

The Eto family members Eto-2 and Mtgr1 are thought to be repressors by binding the NCor/Sin3A/HDAC1 (Gelmetti et al., 1998; Lutterbach et al., 1998; Wang et al., 1998) complex. Eto-2 was the most abundant protein in the MS analysis, suggesting it is a direct (and crucial) binding partner of Ldb1 (confirmed by immunoprecipitations). We also observed that Eto-2 and its family member Mtgr1 interact with each other. Moreover Eto-2 appears to be the bridging factor for Ldb1 to interact with Mtgr1 (Fig. 3, and see Fig. S1A in the supplementary material). Eto proteins and the *D. melanogaster* orthologue *nervy* have four highly conserved protein interaction domains (Davis et al., 1999;

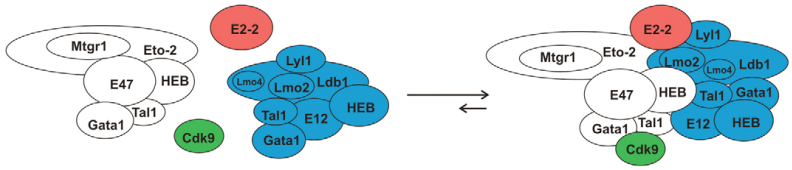


Fig. 7. Model of Ldb1 complexes in uninduced MEL cells. The horizontal arrows indicate that the balance of interaction is towards the large complex in proliferating noninduced cells. Upon the induction of differentiation and termination of proliferation the level of Eto-2 drops whereas the level of Lmo4 rises, hence the equilibrium would shift towards the smaller complexes. The presence of several DNA binding proteins in a single complex may explain the role of Ldb1 as a facilitator of long-range interactions.

Davis et al., 2003). The DNA binding of the Eto-2 complexes we describe probably occurs through its bHLH binding partners and/or Gata1.

Interestingly, with induction Eto-2 and the Eto-2/Ldb1 complex decrease, while the level of Lmo4 increases. Thus, the level of the large Ldb1-Eto-2 complex drops to be replaced by one with Lmo4. Complementary results were obtained by other groups through characterisation of Tal1 complexes (Schuh et al., 2005; Goardon et al., 2006). As indicated by the chromatin immunoprecipitation experiments, the repressive role of Eto-2 may very well explain how late erythroid genes become activated. An alternative or complementary explanation may be post-translational protein modifications of the Eto-2 complex members, a possibility that cannot be excluded at present.

A good candidate to carry out modifications and change specific interactions is Cdk9. This CDC2 orthologue precipitated differentially with Ldb1 in proliferating and differentiating C88 cells. The invariance of the main isoform independent of the cell-cycle stage has been reported (Garriga et al., 2003). Cdk9 is thought to have two functions: regulating RNA polymerase II by phosphorylating its C-terminal domain (Marshall et al., 1996; Zhu et al., 1997); and regulating the cell cycle. A 'knockdown' of the *D. melanogaster* orthologue of *Cdk9* causes an arrest at the G1 to S transition (Bettencourt-Dias et al., 2004), congruent with the in vitro data that the human *CDK9* orthologue phosphorylates Rb (Graña et al., 1994). Interestingly, Gata1 has been found to interact with Rb in vitro (Whyatt et al., 1997). We detect an interaction between Cdk9 and Ldb1 only in noninduced, proliferating C88 cells, suggesting Ldb1 and its partners might be involved in transcriptional control of the cell cycle through the dual function of Cdk9. Putative target genes could be E2F family members expressed in haematopoietic cells. Cdk9 also interacts with BRG1 and STAT3 to activate transcription of the cell-cycle inhibitor p21waf1 (Giraud et al., 2004), suggesting that Cdk9 has different functions in the cell cycle depending on its partners. In this context it is interesting that we detect Eto-2 protein in the cytoplasm of haematopoietic stem cell precursors (Fig. 6) and of developing neurons (N.M., unpublished). Eto-2 may respond to extracellular signals by translocating from the cytoplasm into the nucleus, while it is downregulated in terminally differentiating cells. This suggests that Eto-2 plays a key role in the development of the definitive HSCs, and is required for the subsequent phases of expansion of the different lineages. It then needs to be downregulated to allow the activation of late genes such as *Eklf* for terminal differentiation.

Late erythroid differentiation complexes and early haematopoiesis

It is remarkable that the analysis of interacting proteomes of Gata1 (Rodriguez et al., 2005) and Ldb1 (this paper) in late erythroid cells has resulted in the identification of a number of proteins essential for early haematopoiesis (as found in the zebrafish experiments), in particular Lmo4, Cdk9 and Eto-2.

Lmo4 expression in mice has been detected at E9 from the caudal region of the dorsal and lateral paraxial mesoderm up to the direct vicinity of the dorsal aorta, suggesting that Lmo4 could play a role in HSC formation (Kenny et al., 1999). Lmo4 is also upregulated in late T cell differentiation (Kenny et al., 1999) similar to that observed in C88 cells. In *Xenopus*, Xlmo4 and Gata-2 act synergistically in ventral mesoderm formation. However, an Xlmo4-MO did not prevent ventral mesoderm formation (de la Calle-Mustienes et al., 2003). In zebrafish, *lmo4* is expressed at gastrulation but not during ventral mesoderm formation (Lane et al., 2002). In accordance with this, we found embryonic haematopoiesis in *lmo4*-MO-treated zebrafish embryos to be normal. We show the expression of *lmo4* at later stages to be important for the formation of definitive HSCs in zebrafish. Whether the same phenotype is observed in mice remains to be determined.

MO-mediated repression of *eto-2* had the most drastic phenotype in zebrafish – absence of *runx1* expression in embryonic erythrocytes and lack of definitive haematopoiesis, suggesting it is essential for *runx1* expression at that stage. *Cdk9*-MO-treated embryos lacked definitive cells. However in embryonic blood cells *runx1* expression was detected, suggesting *Cdk9* does not regulate *runx1* directly.

The fact that these factors are already co-expressed in the same cells in the murine P-Sp before the generation of the definitive HSCs in the AGM (de Bruijn et al., 2000) suggests that they are also essential for definitive haematopoiesis in the mouse. Eto-2 and Cdk9 may be more important in this process than Gata1, because loss of Gata1 is not deleterious to these very early stages (with cytoplasmic Eto-2). This in turn suggests that Tal1, Ldb1 and Lmo2 are dominant at this stage. Because all of the Ldb1-bound Gata1 is complexed with Eto-2, we postulate that when Eto-2 translocates from the cytoplasm to the nucleus, the balance of interactions is changed causing the transition to a more proliferative Gata1-driven phase followed by terminal differentiation.

We are grateful to Drs S. Orkin, R. Baer and G. Gill for the α-Lmo2, α-Tal1 and α-Ldb1 antibodies and Dr S. Schulte-Merker for help with the zebrafish injections. N.M. was supported by a DFG Fellowship, A.H. was supported by a Marie Curie Fellowship. This work is supported by the NWO (NL), the NIH and the EU (CiO).

Supplementary material
Supplementary material for this article is available at
http://dev.biologists.org/cgi/content/full/133/24/4913/DC1

References

Agulnick, A. D., Taira, M., Breen, J. J., Tanaka, T., Dawid, I. B. and Westphal, H. (1996). Interactions of the LIM-domain-binding factor Ldb1 with LIM homeodomain proteins. *Nature* **384**, 270-272.

Amann, J. M., Chyla, B. J., Ellis, T. C., Martinez, A., Moore, A. C., Franklin, J. L., McGhee, L., Meyers, S., Ohm, J. E., Luce, K. S. et al. (2005). Mtgr1 is a transcriptional corepressor that is required for maintenance of the secretory cell lineage in the small intestine. *Mol. Cell. Biol.* **25**, 9576-9585.

Anguita, E., Hughes, J., Heyworth, C., Blobel, G. A., Wood, W. G. and Higgs, D. (2004). Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA1 and GATA-2. *EMBO J.* **23**, 2841-2852.

Antoniou, M. (1991). Induction of erythroid-specific expression in murine erthroleukemia (MEL) cell lines. In *Methods in Molecular Biology, Gene Transfer and Expression Protocols*, Vol. 7 (ed. E. J. Murray), pp. 421-434. Totowa, NJ: The Humana Press Inc.

Bagella, L., MacLachlan, T. K., Buono, R. J., Pisano, M. M., Giordano, A. and De Luca, A. (1998). Cloning of murine CDK9/PITALRE and its tissue-specific expression in development. *J. Cell Physiol.* **177**, 206-213.

Bettencourt-Dias, M., Giet, R., Sinka, R., Mazumdar, A., Lock, W. G., Balloux, F., Zafiropoulos, P. J., Yamaguchi, S., Winter, S., Carthew, R. W. et al. (2004). Genome-wide survey of protein kinases required for cell cycle progression. *Nature* **432**, 980-987.

Brand, M., Ranish, J. A., Kummer, N. T., Hamilton, J., Igarashi, K., Francastel, C., Chi, T. H., Crabtree, G. R., Aebersold, R. and Groudine, M. (2004). Dynamic changes in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics. *Nat. Struct. Mol. Biol.* **11**, 73-80.

Cantor, A. B. and Orkin, S. (2001). Hematopoietic development: a balancing act. *Curr. Opin. Genet. Dev.* **11**, 513-519.

Chen, L., Segal, D., Hukriede, N. A., Podtelejnikov, A. V., Bayarsaihan, D., Kennison, J. A., Ogryzko, V. V., Dawid, I. B. and Westphal, H. (2002). Ssdp proteins interact with the LIM-domain-binding protein Ldb1 to regulate development. *Proc. Natl. Acad. Sci. USA* **99**, 14320-14325.

Cumano, A., Ferraz, J. C., Klaine, M., Di Santo, J. P. and Godin, I. (2001). Intraembryonic, but not yolk sac hematopoietic precursors, isolated before circulation, provide long-term multilineage reconstitution. *Immunity* **15**, 477-485.

Davis, J. N., Williams, B. J., Herron, J. T., Galiano, F. J. and Meyers, S. (1999). ETO-2, a new member of the ETO-family of nuclear proteins. *Oncogene* **18**, 1375-1383.

Davis, J. N., McGhee, L. and Meyers, S. (2003). The ETO (MTG8) gene family. *Gene* **303**, 1-10.

de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., Grosveld, F. and Strouboulis, J. (2003). Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc. Natl. Acad. Sci. USA* **100**, 7480-7485.

de Bruijn, M. F., Speck, N. A., Peeters, M. C. and Dzierzak, E. (2000). Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo. *EMBO J.* **19**, 2465-2474.

de Bruijn, M. F., Ma, X., Robin, C., Ottersbach, K., Sanchez, M. J. and Dzierzak, E. (2002). Hematopoietic stem cells localize to the endothelial cell layer in the midgestation mouse aorta. *Immunity* **16**, 673-683.

de la Calle-Mustienes, E., Lu, Z., Cortes, M., Andersen, B., Modolell, J. and Gomez-Skarmeta, J. (2003). Xenopus Xlmo4 is a GATA cofactor during ventral mesoderm formation and regulates Ldb1 availability at the dorsal mesoderm and the neural plate. *Dev. Biol.* **264**, 564-581.

Durand, C. and Dzierzak, E. (2005). Embryonic beginnings of adult hematopoietic stem cells. *Haematologica* **90**, 100-108.

Garriga, J., Bhattacharya, S., Calbo, J., Marshall, R. M., Truongcao, M., Haines, D. S. and Graña, X. (2003). CDK9 is constitutively expressed throughout the cell cycle, and its steady-state expression is independent of SKP2. *Mol. Cell. Biol.* **23**, 5165-5173.

Gekas, C., Dieterlen-Lievre, F., Orkin, S. H. and Mikkola, H. (2005). The placenta is a niche for hematopoietic stem cells. *Dev. Cell* **8**, 365-375.

Gelmetti, V., Zhang, J., Fanelli, M., Minucci, S., Pelicci, P. G. and Lazar, M. (1998). Aberrant recruitment of the nuclear receptor corepressor-histone deacetylase complex by the acute myeloid leukemia fusion partner ETO. *Mol. Cell. Biol.* **18**, 7185-7191.

Gering, M. and Patient, R. (2005). Hedgehog signaling is required for adult blood stem cell formation in zebrafish embryos. *Dev. Cell* **8**, 389-400.

Giraud, S., Hurlstone, A., Avril, S. and Coqueret, O. (2004). Implication of BRG1 and cdk9 in the STAT3-mediated activation of the p21waf1 gene. *Oncogene* **23**, 7391-7398.

Goardon, N., Lambert, J. A., Rodriguez, P., Nissaire, P., Herblot, S., Thibault, P., Dumenil, D., Strouboulis, J., Romeo, P. H. and Hoang, T. (2006). ETO2 coordinates cellular proliferation and differentiation during erythropoiesis. *EMBO J.* **25**, 357-366.

Graña, X., De Luca, A., Sang, N., Fu, Y., Claudio, P. P., Rosenblatt, J., Morgan, D. O. and Giordano, A. (1994). PITALRE, a nuclear CDC2-related protein kinase that phosphorylates the retinoblastoma protein in vitro. *Proc. Natl. Acad. Sci. USA* **91**, 3834-3838.

Guyot, B., Valverde-Garduno, V., Porcher, C. and Vyas, P. (2004). Deletion of the major GATA1 enhancer HS 1 does not affect eosinophil GATA1 expression and eosinophil differentiation. *Blood* **104**, 89-91.

Hahm, K., Sum, E. Y., Fujiwara, Y., Lindeman, G. J., Visvader, J. E. and Orkin, S. H. (2004). Defective neural tube closure and anteroposterior patterning in mice lacking the LIM protein LMO4 or its interacting partner Deaf-1. *Mol. Cell. Biol.* **24**, 2074-2082.

Jowett, T. and Yan, Y. L. (1996). Double fluorescent in situ hybridization to zebrafish embryos. *Trends Genet.* **12**, 387-389.

Kalev-Zylinska, M. L., Horsfield, J. A., Flores, M. V., Postlethwait, J. H., Vitas, M. R., Baas, A. M., Crosier, P. S. and Crosier, K. (2002). Runx1 is required for zebrafish blood and vessel development and expression of a human RUNX1-CBF2T1 transgene advances a model for studies of leukemogenesis. *Development* **129**, 2015-2030.

Kenny, D. A., Jurata, L. W., Saga, Y. and Gill, G. (1999). Identification and characterization of LMO4, an LMO gene with a novel pattern of expression during embryogenesis. *Proc. Natl. Acad. Sci. USA* **95**, 11257-11262.

Korinek, V., Barker, N., Moerer, P., van Donselaar, E., Huls, G., Peters, P. J. and Clevers, H. (1998). Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nat. Genet.* **19**, 379-383.

Lahlil, R., Lecuyer, E., Herblot, S. and Hoang, T. (2004). SCL assembles a multifactorial complex that determines glycophorin A expression. *Mol. Cell. Biol.* **24**, 1439-1452.

Lane, M. E., Runko, A. P., Roy, N. M. and Sagerström, C. (2002). Dynamic expression and regulation by Fgf8 and Pou2 of the zebrafish LIM-only gene, lmo4. *Gene Expr. Patterns* **2**, 207-211.

Lee, S. K., Jurata, L. W., Nowak, R., Lettieri, K., Kenny, D. A., Pfaff, S. L. and Gill, G. (2005). The LIM domain-only protein LMO4 is required for neural tube closure. *Mol. Cell. Neurosci.* **28**, 205-214.

Liu, H. and Herrmann, C. (2005). Differential localization and expression of the Cdk9 42k and 55k isoforms. *J. Cell Physiol.* **203**, 251-260.

Lutterbach, B., Westendorf, J. J., Linggi, B., Patten, A., Moniwa, M., Davie, J. R., Huynh, K. D., Bardwell, V. J., Lavinsky, R. M., Rosenfeld, M. G. et al. (1998). ETO, a target of t(8;21) in acute leukemia, interacts with the N-CoR and mSin3 corepressors. *Mol. Cell. Biol.* **18**, 7176-7184.

Marshall, N. F., Peng, J., Xie, Z. and Price, D. (1996). Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J. Biol. Chem.* **271**, 27176-27183.

McDevitt, M. A., Fujiwara, Y., Shivdasani, R. A. and Orkin, S. (1997). An upstream, DNase I hypersensitive region of the hematopoietic-expressed transcription factor GATA1 gene confers developmental specificity in transgenic mice. *Proc. Natl. Acad. Sci. USA* **94**, 7976-7981.

Medvinsky, A. and Dzierzak, E. (1996). Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell* **86**, 897-906.

Morcillo, P., Rosen, C., Baylies, M. K. and Dorsett, D. (1997). Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in Drosophila. *Genes Dev.* **11**, 2729-2740.

Mukhopadhyay, M., Teufel, A., Yamashita, T., Agulnick, A. D., Chen, L., Downs, K. M., Schindler, A., Grinberg, A., Huang, S. P., Dorward, D. et al. (2003). Functional ablation of the mouse Ldb1 gene results in severe patterning defects during gastrulation. *Development* **130**, 495-505.

Nasevicius, A. and Ekker, S. (2000). Effective targeted gene knockdown in zebrafish. *Nat. Genet.* **26**, 216-220.

Nishioka, N., Nagano, S., Nakayama, R., Kiyonari, H., Ijiri, T., Taniguchi, K., Shawlot, W., Hayashizaki, Y., Westphal, H., Behringer, R. R. et al. (2005). Ssdp1 regulates head morphogenesis of mouse embryos by activating the Lim1-Ldb1 complex. *Development* **132**, 2535-2546.

Onodera, K., Takahashi, S., Nishimura, S., Ohta, J., Motohashi, H., Yomogida, K., Hayashi, N., Engel, J. D. and Yamamoto, M. (1997). GATA1 transcription is controlled by distinct regulatory mechanisms during primitive and definitive erythropoiesis. *Proc. Natl. Acad. Sci. USA* **94**, 4487-4492.

Ottersbach, K. and Dzierzak, E. (2005). The murine placenta contains hematopoietic stem cells within the vascular labyrinth region. *Dev. Cell* **8**, 377-387.

Palstra, R. J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F. and de Laat, W. (2003). The beta-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* **35**, 190-194.

Patrinos, G. P., de Krom, M., de Boer, E., Langeveld, A., Imam, A. M., Strouboulis, J., de Laat, W. and Grosveld, F. (2004). Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. *Genes Dev.* **18**, 1495-1509.

Quinkertz, A. and Campos-Ortega, J. (1999). A new beta-globin gene from the zebrafish, betaE1, and its pattern of transcription during embryogenesis. *Dev. Genes Evol.* **209**, 126-131.

Ramain, P., Khechumian, R., Khechumian, K., Arbogast, N., Ackermann, C. and Heitzler, P. (2000). Interactions between chip and the achaete/scute-daughterless heterodimers are required for pannier-driven proneural patterning. *Mol. Cell* **6**, 781-790.

Robb, L., Lyons, I., Li, R., Hartley, L., Kontgen, F., Harvey, R. P., Metcalf, D. and Begley, C. G. (1995). Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene. *Proc. Natl. Acad. Sci. USA* **92**, 7075-7079.

Rodriguez, P., Bonte, E., Krijgsveld, J., Kolodziej, K. E., Guyot, B., Heck, A. J., Vyas, P., de Boer, E., Grosveld, F. and Strouboulis, J. (2005). GATA1 forms distinct activating and repressive complexes in erythroid cells. *EMBO J.* **24**, 2354-2366.

Schuh, A. H., Tipping, A. J., Clark, A. J., Hamlett, I., Guyot, B., Iborra, F. J., Rodriguez, P., Strouboulis, J., Enver, T., Vyas, P. et al. (2005). ETO-2 associates with SCL in erythroid cells and megakaryocytes and provides repressor functions in erythropoiesis. *Mol. Cell. Biol.* **25**, 10235-10250.

Shivdasani, R. A., Mayer, E. L. and Orkin, S. H. (1995). Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. *Nature* **373**, 432-434.

Shore, S. M., Byers, S. A., Maury, W. and Price, D. (2003). Identification of a novel isoform of Cdk9. *Gene* **307**, 175-182.

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell. Biol.* **10**, 1453-1465.

Torigoi, E., Bennani-Baiti, I. M., Rosen, C., Gonzalez, K., Morcillo, P., Ptashne, M. and Dorsett, D. (2000). Chip interacts with diverse homeodomain proteins and potentiates bicoid activity in vivo. *Proc. Natl. Acad. Sci. USA* **97**, 2686-2691.

Tse, E., Smith, A. J., Hunt, S., Lavenir, I., Forster, A., Warren, A. J., Grutz, G., Foroni, L., Carlton, M. B., Colledge, W. H. et al. (2004). Null mutation of the Lmo4 gene or a combined null mutation of the Lmo1/Lmo3 genes causes perinatal lethality, and Lmo4 controls neural tube development in mice. *Mol. Cell. Biol.* **24**, 2063-2073.

van Meyel, D. J., Thomas, J. B. and Agulnick, A. (2003). Ssdp proteins bind to LIM-interacting co-factors and regulate the activity of LIM-homeodomain protein complexes in vivo. *Development* **130**, 1915-1925.

Wadman, I. A., Osada, H., Grutz, G. G., Agulnick, A. D., Westphal, H., Forster, A. and Rabbitts, T. (1997). The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA1 and Ldb1/NLI proteins. *EMBO J.* **16**, 3145-3157.

Wang, J., Hoshino, T., Redner, R. L., Kajigaya, S. and Liu, J. (1998). ETO, fusion partner in t(8;21) acute myeloid leukemia, represses transcription by interaction with the human N-CoR/mSin3/HDAC1 complex. *Proc. Natl. Acad. Sci. USA* **95**, 10860-10865.

Warren, A. J., Colledge, W. H., Carlton, M. B., Evans, M. J., Smith, A. J. and Rabbitts, T. H. (1994). The oncogenic cysteine-rich LIM domain protein rbtn2 is essential for erythroid development. *Cell* **78**, 45-57.

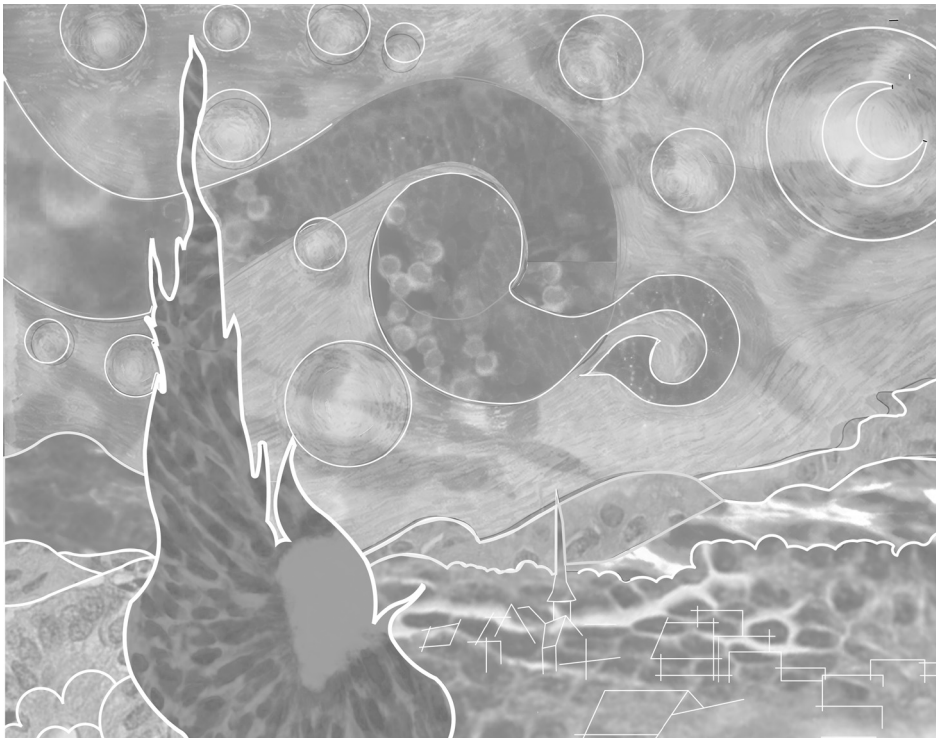
Westerfield, M. (1993). *The Zebrafish Book: A Guide for the Laboratory Use of Zebrafish (Brachydanio rerio).* Eugene, Oregon, USA: University of Oregon Press.

Whyatt, D. J., Karis, A., Harkes, I. C., Verkerk, A., Gillemans, N., Elefanty, A. G., Vairo, G., Ploemacher, R., Grosveld, F. and Philipsen, S. (1997). The level of the tissue-specific factor GATA1 affects the cell-cycle machinery. *Genes Funct.* **1**, 11-24.

Yoder, M. C., Hiatt, K., Dutt, P., Mukherjee, P., Bodine, D. M. and Orlic, D. (1997). Characterization of definitive lymphohematopoietic stem cells in the day 9 murine yolk sac. *Immunity* **7**, 335-344.

Zhu, Y., Pe'ery, T., Peng, J., Ramanathan, Y., Marshall, N., Marshall, T., Amendt, B., Mathews, M. B. and Price, D. H. (1997). Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation in vitro. *Genes Dev.* **11**, 2622-2632.

Zhuang, Y., Cheng, P. and Weintraub, H. (1996). B-lymphocyte development is regulated by the combined dosage of three basic helix-loop-helix genes, E2A, E2-2, and HEB. *Mol. Cell. Biol.* **16**, 2898-2905.



Chapter 3 - *Delangin* deficient mice exhibit developmental abnormalities similar to Cornelia de Lange Syndrome

Sanja Krpic, Arnd Hostert, Marcel Vermeij and Frank Grosveld

work in progress

ABSTRACT

Delangin (NIPBL in humans, Nipped-B in *Drosophila*, Scc2 in yeast) is a protein involved in the loading of Cohesin onto chromatin and hence thought to be important for sister chromatid cohesion. Mutations in *Delangin* and the *Cohesin* complex are associated with the developmental disorder Cornelia de Lange Syndrome (CdLS) in humans. It is a rare dominantly inherited multisystem disorder affecting both physical and mental development. Heterozygous mutations in the *Delangin* gene were found in about half of CdLS cases. To understand the role of Delangin in mammals, we generated mice lacking Delangin. Homozygous mutants are lethal early in development. Heterozygous mice are born exhibiting the characteristics of CdLS that is heterogeneous and not fully penetrant. Affected mice have severely reduced body size and abnormalities in many organs and show abnormal behavior. This suggests that a reduced level of Delangin can cause chromatin defects very early in development. We did not, however, detect cohesion defects *in vitro* experiments. These results suggest that Delangin have important functions in chromosomal dynamics and gene regulation. In addition, we think that our mouse model can be a valuable model for understanding the molecular mechanism behind Cornelia de Lange Syndrome.

INTRODUCTION:

Cell proliferation depends on genome duplication and segregation of sister chromatids to opposite poles of the cell prior to division. Cohesin is a protein complex that keeps sister chromatids together from the time of replication in S phase until their separation at the onset of anaphase, a process that is essential for the correct segregation of chromosomes during mitosis. The Cohesin proteins are evolutionarily conserved across species (Hirano, 2006; Nasmyth and Haering, 2005). The core complex consists of SMC1 and SMC3, members of the SMC (Structural Maintenance of Chromosomes) protein family, and two non-SMC subunits SCC1/Rad21 and SCC3/SA (Nasmyth and Haering, 2005). Together, these proteins form a large ring-shaped structure of about 30-40 nm in diameter that encircles DNA. Several models have been suggested for the mechanism by which Cohesin rings may hold the two sister chromatids together but the precise topology is still unclear (Losada, 2007). Cohesin is required but not sufficient for sister chromatid cohesion. The association of Cohesin with chromatin requires the SMC proteins to have a functional hinge domain, ATPse activity and the activity of SCC2/SCC4 loading complex (Ciosk et al., 2000; Gruber et al., 2006; Weitzer et al., 2003). Cohesin activity is regulated by its regulatory factors SCC2, SCC4, PDS5 and ECO1 (Hartman et al., 2000; Nasmyth and Haering, 2005). In addition to sister chromatid cohesion, Cohesin is important for meiosis, DNA repair, gene expression and regulation of developmental processes (Dorsett, 2007; Hirano and Hirano, 2006; Nasmyth and Haering, 2005; Revenkova

and Jessberger, 2006).

Early evidence that the sister chromatid cohesion apparatus plays a role in gene regulation arose from studies on gene silencing in yeast and on long-range gene activation in *Drosophila* (Donze et al., 1999; Rollins et al., 1999). In more recent studies it was shown that Cohesin has a role in neuronal development in *Drosophila* (Pauli et al., 2008; Schuldiner et al., 2008). Loss of Cohesin binding in dividing cells resulted in Precocious Sister-Chromatid Separation (PSCS) while in non-dividing cells it resulted in defects in axonal pruning (Pauli et al., 2008). In vertebrates, early studies that identified Cohesin as a factor in the regulation of the particular genes focused on the expression of *runx* genes in zebrafish. The combination of a genetic screen and morpholino experiments revealed that Rad21 and SMC3 are essential for proper *runx1* expression (Horsfield et al., 2007). Genome-wide chromatin immunoprecipitation experiments in human and mouse cells subsequently showed that Cohesin accumulates at sides bound by CTCF, a sequence specific DNA-binding protein that has been implicated as a transcriptional regulator, insulator and organizer of higher-order chromatin structure (Parelho et al., 2008; Stedman et al., 2008; Wendt et al., 2008).

The deletion of the *Pds5b* gene, a regulatory component of Cohesin, in mouse, is embryonic lethal, with apparent developmental defects, indicating that Pds5 is important for the role of Cohesin in development (Zhang et al., 2007). However there was no cohesion defect in *PDS5B*-/- cells which suggest that PDS5B and the Cohesin complex have important other functions beyond their role in chromosomal dynamics.

The loading of Cohesin onto chromatin requires the activity of the SCC2/SCC4 complex (Ciosk et al., 2000; Watrin et al., 2006). Consistent with its essential function in regulating sister chromatid cohesion, *Delangin* has been well conserved during evolution. Its orthologues, including *Mis4* (*S.pombe*) and *Rad9* (*C.cinereus*), are known to be involved in various aspects of chromosome function and double-strand DNA repair (Kaur et al., 2005; Michaelis et al., 1997; Rollins et al., 2004; Seitz et al., 1996; Sjogren and Nasmyth, 2001; Strom et al., 2004; Tomonaga et al., 2000; Toyoda et al., 2002; Unal et al., 2004).

Physical linkage of sister chromatids by the cohesin complex is essential for correct chromosome segregation but is also vital for DNA double-strand break (DSB) repair by homologous recombination (HR) during the S- and G2 phase of the cell cycle (Sjogren and Nasmyth, 2001). The inactivation of Cohesin subunits or SCC2/SCC4 results in a reduced efficiency of post-replicative DSB in the G2 phase of the cell cycle. The local enrichment of cohesin at damaged sites depends on the SCC2/SCC4 complex (Strom et al., 2007). Additionally, a homologue of the *Delangin, Pqn-85*, was identified in genome wide screen for genes required for resistance to ionizing radiation in *C.elegans* (van Haaften et al., 2006). Furthermore, there is a evidence for the importance of Delangin in DNA repair in humans, where cells with a mutation in *Delangin* gene derived from Cornelia de Lange patients, have reduced capacity to tolerate DNA damage (Vrouwe et al., 2007).

Delangin (NIPBL, Nipped-B, SCC2) is a large chromatin associated protein

that contains multiple HEAT repeats which mediate protein-protein interactions. The *Delangin* gene is predicted to code for two different isoforms of 2804 and 2697 amino acids, termed Delangin-A and Delangin-B, respectively. The human Delangins share a high degree of homology with *Drosophila* Nipped-B and Scc2 from *S.cerevisiae*.

The N-terminus of the metazoan Delangin contains a Heterochromatin Protein1 (HP1) binding domain that has been shown to interact with HP1α protein (Lechner et al., 2005). Mammalian HP1α heterochromatin proteins have been shown to bind to Delangin through an interaction motif located in a poorly conserved terminal region of the protein (Lechner et al., 2005; Thiru et al., 2004). This interaction raises the possibility that Scc2 has a direct involvement in the establishment and maintenance of heterochromatic domains (Lechner et al., 2005). Delangin homologue in *Drosophila* (*Nipped-B*) is an essential gene that was identified in a genetic screen for mutants interfering with long-range gene activation of the *cut* and *Ultrabithorax (Ubx*) homeobox genes by enhancers positioned some 80 and 50 kbp away from the gene promoters (Rollins et al., 2004; Rollins et al., 1999). Nipped-B facilitates activation of the *cut* gene by the distant wing-margin enhancer and mutations that result in reduced Nipped-B levels result in a lower expression of *cut*. In contrast to other *cut* activators, mutations in *Nipped-B* have a much more striking effect when *cut* activation is hindered by a *gypsy* insulator insertion rather than by loss of important components of the enhancer sequence (Rollins et al., 1999). Subsequent work revealed that, as expected, Nipped-B is also required for sister chromatid cohesion (Rollins et al., 2004). The co-operation with Cohesin subunits in regulating sister chromatid cohesion does not extend to gene regulation. Nipped-B and Cohesin subunits have an opposite effect on the *cut* gene in *Drosophila*: Cohesin inhibits long-range activation of the *cut* gene while Nipped-B activates it (Rollins et al., 2004).

Heterozygous mutations in the *Delangin* gene were found in about half of Cornelia de Lange syndrom patients (Bhuiyan et al., 2006; Gillis et al., 2004; Krantz et al., 2004; Tonkin et al., 2004). In other cases, mutations in the X-linked *SMC1A* and *SMC3* genes were responsible for the same syndrome (Borck et al., 2007; Deardorff et al., 2007; Musio et al., 2006). Cornelia de Lange syndrome (CdLS, also called Brachmann de Lange syndrome; OMIM 122470) is characterized by pre- and postnatal growth retardation, microcephaly, severe mental retardation with speech delay, feeding problems, major malformations including limb defects and characteristic facial features (Jackson et al., 1993). There is a profound heterogeneity of the phenotype in the patients. Individuals who show growth and neuro-developmental delay exhibit classical CdLS features. In mildly affected patients developmental delay and growth retardation are less severe (Ireland et al., 1993; Jackson et al., 1993). Affected individuals are typically identified as sporadic cases though several reports have documented families with multiple affected members (Russell et al., 2001).

Here we describe phenotypic defects resulting from a homozygous or heterozygous deletion of *Delangin* in mice, resulting in a lethal embryonic defect or a classical CdLS phenotype respectively.

RESULTS:

Generation of Delangin-deficient mice

The role of Delangin was examined in mice carrying a targeted allele shown in Figure 1. Using homologous recombination in ES cells, a GFP-neo cassette was inserted in exon 10 (out of 47) 560 bp downstream of the ATG at a unique *Eco*RV site. The targeting of the construct was carried out before the most of the exons were characterized at the 5'end of the gene. Since N-terminus of the *Delangin* gene is not conserved among the species a disruption at exon 10 is highly unlikely to yield a functional protein. Homologous integration of the construct in the genome results in the production of GFP-protein from the endogenous promoter. Three heterozygous independent ES cell clones (Delangin KO ES104/171; Delangin KO ES104B2; Delangin KO ES104B23) were injected into C57BL/6 blastocysts to produce chimeric mice that transmitted the targeted allele through the germline. Southern blot (Figure 1B) shows one positive clone. Chimeric mice were then crossed with CAG-Cre (Sakai and Miyazaki, 1997) mice and the neomycin gene was deleted from the genome. Deletion of the neomycin resistance gene is essential since the promoter included in the selection cassette may influence neighboring genes. The male chimeras were crossed with FVB females to generate mice that are heterozygous for the targeted allele. No differences between 3 independent mouse lines were detected. This indicates that the resulting mouse mutants are indeed the outcome of the *Delangin* genetic modification. When heterozygous mice were crossed, no viable homozygous embryos were born.

To characterize the timing of embryonic lethality caused by homozygous deletion of *Delangin*, timed pregnancies between *Delangin* heterozygous mice were set up. Heterozygous mice exhibit difficulties in breeding making it very difficult to obtain homozygous null mice. However no viable homozygous embryos were detected during embryonic development at different time points starting from 6.5 E days (25 embryos were analyzed) concluding that the homozygous deletion is an early embryonic lethal event. To generate heterozygous mice for analysis we used crosses between heterozygous males and FVB females. Heterozygous mice are born in accordance to the Mendelian ratio.

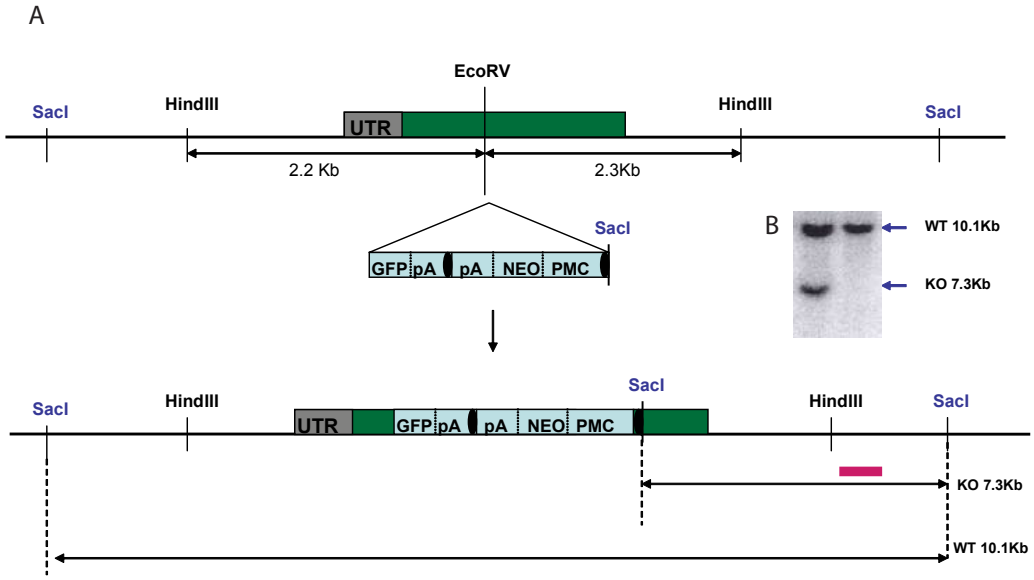


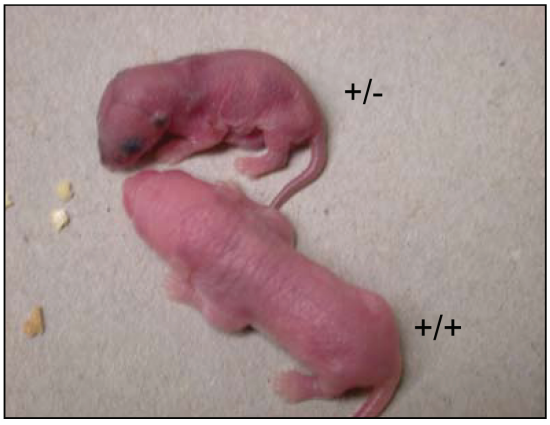
Figure 1
Targeting construct and analysis of a targeted ES clone

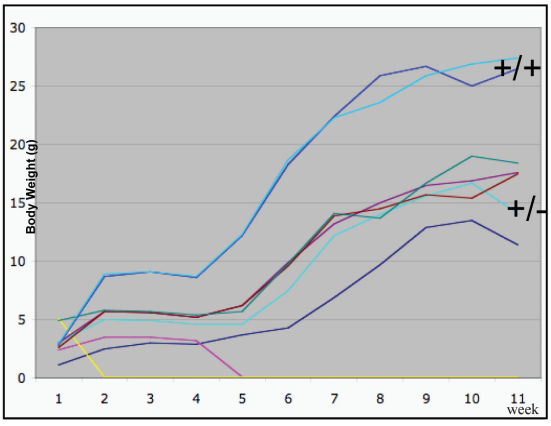
(A) Scheme of a portion of the mouse Delangin gene and the targeting construct
(Top) The genomic region before targeting around exon 10 is shown with the position of a unique EcoRV site where the neo-cassette is inserted and the positions of *Hind*III and *Sac*I restriction sites used to identify legitimate recombination events.
(Middle) The targeting construct (in the 5' to 3' direction) consisting of neomycin resistance gene flanked by lox P sites linked (black ellipses) to a GFP reporter gene and NEO-polyA under PMC (mutated polyoma enhancer).
(Bottom) The genomic arrangement after targeting with the positions of the restriction fragments and the position of the probe used for Southern hybridization (red bar). The *Sac*I fragment in wild type allele is 10.1Kb, and after targeting a novel *Sac*I fragment of 7.3Kb is generated because of the insertion of the neo-cassette.
(B) Analysis of WT and 1 positive clone by Southern hybridization. Note the expected recombinants bands after *Sac*I restriction cuts and hybridization with the probe depicted in red.

Delangin deficiency results in size reduction of the embryos and aberrant postnatal growth

The WT (Wild type) embryos were indistinguishable from mutant once until embryonic day 16. From day E16.5 the mutants began to show signs of progressive growth retardation. Newborn Delangin-deficient mice are drastically smaller from WT littermates with up to 50% reduction in body weight (Figure 2A). In Figure 2B, the weight difference within a litter is followed during 11 weeks. The growth curves show a similar ratio of weight gain between the two groups but the mutants do not draw level with WT littermates resulting in substantial weight differences. Surviving littermates

achieve, on average, only 2/3 of the body weight of WT (Figure 2C). This observation is in consistent with the most commonly observed clinical feature of CdLS patients. i. e. small body size.





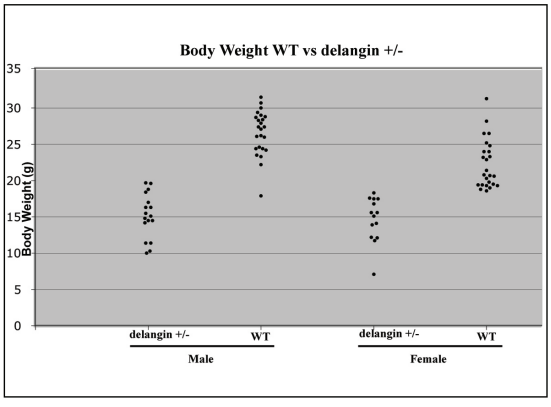


Figure 2

The size reduction of the newborns and aberrant postnatal growth in Delangin- deficient mice

(A) Newborn Delangin heterozygous mice are smaller from WT littermates with up to 50% reduction in body weight.

(B) The weight difference of one litter followed during 11 weeks showing a substantial difference but similar ratio of weight gain.

(C) The weight distribution of Delangin-deficient mice and WT (2 months old). Note a significant difference between two groups for either males or females.

High lethality and a complex, heterogeneous and not fully penetrant phenotype in Delangin-deficient mice

The heterozygous Delangin mice are born exhibiting a phenotype that is complex, heterogeneous and not fully penetrant. They display profound variations from normal morphology that is indistinguishable from WT, to severe body weight reductions and complex developmental abnormalities. In humans, a similar variation in phenotype occurs, for example, the same mutation occurring in different patients was associated with substantially different degrees in pathology. Conversely, patients with the same pathology will in many cases have different mutations within *Delangin* gene (Tonkin et al., 2004). The lethality rate is very high in the post-natal period. The majority of the affected mice die around week 4 usually from heart failure and breathing difficulties.

Since Delangin is ubiquitously expressed in mouse tissues during development, the alteration in Delangin function could potentially have a significant impact on every tissue in the embryo. Various tissues were examined and severe abnormalities in the heart, lungs, stomach, thymus, bone marrow, kidney and the endocrine system were detected.

Heart and lung abnormalities

Since Delangin-deficient mice suffer from respiratory distress and perinatal lethality, we examined cardiac malformation in affected animals. In many cases, we could detect severely dilated cardiomyopathy that is not seen in the control animals (Figure 3 A and B). The diameter of the myocytes is less than normal in many instances. Fibrosis is also detected in the heart muscle (Figure 3 C and D). These abnormalities can cause the respiratory distress and contribute to early postnatal lethality. Patients with CdLS often have congenital heart defects.

Young Delangin-deficient mice developed respiratory insufficiency whereas littermate controls were asymptomatic. Histopathology analysis of lung sections revealed bronchopneumonia with lung emphysema that is not detected in the controls (data not shown). In addition to this pathology of the lungs of older mice had atrophic, or disrupted alveolar walls (Figure G and K). The size of the alveoli and alveolar sacs were drastically different when compared to the control animals resulting in dilation.

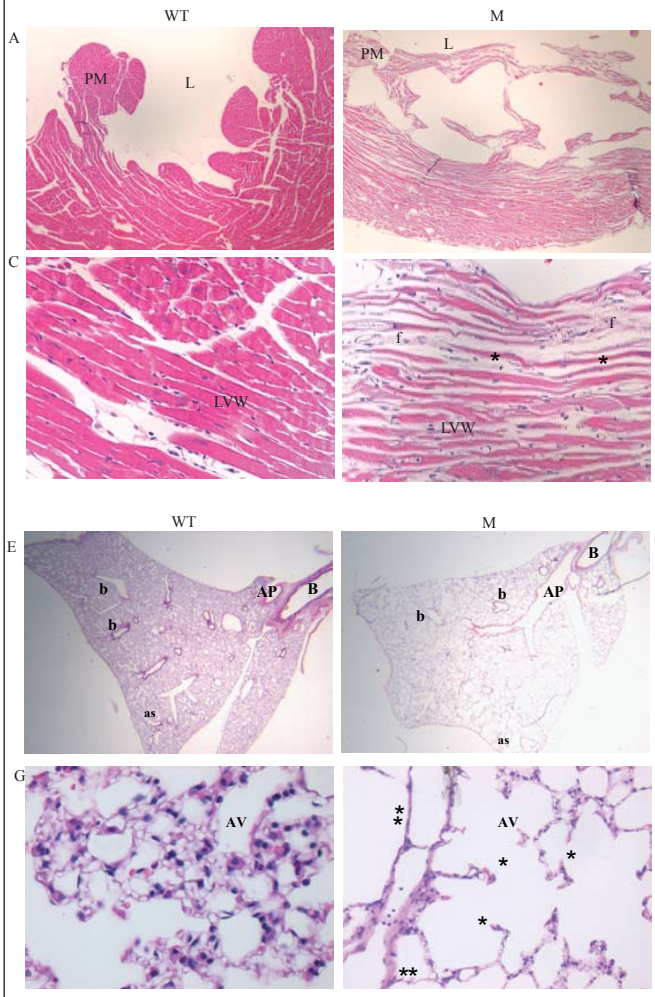


Figure 3

Heart and lung abnormalities in Delangin- deficient mice

(A-D) Heart abnormalities
(A-B) Severe cardiomyopathy in Delangin-deficient mice (B) that is not seen in WT (Wild Type) animals (A).
(C-D) Fibrosis and hyperatrophic heart muscle in Delangin-deficient mice (D) Note the reduced diameter of the myocytes in Delangin-deficient mice compared to normal (C). WT wild type; M mutant, Delangin- deficient
PM papillary muscle; LVW left ventricle wall; L lumen
*atrophic heart muscle f fibrosis

(E-K) Lung abnormalities
(E-K) The lungs of Delangin-deficient mice are atrophic with disrupted alveolar walls (F and K) compared to WT (E and G).
WT wild type; M mutant, Delangin-deficient
B bronchus; AP arteria pulmonaris; b bronchiole; as alveolar sac; AV alveoli
*disrupted alveolar wall
**atrophic alveolar wall

Stomach and Thymus

The liver, small intestine, cecum and colon tissues of Delangin-deficient mice were histologically similar to those in controls animals apart from some sporadic cases. However, the stomach had the most striking defects which were consistent among mutants. The overall mucosa height is drastically decreased in Delangin-deficient mice (Figure 4A and B) and they had an irregular arrangement and reduction of the gastric gland layer that showed a decrease in the number of glands, especially fundic gastric glands. This ultimately leads to atrophy of the gastric mucosa. In addition, the thickness of the muscularis mucosa and muscularis externa was decreased (Figure 4A, 4B, 4C and 4D).

The thymic architecture in Delangin-deficient mice is also disturbed (Figure E

and F). Histological analysis of thymi of young mice revealed progressive changes that result in poorly defined cortical and medular structures. The thymic cortex of Delangin-deficient mice is atrophic and depleted of T lymphocytes resulting in a loss of corticomedullary distinction.

Delangin-deficient mice do not display limb or digit truncations, or loss of any other bone elements features that can be observed in a subset of individuals with CdLS. Upon staining the skeletons of young and adult animals with Alcian Blue and Alizarin red, there were not obvious delays in ossification in Delangin-deficient animals. Also, no apparent lesions were found in the brain in any of the Delangin-deficient animals.

Chromosomal analysis show normal sister chromatid separation in

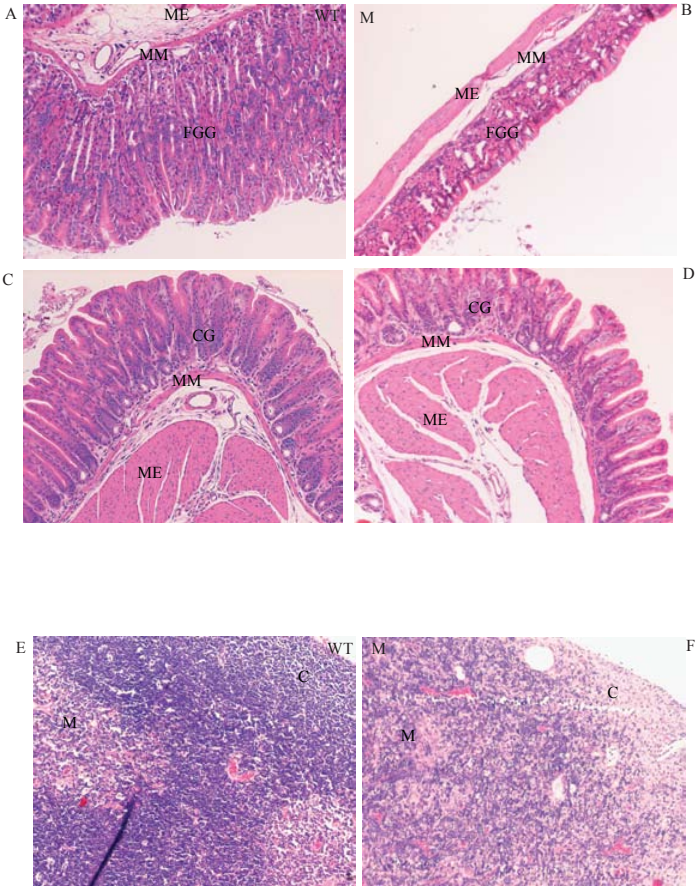


Figure 4
Stomach and Thymus abnormalities

(A-D) Stomach abnormalities
The overall mucosa height is drastically decreased in Delangin-deficient mice (B and D) compared to WT (A and C). Note the difference in the thickness of muscularis mucosa (MM) and muscularis externa (ME).
WT wild type; M mutant, Delangin- deficient
FGG fundic gastric glands
CG cardiac glands
MM Muscularis mucosa
ME Muscularis externa

(E-F) Thymus abnormalities
The thymic architecture is disturbed with poorly defined cortical and medullar structures (F) that are distinctive in WT (Wild type) thymi (E).
M medulla
C cortex

Delangin-deficient mice

The yeast homologues of *Delangin (Scc2)* is important for sister chromatid cohesion. The evidence that absence of *Delangin* results in a similar cohesion abnormality in *Drosophila* (Rollins et al., 2004), and in some cases in CdLS patients (Kaur et al., 2005), prompted us to evaluate Delangin's role in sister chromatid cohesion in our mouse model. Metaphase spreads from undifferentiated ES clones as well as mouse embryonic fibroblasts (MEFs) and B- and T cells were scored for prematurely separated 1, 2 or 3 chromosomes (PSCS1, PSCS2 and PSCS3) as shown in the Figure 5A and B. Western blot analysis shows that the cells are haploinsufficient for Delangin protein and 100 metaphase spreads from each group was evaluated and compared to WT controls. There were no cases where more than 3 chromosomes were separated prematurely. Comparison of metaphase spreads from Delangin-deficient cells with controls led as to conclusion that Delangin-deficient cells did not show a significant increase in PSCS. Other types of chromosomal aberrations such as breaks or fragmentation were not analyzed but an enhanced aneuploidy in mutant groups was also not seen. In accordance with this finding, mice carrying a mutated allele do not develop cancer suggesting that there is no obvious chromosome instability in these mice. We therefore conclude that the phenotype observed in mutant mice is not caused by a function of Delangin maintaining chromosome stability since the cell cycle and proper chromosome segregation were not effected.

A

Cell line		Abnormal cells	PSCS 1	PSCS2	PSCS3
ES	ES B2	7/100	3/100	4/100	0/100
	ESB23	6/100	4/100	1/100	1/100
	WT	8/100	4/100	2/100	2/100
MEF	B2	3/100	2/100	0/100	1/100
	WT	4/100	1/100	3/100	0/100
T-cells	B2	2/100	1/100	1/100	0/100
	WT	2/100	2/100	0/100	0/100
B-cells	B2	1/100	1/100	0/100	0/100
	WT	2/100	2/100	0/100	0/100

B

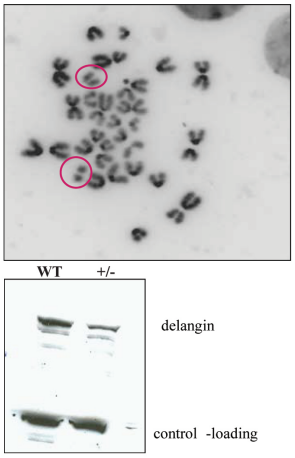


Figure 5
Chromosomal analysis of Premature Sister Chromatid Separation (PSCS) in Delangin-deficient mice

(A) Metaphase spreads from 2 ES clones (ESB2 and ESB23), MEFs (B2 clone), T-cells (B2 clone) and B-cells (B2 clone) were scored for prematurely separated 1, 2 or 3 chromosomes (PSCS1, PSCS2 and PSCS3) and compared to relevant WT(Wild Type) controls.

(B) The example of PSCS2 (prematurely separated chromosomes are encircled) and Western Blot showing reduction in the Delangin protein level.

Delangin heterozygous ES cells are sensitive to γ-radiation and slightly sensitive to UV light

The Delangin homologues SCC2 in yeast, and Pqn-85 in *C. elegans,* have been shown to be required for resistance to double-stranded DNA breakage and cells from CdLS patients have a reduced capacity to tolerate DNA damage (Vrouwe et al., 2007). We therefore investigated the sensitivity of heterozygous Delangin cells to the DNA-damaging agents: γ-radiation and UV light.

Cell survival of the Delangin-deficient ES cells after exposure to damaging agents was tested. Cells were exposed to the increased dosage of UV light and γ-radiation. The sensitivity to damage was determined by measuring their colony-forming ability. Exposure of Delangin-deficient ES cells to increasing doses of UV radiation did not result in an increase in radiation sensitivity. Delangin-deficient ES are slightly sensitive to UV-light only at higher doses of 6 and 9J when compared to WT cells and the UV sensitive Xpa-/- ES positive control cells, which are deficient in nucleotide excision repair (Figure 6A). However Delangin-deficient ES cells a 2 to 3

fold increased sensitivity to γ-radiation (both B2 and B23 clones, Figure 6B). The Rad 54 repair protein deficient cells were used as a positive control (Essers et al., 1997). In the Rad54 deficient cells, both copies of the gene are inactivated while in Delangin deficient clones there is still one WT allele. This makes the sensitivity of Delangin-deficient cells even more striking.

.

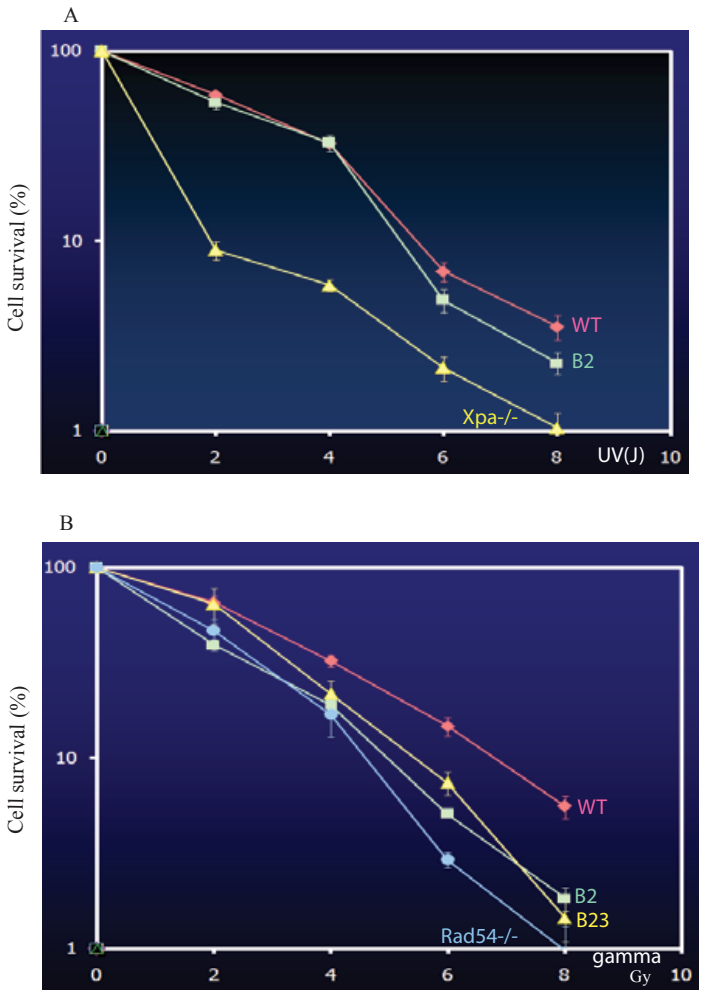


Figure 6

Effect of UV-light and ionizing radiation on survival of Delangin- deficient ES cells
Clonogenic survival of ES cells: WT (Wild Type), Delangin-deficient clones (B2 clone and B23 clone), Xpa-/- and Rad54 -/- treated with:
(A) UV-light
(B) Ionizing radiation
After treatment with increasing doses of DNA-damaging agents, cells were grown for 7 days, fixed, stained and colonies were counted. The ability to form colonies is plotted as a function of increasing dose of DNA damage. Error bars represent standard error of the mean (SEM). Genotypes of analyzed cell lines are indicated.

MATERIAL AND METHODS

Generation of Delangin heterozygous mice

Delangin-deficient mice were generated by homologous recombination and genotypes were determined by Southern and PCR-analysis. A neomycine resistance gene flanked by LoxP sites linked to a GFP reporter gene was introduced into 10th exon of the murine Delangin gene. The Delangin recombination vector, flanked by two homologous sequences of 2.2kb on the 5' end and 2.3kb on the 3'end, was constructed in pBluescript SK+. The targeting vector was linearized with *Hind*III sites and transfected by electroporation into to IB10 cells. Homologous recombinants were screened by Southern blot hybridization using genomic DNA that was digested with *Sac*I restriction enzyme giving two different fragments: 10.1 kb for WT allele and 7.3kb for targeted allele. The probe is situated at the 3' end of the construct. Three targeted ES clones with the correct karyotype were injected into blastocyst Male chimeras were crossed with FvB females for generation of mice that were heterozygous for delangin. The screening of offspring was done with Southern blot analysis using the same probe for generation of ES targeted clones. For deletion of PMC-Neo cassette, heterozygous Delangin mice were bred with CAG-Cre mice that express the Cre recombinase under the control of the cytomegalovirus early enhancer chicken β-hybrid (CAG) promoter (Sakai and Miyazaki, 1997).

In order to determine the genotype following primers were used for PCR reaction designed to amplify GFP fragment in targeted allele:

F 5': TTTTCTGCAGGTGAGCAAGGGCGAGGAGCTG 3' and

R 5': TTTTCCCGGGTTACTTGTACAGCTCGTCCATG 3'

Isolation of the embryos

Timed pregnancies were obtained by crossing heterozygous *Delangin* mice. The plugged females were sacrificed at a specific time points using the day of plugging as 0.5E. Embryos were isolated and genotyped by PCR.

Cells and cell culture

Mouse ES cells used were wt (IB10) and Delangin heterozygous ES cells. The cells were cultured on gelatinized dishes in a 1:1 mixture of DMEM and BRL (Buffalo Rat Liver) conditioned medium, supplemented with 10% fetal calf serum, 0.1M non-essential amino acids, 100U/ml penicillin/streptomycin, 50 μM β-mercaptoethanol

and 500 U/ml leukemia inhibitory factor.

Delangin heterozygous and Wild type MEFs were prepared from embryonic day 13.5 (E13.5). Single torsos were minced, and cells dispersed by rotation for 1 hr in 0.1% trypsin-EDTA solution at 37°C. Cells were washed once in phosphate-buffered saline (PBS), taken up in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal calf serum (FCS). The explants that were observed to grow out from the tissue pieces were passaged until the dish is full and then frozen.

B- and T- cells were derived from mouse peripheral blood from WT and Delangin-deficient littermates. Blood was first lysed in Gey's lysis solution for removal of erythrocytes and then cultured for 7 days in IMDM medium supplemented with 4 mM glutamine, 10% Fetal Calf Serum, 5 10^{-5} M β-mercaptoethanol and penicillin/streptomycin .

B-cells were stimulated with LPS (5μg/ml final concentration) for 5 days without splitting. After splitting no LPS was added, cells were counted, lysed and spotted on the slides for metaphase preparation. T-cells were cultured in complete IMDM medium on anti-CD3 and CD-28 coated plates for 3 days. After one day culturing without stimulation, they were split and cultured with IL-2 (5μg/ml final concentration).

Metaphase spreads

Metaphase spreads of ES and MEFs were prepared by adding Colcemid (Gibco Karyomax, 10μg/ml) to the single cell suspension in a final concentration of 0.04 μg ml-1 for 30 min at 37°C. After hypotonic treatment (0.075 M KCl) for 10 min, the pellet was fixed in Carnoy's Fix (3:1 methanol/acetic acid). Cells were then dropped onto Superfrost Plus glass slides, air-dried and kept at -20°C.

For B-cells, 2 μl/ml of Colcemid (Gibco Karyomax, 10μg/ml) was added one day before ending incubation, followed by 5 μl/ml for 1hour at 37°C. After hypotonic treatment (0.075M KCl) for 10 min., the pellet was fixed in Carnoy's Fix (3:1 methanol/acetic acid). The cells were dropped onto moist, chilled Superfrost Plus glass slides. DNA was counterstained with 100ng/ml of DAPI and mounted in Vectashield.

Colony survival curves

The sensitivity of ES cells to increasing doses of DNA-damaging agents was determined by measuring their colony-forming ability. The cells were trypsinized, counted and plated onto gelatinized 6 cm dishes. After 16 hours, cells were irradiated with a single dose in the range 0-12 Gy using a ^{137}Cs source or UV (0-8J). Cells were grown for 6 days, fixed, stained and colonies were counted. All the experiments were done in triplicate.

Histopathology analysis

The mice were euthanized with O_2 and CO_2 and the body weights were measured. The perfusion was done with 10% formalin in PBS. They are post-fixed with the same solution for 48hours at 4°C. The necropsy is completed the organs weight were measured. The organs were embedded in paraffin and were serially sectioned at 6μm thickness, stained with Hematoxylin and Eosin and examined microscopically.

DISCUSSION

In eukaryotic cells, replicated DNA molecules remain physically connected from the moment of synthesis in S-phase until their separation during anaphase. The Cohesin complexes are essential for the proper separation of the duplicated genome. Cohesin function is also important in other biological processes, like DNA repair and transcriptional regulation. In each, Cohesin is required but not sufficient for proper function. It is therefore important to understand the roles of the additional factors that are necessary for proper Cohesin function.

Mutations in different Cohesin proteins (Delangin, SMC1A and SMC3) have been found in the patients with the Cornelia de Lange Syndrome causing a broad spectrum of prenatal and postnatal abnormalities (Krantz et al., 2004; Musio et al., 2006; Tonkin et al., 2004). The complexity of the phenotype in CdLS patients suggests the involvement of Cohesin and Delangin in early developmental processes either through chromosomal dynamics or through alternative functions in DNA damage repair or transcriptional regulation. It is very well possible that a combination of all these processes is involved in the observed phenotypes.

In this work, we have generated *Delangin*-deficient mice that manifest a spectrum of developmental abnormalities, similar to those seen in CdLS patients. In CdLS, mutations in the *Delangin* gene, even in a presence of a wild-copy of the gene, triggers a haploinsuffiency in the Cohesin pathway. Mutations in the *Cohesin* pathway cause severe developmental defects in different model systems. For example, in *Drosophila*, mutations in *Cohesin* trigger a decrease in the levels of the ecdysone B receptor that effects neuronal development (Schuldiner et al., 2008).

The phenotype of the affected (heterozygous) mice resembles to some of the abnormalities seen in human patients. The most evident are pre- and postnatal growth retardation, heart abnormalities, gastro-intestinal problems and hearing loss. The abnormal skeletal patterning, a common defect occurring in severely affected human patients, was not seen in Delangin-deficient mice.

One of the possibilities that may explain the lack of a defect in skeletal patterning in Delangin-deficient mice is that in some patients mutations cause the production of a truncated protein that displays a dominant-negative effect on the WT allele

resulting in a more severe phenotype. It can, of course, not be excluded that the limb patterning pathways are affected in a different ways in mouse and humans. More importantly, there is a significant overlap in the phenotypic characteristics between Delangin-deficient mouse model and the CdLS patients. The complex and broad spectrum of symptoms observed in the mice has an incomplete penetrance and variable expression which is in agreement with features seen in human patients. In the CdLS patients symptoms vary from mild with no obvious defects, through moderate ones where growth and mental health is affected, to very prominent cases with a combination of limb truncation, severe mental and growth retardation.

The behavior of the Delangin-deficient mice is also impaired. Although we did not perform extensive analysis it was obvious that the mice have difficulties in coping with new environments. Overall, our results suggest that Delangin-deficient mice can be used as a valuable model for investigating the mechanisms underlying Cornelia de Lange Syndrome.

Even though the understanding of the function of Delangin in higher organisms is improving, the molecular mechanism through which Delangin acts is still not clear. In mammalian cells, the function of most Cohesin factors, including Cohesin and Delangin were investigated by RNA interference. However, the most desirable alternative system to study gene function and pathogenesis of the syndromes in mammalian organism is undoubtedly KO mouse. To date, the only mouse models that were generated are null mutants of the meiosos specific subunits Rec8 and SMC1B, which are both sterile (Revenkova et al., 2004; Xu et al., 2005). With respect to Cohesin interacting factors, PDS5B-/- mice are reported to have overlap with congenital anomalies also found in CdLS patients (Zhang et al., 2007).

The analysis of chromosome segregation in the different types of cells coming from Delangin-deficient mice did not show obvious abnormalities in sister chromatid cohesion. This result is consistent with reports of the lack of sister chromatid cohesion defect in cells from CdLS patients (Tonkin et al., 2004; Vrouwe et al., 2007) or only mild cohesion defect in 40% of CdLS patients (Kaur et al., 2005). In our experiment, both isoforms of the Delangin (A and B) are inactivated so the possibility of the redundancy with the alternative isoform is excluded. One of the possibilities is that Premature Sister Chromatid Separation (PSCS) occurs in particular cells that we have not analyzed even though we included several types of the cells: ES cells, Mouse Embryonic Fibroblasts (MEFs) from embryonic tissues and B- and T-cells from adult tissues. The lack of PSCS in our mouse model favors a hypothesis that defects present in CdLS patients are related to Cohesin-mediated functions other than sister chromatid cohesion. We also did not detect aneuploidy, often seen as a consequence of premature chromatid segregation. Aneuploidy often leads to tumor formation in organisms with this type of chromosome instability. The incidence of a tumor formation in the Delangin-deficient mice was very low, (to date only in one animal). This is in agreement with the infrequent occurrence of tumors in CdLS individuals, only four cases being described in the literature (DuVall and Walden, 1996; Maruiwa et al., 1988; Sugita et al., 1986).

The Cohesin complex is involved in DNA double-strand break (DSB) repair by

homologous recombination during the S-and G2 phase of the cell cycle (Sjogren and Nasmyth, 2001). Inactivation of SCC2 (Delangin) or one of the Cohesin subunits results in a reduced efficiency of post-replicative DSB repair in G2 /M. In the studies presented by Strom et al. (Strom et al., 2004) and Unal et al. (Unal et al., 2004) was demonstrated that the local enrichment of Cohesin depends on the Delangin (SCC2/SCC4) complex. More recent studies in cells from CdLS patients showed increased DNA damage sensitivity (Vrouwe et al., 2007). The implication of Delangin and Cohesin homologues in the DNA damage responses raises the question whether the cells derived from the Delangin-deficient mice have a similar defect and such a defect is indeed found.

Nipped-B, the Delangin homologue in *Drosophila*, participates in mitotic sister chromatid cohesion and regulates the transcriptional control of cut and *ultrabithorax* genes, which are involved in development (Dorsett et al., 2005; Rollins et al., 2004; Rollins et al., 1999). Since Delangin acts through Cohesin, the mutations that affect Delangin are reflected on Cohesion targets. Misregulation of a single gene can have severe downstream effect especially during development. The expression of a number of genes crucial for development, such as homeotic genes, could be very sensitive to the presence of Cohesin around the promoters. Recent work strongly suggests that Cohesin is involved in gene regulation during developmental processes in higher organisms. For instance, Cohesin enhances gene expression in the case of the zebrafish transcriptional factor *runx* involved in cell fate determination of hematopoietic and other lineages (Horsfield et al., 2007). If we think about the possibility that Delangin mutation causes inappropriate loading or positioning of Cohesin on its target, in this case *runx* gene, the consequence is deregulation of *runx* gene expression and *runx* downstream targets.

In another scenario, chromatin organization can also be affected. Studies in human and mouse cells have shown that Cohesin accumulates at sites bound by CTCF, a protein with insulator activity that is important for the communication between enhancers and promoters (Parelho et al., 2008; Stedman et al., 2008; Wendt et al., 2008). Although functional studies showing this directly are missing, it may be possible that Cohesin assists in the process of holding the chromatin loops together and attracting particular sequences to affect gene regulation. In that case, and assuming that *Delangin* mutations affect Cohesin dynamics of regulatory sequences that organize chromatin domains, it would cause abnormal regulation of different sets of the genes. Nevertheless the observation that mutations in SMC1 and SMC3 are associated with milder abnormalities in comparison to *Delangin* mutations suggests that it may have additional roles to those discussed above.

REFERENCES

Bhuiyan, Z. A., Klein, M., Hammond, P., van Haeringen, A., Mannens, M. M., Van Berckelaer-Onnes, I. and Hennekam, R. C. (2006). Genotype-phenotype correlations of 39 patients with Cornelia De Lange syndrome: the Dutch experience. J Med Genet 43, 568-75.

Borck, G., Zarhrate, M., Bonnefont, J. P., Munnich, A., Cormier-Daire, V. and Colleaux, L. (2007). Incidence and clinical features of X-linked Cornelia de Lange syndrome due to SMC1L1 mutations. Hum Mutat 28, 205-6.

Ciosk, R., Shirayama, M., Shevchenko, A., Tanaka, T., Toth, A., Shevchenko, A. and Nasmyth, K. (2000). Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. Mol Cell 5, 243-54.

Deardorff, M. A., Kaur, M., Yaeger, D., Rampuria, A., Korolev, S., Pie, J., Gil-Rodriguez, C., Arnedo, M., Loeys, B., Kline, A. D. et al. (2007). Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of cornelia de Lange syndrome with predominant mental retardation. Am J Hum Genet 80, 485-94.

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

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

Dorsett, D., Eissenberg, J. C., Misulovin, Z., Martens, A., Redding, B. and McKim, K. (2005). Effects of sister chromatid cohesion proteins on cut gene expression during wing development in Drosophila. Development 132, 4743-53.

DuVall, G. A. and Walden, D. T. (1996). Adenocarcinoma of the esophagus complicating Cornelia de Lange syndrome. J Clin Gastroenterol 22, 131-3.

Essers, J., Hendriks, R. W., Swagemakers, S. M., Troelstra, C., de Wit, J., Bootsma, D., Hoeijmakers, J. H. and Kanaar, R. (1997). Disruption of mouse RAD54 reduces ionizing radiation resistance and homologous recombination. Cell 89, 195-204.

Gillis, L. A., McCallum, J., Kaur, M., DeScipio, C., Yaeger, D., Mariani, A., Kline, A. D., Li, H. H., Devoto, M., Jackson, L. G. et al. (2004). NIPBL mutational analysis in 120 individuals with Cornelia de Lange syndrome and evaluation of genotype-phenotype correlations. Am J Hum Genet 75, 610-23.

Gruber, S., Arumugam, P., Katou, Y., Kuglitsch, D., Helmhart, W., Shirahige, K. and Nasmyth, K. (2006). Evidence that loading of cohesin onto chromosomes involves opening of its SMC hinge. Cell 127, 523-37.

Hartman, T., Stead, K., Koshland, D. and Guacci, V. (2000). Pds5p is an essential chromosomal protein required for both sister chromatid cohesion and condensation in Saccharomyces cerevisiae. J Cell Biol 151, 613-26.

Hirano, M. and Hirano, T. (2006). Opening closed arms: long-distance activation of SMC ATPase by hinge-DNA interactions. Mol Cell 21, 175-86.

Hirano, T. (2006). At the heart of the chromosome: SMC proteins in action. Nat Rev Mol Cell Biol 7, 311-22.

Horsfield, J. A., Anagnostou, S. H., Hu, J. K., Cho, K. H., Geisler, R., Lieschke, G., Crosier, K. E. and Crosier, P. S. (2007). Cohesin-dependent regulation of Runx genes. Development 134, 2639-49.

Ireland, M., Donnai, D. and Burn, J. (1993). Brachmann-de Lange syndrome. Delineation of the clinical phenotype. Am J Med Genet 47, 959-64.

Jackson, L., Kline, A. D., Barr, M. A. and Koch, S. (1993). de Lange syndrome: a clinical review of 310 individuals. Am J Med Genet 47, 940-6.

Kaur, M., DeScipio, C., McCallum, J., Yaeger, D., Devoto, M., Jackson, L. G., Spinner, N. B. and Krantz, I. D. (2005). Precocious sister chromatid separation (PSCS) in Cornelia de Lange syndrome. Am J Med Genet A 138, 27-31.

Krantz, I. D., McCallum, J., DeScipio, C., Kaur, M., Gillis, L. A., Yaeger, D., Jukofsky, L., Wasserman,

N., Bottani, A., Morris, C. A. et al. (2004). Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B. Nat Genet 36, 631-5.

Lechner, M. S., Schultz, D. C., Negorev, D., Maul, G. G. and Rauscher, F. J., 3rd. (2005). The mammalian heterochromatin protein 1 binds diverse nuclear proteins through a common motif that targets the chromoshadow domain. Biochem Biophys Res Commun 331, 929-37.

Losada, A. (2007). Cohesin regulation: fashionable ways to wear a ring. Chromosoma 116, 321-9.

Maruiwa, M., Nakamura, Y., Motomura, K., Murakami, T., Kojiro, M., Kato, M., Morimatsu, M., Fukuda, S. and Hashimoto, T. (1988). Cornelia de Lange syndrome associated with Wilms' tumour and infantile haemangioendothelioma of the liver: report of two autopsy cases. Virchows Arch A Pathol Anat Histopathol 413, 463-8.

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

Musio, A., Selicorni, A., Focarelli, M. L., Gervasini, C., Milani, D., Russo, S., Vezzoni, P. and Larizza, L. (2006). X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. Nat Genet 38, 528-30.

Nasmyth, K. and Haering, C. H. (2005). The structure and function of SMC and kleisin complexes. Annu Rev Biochem 74, 595-648.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H. C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T. et al. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell 132, 422-33.

Pauli, A., Althoff, F., Oliveira, R. A., Heidmann, S., Schuldiner, O., Lehner, C. F., Dickson, B. J. and Nasmyth, K. (2008). Cell-type-specific TEV protease cleavage reveals cohesin functions in Drosophila neurons. Dev Cell 14, 239-51.

Revenkova, E., Eijpe, M., Heyting, C., Hodges, C. A., Hunt, P. A., Liebe, B., Scherthan, H. and Jessberger, R. (2004). Cohesin SMC1 beta is required for meiotic chromosome dynamics, sister chromatid cohesion and DNA recombination. Nat Cell Biol 6, 555-62.

Revenkova, E. and Jessberger, R. (2006). Shaping meiotic prophase chromosomes: cohesins and synaptonemal complex proteins. Chromosoma 115, 235-40.

Rollins, R. A., Korom, M., Aulner, N., Martens, A. and Dorsett, D. (2004). Drosophila nipped-B protein supports sister chromatid cohesion and opposes the stromalin/Scc3 cohesion factor to facilitate long-range activation of the cut gene. Mol Cell Biol 24, 3100-11.

Rollins, R. A., Morcillo, P. and Dorsett, D. (1999). Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. Genetics 152, 577-93.

Russell, K. L., Ming, J. E., Patel, K., Jukofsky, L., Magnusson, M. and Krantz, I. D. (2001). Dominant paternal transmission of Cornelia de Lange syndrome: a new case and review of 25 previously reported familial recurrences. Am J Med Genet 104, 267-76.

Sakai, K. and Miyazaki, J. (1997). A transgenic mouse line that retains Cre recombinase activity in mature oocytes irrespective of the cre transgene transmission. Biochem Biophys Res Commun 237, 318-24.

Schuldiner, O., Berdnik, D., Levy, J. M., Wu, J. S., Luginbuhl, D., Gontang, A. C. and Luo, L. (2008). piggyBac-based mosaic screen identifies a postmitotic function for cohesin in regulating developmental axon pruning. Dev Cell 14, 227-38.

Seitz, L. C., Tang, K., Cummings, W. J. and Zolan, M. E. (1996). The rad9 gene of Coprinus cinereus encodes a proline-rich protein required for meiotic chromosome condensation and synapsis. Genetics 142, 1105-17.

Sjogren, C. and Nasmyth, K. (2001). Sister chromatid cohesion is required for postreplicative double-strand break repair in Saccharomyces cerevisiae. Curr Biol 11, 991-5.

Stedman, W., Kang, H., Lin, S., Kissil, J. L., Bartolomei, M. S. and Lieberman, P. M. (2008). Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. Embo J 27, 654-66.

Strom, L., Karlsson, C., Lindroos, H. B., Wedahl, S., Katou, Y., Shirahige, K. and Sjogren, C. (2007). Postreplicative formation of cohesion is required for repair and induced by a single DNA break. Science 317, 242-5.

Strom, L., Lindroos, H. B., Shirahige, K. and Sjogren, C. (2004). Postreplicative recruitment of cohesin to double-strand breaks is required for DNA repair. Mol Cell 16, 1003-15.

Sugita, K., Izumi, T., Yamaguchi, K., Fukuyama, Y., Sato, A. and Kajita, A. (1986). Cornelia de Lange syndrome associated with a suprasellar germinoma. Brain Dev 8, 541-6.

Thiru, A., Nietlispach, D., Mott, H. R., Okuwaki, M., Lyon, D., Nielsen, P. R., Hirshberg, M., Verreault, A., Murzina, N. V. and Laue, E. D. (2004). Structural basis of HP1/PXVXL motif peptide interactions and HP1 localisation to heterochromatin. Embo J 23, 489-99.

Tomonaga, T., Nagao, K., Kawasaki, Y., Furuya, K., Murakami, A., Morishita, J., Yuasa, T., Sutani, T., Kearsey, S. E., Uhlmann, F. et al. (2000). Characterization of fission yeast cohesin: essential anaphase proteolysis of Rad21 phosphorylated in the S phase. Genes Dev 14, 2757-70.

Tonkin, E. T., Wang, T. J., Lisgo, S., Bamshad, M. J. and Strachan, T. (2004). NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. Nat Genet 36, 636-41.

Toyoda, Y., Furuya, K., Goshima, G., Nagao, K., Takahashi, K. and Yanagida, M. (2002). Requirement of chromatid cohesion proteins rad21/scc1 and mis4/scc2 for normal spindle-kinetochore interaction in fission yeast. Curr Biol 12, 347-58.

Unal, E., Arbel-Eden, A., Sattler, U., Shroff, R., Lichten, M., Haber, J. E. and Koshland, D. (2004). DNA damage response pathway uses histone modification to assemble a double-strand break-specific cohesin domain. Mol Cell 16, 991-1002.

van Haaften, G., Romeijn, R., Pothof, J., Koole, W., Mullenders, L. H., Pastink, A., Plasterk, R. H. and Tijsterman, M. (2006). Identification of conserved pathways of DNA-damage response and radiation protection by genome-wide RNAi. Curr Biol 16, 1344-50.

Vrouwe, M. G., Elghalbzouri-Maghrani, E., Meijers, M., Schouten, P., Godthelp, B. C., Bhuiyan, Z. A., Redeker, E. J., Mannens, M. M., Mullenders, L. H., Pastink, A. et al. (2007). Increased DNA damage sensitivity of Cornelia de Lange syndrome cells: evidence for impaired recombinational repair. Hum Mol Genet 16, 1478-87.

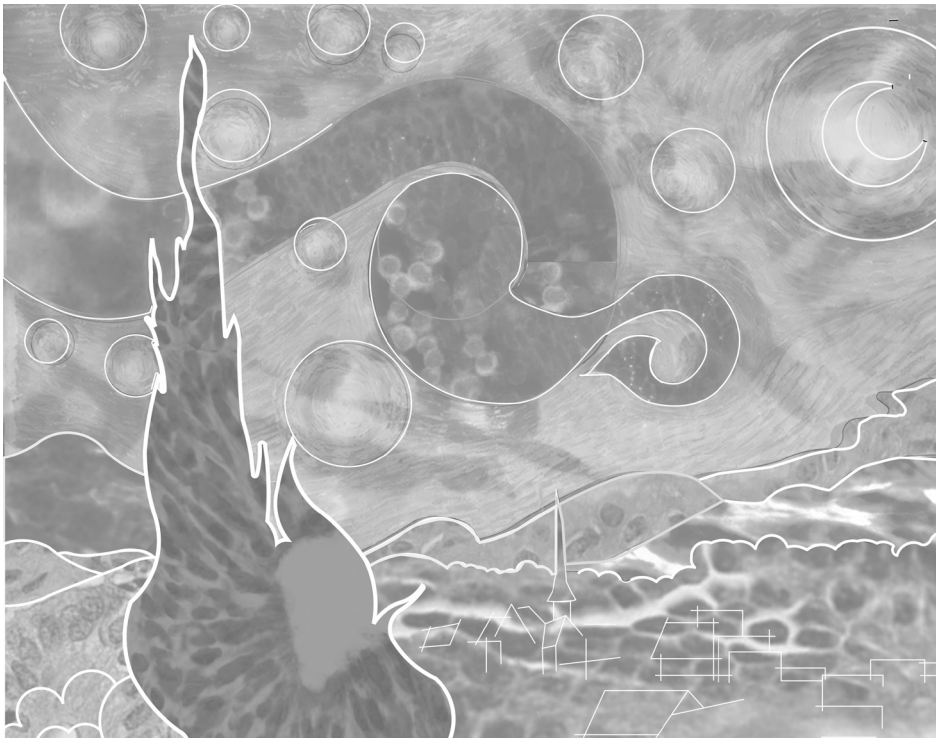
Watrin, E., Schleiffer, A., Tanaka, K., Eisenhaber, F., Nasmyth, K. and Peters, J. M. (2006). Human Scc4 is required for cohesin binding to chromatin, sister-chromatid cohesion, and mitotic progression. Curr Biol 16, 863-74.

Weitzer, S., Lehane, C. and Uhlmann, F. (2003). A model for ATP hydrolysis-dependent binding of cohesin to DNA. Curr Biol 13, 1930-40.

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

Xu, H., Beasley, M. D., Warren, W. D., van der Horst, G. T. and McKay, M. J. (2005). Absence of mouse REC8 cohesin promotes synapsis of sister chromatids in meiosis. Dev Cell 8, 949-61.

Zhang, B., Jain, S., Song, H., Fu, M., Heuckeroth, R. O., Erlich, J. M., Jay, P. Y. and Milbrandt, J. (2007). Mice lacking sister chromatid cohesion protein PDS5B exhibit developmental abnormalities reminiscent of Cornelia de Lange syndrome. Development 134, 3191-201.



Chapter 4 Interactions in β-globin locus: the role of Cohesin

Sanja Krpic, Erik Splinter, Thamar Van Dijk, Wilfred van IJcken, Erik-Jan Rijkers, Christel Kockx and Frank Grosveld

Manuscript in preparation

.

INTRODUCTION:

The genome appears to be organized in chromatin loops that demarcate "domains". Such an organization allows efficient packaging and interactions in the constricted space of the nucleus. It would also separate regulatory elements in *cis*, although recent data indicate that there is extensive interaction between very distant regions in *cis* and even in *trans*. Nevertheless an organization in loops would favor interactions between *cis*-regulatory elements on the same loop and diminishing the influence of more distant elements on other loops thereby maintaining proper expression domains and patterns in a lineage specific manner. This task is accomplished by chromatin elements that demarcate expression domains (Dillon and Sabbattini, 2000) where functional gene expression domains define the functional unit of eukaryotic gene regulation. These elements are often referred to as insulators due to their ability to block enhancer or silencer signals when physically positioned between the *cis*-regulatory element and the promoter of the gene (Bell et al., 2001; Ohlsson et al., 2001).

Insulators are best exemplified by the *gypsy* element in *Drosophila* (Capelson and Corces, 2005). The *Gypsy* insulators are occupied in a sequence specific manner by SUHW and Mod (Mdg4) which recruit other factors (Pai et al., 2004). Multiple *gypsy* sites and associated proteins cluster together to form "insulator bodies" with the effect of organizing nearby chromatin into loop domains (Capelson and Corces, 2006). In vertebrates, a similar function is associated with the chromatin binding protein CTCF (CCCTC-binding factor). It is the prototype vertebrate protein exhibiting insulator activity as an enhancer blocker or as a barrier against inactivation from nearby heterochromatin in transfection assays (Bell and Felsenfeld, 2000; Bell et al., 1999; Recillas-Targa et al., 2002; West et al., 2002). *In vivo*, CTCF appears to be involved in the formation of chromatin loops (Burke et al., 2005; Splinter et al., 2006; Tolhuis et al., 2002). It is very stably bound to the chromatin (Galjart et al personal comm.,) and appears to be essential for the maintenance of chromatin structure. This property is used to achieve allele specific gene expression at imprinted loci. CTCF binds to the imprinting control region (ICR) of the H19/insulin-like growth factor (Igf2) locus on the maternally inherited allele. On the paternal copy the ICR is methylated preventing CTCF binding resulting in the formation of a different loop (Szabo et al., 2004), which is in effect abolishing its insulator function (Bell and Felsenfeld, 2000).

The mouse β-globin locus is situated on chromosome 7 and contains four functional genes: εγ, β_{h1}, β_{maj} and β_{min} (from 5'to 3'). During primitive eryhropoiesis, yolk sac derived erythroid cells express mostly εy and β_{h1}, while the expression of β_{maj} and β_{min} is low. In the murine β-globin locus, three CTCF-binding sites have been identified upstream (HS-85, HS-62 and HS5) and one downstream (3'HS1)

of the β-globin locus (Bulger et al., 2003; Farrell et al., 2002; Splinter et al., 2006). All the CTCF sites are in contact with each other (Palstra et al., 2003; Splinter et al., 2006) maintaining a compact domain structure that is dependent on CTCF. Depletion of CTCF, or the disruption one of its DNA-binding site destabilize these long-range interactions (Splinter et al., 2006).

Genome-wide approaches have resulted in the identification most or all the CTCF binding sites in human genome. The sites correlate strongly with regions containing genes in line with an important role of CTCF in gene regulation in *cis* (Kim et al., 2007) and perhaps even in trans (Ling et al., 2006).

Other proteins known to "organize" DNA are the Cohesins. The Cohesin complex is best known for its role in holding sister chromatids together after eukaryotic DNA replication. This allows the spindle to recognize pairs of replication products for segregation into opposite directions during mitosis. The complex is conserved from yeast to humans and is composed of two SMC (Structural Maintenance of Chromosomes) subunits SMC1 and SMC3, and two non-SMC subunits, Scc1/Rad21 and Scc3, which together form a ring-shape complex that is essential for cohesion between sister chromatids (Gruber et al., 2003).

Cohesin is initially loaded onto chromosomes at separate places along the genome by Scc2 (Lengronne et al., 2004) and its orthologs (Gillespie and Hirano, 2004; Seitan et al., 2006). The genomic location of Cohesin have been studied in several species by chromatin-immunoprecipitation followed by hybridization to DNA microarrays (ChIP-on-chip). In yeast, Cohesin binds approximately every 10 kbp and is localized almost exclusively between genes that are transcribed in opposite directions. Changes in transcription lead to repositioning of Cohesin (Lengronne et al., 2004). In contrast, ChIP-on-chip data from *D.melanogaster* cell lines indicate that Cohesin has a clear preference for active over inactive genes, with Cohesin binding occurring over active transcriptional units as well as untranscribed regions (Misulovin et al., 2008). Cohesin binds between a remote wing margin enhancer and the promoter at the *cut* locus. Although it is not known how Cohesin interacts with DNA, reduction of the SMC1 level increases *cut* expression in the developing wing margin (Dorsett et al., 2005). Recently, four groups mapped numerous Cohesin-binding sites in mammalian chromosomes and found substantial overlap with CTCF (Parelho et al., 2008; Rubio et al., 2008; Stedman et al., 2008; Wendt et al., 2008). After important finding of overlapping binding sites of CTCF and Cohesin, one of the key questions that follow is if there is functional significance of shared binding sites with respect to gene regulation.

Here we show that subsets of the CTCF binding sites are shared with Cohesin in the mouse genome. We have focused specifically on CTCF and Cohesin binding sites in the mouse β-globin locus and surrounding sequences in erythroid cells. Previously, it was shown that CTCF function in the formation of chromatin loops using two independent lines of evidence. The removal of most CTCF protein, as well as targeted disruption of a CTCF-binding site, resulted in destabilization of long-range contacts between cognate binding sites in the β-globin locus (Splinter et al., 2006). Since Cohesin and CTCF share binding sites in β-globin locus we were interested

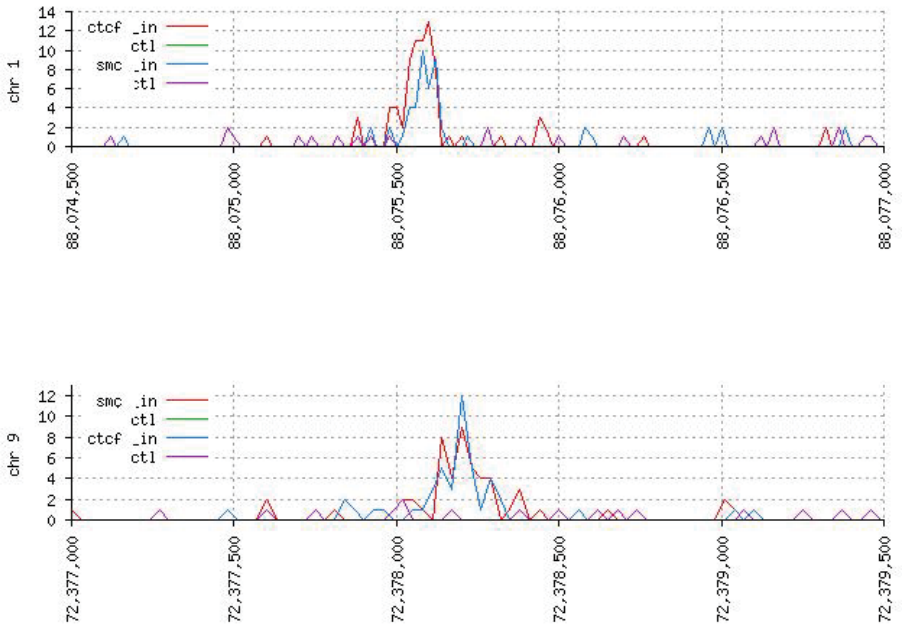
to know whether Cohesin is required for the long-range interactions in this locus at the shared CTCF/Cohesin binding sites. We show that Cohesin binding is indeed required in a CTCF dependent manner.

RESULTS

Identification of Cohesin and CTCF binding sites by ChIP-on chip

Cohesin and CTCF binding sites were identified by Chromatin Immunoprecipitation (ChIP) in mouse I/11 cells. I/11 cells are erythroid progenitor cells established from fetal livers of p53-/- mice. The cells can be expanded indefinitely *in vitro* and after exposure to physiologically relevant stimuli such as erythropoietin, they undergo terminal differentiation into enucleated erythrocytes in a synchronized manner (von Lindern et al., 2001). The choice of I/11 cells was based on their ability to faithfully recapitulate the erythroid differentiation program *in vivo*. Analysis of this erythroid line show that upon induction (maturation) β-globin gene expression increases to reach the high transcription level *in vivo* (Dolznig et al., 2005). Furthermore, increased expression rates of β-major in I/11 cells coincides well with increased LCR-β-major interaction frequencies (Kooren et al., 2007) supporting this line to be a good model to study the conformational dynamics in the β-globin locus. For ChIP analysis, we initially used oligonucleotide tiling arrays for mouse chromosomes 3, 7 and 13 (an Affymetrix set of arrays) since we are particularly interested in binding sites in mouse β-globin locus on chromosome 7.

We performed a ChIP-on-chip analysis on non-induced and induced I/11 cells using an SMC1 antibody, to bring down the Cohesin complex, and a CTCF antibody, to bring down CTCF. ChIP-on-chip analysis revealed a nonrandom distribution of SMC1 (Cohesin) and CTCF binding sites across 3 mouse chromosomes. Detailed analysis showed as expected that there is an overlap of CTCF and SMC1 binding sites. Since we are particularly interested in the β-globin gene and surrounding sequences, we focused on binding sites on chromosome 7 (110,000,000-112,000,000 coordinates build Build 37.1 assembly by NCBI). For visualization of the data we have used IGB (Integrated Genome Browser) from Affymetrix (data not shown) or UCSC genome browser.



Supplementary Figure 1

The graph shows a typical example of Solexa-sequencing analysis of the overlap of 2 randomly chosen CTCF and SMC3 (Cohesin) binding sites with relevant controls on the chromosome 1 and chromosome 9. The data are presented like profile tracks in UCSC genome browser. The SMC and its control track are shown in blue and purple, the CTCF track and its control in red and green. The horizontal axis shows the coordinates on chromosomes 1 and 9 (Mouse Build 37.1). The vertical axis shows the number of overlapping sequences scored at a particular position.

Identification of Cohesin and CTCF binding genome-wide by Solexa sequencing

The ChIP-on-chip experiments were not pursued further since this method has several limitations including the large sets of arrays needed to cover the mammalian genome, its potential bias introduced by DNA amplification (Bernstein et al., 2007) and its limited quantitative range. The technique gives many false positive and/or negative results. For example, we screened more than 25 sites that appeared to be positive for CTCF and/or Cohesin by conventional qPCR and only 2 showed enrichment. To overcome this problem we switched to Chip-sequencing using a Solexa 1G Genome Analyzer. It performs massive parallel sequencing to directly sequence the ends of ChIP-DNA after sonication. It is clearly a more comprehensive, quantitative and reproducible method to analyze protein target sites in large mammalian genomes.

The identification of Cohesin and CTCF binding sites by ChIP-sequencing was done as described above for the ChIP-on-chip analysis. The quality of the chromatin precipitation was confirmed by using real-time quantitative PCR analysis of known target sites before further analysis using the Solexa 1G Genome Analyzer. The sequencing procedure requires a one-step adaptor ligation and limited PCR amplification (18 cycles) of ChIP DNA molecules followed by cluster generation and sequencing-by-synthesis (Described in Materials and Methods). One sequencing run produced between 2-6 million sequence tags of 36bp each. Since the ChIP-Sequencing is analogous to direct counting of the molecules in the ChIP-DNA sample, it requires minimal normalization.

Again, similar to ChIP-on-chip analysis we could detect overlap between CTCF and Cohesin binding sites. In Supplementary Figure 1, an example is given of two overlapping sites situated at the chromosome 1 and 9 with relevant controls and the positions along the chromosomes (Mouse Build 37.1). The statistical analysis showing the degree of CTCF and Cohesin overlap in the mouse genome is in progress. We also identified many sites that were occupied by one or the other protein. Interestingly, there was extensive overlap of CTCF binding sites with an independent study performed on mouse ES cells using the same Chip-sequencing method (Chen et al., 2008).

In the β-globin locus we detected the previously identified CTCF binding sites. They are present at the 3'HS1, in the LCR (5'HS5/cHS4) and the 5' sites (HS-62.5 and HS-85) (Bulger et al., 2003; Farrell et al., 2002) and two new sites B1 and B2 (see below). For simplicity, only 3'HS1 and HS-85 sequence are depicted in the Figure 1A. We could detect CTCF at all 4 known binding sites while SMC3 was enriched only at 3'HS1. The validation of these binding sites is presented in Figure 1B, which shows the relative enrichments of SMC1 and SMC3, and CTCF and Rad21 at CTCF binding sites. Clearly, all 3 Cohesin subunits show enrichment only at 3'HS1 while CTCF is enriched at all 4 sites. The negative control, the β-globin promoter shows no binding for all 4 proteins. As mentioned above, we detected positive signal for SMC3 and CTCF at two novel locations in the β-globin locus that have not been reported before. Interestingly, both "new" CTCF binding sites also bind Cohesin. The position and relative enrichment for CTCF and Cohesin at these sites is shown in Figures 1A and C. Interestingly, B1 is situated within the β-globin (major) gene in the third exon. The additional site B2, was mapped 13 kilobases upstream of 3'HS1.

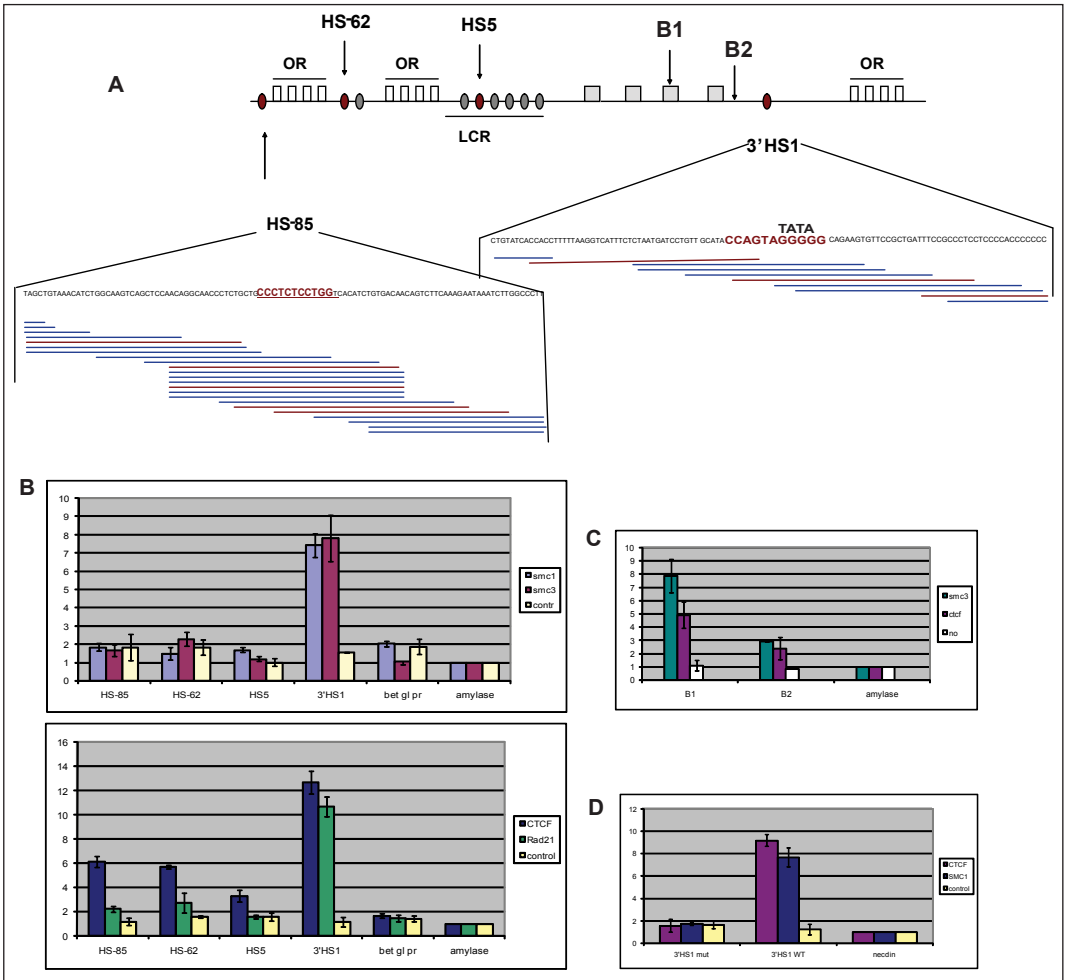


Figure 1

A _ Schematic of CTCF binding sites in β-globin locus: HS-85, HS-62, HS5 and 3'HS1.
B1 and B2 (newly identified binding sites) are also shown with their relative positions to the other sites.
At the HS-85 and 3'HS1 binding sites, the sequences identified by Solexa-sequencing are depicted by simple lines (instead of real sequences). The CTCF consensus binding motif, with surrounding sequences is also included. At the 3'HS1 binding motif, the upper sequence (TATA) represents the position of the mutated CTCF consensus sequence (Splinter, Heath et al. 2006).
B _ Chromatin immunoprecipitation (ChIP) using antibodies against: SMC1, SMC3, Rad21, CTCF and an antibody control. Plotted are relative enrichments measured for the sites in β-globin locus.
C _ Chromatin immunoprecipitation (ChIP) using antibodies against SMC3, CTCF and an Ab control. Plotted are relative enrichments measured for the sites in the β-globin locus.
D _ Chromatin immunoprecipitation (ChIP) using antibodies against CTCF, SMC3 and a non relevant Ab (control). Plotted are the relative enrichments measured for the two different alleles 3'HS Wild Type and 3'HS1 mutated.

The relative enrichments are normalized to amylase.
In all panels the standard error is indicated (from 3 independent ChIP experiments, 2 Q-PCR each).

Cohesin binding to 3'HS1 in the β-globin locus is dependent on CTCF

Since CTCF and Cohesin binding sites have a high degree of overlap, we examined whether CTCF is required for Cohesin binding at a shared site. To answer this, we used a disrupted 3'HS1 CTCF binding site of which 4 conserved nucleotides in the in the core CTCF-binding site were mutated (Splinter et al., 2006). Chromatin immunoprecipitation was performed on fetal livers from embryos carrying a mutation in the 3'HS1 site in one of its alleles while the other allele was intact. The two alleles can be distinguished by different sets of primers that specifically amplify each allele. The binding of CTCF to the mutated allele was completely abolished and interestingly, Cohesin was also no longer able to bind (Figure 1D). Both CTCF and Cohesin were enriched normally at the normal allele with the preserved CTCF binding site.

These data demonstrate that Cohesin binding at the 3'HS1 site directly depends on the presence of CTCF, suggesting that CTCF is required for the enrichment of Cohesin at its binding sites. However we were unable to detect a direct interaction between CTCF and any of the Cohesin subunits (data not shown). Possibly the interaction is weak or Cohesin and CTCF interact only transiently. Alternatively they may interact indirectly via other proteins or DNA.

The role of Cohesin in the organization of β-globin locus

To address the question whether Cohesin is required for the organization of β-globin locus we performed a set of experiments where non–induced I/11 cells were depleted for CTCF or Cohesin by lentiviral transduction with shRNA. We used several constructs for each protein that were successful in reducing the amount of the proteins and each gave a similar cell phenotype/result. We followed the dynamics of the knockdown during 5 days and we were able to deplete CTCF protein to more than 95% as shown by Western blot (Figure 2A, left panel, day 3 and 5 time points). Interestingly, the cells were still viable and progressed through the cell cycle although at a slower rate when compared to mock transfected control cells (data not shown). Cohesin was depleted up to 70% by transduction with shSMC3 RNA (Figure 2A, right panel) when the cells were still viable and dividing. Depletion of more then 70% caused massive cell death. The simultaneous depletion of both of the proteins was not successful since it gave rise to massive cell death (not shown), hence we have used incomplete knock down of Cohesin (Figure 2A, right panel).

As expected in Cohesin depleted samples Cohesin is not bound to 3'HS1 it is in (Figure 2B, left panel). In the same sample, CTCF was enriched at 3'HS1 to a level similar to that seen in the WT control suggesting that Cohesin does not influence CTCF binding to its site. In contrast, CTCF depletion disturbs Cohesin binding (Figure 2B, right panel), it is no longer bound to 3'HS1 even though the level of SMC protein is not reduced (not shown). This result agrees with the mutation result above and

shows that CTCF is necessary for proper positioning of Cohesin at its binding sites.

Since previous experiments have shown that depletion of CTCF, or the disruption of its DNA binding site destabilize specific long-range interactions (Splinter et al., 2006) we asked the question what the contribution of Cohesin is to range interactions in the β-globin locus using a 3C interaction analysis (Splinter, Heath et al. 2006). We were unfortunately not able to include the 2 newly identified CTCF/ Cohesin sites (B1 and B2) in this analysis since they are surrounded by repetitive DNA sequences that prevent a proper primer design. We analyzed the known contacts within β-globin DNA in erythroid cells lacking CTCF (not shown, Splinter, Heath et al 2006) or SMC3 (Cohesin) protein. In case of the SMC3 Knock Down CTCF is still bound to 3'HS1 (Figure 2B) and hence any measured effects will be exclusively due to Cohesin depletion. The 3C relative crosslinking frequencies of the different interacting fragments are shown in Figure 2C. All the data are corrected for HS2/HS3 interactions (Figure 2C, upper right), which are neighboring hypersensitive sites situated in the LCR of β-globin locus and which do not bind CTCF or Cohesin. Calreticulin, a non-related locus that is irrelevant for globin gene expression, behaves the same in all the samples (Figure 2C, upper left panel). Surprisingly, without Cohesin, HS-85 seems to lose its interaction with other CTCF binding sites both when tested for interactions with 3'HS1 (Figure 2C, middle left) and HS5 (Figure 2C, middle right). The remainder of the chromatin hub interactions appear not to be affected or are even slightly increased their relative crosslinking frequencies as shown in Figure 2C, (bottom left, 3'HS1-HS5 interactions or bottom right 3'HS1-HS-65). We conclude that Cohesin is essential for long range interactions at the shared CTCF sites.

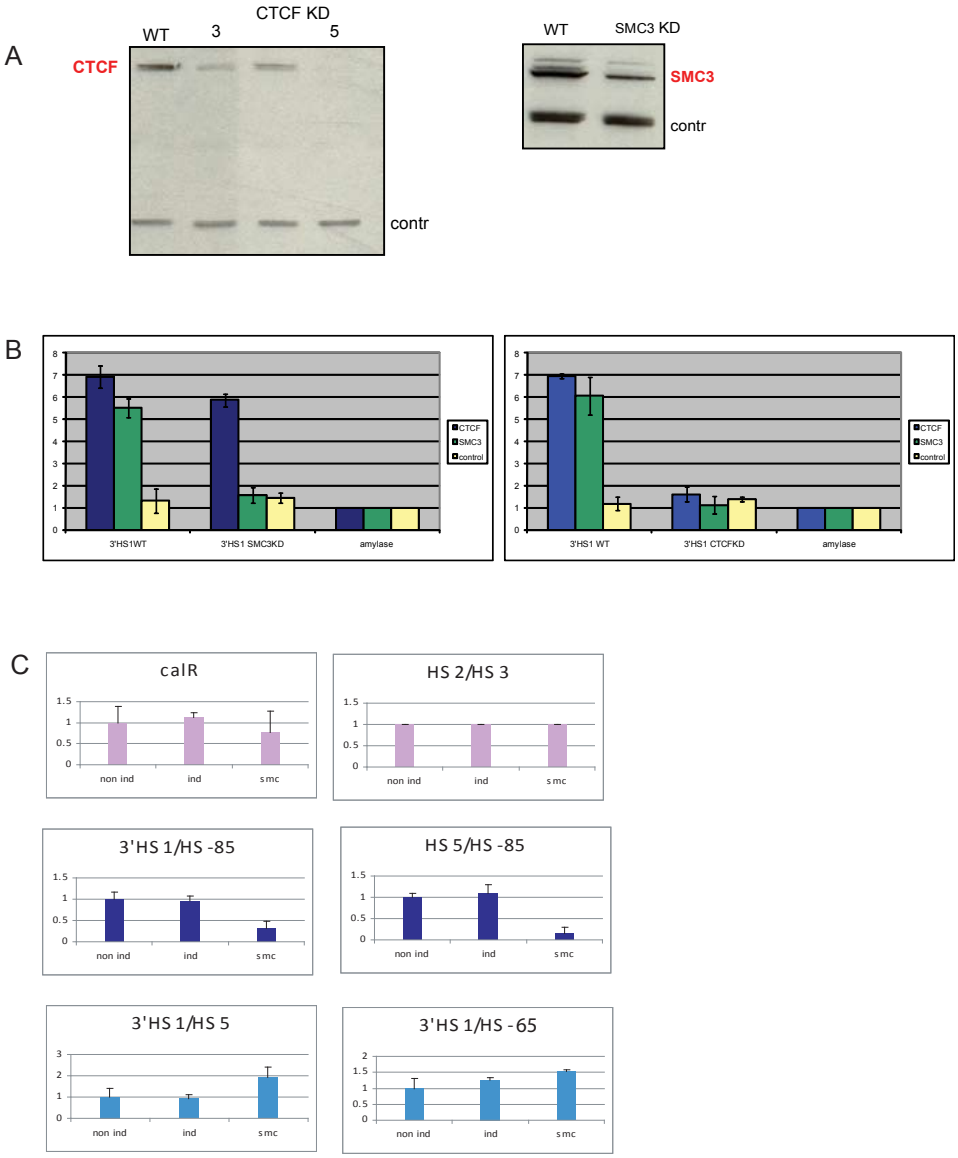


Figure 2

A _ Western blot from I/11 CTCF and SMC3 Knock Down cells. In the left panel, 2 time points of 3 (2 lines) and 5 days are shown. SMC3 KD after 5 days is shown in the right panel.
B _ Chromatin immunoprecipitation (Chip) using antibodies against SMC3, CTCF and a non relevant Ab (control). Plotted are the relative enrichments measured for the 3'HS1 site in WT cells and SMC3KD (left) and CTCFKD (right) cells. The relative enrichments are normalized to amylase. In all panels the standard error (from 2 independent ChIP experiments, 2 Q-PCR each) is indicated.
C _ 3C analysis showing crosslinking frequencies in non-induced, induced I/11 cells and SMC3 KD cells (non induced) between different CTCF binding sites. The interactions are normalized to HS2/HS3 primers, neighboring hypersensitive sites in LCR that do not bind any of these proteins. CalR-calreticulin non-related locus that is irrelevant for globin gene expression

DISCUSSION

The present study investigated the role of Cohesin in long range interactions in the β-globin locus. It was previously shown that CTCF binds a number of sites in the mouse β-globin locus and that CTCF is directly involved in the chromatin architecture of this locus (Splinter et al., 2006). One of the CTCF binding sites, 3'HS1 is also highly enriched for Cohesin. Cohesin complexes are known to mediate sister chromatid cohesion which is essential for proper chromosome separation and post-replicative DNA repair. In addition, it was recently suggested to also regulate gene expression and long-range gene interaction through its enrichment at CTCF binding sites. CTCF, the prototype vertebrate insulator protein, had already been strongly implicated in mediating gene regulation, chromatin structure and organization of gene loci before this discovery. Thus its link with Cohesin suggests that cohesion may also have a gene regulatory function.

The requirement of Cohesin for sister chromatid cohesion after DNA replication is well documented, but a function of Cohesin in gene regulation is less clear even though several studies in yeast, *Drosophila* and zebrafish have indicated that Cohesin directly influences transcription. For example, in Zebrafish, a genetic screen revealed that embryos lacking Rad21, one of the Cohesin subunits, fail to develop differentiated blood cells because of a lack of *runx3* expression (Horsfield et al., 2007).

In human cells, Cohesin sites are present in introns (35%), in intergenic regions (50%) or in the regions that are directly up-stream or down-stream from the genes (13%). Compared with an overall frequency, Cohesin sites are enriched in the regions adjacent to the genes (Wendt et al., 2008), favoring the possibility that these sites have a role in gene regulation. Since Cohesin accumulates at CTCF-binding sites, it is tempting to speculate that Cohesin stabilizes the CTCF-mediated interaction probably by encircling CTCF formed loops. In our study, we clearly show that depletion of CTCF or mutation in a CTCF binding motif completely abolishes Cohesin association to common binding sites. This observation suggests that CTCF binding sites are "meeting points" for those two factors although it is still not clear what the functional relevance of the shared sites is. It is also not clear what keeps Cohesin at CTCF binding sites since there are also nonshared sites. Most of the studies (including ours) show that there is no direct interaction between CTCF and Cohesin. If correct, it suggests other proteins are involved in keeping Cohesin at CTCF sites and their identification would be the next step in revealing how Cohesin and CTCF functionally interact. It would also be very interesting to understand why Cohesin is found at a subset of the CTCF binding sites and not at the other sites despite some discrepancies among different studies about the percentage of shared binding sites. What is so special about, for example 3'HS1 in β-globin locus that binds Cohesin, compared to other CTCF binding sites. If we exclude the possibility that Cohesin is found at this site only because of an experimental set up (example's. better cross-linking at 3'HS1 when compared to other sites), there is a clear difference in Cohesin binding to 3'HS1 compared to, for example HS5 and HS-85.

Another serious obstacle to understanding the functional relevance of Cohesin

and CTCF shared binding sites is the lack of knowledge about the interaction of Cohesin with DNA. The commonly most accepted "ring" model, proposes that a tripartite ring formed by SMC1, SMC3 and SCC1/Rad21 encircles two 10-nm chromatin fibers but is not contacting the DNA directly (Haering et al., 2002). Perhaps such a ring is formed during the transient interaction of DNA strands in *cis* (CTCF mediated?) and slides along the chromatin fiber until is "locked" by CTCF to stabilize specific interactions.

If we postulate that Cohesin stabilizes CTCF mediated interactions, one would then expect that removal of Cohesin would influence the conformation of the locus by decreasing the interactions between CTCF sites. In our 3C experiments, Cohesin depletion appears to influence the interaction between CTCF binding sites in two opposing ways. First Cohesin depletion prevents the HS-85 site from contacting with the remainder of the CTCF binding sites. However, and perhaps surprisingly, 3'HS1 which is the actual Cohesin binding site, does not lose its contacts but it even increases relative cross-linking frequencies with HS5 while that with HS-65 remains more or less the same. It is important to note that Cohesin was still present (around 30%) in this experiment. One possible explanation is that a decreased concentration of Cohesin results in the selective loss of "weaker" interactions, i.e. the HS-85. Although the experiment should be extended into a more detailed analysis that also takes the novel B1 and B2 sites into account, it is tempting to speculate about other possibilities. For example other factors might stabilize loops formed by CTCF and these compete with Cohesin mediated interactions. This would also explain why when the interactions with HS-85 are lost other interaction frequencies increase (Figure 3C). Such interactions may be stabilized by factors such as the Ldb1 complex (Meier, Krpic et al. 2006; Song, Hou et al 2007) independent of Cohesin.

It should also be noted that the newly found sites B1 and B2 may provide an explanation why in our previous experiments of mutating the 3'HS1 CTCF site did not result in a change of gene expression of the β-globin expression, its function in β-gene expression may be minor compared to these novel sites.

Finally perhaps the most interesting conclusion from this work is that a cohesion mediated loop i.e. between 5'HS -85 and 3'HS1, is formed by sites where only one of the partner sites (3'HS1) binds Cohesin.

MATERIALS AND METHODS

Culturing I/11 cells:

Primary fetal liver cells were isolated from p53-/- mice as described (Splinter, Heath et al 2006). We maintained proliferating I/11 cells in (StemPro-34 SFM, Invitrogen) containing 0.5 units/ml of erythropoietin, 100 ng/ml stem cell factor and 1mM dexamethasone. Proliferating cells were expanded and kept at a density between $1.5x10^6$ and $3x10^6$ cells/ml. To induce differentiation, proliferating cells were washed twice with Hank's balanced salt solution and seeded at $2x10^6$ cells/ml in differentiation medium containing 5 units/ml of erythropoietin and 1mg/ml transferrin (Sigma-Aldrich). During differentiation, cells were kept at densities of 2-6 $x10^6$ cells/ml. Differentiation status was monitored by measuring cellular size distribution using an electronic cell counter. Differentiated cells were harvested after 72h to 120h. The size distribution ranged between 7-8 mm when differentiated and 9-11mm when not.

Chromatin Immunoprecipitation (ChIP):

ChIP analyses were performed as described in the Upstate protocol (www.upstate.com) except that cells were cross-linked using 2% formaldehyde for 5 min at room temperature. Real-time PCR quantification of precipitated DNA sequences (average fragment size 300bp) was performed on a BioRad MyIQ machine using Platinum TaqDNA polymerase (Invitrogen) and SYBR Green (Sigma-Aldrich) under the following cycling conditions:

2 min at 94 ^oC, 44 cycles of 30s at 94 ^oC, 30s at 55 ^oC, 1min at 72 ^oC during which measurements were taken, and 79 times 30s at 55 ^oC for melting curves. Enrichment was calculated relative to 2 control genes (amylase and necdin) and all values were normalized to input measurements. The ChIP DNA was used for real-time PCR, ChIP-on-Chip analysis and Illumina sequencing.

Antibodies

For ChIP and Western blot analysis the following antibodies were used: CTCF N3 polyclonal were generated as described (Hoogstraten et al., 2002), SMC1 (Bethyl A300-055A), SMC3 (Bethyl A300-060A), Rad21 (Abcam ab992).

ChIP-on-Chip:

Probe Labeling and Microarray Hybridization

ChIP DNA was amplified and labeled according to the standard Affymetrix protocol (www.affymetrix.com/products/arrays/specific/mouse _tiling_2.affx). Ten micrograms of labeled products were hybridized to Affymetrix mouse tiling 2.0R A and C arrays (900897) (Affymetrix, Santa Clara, CA). These tiled arrays are designed to contain 25-bp probes located at 35 nucleotide resolution.

Affymetrix Data Analysis

The scanned output files were analyzed with Tiling Analysis Software version 1.1 (Affymetrix, Santa Clara, CA) as described in (Gao et al., 2008). Probes were mapped to mouse chromosomes according to NCBIv33 (mm5) genome assembly. The samples (CTCF-non induced, CTCF-induced, SMC3-non induced, SMC3-induced, IgG-non induced, IgG-induced and 2 genomic input samples from non induced and induced I/11 cells) were normalized in groups. The genomic DNA corresponding to every ChIP-enriched region identified by tiling array was retrieved from the UCSC genome browser with a suitable version or IGB (Integrated Genome Browser) from Affymetrix.

ChIP DNA sample prep for Illumina sequencing

ChIP DNA was prepared for Illumina massive parallel sequencing according to the Illumina ChipSeq protocol (www.illumina.com). In brief, 10 ng of ChIP DNA was end repaired and purified with a QiaQuick PCR purification kit (Qiagen). An A base was added to the 3' end of the end repaired DNA fragments and the A-tailed fragments were purified once using a MinElute PCR purification Kit (Qiagen) in 10 ul of EB. Illumina provided single read adaptors were ligated to the end of the DNA fragments followed by a MinElute PCR purification (Qiagen). Ligation products were run on a 2% agarosegel and 200 bp (+/- 35 bp) fragments were isolated from gel by the GEL extraction Kit (Qiagen) and eluted in 36 ul EB. The size-selected fragments were PCR amplified by Phusion polymerase and subjected to the following conditions: 98 °C for 30 sec, (98 °C for 10 sec, 65 °C for 30 sec and 72 °C for 30 sec) for 18 cycles, and 72 °C for 5 min. PCR fragments were purified using by MinElute PCR purification Kit in 15 ul EB (Qiagen). One microliter of each sample library was quantified on Bioanalyzer (Agilent Technologies) using a DNA1000 assay. A 10 nM DNA stock in EB was prepared for each sample.

Cluster generation was performed according to the Illumina Cluster Reagents preparation protocol (www.illumina.com). In brief, 1ul of the stock was denatured with

NaOH, diluted to 3 pM and hybridized onto the flowcell. The hybridized fragments are sequentially amplified, linearized and end-blocked according to the Single read Sequencing user guide protocol (www.illumina.com). Sequencing reagents were prepared as described in the Genome Analyzer reagents protocol (www.illumina.com). After hybridization of the seq-primer sequencing by synthesis was performed using the Genome Analyzer and a 36-cycle protocol according to manufacturer's instructions. Images are analyzed by the Illumina GAP pipeline.

The raw data from the Illumina Genome Analyzer was processed using a combination of the IPAR (Integrated Primary Analysis Reporting Software) using version 1.0 of the GAP (Genome Analyzer Pipeline), Inc. The resultant sequences were mapped against NCBI build 37.1 of the mouse genome using the ELAND alignment software (Anthony J. Cox, Illumina Inc., unpublished work).

Chromosome Conformation Capture (3C)

3C was carried out as described (Tolhuis et al., 2002). Shortly, formaldehyde cross-linked chromatin from non-induced, induced, CTCF depleted and SMC3 depleted cells was digested with *Hind*III restriction enzyme overnight, followed by ligation with T4 DNA ligase at 16 ^{0}C for 4hr. After reversing the crosslinks, genomic DNA was purified by phenol extraction and ethanol precipitation. For PCR analysis, the linear range of amplification was for the samples by serial dilution. An appropriate of DNA within the linear range was subsequently used for the experiments. PCR products were run on 2% agarose gels and quantified on a Typhoon imager. All data points were generated from an average of 3 different experiments performed in duplo.

Lentivirus-Mediated CTCF and SMC3 RNA interference

Lentiviral hairpin RNA interference plasmids (pLK0.1-CTCF TRCN0000039019, pLK0.1-SMC3 TRCN00001090 05) were obtained from The RNAi Consortium (Mission Sigma Aldrich). The sequence of short hairpin RNA targeting the mouse CTCF gene (GenBank accession no. NM_007794.1, NM_181322.2) was 5'CCGGGCAGAGAAAGTAGTTGGTAATCTCGAGATTACCAACTACTTTCTCTGCTTTTTG 3'

in CDS region and SMC3 gene (GenBank accession no. NM_007790.2) was 5'CCGGCCCTGTAATGTTACATTTCTACTCGAGTAGAAATGTAACATTACAGGGTTTTTG 3'

in 3'UTR region.

Each of these vectors had been sequence-verified. Vectors were expanded in E.*coli* and purified. Lentivirus was produced by transient transfection of 293T cells according to standard protocols (Zufferey et al., 1997). 293T cells were transfected with a 3:1:4 mixture of psPAX-2, pMD2G-VSVG and a transfer vector construct using poly (ethylenimine)(PEI). After 24 hours medium was refreshed and virus-containing medium was harvested 48h and 72hr after transfection. After filtering, the virus stock

was concentrated 1000 times Virus activity/functionality was tested by serial dilutions on I/11 cells by Western blot. GFP control construct was used for measuring of transduction efficiency. Transduced cells were selected with puromycin and harvested every day for WB analysis, and 4 and 5 day for additional Chip analysis and 3C.

Primer sequences used for ChIP

(547) HS-85-f: GAGACTAAGTAATTCACCATGGG
(562) HS-85-r: GGATCTATCTTGATTGTCCTCC
(525) HS-62-f: GCACATGCCGTAGTTCTC
(526) HS-62-r: TCTGGAGTTCTCAGTTGTATGAC
(805) HS5-f: AACTAGAGAAAAAGAATGAGGCGTTT
(804) HS5-r: CTGGAGAATCCACACACCTAGGT
(529) 3'HS1-f: AATCAGTGGAACACTTCTGC
(530) 3'HS1-r: GTCTCAGGTTGTCAACTAAAGC
(654) necdin-f: GGTCCTGCTCTGATCCGAAG
(655) necdin-r: GGGTCGCTCAGGTCCTTACTT
(99) amylase-f: CTCCTTGTACGGGTTGGT
(100) amylase-r: AATGATGTGCACAGCTGAA
B1-f: CAGATGCTCTCTTGGGAAC
B1-r: CCTGGGCAATATGATCGTG
B2-f: AACTCTGGGAAACTCTGCC
B2-r: GCTCTGAGGCATGTTCTC

Allele-specific primers at 3'HS1

1:(766) 3'HS1-f; AGAGGAGGGCGGAAATCAGT
2:(530) 3'HS1-r:GTCTCAGGTTGTCAACTAAAGC
3:(765) 3'HS1 mut-f: GGAGGAGGGCGGAAATCAGC
4: (642) 3'HS1 mut-r: GTCTCAGGTTGTCAA CTAAACG

REFERENCES:

Bell, A. C. and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature 405, 482-5.

Bell, A. C., West, A. G. and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell 98, 387-96.

Bell, A. C., West, A. G. and Felsenfeld, G. (2001). Insulators and boundaries: versatile regulatory elements in the eukaryotic. Science 291, 447-50.

Bernstein, B. E., Meissner, A. and Lander, E. S. (2007). The mammalian epigenome. Cell 128, 669-81.

Bulger, M., Schubeler, D., Bender, M. A., Hamilton, J., Farrell, C. M., Hardison, R. C. and Groudine, M. (2003). A complex chromatin landscape revealed by patterns of nuclease sensitivity and histone modification within the mouse beta-globin locus. Mol Cell Biol 23, 5234-44.

Burke, L. J., Zhang, R., Bartkuhn, M., Tiwari, V. K., Tavoosidana, G., Kurukuti, S., Weth, C., Leers, J., Galjart, N., Ohlsson, R. et al. (2005). CTCF binding and higher order chromatin structure of the H19 locus are maintained in mitotic chromatin. Embo J 24, 3291-300.

Capelson, M. and Corces, V. G. (2005). The ubiquitin ligase dTopors directs the nuclear organization of a chromatin insulator. Mol Cell 20, 105-16.

Capelson, M. and Corces, V. G. (2006). SUMO conjugation attenuates the activity of the gypsy chromatin insulator. Embo J 25, 1906-14.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., Wong, E., Orlov, Y. L., Zhang, W., Jiang, J. et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell 133, 1106-17.

Dillon, N. and Sabbattini, P. (2000). Functional gene expression domains: defining the functional unit of eukaryotic gene regulation. Bioessays 22, 657-65.

Dolznig, H., Kolbus, A., Leberbauer, C., Schmidt, U., Deiner, E. M., Mullner, E. W. and Beug, H. (2005). Expansion and differentiation of immature mouse and human hematopoietic progenitors. Methods Mol Med 105, 323-44.

Dorsett, D., Eissenberg, J. C., Misulovin, Z., Martens, A., Redding, B. and McKim, K. (2005). Effects of sister chromatid cohesion proteins on cut gene expression during wing development in Drosophila. Development 132, 4743-53.

Farrell, C. M., West, A. G. and Felsenfeld, G. (2002). Conserved CTCF insulator elements flank the mouse and human beta-globin loci. Mol Cell Biol 22, 3820-31.

Gao, H., Falt, S., Sandelin, A., Gustafsson, J. A. and Dahlman-Wright, K. (2008). Genome-wide identification of estrogen receptor alpha-binding sites in mouse liver. Mol Endocrinol 22, 10-22.

Gillespie, P. J. and Hirano, T. (2004). Scc2 couples replication licensing to sister chromatid cohesion in Xenopus egg extracts. Curr Biol 14, 1598-603.

Gruber, S., Haering, C. H. and Nasmyth, K. (2003). Chromosomal cohesin forms a ring. Cell 112, 765-77.

Haering, C. H., Lowe, J., Hochwagen, A. and Nasmyth, K. (2002). Molecular architecture of SMC proteins and the yeast cohesin complex. Mol Cell 9, 773-88.

Hoogstraten, D., Nigg, A. L., Heath, H., Mullenders, L. H., van Driel, R., Hoeijmakers, J. H., Vermeulen, W. and Houtsmuller, A. B. (2002). Rapid switching of TFIIH between RNA polymerase I and II transcription and DNA repair in vivo. Mol Cell 10, 1163-74.

Horsfield, J. A., Anagnostou, S. H., Hu, J. K., Cho, K. H., Geisler, R., Lieschke, G., Crosier, K. E. and Crosier, P. S. (2007). Cohesin-dependent regulation of Runx genes. Development 134, 2639-49.

Kim, T. H., Abdullaev, Z. K., Smith, A. D., Ching, K. A., Loukinov, D. I., Green, R. D., Zhang, M. Q., Lobanenkov, V. V. and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell 128, 1231-45.

Kooren, J., Palstra, R. J., Klous, P., Splinter, E., von Lindern, M., Grosveld, F. and de Laat, W. (2007). Beta-globin active chromatin Hub formation in differentiating erythroid cells and in p45 NF-E2 knock-out mice. J Biol Chem 282, 16544-52.

Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G. P., Itoh, T., Watanabe, Y., Shirahige, K.

and Uhlmann, F. (2004). Cohesin relocation from sites of chromosomal loading to places of convergent transcription. Nature 430, 573-8.

Ling, J. Q., Li, T., Hu, J. F., Vu, T. H., Chen, H. L., Qiu, X. W., Cherry, A. M. and Hoffman, A. R. (2006). CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. Science 312, 269-72.

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

Ohlsson, R., Renkawitz, R. and Lobanenkov, V. (2001). CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. Trends Genet 17, 520-7.

Pai, C. Y., Lei, E. P., Ghosh, D. and Corces, V. G. (2004). The centrosomal protein CP190 is a component of the gypsy chromatin insulator. Mol Cell 16, 737-48.

Palstra, R. J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F. and de Laat, W. (2003). The beta-globin nuclear compartment in development and erythroid differentiation. Nat Genet 35, 190-4.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H. C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T. et al. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell 132, 422-33.

Recillas-Targa, F., Pikaart, M. J., Burgess-Beusse, B., Bell, A. C., Litt, M. D., West, A. G., Gaszner, M. and Felsenfeld, G. (2002). Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. Proc Natl Acad Sci U S A 99, 6883-8.

Rubio, E. D., Reiss, D. J., Welcsh, P. L., Disteche, C. M., Filippova, G. N., Baliga, N. S., Aebersold, R., Ranish, J. A. and Krumm, A. (2008). CTCF physically links cohesin to chromatin. Proc Natl Acad Sci U S A 105, 8309-14.

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

Splinter, E., Heath, H., Kooren, J., Palstra, R. J., Klous, P., Grosveld, F., Galjart, N. and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev 20, 2349-54.

Stedman, W., Kang, H., Lin, S., Kissil, J. L., Bartolomei, M. S. and Lieberman, P. M. (2008). Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. Embo J 27, 654-66.

Szabo, P. E., Tang, S. H., Silva, F. J., Tsark, W. M. and Mann, J. R. (2004). Role of CTCF binding sites in the Igf2/H19 imprinting control region. Mol Cell Biol 24, 4791-800.

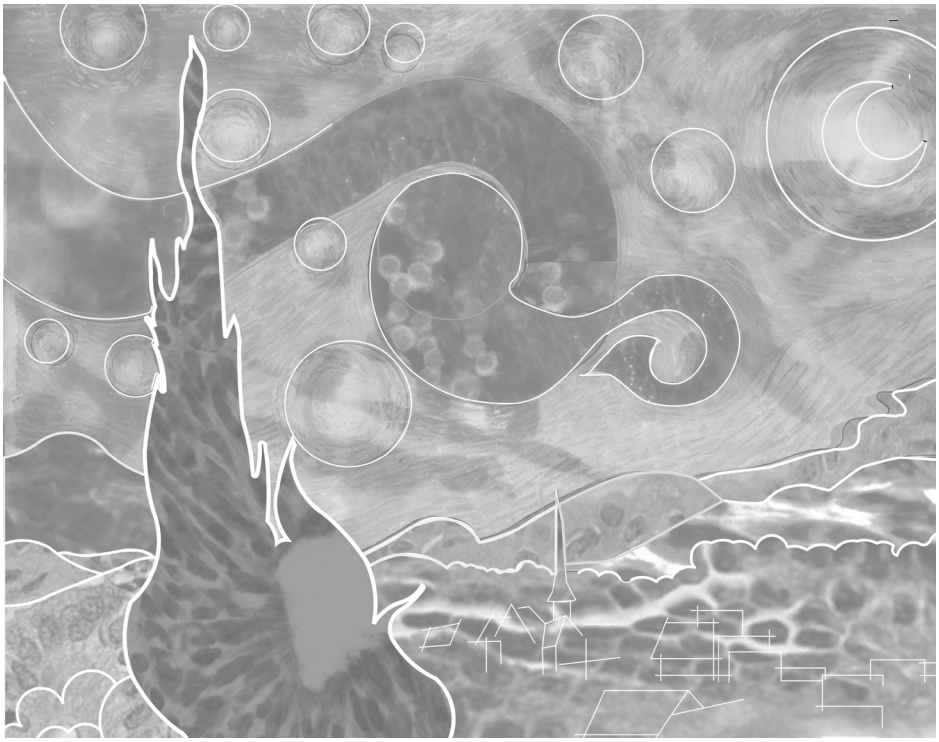
Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell 10, 1453-65.

von Lindern, M., Deiner, E. M., Dolznig, H., Parren-Van Amelsvoort, M., Hayman, M. J., Mullner, E. W. and Beug, H. (2001). Leukemic transformation of normal murine erythroid progenitors: v- and c-ErbB act through signaling pathways activated by the EpoR and c-Kit in stress erythropoiesis. Oncogene 20, 3651-64.

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

West, A. G., Gaszner, M. and Felsenfeld, G. (2002). Insulators: many functions, many mechanisms. Genes Dev 16, 271-88.

Zufferey, R., Nagy, D., Mandel, R. J., Naldini, L. and Trono, D. (1997). Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. Nat Biotechnol 15, 871-5.



Chapter 5 Critical role for the transcription regulator CCCTC-binding factor in the control of Th2 cytokine expression.

Ribeiro de Almeida C*, Heath H*, **Krpic S**, Dingjan GM, van Hamburg JP, Bergen I, van de Nobelen S, Sleutels F, Grosveld F, Galjart N, Hendriks RW.J Immunol. 2009 Jan 15;182(2):999-1010.

Critical Role for the Transcription Regulator CCCTC-Binding Factor in the Control of Th2 Cytokine Expression[1]

Claudia Ribeiro de Almeida,[2]*† Helen Heath,[2]‡ Sanja Krpic,‡ Gemma M. Dingjan,* Jan Piet van Hamburg,* Ingrid Bergen,† Suzanne van de Nobelen,‡ Frank Sleutels,‡ Frank Grosveld,‡ Niels Galjart,[3]‡ and Rudi W. Hendriks[3]†

Differentiation of naive CD4$^+$ cells into Th2 cells is accompanied by chromatin remodeling at the Th2 cytokine locus allowing the expression of the IL-4, IL-5, and IL-13 genes. In this report, we investigated the role in Th2 differentiation of the transcription regulator CCCTC-binding factor (CTCF). Chromatin immunoprecipitation analysis revealed multiple CTCF binding sites in the Th2 cytokine locus. Conditional deletion of the *Ctcf* gene in double-positive thymocytes allowed development of peripheral T cells, but their activation and proliferation upon anti-CD3/anti-CD28 stimulation in vitro was severely impaired. Nevertheless, when TCR signaling was circumvented with phorbol ester and ionomycin, we observed proliferation of CTCF-deficient T cells, enabling the analysis of Th2 differentiation in vitro. We found that in CTCF-deficient Th2 polarization cultures, transcription of IL-4, IL-5, and IL-13 was strongly reduced. By contrast, CTCF deficiency had a moderate effect on IFN-γ production in Th1 cultures and IL-17 production in Th17 cultures was unaffected. Consistent with a Th2 cytokine defect, CTCF-deficient mice had very low levels of IgG1 and IgE in their serum, but IgG2c was close to normal. In CTCF-deficient Th2 cultures, cells were polarized toward the Th2 lineage, as substantiated by induction of the key transcriptional regulators GATA3 and special AT-rich binding protein 1 (SATB1) and down-regulation of T-bet. Also, STAT4 expression was low, indicating that in the absence of CTCF, GATA3 still operated as a negative regulator of STAT4. Taken together, these findings show that CTCF is essential for GATA3- and SATB1-dependent regulation of Th2 cytokine gene expression. *The Journal of Immunology,* **2009, 182: 999–1010.**

Protection against pathogens relies on the ability of T cells to give rise to various effector cell fates upon activation. Classically, naive CD4$^+$ T cells are thought to undergo programmed differentiation into mainly two functionally distinct subsets, termed Th1 and Th2 (for review, see: Refs. 1 and 2). Th1 cells, which produce IFN-γ, are predominantly involved in cellular immunity against intracellular pathogens. Two major signaling pathways facilitate Th1 development, one involving IL-12/STAT4 and the other involving IFN-γ/STAT1/T-bet (3–5). Th2 cells, producing IL-4, IL-5, and IL-13, mediate humoral immunity and are essential for the eradication of parasitic worms, but also mediate allergic responses. Th2 cytokine production is dependent on the transcription factor GATA3, which is rapidly induced by IL-4 through STAT6 (6–10). Recently, a distinct effector T cell subset has been described, termed Th17. These cells produce IL-17 and control a wide range of infections at mucosal surfaces and are implicated in the pathogenesis of several autoimmune diseases first thought to be caused by deregulated Th1 function (reviewed in Ref. 2). In mice, both TGF-β and IL-6 are required to drive Th17 differentiation through activation of the orphan nuclear receptors RORγt and RORα (11, 12). Differentiation of pathogenic Th17 cells is developmentally related to anti-inflammatory Foxp3$^+$ regulatory T (Treg)[4] cells, which can be generated in vitro by stimulation with TGF-β in the absence of IL-6 (13).

Subset-specific expression of cytokine genes in T cells involves unique transcriptional, epigenetic, and structural mechanisms. When naive T cells are stimulated with Ag, they show low transcription of both IFN-γ and Th2 cytokines (14, 15). Th2 cytokine gene promoters and the Th2 locus control region come into close spatial proximity to form a higher-order chromatin structure, suggesting that early expression of the Th2 cytokines in naive T cells is supported by an initial "poised" chromatin configuration (16). Upon Th2 differentiation, a substantial increase in the transcriptional activity of IL-4, IL-5, and IL-13 and concomitant silencing of IFN-γ are observed. The converse pattern of gene activation and silencing is present in differentiating Th1 cells. Such polarized patterns of cytokine gene expression are achieved through the activation of cell type-specific transcription factors and chromatin remodeling proteins which bind to *cis*-regulatory elements of cytokine genes, thus initiating substantial and reciprocal alterations in the chromatin structure of the IFN-γ and Th2 cytokine loci (reviewed in Refs. 17 and 18). Indeed, both STAT6 and GATA3 are responsible for the establishment and/or maintenance of the

*Department of Immunology, †Department of Pulmonary Medicine, and ‡Department of Cell Biology and Genetics, Erasmus MC, Rotterdam, The Netherlands

Received for publication April 8, 2008. Accepted for publication November 13, 2008.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was supported by the Dutch Cancer Society KWF, the European Science Foundation EUROCORES Programme EuroDYNA (ERAS-CT-2003-980409), Fundação para a Ciência e a Tecnologia, and the Association for International Cancer Research.

[2] C.R.d.A. and H.H. contributed to this study equally.

[3] Address correspondence and reprint requests to Dr. Niels Galjart, Department of Cell Biology and Genetics, Erasmus MC Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands and Dr. Rudi H. Hendriks, Department of Pulmonary Medicine, Erasmus MC, Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands. E-mail addresses: n.galjart@erasmusmc.nl and r.hendriks@erasmusmc.nl

[4] Abbreviations used in this paper: Treg, regulatory T; SATB1, special AT-rich binding protein 1; CTCF, CCCTC-binding factor; CBS, CTCF binding site; ChIP, chromatin immunoprecipitation; BM-DC, bone marrow-derived dendritic cell; TNP-KLH, 2,4,6-trinitrophenyl-keyhole limpet hemocyanin; DN, double negative; DP, double positive; ISP, immature single positive; SP, single positive; WT, wild type.

chromatin conformation changes in the Th2 cytokine locus of developing Th2 cells (16, 19). Particularly, because GATA3 can induce chromatin-remodeling activity (20), it may facilitate the interaction between the Th2 locus control region and the cytokine gene promoters and thereby coregulate IL-4, IL-5. and IL-13 expression. In addition, Th2 cytokine expression requires special AT-rich binding protein 1 (SATB1), which mediates the formation of a densely looped, transcriptionally active chromatin structure at the Th2 locus containing GATA3, STAT6, c-Maf, the chromatin-remodeling enzyme Brg1, and RNA polymerase II (21).

The 11-zinc finger protein CCCTC-binding factor (CTCF) is a ubiquitously expressed and highly conserved transcriptional regulator implicated in many key processes within the nucleus, including promoter activation and repression, hormone-responsive gene silencing, and genomic imprinting (for review, see Ref. 22). CTCF often binds in the vicinity of insulators, elements that affect gene expression by preventing the spread of heterochromatin (acting as "barrier") and inhibiting inappropriate interactions between regulatory elements on adjacent chromatin domains (acting as "enhancer blocker") (23). It has been shown that CTCF is required for the enhancer-blocking activity of insulators (24). Consistent with a role for CTCF as an insulator protein, we have shown that in the mouse β-globin locus, CTCF mediates long-range chromatin looping and regulates local histone modifications (25). However, CTCF binding is not always required for chromatin insulation (26, 27). Combined, these data establish CTCF as an important protein involved in long-range DNA interactions and the regulation of active and repressive chromatin marks.

Genome-wide mapping of CTCF binding sites (CBS) in the human genome identified ~14,000 sites, whose distribution correlated with genes but not with transcriptional start sites (28, 29). Domains with few or no CTCF sites tend to include clusters of transcriptionally coregulated genes, whereby these regions are often flanked by CTCF binding sites (28, 30). The genome-wide analyses also revealed CTCF binding sites near genes displaying extensive alternative promoter usage, including protocadherin γ, the Ig λ L chain, and the TCR α/δ- and β-chain loci. In mice, CTCF binding was observed downstream of the TCR α/δ and the Ig H chain loci (31, 32). Very recently, CTCF was found to control MHC class II gene expression and long-range chromatin interactions between MHC class II promoter regions (33). These data imply an important role for CTCF in lymphocytes, in particular in the regulation of gene transcription in complex loci. We have recently found that conditional inactivation of *Ctcf* early in thymocyte development resulted in a severe arrest of early T cell development (34). Our findings indicated that CTCF regulates cell cycle progression of $\alpha\beta$ T cells in the thymus (34).

In this report, we investigated whether CTCF is important for Th2 cytokine expression. Chromatin immunoprecipitation (ChIP) assays revealed the presence of multiple CTCF binding sites in the Th2 cytokine locus. We show that conditional deletion of the *Ctcf* gene in the thymus, using CD4-Cre mice, allowed the generation of peripheral T cells, albeit with reduced numbers. In vitro polarization cultures of CTCF-deficient $CD4^+$ T cells revealed a Th2 cytokine expression defect, despite normal induction of the transcription factors GATA3 and SATB1.

Materials and Methods

Mice

T cell-specific deletion of *Ctcf* was achieved by breeding $Ctcf^{f/f}$ mice (34), which were crossed on the C57BL/6 background for >10 generations to CD4-Cre mice (35), which were provided by Dr. C. Wilson (University of Washington, Seattle, WA). OT-II-transgenic mice have been described previously (36). Mice were genotyped by Southern blotting or by PCR using Cre-specific primers. Mice were bred and maintained in the Erasmus MC animal care facility under specific pathogen-free conditions and analyzed at 6–10 wk of age. Experimental procedures were reviewed and approved by the Erasmus University committee of animal experiments.

RNA and protein analyses

Total RNA was extracted using the GeneElute mammalian total RNA miniprep system (Sigma-Aldrich). Primers spanning at least one intron-exon junction were designed either manually or using the ProbeFinder software (Roche Applied Science). Probes were chosen from the universal probe library (Roche Applied Science) or designed manually (GATA3, Gapdh) and purchased from Eurogentec. Quantitative real-time PCR was performed using an Applied Biosystems Prism 7700 sequence detection system. To confirm the specificity of the amplified products, samples were separated by standard agarose gel electrophoresis. Threshold levels were set and further analysis was performed using the SDS version 1.9 software (Applied Biosystems). Obtained cycle threshold values were normalized to cycle threshold values of Gapdh or β-actin. Each PCR was performed at least in triplicate. Primer sequences and PCR conditions used are available on request.

Nuclear extracts were prepared and analyzed by Western blot (37). Abs specific for CTCF (N3) and fibrillarin have been previously described (34). Alternatively, anti-CTCF antiserum was purchased from Millipore. Anti-DNMT1 was from Abcam and anti-UBF, anti-SATB1, and anti-fibrillarin were from Santa Cruz Biotechnology. Primary Ab incubation was done overnight at 4°C in TBS containing 5% (w/v) BSA and 0.15% (v/v) Nonidet P-40. Blots were incubated with secondary goat anti-rabbit or mouse Abs coupled to HRP (GE Healthcare). Signal detection was performed using ECL (Amersham Biosciences). Western blots were scanned and quantified using the gel macro function in ImageJ (W. S. Rasband, National Institutes of Health, Bethesda, MD; http://rsb.info.nih.gov/ij/).

ChIP assay

ChIP followed by ultrahigh-throughput DNA sequencing on I11 erythroid cells was performed using a Solexa 1G Genome Analyser (38). ChIP analysis of CTCF binding in the Th2 cytokine locus was performed as described in the ChIP Assay Kit protocol (Upstate Biotechnology) using the anti-CTCF Abs listed above or IgG as control. Quantitative real-time PCR (Bio-Rad IQ5) on immunoprecipitated DNA was performed using SYBR Green (Sigma-Aldrich) and Platinum *Taq* DNA Polymerase (Invitrogen). Enrichment was calculated relative to Necdin and values were normalized to input measurements. The sequences of the primers used were as follows: CBS-1F, 5′-GGTCTTAGCAGGTTCCCAA-3′; CBS-1R, 5′-CGTTCGG TAAGACAAGCAC-3′; CBS-2F, 5′-CACTCAGCACCTTACCTG-3′; CBS-2R, 5′-CCTGGGCTAAATGAATCAGT-3′; CBS-3F, 5′-AGGCA CAGTGTAGAAGTGT-3′; CBS-3R, 5′-GTCTCTCTTCCAGTCCAGTT-3′; CBS-4F, 5′-GGCACTTGTAACGCTCTAA-3′; CBS-4R, 5′-CCCTG ACCAACATCTCCAA-3′; CBS-5F, 5′-ATTGTGGAGGCTGGCAAG-3′; CBS-5R, 5′-GGTGACAGCCCAAATAAGT-3′; CBS-6F, 5′-CCA CATCCACCTGTCACTT-3′; CBS-6R, 5′-CTGTTTCACATCCATC GCA-3′; CBS-7F, 5′-CAGGCTTGTATCATCACCA-3′; and CBS-7R, 5′-TTCTTGAGGGACAGCACT-3′.

Flow cytometric analyses

Preparation of single-cell suspensions and mAb incubations for four-color cytometry has been previously described (39). All mAbs were purchased from BD Biosciences except for PE-conjugated anti-granzyme B (GB12; Caltag Laboratories), anti-GATA3 (Hg-3-31; Santa Cruz Biotechnology), allophycocyanin-conjugated anti-Foxp3 (FJK-16s; eBioscience), biotinylated anti-IL-13 (R&D Systems), and allophycocyanin-conjugated IL-10 (JES5-16E3; eBioscience).

For intracellular detection of cytokines, cells were restimulated with plate-bound anti-CD3 (10 μg/ml in PBS; 145-2C11) or PMA (50 ng/ml; Sigma-Aldrich) plus ionomycin (500 ng/ml; Sigma-Aldrich) in the presence of GolgiStop (BD Pharmingen) for 4 h. Cells were harvested and stained extracellularly, followed by standard intracellular staining using paraformaldehyde and saponin. Foxp3 expression was evaluated by intracellular staining using a Foxp3 buffer set (eBioscience).

Cell cycle status of T cell cultures was determined after fixing in ice-cold ethanol and subsequent staining in PBS containing 0.02 mg/ml propidium iodide, 0.1% v/v Triton X-100, and 0.2 mg/ml RNase. Doublet cells were excluded by measuring peak area and width. CFSE labeling of cells was performed as described elsewhere (40).

Samples were acquired on a FACSCalibur or FACS LSRII flow cytometer and analyzed using CellQuest (BD Biosciences) or FlowJo (Tree Star) research software.

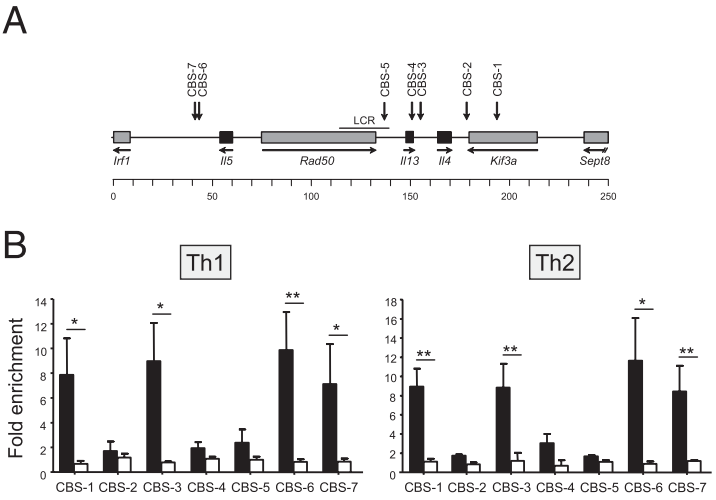


FIGURE 1. CTCF binding sites in the murine Th2 cytokine locus. *A*, Schematic representation of the Th2 cytokine locus and CTCF binding sites (CBS-1 to CBS-7). CBS were identified by a bioinformatics approach in combination with ChIP assays. *B*, ChIP analysis in Th1 and Th2 cells. Analysis was performed with CTCF (■) or IgG (□) Abs in 7-day polarized Th1 or Th2 cells. Mean values and SD are given for two ChIP experiments; *, $p < 0.05$ and **, $p < 0.01$.

In vitro T cell cultures

For in vitro T cell stimulations and Th1/Th2 polarization cultures, naive $CD62L^+CD4^+$ or $CD8^+$ T cells were purified by cell sorting using a FACSVantage VE equipped with Diva Option and BD FACSDiva software (BD Biosciences). The purity of obtained fractions was >98%.

For ChIP experiments, $CD4^+$ T cells from C57BL/6 mice were obtained through incubation with biotinylated mAbs (BD Pharmingen) specific for CD11b (M1/70), Gr-1 (RB6-8C5), Ter119 (Ly-76), TCR$\gamma\delta$ (GL3), B220 (RA3-6B2), NK1.1 (PK136), and CD8 (53-6.7), followed by streptavidin-conjugated microbeads and autoMACS purification according to the manufacturer's instruction (Miltenyi Biotec). The purity of $CD4^+$ T cell fractions was confirmed by FACS to be >95%.

T cell fractions were cultured at a concentration of 1×10^6 cells/ml in IMDM (BioWhittaker) containing 10% heat-inactivated FCS, 5×10^{-5} M 2-ME, 100 U/ml penicillin, and 100 μg/ml streptomycin. Stimulation was with plate-bound anti-CD3 (145-2C11) and anti-CD28 (37.51) mAbs (coated at 10 μg/ml each at 4°C overnight) or PMA (50 ng/ml) plus ionomycin (300 ng/ml). For Th1-polarizing conditions, anti-IL-4 (10 μg/ml; 11B11) and IL-12 (10 ng/ml) were added to the medium. Th2-polarizing cultures included anti-IFN-γ (5 μg/ml; R4-6A2,), anti-IL-12 (5 μg/ml; C17.8), and IL-4 (10 ng/ml). Treg- and Th17-polarizing conditions included TGF-β (3 ng/ml), anti-IL-4, and anti IFN-γ. Th17-polarizing conditions additionally contained IL-6 (20 ng/ml). For Th0 conditions, no cytokines or mAbs were added. For differentiation of $CD8^+$ effector T cells, only rIL-2 (5 ng/ml) was added to the medium. For Th0, Th1, and Th2 cultures, cells were supplemented with IL-2 (5 ng/ml) on day 3 after activation and expanded up to day 7 under the same cytokine conditions as the primary cultures. In Th17 cultures, cells were restimulated with PMA plus ionomycin at day 3, supplemented with TGF-β and IL-6 and expanded up to day 5. All cytokines were from R&D Systems.

Stimulation of OT-II-transgenic $CD4^+$ T cells was conducted in the presence of bone marrow-derived dendritic cells (BM-DC). Briefly, BM single-cell suspensions were prepared from C57BL/6 femurs and seeded at 2×10^6 per petri dish in complete IMDM and 200 ng/ml murine GM-CSF (BioSource International). On days 3 and 6, 200 ng/ml murine GM-CSF was added in 10 ml of fresh IMDM. On day 8, the nonadherent cells consisting of immature and mature BM-DC were harvested. For in vitro T cell proliferation studies, 0.2×10^6 CFSE-labeled OT-II-transgenic naive $CD4^+$ T cells were cocultured with 0.2×10^6 BM-DC previously pulsed with OVA $peptide_{323–339}$ (50 μg/ml) in complete IMDM. At day 4, cultured T cells were harvested for proliferation analysis.

ELISA

In vitro-polarized Th2 cells were harvested after 7 days in culture and washed twice with culture medium. Cells were resuspended (1.5×10^6 cells/ml) in fresh culture medium containing PMA (50 ng/ml) plus ionomycin (300 ng/ml). Three days later, supernatants were harvested and analyzed for the presence of cytokines using eBioscience (IL-4 and IL-5) and R&D (IL-13) ELISA systems.

Total serum Ig levels were determined by subclass-specific sandwich ELISA, as described; IgE was induced by i.p. injection of 10 μg of 2,4,6-trinitrophenyl-keyhole limpet hemocyanin (TNP-KLH) precipitated on alum (41).

Statistical evaluations

All statistical evaluations were done with Student's *t* test.

Results

CTCF binding sites in the Th2 cytokine locus

CTCF binding sites in the human Th2 locus have recently been identified in $CD4^+$ T cells (30). Taking into account that CTCF binding sites are largely invariant between cell types (28), we first used ChIP coupled to ultrahigh-throughput DNA sequencing data obtained in mouse I11 erythroid cells (S. Krpic and F. Grosveld, manuscript in preparation) to gain insight into CTCF binding in the murine Th2 cytokine locus. We identified four CTCF binding sites (CBS-1, CBS-3, CBS-6, and CBS-7) encompassing the 200-kb region containing the *Il5*, *Rad50*, *Il13*, *Il4,* and *Kif3a* genes (Fig. 1*A*). These sites, as well as three other CTCF binding sites (CBS-2, CBS-4, and CBS-5), have been reported to be occupied in mouse embryonic stem cells (42). We subsequently analyzed CTCF-binding to CBS-1 to CBS-7 in cultured Th1 and Th2 cells by ChIP. For these experiments, MACS-purified $CD4^+$ T cells from spleen and lymph nodes were stimulated with anti-CD3/CD28 under Th1-polarizing conditions (with IL-12 and anti-IL-4 Abs) or Th2-polarizing conditions (with IL-4, anti-IL-12, and anti-IFN-γ Abs) for 7 days. Quantitative real-time PCR analyses showed that CTCF binding to CBS-1, CBS-3, CBS-6, and CBS-7 was significantly increased when compared with IgG control, both in Th1- and Th2-polarized cells (Fig. 1*B*). A similar enrichment at these sites was also found in PMA/ionomycin- activated Th1- and Th2-polarized cells (data not shown). Thus, the *Il-4, Il-5*, and *Il-13* genes in the Th2 locus are flanked by CBS-6 and CBS-7 upstream of the *Il-5* gene and CBS-1 downstream of the *Il-4* gene, within the Kif3a gene. Interestingly, CBS-3 is located in the intergenic region between the *Il-13* and *Il-4* genes, close to a conserved noncoding sequence, designated CNS-1, which has been shown to be critical

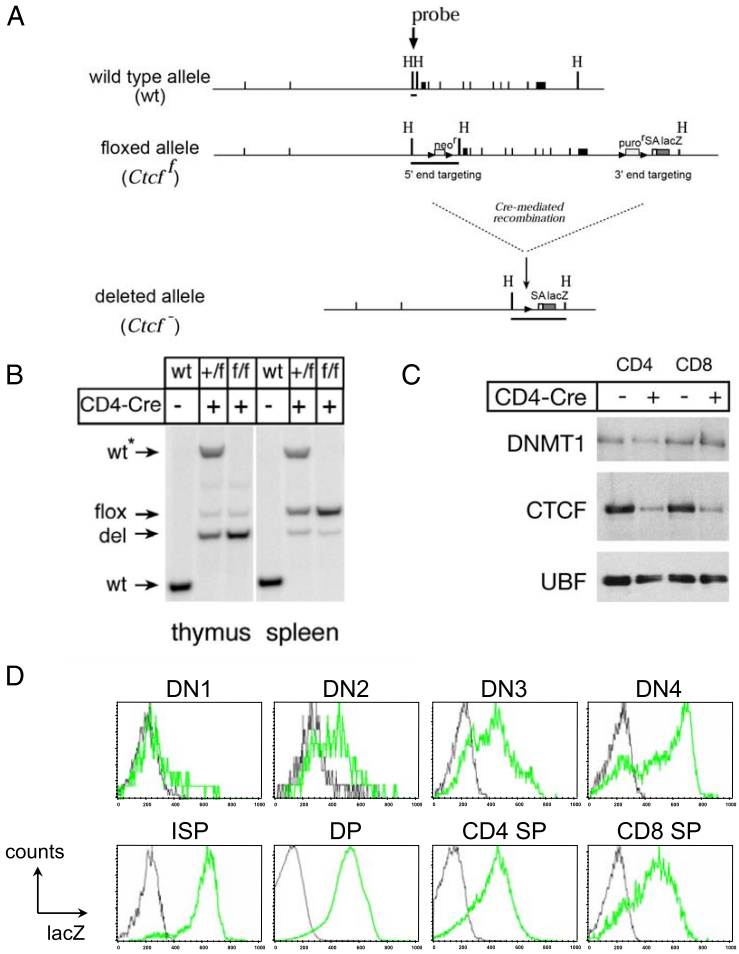


FIGURE 2. Conditional targeting of the mouse *Ctcf* gene. *A*, Simplified map of the WT murine *Ctcf* locus and modified alleles. *Top line*, WT *Ctcf* allele, with exons shown as solid boxes, and position of *Hin*dIII sites (H) and the Southern blot probe (arrow) indicated. *Middle line*, Floxed *Ctcf* allele (*Ctcf*^{f}). Two targeting cassettes were inserted into the *Ctcf* gene via consecutive rounds of homologous recombination. The first cassette contained LoxP sites (▲) flanking a PMC1-driven neomycin resistance gene (neo^{r}), the second cassette contained LoxP sites flanking a PGK-driven puromycin resistance gene (puro^{r}), followed by a splice acceptor-lacZ cassette (SA-LacZ). *Bottom line*, deleted Ctcf gene (*Ctcf*$^{-}$) generated after complete Cre-mediated recombination at the outermost loxP sites. For details on constructs and targeting, see Heath et al. (34). The small horizontal lines underneath each allele represent *Hin*dIII fragments recognized by the probe in the Southern blot analysis of *B* below. *B*, Southern blot analysis of CD4-Cre recombinase activity. *Hin*dIII-digested genomic DNA from thymus and spleen of mice of the indicated genotypes was analyzed by hybridization with a CTCF locus-specific probe (see *A*). The positions of the WT, *Ctcf*$^{f/f}$ (flox), and *Ctcf*$^{-/-}$ (del) alleles are indicated (asterisk indicates a polymorphic WT allele from the FVB background). *C*, Western blot analysis of sorted naive CD62L^{+}CD4^{+} and CD8^{+} T cell fractions from nontransgenic (−) or CD4-Cre-transgenic (+) *Ctcf*$^{f/f}$ mice were analyzed for CTCF. DNMT1 and UBF were used as loading controls. *D*, Flow cytometric analysis of *lacZ* expression in CTCF conditionally deleted mice. *LacZ* expression was analyzed in conjunction with cell surface markers. The indicated cell populations were gated and *lacZ* expression data are displayed as histogram overlays of CD4-Cre *Ctcf*$^{f/f}$ mice (green) on top of the background signals in WT mice (black). ISP, Immature SP.

for GATA3 binding and Th2 cytokine expression (43, 44). Our data indicate that the CTCF protein binds to CBS-1, CBS-3, CBS-6, and CBS-7, irrespectively, of Th1 or Th2 polarization.

Conditional deletion of the Ctcf *gene in T lymphocytes*

To study CTCF function in vivo, we generated a conditional *Ctcf* allele (*Ctcf*^{f}) by inserting a *loxP* site upstream of exon 3 and a *loxP* site along with a *lacZ* reporter downstream of exon 12 (Fig. 2*A*) (34). We bred *Ctcf*$^{f/f}$ mice to mice carrying a Cre-encoding transgene under the control of the CD4 promoter (35). Southern blotting showed almost complete deletion of the *Ctcf* gene in the thymus of CD4-Cre *Ctcf*$^{f/f}$ mice, whereas in the spleen *Ctcf* deletion was only partial, reflecting the presence of many non-T lineage cells (Fig. 2*B*). To evaluate the onset of *Ctcf*^{f} gene deletion, we analyzed thymocyte subpopulations for expression of the Ctcf-*lacZ* fusion transcript, using fluorescein-di-β-D-galactopyranoside as a substrate

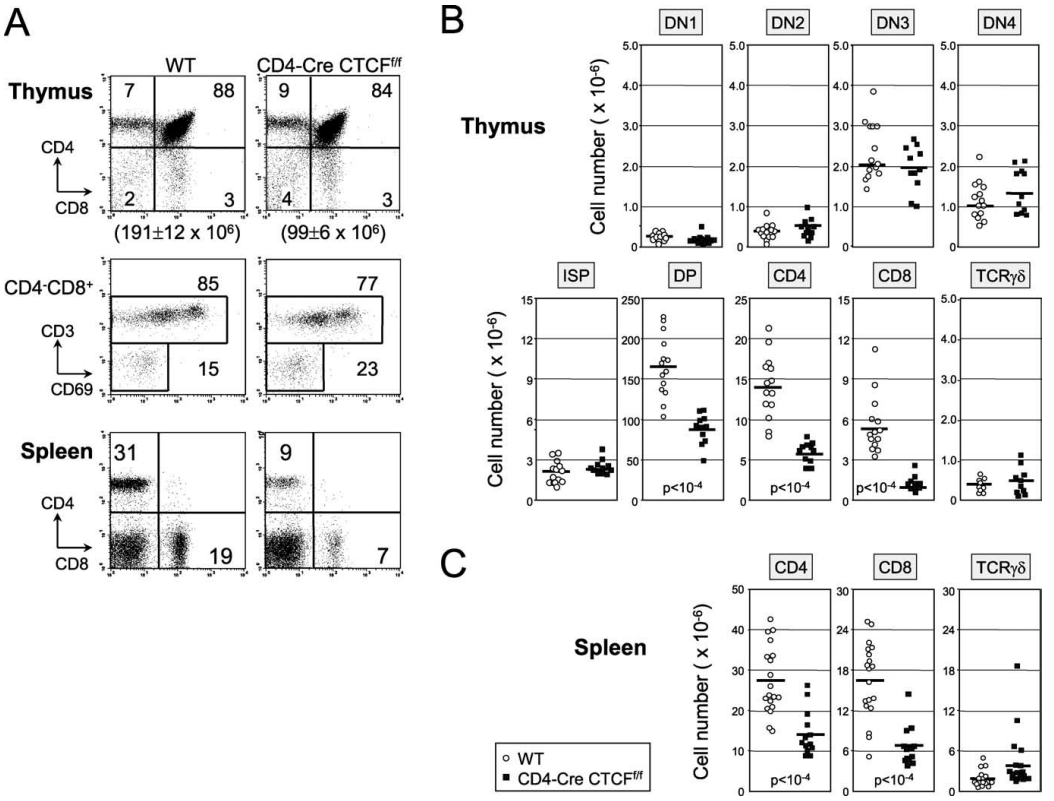


FIGURE 3. Defective TCR$\alpha\beta$ lineage development in CTCF-deficient mice. *A*, Flow cytometric analyses of the indicated cell populations in thymus or spleen from WT and CD4-Cre *Ctcf*$^{f/f}$ mice. Expression profiles of surface markers are shown as dot plots and the percentages of cells within the quadrants or gates are given. *B* and *C*, Absolute numbers of the indicated thymic and splenic T cell subpopulations. Each symbol represents one individual animal and lines indicate average values. Values of *p* of significant differences between WT and CD4-Cre *Ctcf*$^{f/f}$ mice are indicated.

in conjunction with cell surface markers. Consistent with the reported CD4-Cre activity at the double-negative (DN) stage (45), *Ctcf* deletion was initiated in DN cells (DN2 to DN4) and completed from the immature single-positive (ISP) cell stage onward (Fig. 2*D*). Despite efficient deletion of the *Ctcf* gene in the thymus, residual CTCF protein was still detectable in purified fractions of peripheral naive CD62L$^+$CD4$^+$ and CD8$^+$ T cells (~25% of control; Fig. 2*C*), indicating that CTCF is a remarkably stable protein in resting naive T cells.

To examine the effects of *Ctcf* deletion on T cell development, thymocyte subpopulations from 6- to 8-wk-old CD4-Cre *Ctcf*$^{f/f}$ mice and wild-type (WT) littermates were analyzed by flow cytometry. CD4-Cre *Ctcf*$^{f/f}$ mice displayed low thymic cellularity, with reduced numbers of double-positive (DP), CD4, and CD8 single-positive (SP) cells, when compared with WT controls (Fig. 3, *A* and *B*). The CD4$^-$CD8$^+$ thymocyte fraction had a relative increase of CD3^{low}CD69^{low} ISP cells and a decrease of CD3$^+$CD8$^+$ SP cells. Consistent with impaired thymic SP cell production, the numbers of mature CD4$^+$ and CD8$^+$ T cells in spleen and lymph nodes of CD4-Cre *Ctcf*$^{f/f}$ mice were significantly reduced (Fig. 3, *A* and *C*, and data not shown). Furthermore, the numbers of $\gamma\delta$ T cells in the spleens of CD4-Cre *Ctcf*$^{f/f}$ mice were increased (Fig. 3*C*).

In summary, in CD4-Cre *Ctcf*$^{f/f}$ mice, $\alpha\beta$ T cell development is partially arrested at the DP stage, resulting in a significant reduction of the numbers of peripheral CD4$^+$ and CD8$^+$ T cells.

Defective TCR/CD3-mediated proliferation of CTCF-deficient CD4$^+$ and CD8$^+$ T cells

To investigate cellular activation of CTCF-deficient T cells, we performed in vitro stimulation experiments with highly purified naive CD62L$^+$CD4$^+$ and CD8$^+$ peripheral T cells. We evaluated their ability to go through sequential cell divisions by CFSE labeling and observed severely reduced proliferation of anti-CD3/CD28-activated CTCF-deficient CD4$^+$ T cells at day 3 (Fig. 4*A*). Proliferation was not only defective when standard conditions of 10 μg/ml anti-CD3/anti-CD28 were used, but also when we increased either of the two (or both) Ab concentrations to 50 μg/ml (data not shown). Next, the capacity of CTCF-deficient CD4$^+$ T cells to proliferate upon a more physiological, Ag-specific stimulation was investigated by crossing CD4-Cre-*Ctcf*$^{f/f}$ mice with OT-II-transgenic mice, which harbor a TCR specific for OVA peptide. Upon activation by OVA peptide$_{323–339}$-pulsed APC, survival of nondividing (WT and CTCF-deficient OT-II-transgenic CD4$^+$ T cells was similar, but we noticed a severe proliferation defect in the absence of CTCF (Fig. 4*A*). Although anti-CD3/CD28-activated

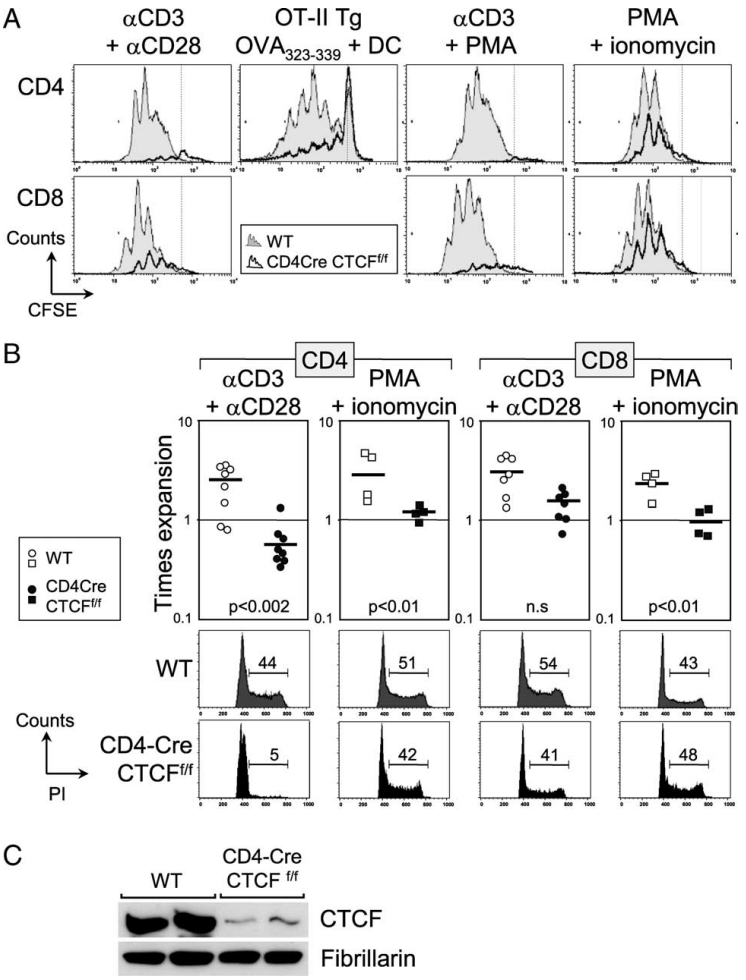


FIGURE 4. Impaired anti-CD3ε-mediated proliferation of CTCF-deficient T cells. *A*, CFSE profiles of T cell cultures of sorted CD62L$^+$CD4$^+$ and CD8$^+$ T cell populations from WT mice (gray histograms) in comparison to cultures from CD4-Cre *Ctcf$^{f/f}$* mice (black line). Cells were activated by the indicated stimuli and cultured for 3 days (or 4 days for OT-II CD4 T cells). Dotted lines indicate the fluorescence intensity of unstimulated cells. *B*, Cellular expansion in 3-day cultures upon anti-CD3/CD28 or PMA/ionomycin stimulation (*upper part*). Symbols represent the expansion values of cultures from individual mice, whereby cell numbers at the start of the culture were set to one; lines indicate average values. Values of *p* of significant differences between WT and CD4-Cre *Ctcf$^{f/f}$* mice are indicated; n.s., not significant. The *lower part* shows the cell cycle status of the indicated cultures, whereby DNA content was examined by propidium iodide (PI) staining. The percentages of cycling cells (S-G_2-M phase) are shown. *C*, Western blotting analysis of naive CD62L$^+$CD8$^+$ T cells from WT or CD4-Cre-transgenic *Ctcf$^{f/f}$* mice cultured for 5 days and examined for CTCF levels. Fibrillarin was used as a loading control.

CTCF-deficient CD8$^+$ T cells were able to undergo cell division, they lagged behind WT cells by approximately one cell cycle, and cell recovery was reduced when compared with WT CD8$^+$ T cells (Fig. 4*A*).

PMA bypasses proximal TCR signaling events and directly activates protein kinase C signaling (46). Under conditions where PMA was added as a costimulatory signal with anti-CD3, CTCF-deficient cells showed defective proliferation. However, when T cells were stimulated by PMA and the calcium ionophore ionomycin (which raises the intracellular level of Ca^{2+}), we observed significant proliferation of CTCF-deficient CD4$^+$ and CD8$^+$ T cells (Fig. 4*A*).

Consistent with limited cell division observed in the CFSE experiments, anti-CD3/CD28-stimulated CTCF-deficient CD4$^+$ T cell cultures showed diminished cell recovery and an almost complete lack of cells in the S-G_2-M phase of the cell cycle (Fig. 4*B*). Although CTCF-deficient PMA/ionomycin-stimulated CD4$^+$ or CD8$^+$ T cell cultures exhibited lower expansion rates, their cell cycle profiles at day 3 were similar to those from WT cells (Fig. 4*B*). In vitro proliferation of CD4-Cre *Ctcf$^{f/f}$* T cells did not reflect a specific expansion of rare cells that have escaped CTCF deletion, as CTCF protein levels were severely reduced both in CD8$^+$ (Fig. 4*C*) and in CD4$^+$ T cell cultures (see Fig. 9C).

Taken together, these findings demonstrate that conditional deletion of the *Ctcf* gene in DP thymocytes allows development of peripheral T cells, but their activation and proliferation upon anti-CD3/anti-CD28 stimulation in vitro is severely impaired. Nevertheless, when TCR signaling is circumvented with phorbol ester and ionomycin, CTCF-deficient T cells have the capacity to proliferate, indicating that in this context loss of CTCF can be compensated for by signaling molecules or nuclear factors induced by PMA/ionomycin. These results therefore indicate that CTCF is not absolutely required for cell proliferation.

Defective TCR/CD3-mediated activation of CTCF-deficient CD4$^+$ and CD8$^+$ T cells

Next, we investigated whether defective proliferation of CTCF-deficient T cells was caused by impaired cellular activation. Binding of IL-2 to its receptor is a critical event in the initiation of T cell proliferation, since it regulates transition of the cell cycle from G_1 into S phase (47). Therefore, we investigated

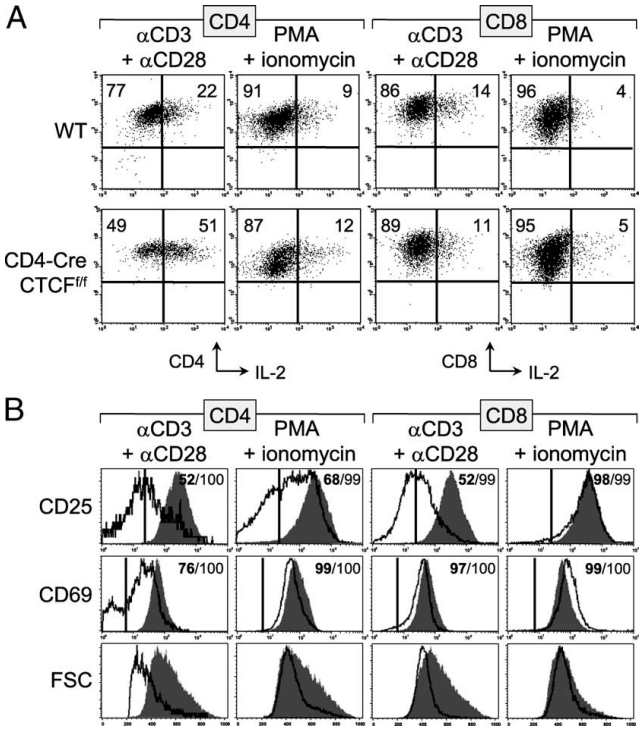


FIGURE 5. Impaired cellular activation of CTCF-deficient T cells. *A*, Analysis of IL-2 expression in anti-CD3/CD28 and PMA/ionomycin-stimulated cultures of sorted CD62L$^+$CD4$^+$ and CD8$^+$ T cell fractions from WT and CD4-Cre *Ctcf$^{f/f}$* mice. At day 3, cells were restimulated for 4 h before intracellular flow cytometric analysis. Total living cells were gated and CD4/IL-2 and CD8/IL-2 profiles are displayed as dot plots and the percentages of cells within the quadrants are given. Data shown are representative of four mice per group. *B*, Phenotypic characteristics of anti-CD3/CD28 or PMA/ionomycin-stimulated CD4$^+$ and CD8$^+$ T cells. CD25, CD69, and forward scatter profiles are displayed as histogram overlays of WT (gray histograms) and CD4-Cre *Ctcf$^{f/f}$* cultures (bold lines). The percentages shown represent the fractions of the cells within the indicated marker in WT (gray) or CD4-Cre *Ctcf$^{f/f}$* (black, bold) cultures. Data shown are representative of four to six mice per group.

whether CTCF-deficient T cells had an IL-2 production defect or impaired IL-2R induction. We found that upon stimulation with anti-CD3/CD28 or PMA/ionomycin, the absence of CTCF did not affect IL-2 production in CD4$^+$ or CD8$^+$ T cells, as analyzed by intracellular cytokine staining at day 3 (Fig. 5*A*). By contrast, induction of the IL-2 receptor CD25 on CTCF-deficient CD4$^+$ and CD8$^+$ T cells was severely impaired, when activated by plate-bound anti-CD3/CD28 (Fig. 5*B*). When activated by PMA/ionomycin, CTCF-deficient CD4$^+$ T cells displayed a partial defect in CD25 upregulation, while in CD8$^+$ T cells CD25 induction was normal.

Expression of CD69, an ~30-kDa glycoprotein induced in activated T cells, was hampered in CTCF-deficient CD4$^+$ T cells upon TCR stimulation, but was normal in CD8$^+$ T cells or in PMA/ionomycin-stimulated CTCF-deficient T cells. Finally, analyses of forward scatter values of stimulated cells showed that cell size increases in CTCF-deficient cells were limited, particularly upon anti-CD3/CD28 stimulation, whereby CD4$^+$ cells were more affected than CD8$^+$ cells (Fig. 5*B*).

From these findings, we conclude that anti-CD3/CD28 treatment does not elicit proper activation of CTCF-deficient peripheral T cells, in terms of induction of CD25, CD69, and cell size increase, whereby CD4$^+$ T cells are somewhat more affected than CD8$^+$ T cells. However, when T cells are stimulated by PMA/ionomycin, expression levels of CD25 and CD69 are largely in the normal ranges.

CD4-Cre Ctcf$^{f/f}$ *mice have severely reduced serum levels of IgG1 and IgE*

Concentrations of Ig subclasses in the serum of CD4-Cre *Ctcf$^{f/f}$* mice were reduced, except for IgM and IgG3, which are T cell

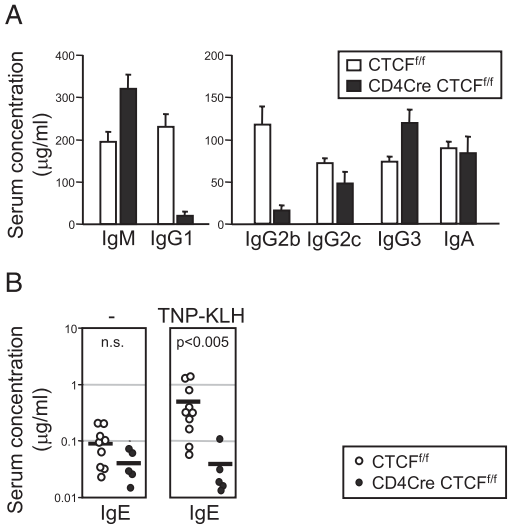


FIGURE 6. Serum Ig analysis in CTCF-deficient mice. *A*, Serum concentrations of Ig isotypes displayed as average values plus SD. Mice were 2 mo of age. *B*, Total IgE serum concentrations in nonimmunized mice (*left*) and in immunized mice 7 days after i.p. injection with 10 μg TNP-KLH (*right*).

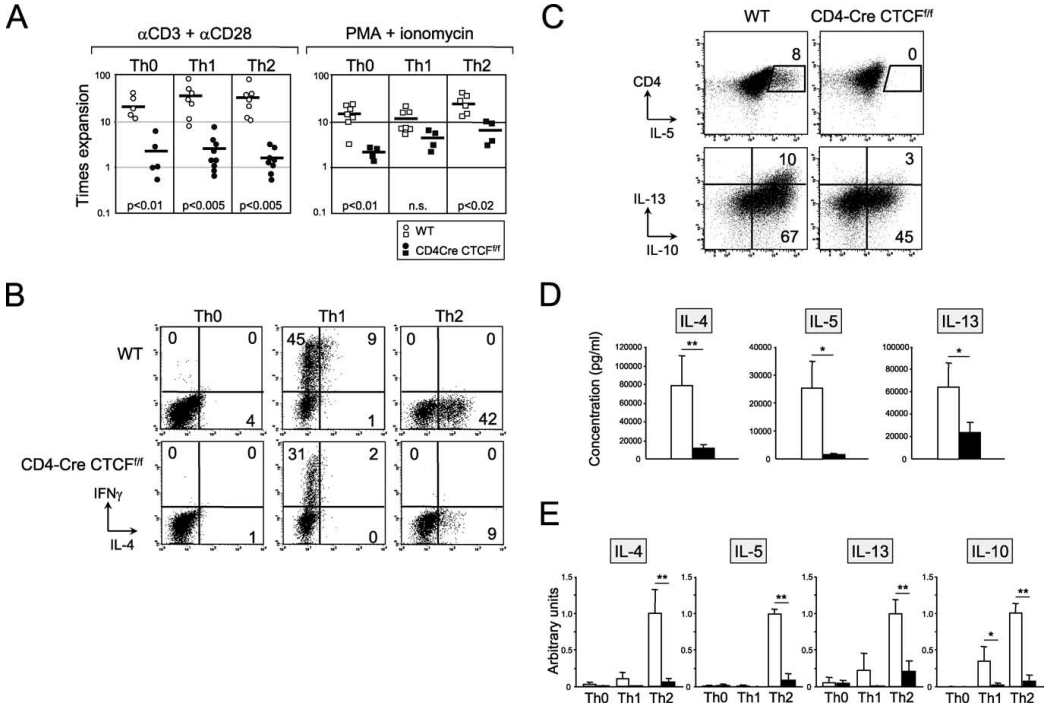


FIGURE 7. CTCF regulates Th2 cytokine expression. *A*, Expansion of Th0, Th1, and Th2 cultures 7 days after stimulation with anti-CD3/CD28 or PMA/ionomycin. Symbols represent the expansion values of cultures from individual mice, whereby cell numbers at the start of the culture were set to 1. Lines indicate average values. Values of p of significant differences between WT and CD4-Cre *Ctcf*$^{f/f}$ mice are indicated. *B*, Flow cytometric analysis for intracellular expression of IFN-γ and IL-4 in the indicated T cell cultures after stimulation with PMA/ionomycin. CD4$^+$ T cells were gated and expression profiles are displayed as dot plots. The percentages of cells within the quadrants are given. Mean fluorescence values for IL-4 were 138 and 75 for WT and CD4-Cre *Ctcf*$^{f/f}$ IL-4$^+$ Th2 cells, respectively. *C*, Flow cytometric analysis for intracellular expression of the indicated cytokines after stimulation with PMA/ionomycin. Total cells and CD4$^+$ T cells were gated and expression profiles are displayed as CD4/IL-5 and IL-10/IL-13 dot plots, respectively. *D*, Cytokine levels of WT (□) and CD4-Cre *Ctcf*$^{f/f}$ (■) Th2 cultures supernatant after stimulation with PMA/ionomycin. Mean values and SD are given for two seven mice analyzed per group; *, $p < 0.05$ and **, $p < 0.01$. *E*, Quantitative RT-PCR analysis of expression of the indicated cytokines in different T cell cultures (WT, □; CD4-Cre *Ctcf*$^{f/f}$, ■). Expression levels were normalized for GAPDH and are expressed as arbitrary units, whereby the values in WT Th2 cells were set to 1. Mean values and SD are given for four mice analyzed per group; *, $p < 0.05$ and **, $p < 0.01$.

independent and IgA which requires TGF-β expression (Fig. 6*A*). This does not necessarily mean that CTCF-deficient CD4$^+$ T cells produce normal levels of TGF-β, since this cytokine is secreted by many cell types.

Interestingly, serum levels of the IL-4-dependent isotype IgG1 were more affected than those of the IFNγ-dependent isotype IgG2c (~10 and ~60% of WT, respectively). Serum concentrations of the IL-4-dependent isotype IgE were also lower in the absence of CTCF (Fig. 6*B*). Moreover, when Th2-mediated responses were tested in vivo by injection of a low dose of TNP-KLH in alum (10 μg), we observed an increase in the concentration of total IgE in WT mice at day 10, but not in CD4-Cre *Ctcf*$^{f/f}$ animals (Fig. 6*B*). Thus, the absence of CTCF resulted in a severe deficiency for the Th2-dependent subclasses IgG1 and IgE in the serum, whereas the Th1-dependent subclass IgG2c was only moderately affected. CD4-Cre *Ctcf*$^{f/f}$ mice also had very low levels of IgG2b in the serum. Because class switch recombination to IgG2b is thought to be regulated by various cytokines including IFN-γ and TGF-β, but not by Th2 cytokines, the IgG2b deficiency cannot be explained by a selective Th2 defect.

Th2 cytokine defect in CD4-Cre Ctcf$^{f/f}$ *mice*

Next, we performed in vitro polarization cultures to investigate whether CTCF is specifically required for differentiation of Th2 effector cells. Sorted naive CD62L$^+$CD4$^+$ T cells from spleen and lymph nodes were stimulated with anti-CD3/CD28 or PMA/ionomycin under Th0 conditions (without additional cytokines or Abs), Th1-polarizing conditions (with IL-12 and anti-IL-4), or Th2-polarizing conditions (with IL-4, anti-IL-12, and anti-IFN-γ) for 7 days. Under these conditions, anti-CD3/CD28-stimulated WT CD4$^+$ T cell fractions manifested a ~10- to 100-fold expansion in 7 days (Fig. 7*A*). Consistent with the observed severely defective proliferation of anti-CD3/CD28-stimulated CTCF-deficient CD4$^+$ T cells, we found that the expansion of these cells at day 7 was negligible. Importantly, replacing anti-CD3/CD28 by PMA/ionomycin stimulation resulted in significant expansion of CTCF-deficient T cell cultures (~2–10 times at day 7; Fig. 7*A*), enabling the analysis of Th1 and Th2 development in vitro. Under these conditions, WT Th cultures showed an ~10- to 30-fold expansion (Fig. 7*A*).

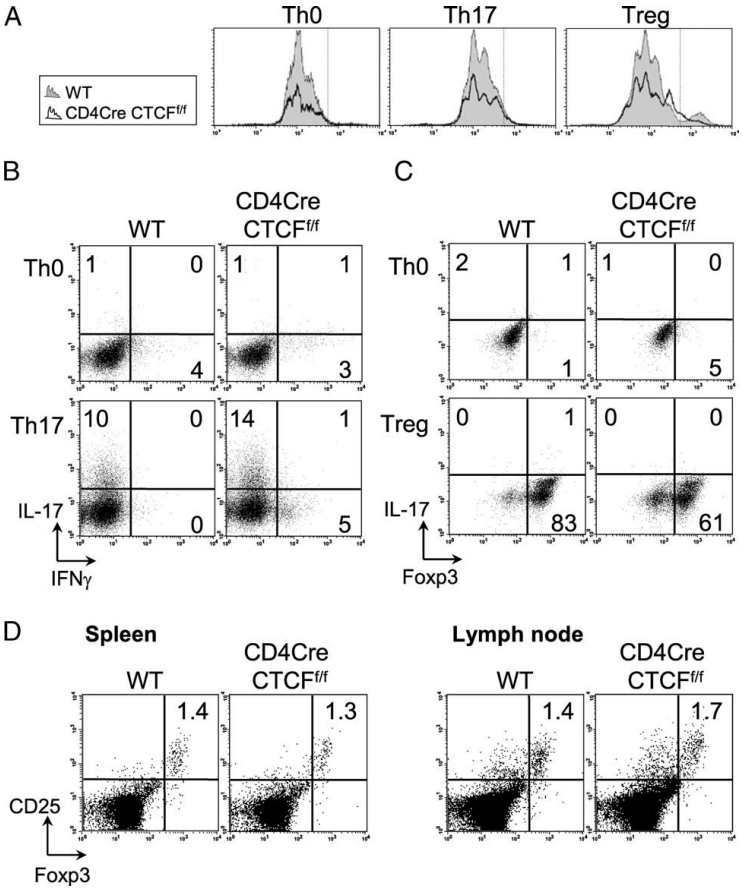


FIGURE 8. CTCF is not essential for Th17 or Treg differentiation. *A*, CFSE profiles of Th0, Th17, and Treg cultures of sorted CD62L$^+$CD4$^+$ T cell populations from WT mice (gray histogram) in comparison to cultures from CD4-Cre *Ctcf$^{f/f}$* mice (black line). Cells were activated with PMA/ionomycin and cultured for 3 days. Dotted lines indicate the fluorescence intensity of unstimulated cells. *B*, Flow cytometric analysis for intracellular expression of IFN-γ and IL-17 in Th0 and Th17 cultures after stimulation with PMA/ionomycin. CD4$^+$ T cells were gated and expression profiles are displayed as dot plots. The percentages of cells within the quadrants are given. *C*, Flow cytometric analysis for intracellular Foxp3 and IL-17 expression in PMA/ionomycin-stimulated Treg cultures. CD4$^+$ T cells were gated and expression profiles are displayed as dot plots. *D*, Flow cytometric analysis for intracellular Foxp3 and membrane CD25 with quantification of CD25$^+$Foxp3$^+$ Treg cells in spleen and mesenteric lymph nodes in vivo. Data shown are representative of four to six mice per group.

In Th1-polarized cultures, CTCF-deficient T cells produced moderately reduced levels of IFN-γ when compared with WT T cells, as determined by intracellular flow cytometry (Fig. 7*B*). The proportions of IFN-γ$^+$ cells were 55 ± 15% ($n = 7$) and 23 ± 6% ($n = 4$; $p < 0.005$) in WT and CTCF-deficient Th1 cultures, respectively. Remarkably, CTCF-deficient T cells showed a more severe IL-4 production defect in Th2-polarized cultures. Both the frequency of IL-4$^+$ T cells (51 ± 11% ($n = 7$) in WT and 9 ± 4% ($n = 4$; $p < 0.00005$) in CTCF-deficient cultures) and intracellular IL-4 signals per cell were significantly reduced in the absence of CTCF (Fig. 7*B*). Additional intracellular flow cytometry analyses showed that in CTCF-deficient Th2 cell cultures IL-5 production was not detectable and IL-10 and IL-13 were severely reduced (Fig. 7*C*). Consistent with these findings, in the supernatants of Th2-polarized cultures, production of IL-4, IL-5, and IL-13 was significantly reduced in the absence of CTCF, as determined by ELISA (Fig. 7*D*). Finally, quantitative RT-PCR analysis of day 7 cultures showed that in CTCF-deficient Th2 cultures transcription of the IL-4, IL-5, and IL-13 cytokines was strongly reduced (Fig. 7*E*). We also analyzed transcription of the *Il-10* gene, which is not located within the Th2 locus, and found that in the absence of CTCF IL-10 expression was very low, both in Th1 and Th2 cultures.

Thus, in CTCF-deficient Th2 polarization cultures in vitro, transcription of IL-4, IL-5, and IL-13 was strongly reduced. By contrast, CTCF deficiency had a moderate effect on IFN-γ production in Th1 cultures.

Th17 and Treg differentiation in CD4-Cre Ctcf$^{f/f}$ *mice*

The observed Th2 cytokine defect in CTCF-deficient Th2 cells prompted us to investigate whether differentiation toward alternative CD4 T cell fates, including Th17 and Treg cells, was possible in the absence of CTCF. We activated sorted naive CD62L$^+$CD4$^+$ T cells with PMA/ionomycin and cultured them under Th17 conditions (in the presence of TGF-β, IL-6, anti-IFN-γ, and anti-IL-4) or Treg conditions (with TGF-β, anti-IFN-γ, and anti-IL-4). As examined in CSFE experiments, we observed significant proliferation of CTCF-deficient CD4$^+$ T cells under Th17 and Treg conditions at day 3 (Fig. 8*A*). When Th17 cultures were analyzed for intracellular cytokines at day 5, we found that the expression of IL-17 in CTCF-deficient T cells reached values that were similar to those found in WT T cells (~8%, $n = 6$; Fig. 8*B*). In addition, we noticed a small but consistent population of IFN-γ$^+$ cells in CTCF-deficient Th17 cultures, which was not detected in WT Th17 cultures.

When naive WT CD62L$^+$CD4$^+$ T cells were cultured under Treg conditions for 3 days, a large majority of cells (85 ± 2%,

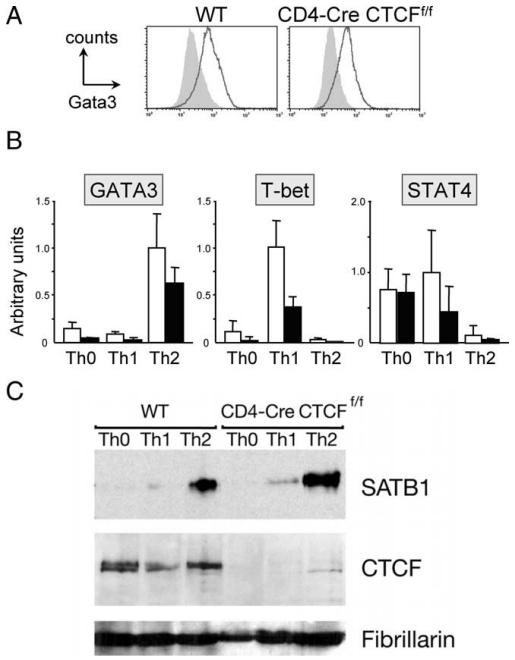


FIGURE 9. GATA3 and SATB1 are induced in CTCF-deficient Th2 cultures. *A*, Flow cytometric analysis for intracellular GATA3 protein expression in PMA/ionomycin-stimulated T cell cultures. $CD4^+$ T cells were gated and for the indicated mice expression profiles are displayed as histograms overlays of Th1 (gray histograms) and Th2 cultures (bold lines). *B*, Quantitative RT-PCR analysis of GATA3, T-bet, and STAT4 expression in different T cell cultures from WT (□) or CD4-Cre *Ctcf*$^{f/f}$ mice (■). Expression levels were normalized for GAPDH and are expressed as arbitrary units, whereby expression in WT Th1 cells (T-bet, STAT4) or Th2 cells (GATA3) was set to 1. Mean values and SD are given for four mice analyzed per group. *C*, Western blotting analysis of SATB1 and CTCF protein levels in PMA/ionomycin-stimulated T effector cell cultures at day 7. Fibrillarin was used as a loading control.

$n = 3$; Fig. 8*C*) expressed Foxp3, the transcription factor associated with Treg differentiation (48). Such $Foxp3^+$ cells were also present in Treg cultures of CTCF-deficient T cells, albeit that the proportions were slightly lower ($62 \pm 2\%$; $n = 3$). Furthermore, the proportions of naturally occurring $CD25^+Foxp3^+CD4^+$ Treg cells in spleen and lymph nodes in vivo were not different between WT and CTCF-deficient mice (Fig. 8*D*).

As a control, we also evaluated IFN-γ and granzyme B expression in day 7 cultures of anti-CD3/CD28- or PMA/ionomycin-stimulated $CD62L^+CD8^+$ T cells. We found that production of IFN-γ and granzyme B was only moderately affected (data not shown).

Collectively, these findings show that differentiation of CTCF-deficient T cells into IL-17-producing Th17 cells is apparently normal and lack of CTCF has limited effects on differentiation of Treg and $CD8^+$ T cells. We therefore conclude that the absence of CTCF does not result in a global defect in effector T cell differentiation.

GATA3 and SATB1 expression in CD4-Cre Ctcf$^{f/f}$ *Th2 cells*

Because Th2 cytokine production depends on the transcription factor GATA3 (6–10), we evaluated its expression in the T cell cultures. As determined by intracellular flow cytometry, GATA3 expression appeared unaffected in CTCF-deficient Th2 cultures (Fig. 9*A*), excluding the possibility that Th2 cytokine production was impaired due to defective GATA3 induction. Furthermore, the CTCF-deficient Th2 cultures displayed clear features of Th2- polarized cells, including low mRNA levels of T-bet and STAT4 (Fig. 9*B*). Since GATA3 has the capacity to inhibit STAT4 transcription (49), the finding of low STAT4 expression levels suggest that in the absence of CTCF GATA3 still operated as a negative regulator of STAT4. In CTCF-deficient Th1 cultures, T-bet expression was reduced when compared with WT (Fig. 9*B*), which is consistent with the observed reduction in IFN-γ expression. Next to GATA3, SATB1 has also been implicated in Th2 locus expression (21). SATB1 was specifically induced, both in WT and CTCF-deficient Th2 cultures (Fig. 9*C*).

Taken together, these data indicate that differentiating CTCF-deficient Th2 cells show impaired expression of Th2 cytokines IL-4, IL-5, and IL-13, but nevertheless up-regulate the Th2-specific factors GATA3 and SATB1 and down-regulate T-bet and STAT4.

Discussion

The differentiation process of naive $CD4^+$ T cells to Th1 or Th2 cells is critically dependent on coordinated transcriptional regulation of cytokine gene loci (18). To investigate whether CTCF regulates transcription of cytokine-encoding genes, we studied mice in which CTCF was conditionally deleted in the T cell lineage. Our data indicate that CTCF deficiency affects differentiation of Th2 effector cells by impairing T cell activation as well as Th2 cytokine production. Activation and proliferation of CTCF-deficient CD4 and CD8 cells upon anti-CD3/anti-CD28 stimulation in vitro was severely hampered. However, when TCR signaling was circumvented with phorbol ester and ionomycin, we observed proliferation of CTCF-deficient T cells, enabling the analysis of Th cell differentiation. We found that in CTCF-deficient Th2 polarization cultures, transcription of IL-4, IL-5, and IL-13 was strongly reduced. By contrast, CTCF deficiency had only a modest effect on IFN-γ production in Th1 cultures and did not appear to affect Th17 differentiation and IL-17 production. In CTCF-deficient Th2 cultures, cells were polarized toward the Th2 lineage, as the key transcriptional regulators GATA3 and SATB1 were induced and T-bet and STAT4 were down-regulated. Nevertheless, expression of IL-4, IL-5, and IL-13 remained strongly inhibited in these otherwise Th2-polarized cells. We therefore propose that CTCF plays a major role in the GATA3- and SATB1-dependent regulation of the expression of genes within the Th2 cytokine locus.

We found that CTCF-deficient $CD4^+$ and $CD8^+$ T cells have severely defective activation and proliferation in response to anti-CD3/CD28 stimulation. Nevertheless, one of the most important downstream events, production of IL-2, was not noticeably affected, ruling out global defects in TCR clustering, actin polymerization, or activation of NFAT, NF-κB, and JNK cascades (50, 51). The finding that PMA/ionomycin stimulation bypassed the CD3/CD28-signaling defect suggests that the presence of CTCF is important for signaling events downstream of the TCR which are necessary for full T cell activation. In this context, we found that PMA/ionomycin partially (in $CD4^+$ T cells) or completely (in $CD8^+$ T cells) rescued the defective induction of IL-2R expression in CTCF-deficient T cells observed upon anti-CD3/CD28 stimulation (Fig. 5*B*). Additional experiments are required to clarify whether CTCF controls gene expression of specific proteins involved in proximal signaling events (induced upon TCR stimulation) that regulate IL-2R induction.

The in vitro Th1 polarization cultures showed that CTCF-deficient Th1 effector cells produced significant amounts of IFN-γ, but expression was reduced when compared with WT. Also, in anti-CD3/CD28-stimulated CTCF-deficient $CD8^+$ T cells, we found that the proportion of IFN-γ^+ cells was ~60% of that in WT cells (C. R. de Almeida, unpublished data). In contrast, we observed a small cloud of IFN-γ^+ cells in Th17 cultures from CTCF-deficient T cells, which was not detectable in WT Th17 cultures, indicating that IFN-γ expression is not reduced under all culture conditions. Therefore, we conclude that the observed reduction of IFN-γ production in CTCF-deficient Th1 cells does not necessarily implicate CTCF in transcriptional regulation of the IFN-γ locus, as decreased IFN-γ production may well result form reduced expression of essential transcription factors such as T-bet and STAT4 (Fig. 9*B*).

Our analysis of CTCF binding in the Th2 cytokine locus revealed four CBS: three sites flanking the Th2 locus (CBS-6 and CBS-7 upstream of the *Il-5* gene, CBS-1 downstream of *Il-4*, within the Kif3a gene) and one site in the intergenic region between *Il-13* and *Il-4* (CBS-3), irrespective of Th1 or Th2 polarization. CBS-3 is located near the conserved noncoding sequence CNS-1, which is critical for Th2 cytokine expression and binds the C-terminal zinc finger of GATA3 (43, 44). However, because the distance between CNS-1 and CBS-3 is ~1 kb, it is not very likely that CTCF binding to CBS-3 influences Th2 cytokine expression by direct interaction with nuclear proteins recruited to the CNS-1 region.

Interestingly, CBS-3 is located at the constitutive hypersensitive site HSS-3, which has been shown to be present both in naive $CD4^+$ cells and Th1 and Th2 cells (43). Although nine SATB1 sites were identified in the mouse Th2 locus (21), none of these is located near the CNS-1 region. Thus, SATB1 and CTCF binding sites in the Th2 locus are interspersed. One of the possible explanations for defective Th2 cytokine expression in the absence of CTCF would be that CTCF is involved in chromatin organization of the Th2 locus. In such a model, loss of CTCF would affect SATB1-mediated looping of the Th2 locus and, consequently, Th2 cytokine expression. Indeed, studies in the chicken β-globin locus have led to a model of CTCF-dependent enhancer-blocking function based on the interaction between CTCF and the nucleolar protein nucleophosmin, whereby tethering of the insulator to a nuclear structure prevents enhancer-promoter communication (52). Moreover, SATB1 was originally identified as a matrix attachment region DNA-binding protein (53), which possibly contributes to chromatin loop organization (54). It would be interesting to investigate whether CTCF binding sites at the Th2 cytokine locus are involved in the tethering of this 200-kb DNA region to subnuclear sites, thereby allowing coordinated expression of Th2 cytokine genes from a SATB1-dependent, transcriptionally active chromatin structure.

However, mechanisms other than formation of chromatin loops can also account for CTCF function and explain the defective Th2 cytokine expression observed in the absence of CTCF. Similar to its role in the mouse β-globin locus (25), CTCF could function to direct local histone modifications at the Th2 cytokine locus. Indeed, high-resolution profiling of histone methylation in the human genome showed that CTCF marks boundaries of histone methylation domains (30) and CTCF interaction with histones or histone-modifying proteins have been reported (30, 52, 55, 56). Additionally, CTCF has also been shown to interact with the large subunit of RNA polymerase II (57), shown to be recruited to the Th2 cytokine locus upon Th2 cell activation (21). Finally, it remains possible that the effect of CTCF on Th2 cytokine expression is indirect. For example, CTCF could be an essential regulator of the expression 1) of nuclear regulators other than GATA3 or SATB1 that are required for Th2 cytokine transcription or 2) of enzymes involved in posttranslational modifications of Th2-specific transcription factors.

The role for CTCF in the GATA3/SATB1-mediated regulation of the Th2 cytokine locus may well parallel its recently described role in the control of MHC class II gene expression and the formation of long-distance chromatin interactions involving the CIITA (33). In contrast, we reported that deletion of one CTCF binding site in the mouse β-globin locus did not affect expression of the β-globin genes (25). Thus, it is clear that CTCF has cell type-specific functions. It was proposed that CTCF remains bound to its ~14,000 cognate binding sites irrespective of cell type (28). It will be interesting to determine how CTCF performs cell type-specific roles while remaining bound to its cognate sites. Equally interesting are the questions how and which chromosomal interactions, both in *cis*-and in *trans*, persist in the absence of CTCF. Importantly, our experiments in mature CTCF-negative T cells show they can proliferate and differentiate under appropriate conditions and it is therefore feasible to address these issues using CTCF knockdown or conditional targeting approaches.

Acknowledgments

We thank Harmjan Kuiper, Janneke Samsom, Lisette van Berkel, Edwin de Haas, Tom Schoonerwille, Marjolein de Bruin (Erasmus MC, Rotterdam The Netherlands), and Louis Boon (Bioceros BV, Utrecht, The Netherlands) for their assistance.

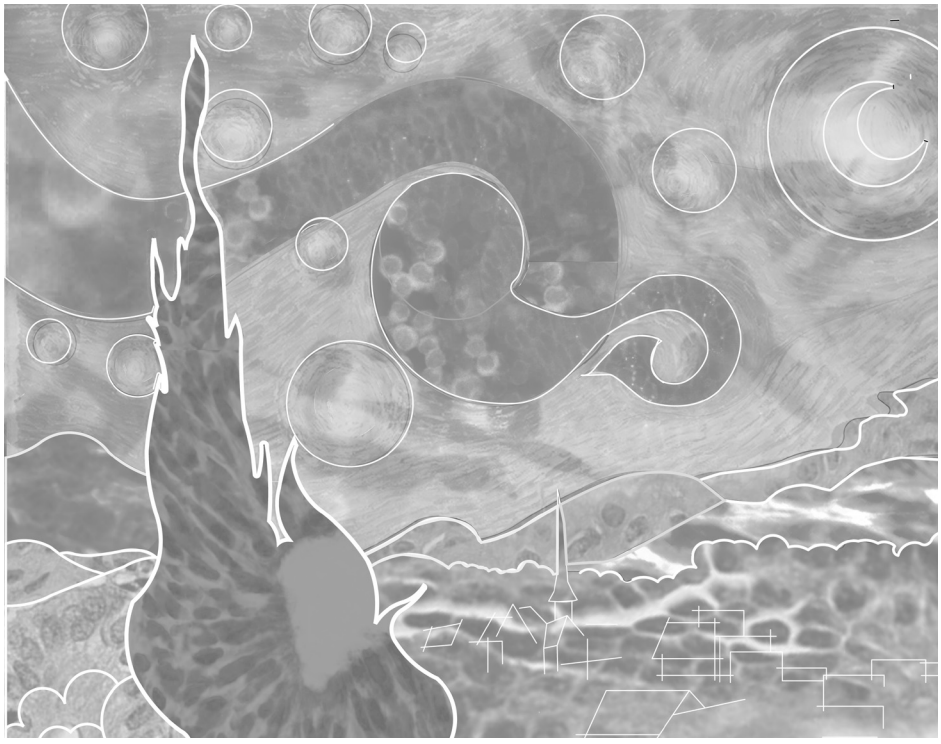
Disclosures

The authors have no financial conflict of interest.

References

1. Reiner, S. L. 2007. Development in motion: helper T cells at work. *Cell* 129: 33–36.
2. Steinman, L. 2007. A brief history of TH17, the first major revision in the TH1/TH2 hypothesis of T cell-mediated tissue damage. *Nat. Med.* 13: 139–145.
3. Szabo, S. J., S. T. Kim, G. L. Costa, X. Zhang, C. G. Fathman, and L. H. Glimcher. 2000. A novel transcription factor, T-bet, directs Th1 lineage commitment. *Cell* 100: 655–669.
4. Afkarian, M., J. R. Sedy, J. Yang, N. G. Jacobson, N. Cereb, S. Y. Yang, T. L. Murphy, and K. M. Murphy. 2002. T-bet is a STAT1-induced regulator of IL-12R expression in naive $CD4^+$ T cells. *Nat. Immunol.* 3: 549–557.
5. Usui, T., J. C. Preiss, Y. Kanno, Z. J. Yao, J. H. Bream, J. J. O'Shea, and W. Strober. 2006. T-bet regulates Th1 responses through essential effects on GATA-3 function rather than on IFNG gene acetylation and transcription. *J. Exp. Med.* 203:755–766.
6. Zheng, W., and R. A. Flavell. 1997. The transcription factor GATA-3 is necessary and sufficient for Th2 cytokine gene expression in CD4 T cells. *Cell* 89: 587–596.
7. Zhang, D. H., L. Cohn, P. Ray, K. Bottomly, and A. Ray. 1997. Transcription factor GATA-3 is differentially expressed in murine Th1 and Th2 cells and controls Th2-specific expression of the interleukin-5 gene. *J. Biol. Chem.* 272: 21597–21603.
8. Ouyang, W., M. Lohning, Z. Gao, M. Assenmacher, S. Ranganath, A. Radbruch, and K. M. Murphy. 2000. Stat6-independent GATA-3 autoactivation directs IL-4-independent Th2 development and commitment. *Immunity* 12: 27–37.
9. Lee, G. R., P. E. Fields, and R. A. Flavell. 2001. Regulation of IL-4 gene expression by distal regulatory elements and GATA-3 at the chromatin level. *Immunity* 14: 447–459.
10. Zhu, J., B. Min, J. Hu-Li, C. J. Watson, A. Grinberg, Q. Wang, N. Killeen, J. F. Urban, Jr., L. Guo, and W. E. Paul. 2004. Conditional deletion of Gata3 shows its essential function in TH1-TH2 responses. *Nat. Immunol.* 5: 1157–1165.
11. Ivanov, I. I., B. S. McKenzie, L. Zhou, C. E. Tadokoro, A. Lepelley, J. J. Lafaille, D. J. Cua, and D. R. Littman. 2006. The orphan nuclear receptor RORγt directs the differentiation program of proinflammatory IL-17^+ T helper cells. *Cell* 126: 1121–1133.
12. Yang, X. O., B. P. Pappu, R. Nurieva, A. Akimzhanov, H. S. Kang, Y. Chung, L. Ma, B. Shah, A. D. Panopoulos, K. S. Schluns, et al. 2008. T helper 17 lineage differentiation is programmed by orphan nuclear receptors RORα and RORγ. *Immunity* 28: 29–39.
13. Bettelli, E., Y. Carrier, W. Gao, T. Korn, T. B. Strom, M. Oukka, H. L. Weiner, and V. K. Kuchroo. 2006. Reciprocal developmental pathways for the generation of pathogenic effector TH17 and regulatory T cells. *Nature* 441: 235–238.
14. Grogan, J. L., M. Mohrs, B. Harmon, D. A. Lacy, J. W. Sedat, and R. M. Locksley. 2001. Early transcription and silencing of cytokine genes underlie polarization of T helper cell subsets. *Immunity* 14: 205–215.

15. Ansel, K. M., R. J. Greenwald, S. Agarwal, C. H. Bassing, S. Monticelli, J. Interlandi, I. M. Djuretic, D. U. Lee, A. H. Sharpe, F. W. Alt, and A. Rao. 2004. Deletion of a conserved Il4 silencer impairs T helper type 1-mediated immunity. *Nat. Immunol.* 5: 1251–1259.
16. Spilianakis, C. G., and R. A. Flavell. 2004. Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat. Immunol.* 5: 1017–1027.
17. Ansel, K. M., I. Djuretic, B. Tanasa, and A. Rao. 2006. Regulation of Th2 differentiation and Il4 locus accessibility. *Annu. Rev. Immunol.* 24: 607–656.
18. Lee, G. R., S. T. Kim, C. G. Spilianakis, P. E. Fields, and R. A. Flavell. 2006. T helper cell differentiation: regulation by cis elements and epigenetics. *Immunity* 24: 369–379.
19. Lee, G. R., C. G. Spilianakis, and R. A. Flavell. 2005. Hypersensitive site 7 of the TH2 locus control region is essential for expressing TH2 cytokine genes and for long-range intrachromosomal interactions. *Nat. Immunol.* 6: 42–48.
20. Lee, H. J., N. Takemoto, H. Kurata, Y. Kamogawa, S. Miyatake, A. O'Garra, and N. Arai. 2000. GATA-3 induces T helper cell type 2 (Th2) cytokine expression and chromatin remodeling in committed Th1 cells. *J. Exp. Med.* 192:105–115.
21. Cai, S., C. C. Lee, and T. Kohwi-Shigematsu. 2006. SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat. Genet.* 38: 1278–1288.
22. Ohlsson, R., R. Renkawitz, and V. Lobanenkov. 2001. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet.* 17: 520–527.
23. Wallace, J. A., and G. Felsenfeld. 2007. We gather together: insulators and genome organization. *Curr. Opin. Genet. Dev.* 17: 400–407.
24. Bell, A. C., A. G. West, and G. Felsenfeld. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98: 387–396.
25. Splinter, E., H. Heath, J. Kooren, R. J. Palstra, P. Klous, F. Grosveld, N. Galjart, and W. de Laat. 2006. CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus. *Genes Dev.* 20: 2349–2354.
26. West, A. G., S. Huang, M. Gaszner, M. D. Litt, and G. Felsenfeld. 2004. Recruitment of histone modifications by USF proteins at a vertebrate barrier element. *Mol. Cell* 16: 453–463.
27. Gomos-Klein, J., F. Harrow, J. Alarcon, and B. D. Ortiz. 2007. CTCF-independent, but not CTCF-dependent, elements significantly contribute to TCR-α locus control region activity. *J. Immunol.* 179: 1088–1095.
28. Kim, T. H., Z. K. Abdullaev, A. D. Smith, K. A. Ching, D. I. Loukinov, R. D. Green, M. Q. Zhang, V. V. Lobanenkov, and B. Ren. 2007. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 128: 1231–1245.
29. Xie, X., T. S. Mikkelsen, A. Gnirke, K. Lindblad-Toh, M. Kellis, and E. S. Lander. 2007. Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. *Proc. Natl. Acad. Sci. USA* 104: 7145–7150.
30. Barski, A., S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev, and K. Zhao. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* 129: 823–837.
31. Garrett, F. E., A. V. Emelyanov, M. A. Sepulveda, P. Flanagan, S. Volpi, F. Li, D. Loukinov, L. A. Eckhardt, V. V. Lobanenkov, and B. K. Birshtein. 2005. Chromatin architecture near a potential 3′ end of the igh locus involves modular regulation of histone modifications during B-Cell development and in vivo occupancy at CTCF sites. *Mol. Cell. Biol.* 25: 1511–1525.
32. Magdinier, F., T. M. Yusufzai, and G. Felsenfeld. 2004. Both CTCF-dependent and -independent insulators are found between the mouse T cell receptor α and Dad1 genes. *J. Biol. Chem.* 279: 25381–25389.
33. Majumder, P., J. A. Gomez, B. P. Chadwick, and J. M. Boss. 2008. The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. *J. Exp. Med.* 205: 785–798.
34. Heath, H., C. Ribeiro de Almeida, F. Sleutels, G. Dingjan, S. van de Nobelen, I. Jonkers, K.-W. Ling, J. Gribnau, R. Renkawitz, F. Grosveld, et al. 2008. CTCF regulates cell cycle progression of α-β T cells in the thymus. *EMBO J.* 27: 2839–2850.
35. Lee, P. P., D. R. Fitzpatrick, C. Beard, H. K. Jessup, S. Lehar, K. W. Makar, M. Perez-Melgosa, M. T. Sweetser, M. S. Schlissel, S. Nguyen, et al. 2001. A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival. *Immunity* 15: 763–774.
36. Barnden, M. J., J. Allison, W. R. Heath, and F. R. Carbone. 1998. Defective TCR expression in transgenic mice constructed using cDNA-based α- and β-chain genes under the control of heterologous regulatory elements. *Immunol. Cell Biol.* 76: 34–40.
37. Hoogenraad, C. C., B. Koekkoek, A. Akhmanova, H. Krugers, B. Dortland, M. Miedema, A. van Alphen, W. M. Kistler, M. Jaegle, M. Koutsourakis, et al. 2002. Targeted mutation of Cyln2 in the Williams syndrome critical region links CLIP-115 haploinsufficiency to neurodevelopmental abnormalities in mice. *Nat. Genet.* 32: 116–127.
38. Robertson, G., M. Hirst, M. Bainbridge, M. Bilenky, Y. Zhao, T. Zeng, G. Euskirchen, B. Bernier, R. Varhol, A. Delaney, et al. 2007. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* 4: 651–657.
39. Hendriks, R. W., M. F. de Bruijn, A. Maas, G. M. Dingjan, A. Karis, and F. Grosveld. 1996. Inactivation of Btk by insertion of lacZ reveals defects in B cell development only past the pre-B cell stage. *EMBO J.* 15: 4862–4872.
40. Lyons, A. B., and C. R. Parish. 1994. Determination of lymphocyte division by flow cytometry. *J. Immunol. Methods* 171: 131–137.
41. Nawijn, M. C., G. M. Dingjan, R. Ferreira, B. N. Lambrecht, A. Karis, F. Grosveld, H. Savelkoul, and R. W. Hendriks. 2001. Enforced expression of GATA-3 in transgenic mice inhibits Th1 differentiation and induces the formation of a T1/ST2-expressing Th2-committed T cell compartment in vivo. *J. Immunol.* 167: 724–732.
42. Chen, X., H. Xu, P. Yuan, F. Fang, M. Huss, V. B. Vega, E. Wong, Y. L. Orlov, W. Zhang, J. Jiang, et al. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133: 1106–1117.
43. Takemoto, N., Y. Kamogawa, H. Jun Lee, H. Kurata, K. I. Arai, A. O'Garra, N. Arai, and S. Miyatake. 2000. Cutting edge: chromatin remodeling at the IL-4/IL-13 intergenic regulatory region for Th2-specific cytokine gene cluster. *J. Immunol.* 165: 6687–6691.
44. Mohrs, M., C. M. Blankespoor, Z. E. Wang, G. G. Loots, V. Afzal, H. Hadeiba, K. Shinkai, E. M. Rubin, and R. M. Locksley. 2001. Deletion of a coordinate regulator of type 2 cytokine expression in mice. *Nat. Immunol.* 2: 842–847.
45. Wolfer, A., T. Bakker, A. Wilson, M. Nicolas, V. Ioannidis, D. R. Littman, P. P. Lee, C. B. Wilson, W. Held, H. R. MacDonald, and F. Radtke. 2001. Inactivation of Notch 1 in immature thymocytes does not perturb CD4 or CD8T cell development. *Nat. Immunol.* 2: 235–241.
46. Baier, G. 2003. The PKC gene module: molecular biosystematics to resolve its T cell functions. *Immunol. Rev.* 192: 64–79.
47. Nelson, B. H., and D. M. Willerford. 1998. Biology of the interleukin-2 receptor. *Adv. Immunol.* 70: 1–81.
48. Hori, S., T. Nomura, and S. Sakaguchi. 2003. Control of regulatory T cell development by the transcription factor Foxp3. *Science* 299: 1057–1061.
49. Usui, T., R. Nishikomori, A. Kitani, and W. Strober. 2003. GATA-3 suppresses Th1 development by downregulation of Stat4 and not through effects on IL-12Rβ2 chain or T-bet. *Immunity* 18: 415–428.
50. Nel, A. E. 2002. T-cell activation through the antigen receptor, part 1: signaling components, signaling pathways, and signal integration at the T-cell antigen receptor synapse. *J. Allergy Clin. Immunol.* 109: 758–770.
51. Nel, A. E., and N. Slaughter. 2002. T-cell activation through the antigen receptor, part 2: role of signaling cascades in T-cell differentiation, anergy, immune senescence, and development of immunotherapy. *J. Allergy Clin. Immunol.* 109: 901–915.
52. Yusufzai, T. M., H. Tagami, Y. Nakatani, and G. Felsenfeld. 2004. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol. Cell* 13: 291–298.
53. Dickinson, L. A., T. Joh, Y. Kohwi, and T. Kohwi-Shigematsu. 1992. A tissue-specific MAR/SAR DNA-binding protein with unusual binding site recognition. *Cell* 70: 631–645.
54. de Belle, I., S. Cai, and T. Kohwi-Shigematsu. 1998. The genomic sequences bound to special AT-rich sequence-binding protein 1 (SATB1) in vivo in Jurkat T cells are tightly associated with the nuclear matrix at the bases of the chromatin loops. *J. Cell Biol.* 141: 335–348.
55. Ishihara, K., M. Oshimura, and M. Nakao. 2006. CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Mol. Cell* 23: 733–742.
56. Lutz, M., L. J. Burke, G. Barreto, F. Goeman, H. Greb, R. Arnold, H. Schultheiss, A. Brehm, T. Kouzarides, V. Lobanenkov, and R. Renkawitz. 2000. Transcriptional repression by the insulator protein CTCF involves histone deacetylases. *Nucleic Acids Res.* 28: 1707–1713.
57. Chernukhin, I., S. Shamsuddin, S. Y. Kang, R. Bergstrom, Y. W. Kwon, W. Yu, J. Whitehead, R. Mukhopadhyay, F. Docquier, D. Farrar, et al. 2007. CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide. *Mol. Cell. Biol.* 27: 1631–1648.



Chapter 6 General discussion and future perspectives

GENERAL DISCUSSION AND FUTURE PERSPECTIVES

The morphological and functional diversity of multi-cellular organisms is a result of the precise and complex regulation of the genetic information at different stages of cellular differentiation. Hence, different levels of control must exist to maintain the proper spatio-temporal expression of genes throughout the process of cellular differentiation. In order to control this process, the eukaryotic genomes evolved several regulatory DNA elements that ensure proper gene regulation. It is important to understand how these *cis*-regulatory elements communicate over large distances to control gene expression. The main focus of the research described in this thesis was the study of several factors that have a role in long-range interactions and chromatin organization.

In this Chapter, I will discuss the specific functional interaction between Delangin and Cohesin in more detail to provide a better insight of their involvement in gene regulation. A growing number of Cohesins and their regulatory proteins have been associated with human developmental disorders that are therefore also referred to as "cohesinopathies". Cornelia de Lange Syndrome, caused by mutations in *Delangin*, or in rare cases in *Cohesin* subunits, is a heterogeneous developmental disorder. The complex phenotype seen in humans and in Delangin mouse model (described in Chapter 3) suggests that Delangin is involved in different developmental processes and provides new clues to the underlying molecular mechanisms of this disorder.

Although Cohesin was discovered as a protein that is required for sister chromatid cohesion in mitotic and meiotic cells, recent work provides clear evidence that at least some of the functions of Cohesin are independent of its role in cohesion. Interestingly, Cohesin shares binding sites in the genome with another multifunctional protein, CTCF. The functional relevance of Cohesin and CTCF binding to the same sites in the genome is still unclear.

The role of Delangin in long-range interactions and more

Delangin interactors

Delangin was initially identified as a developmental regulator in *Drosophila*. A genetic screen aimed to isolate genes involved in long- range interactions between promoters and remote enhancers identified *Nipped-B*, a regulator of the *Ultrabithorax* and *cut* genes (Rollins et al., 1999). Subsequent work revealed that Nipped-B is also required for sister chromatid cohesion (Rollins et al., 2004) and it was shown that Delangin and its orthologues are required for Cohesin to bind to chromosomes (Arumugam et al., 2003; Ciosk et al., 2000; Gillespie and Hirano, 2004; Seitan et al., 2006).

The role of Delangin in *Drosophila* in long-range gene regulation triggered our interest in this protein. The fact that Delangin acts through Cohesin, a protein that is involved in many diverse cellular processes, makes it complicated to reveal a specific function in a distinct process. The complexity of the processes involving Delangin requires both, biochemical and genetic approaches to clarify the function of Delangin in mammals. The genetic approach is described in Chapter 3, while a biochemical approach was not advanced enough to be included and will be described in this Chapter.

Protein-protein interactions are crucial in most biological processes, including gene regulation and cell cycle progression. Most proteins function as multiprotein complexes or interact with multiprotein complexes. Identification of protein-protein interactions in the context of their physiologically relevant complexes is therefore a key to fully understand the cellular machinery. Many experimental approaches, including yeast two-hybrid analysis, mass spectrometry and affinity purification have been developed for the detection and identification of interacting proteins in a cellular context. However all these methods can result in considerable numbers of false positive and negative interactions.

Fractionation profiles of nuclear extracts isolated from murine erythroid leukemia (MEL) cells show Delangin to be present in complexes of high molecular weight. In order to identify Delangin protein partners, we used streptavidin pulldown of a biotin-tagged Delangin fusion protein. The method of de Boer and colleagues (de Boer et al., 2003; Rodriguez et al., 2005) was used to generate cDNA constructs coding for fusion proteins with a 23-amino acid N-terminal tag that can be biotinylated *in vivo*. Because of the large protein size, we were unable to express a full-length bio-tagged Delangin in MEL cells. We therefore designed a series of constructs that collectively specified overlapping fragments from the conserved C-terminal region of Delangin. Both isoforms described of mammalian Delangin are large and only a few functional domains have been identified. A C-terminal domain, that is strongly conserved and expected to be functionally significant, was included in this analysis. This domain contains several HEAT repeats, motifs that have been implicated in protein-protein interactions.

Delangin has been reported to interact with several proteins of different functions. We could also detect heterochromatin protein 1 (HP1), a known interacting partner of Delangin (Lechner et al., 2005). To validate Delangin interactions, we carried out co-immunoprecipitation experiments on the endogenous Delangin and putative interactors. Unfortunately, in the reciprocal screen several of these did not appear to interact with Delangin but two of them did: Topoisomerase II α and Metastasis Tumor Antigen 1 (MTA1). What could be the functional relevance of Delangin interacting with these two proteins?

Metastasis tumor antigen 1 (MTA1), a candidate member of the metastasis-associated gene family, is a component of the nucleosome remodeling and deacetylating (NuRD) complex. The NuRD complex has been implicated in ATP-dependent chromatin remodeling and transcriptional regulation. As a part of the NuRD complex, MTA1 is thought to modulate transcription by influencing chromatin remodeling (Kumar et al., 2003; Toh et al., 1994). We were able to detect an interaction of endogenous Delangin only with MTA1, while the other components of the NurRD chromatin remodeling complex did not interact with Delangin in our experimental setup. A recent study (Jahnke et al., 2008) showed that Delangin forms stable complexes with endogenous HDAC1 and HDAC3 (also components of NuRD complex) in mammalian cells. Our data, in combination with the HDAC1/3 interaction data indicate a molecular connection with the chromatin remodeling machinery, although we still miss functional evidence to that effect.

The discrepancy between an interaction of Delangin with MTA1 (in our experiments) or HDAC1/3 (Jahnke et al., 2008) may be due to different direct interactions within two different systems. Another interesting set of interactions is found between Cohesin and a Chromatin-remodeling complex containing the ATPase SNF2h in human cells (Hakimi et al., 2002). Although the interplay of chromatin structure and histone modification with Cohesin has been described, its precise function is not understood. We can only speculate that chromatin structure may play an important role in determining where Delangin binds to chromosomes in eukaryotic cells. Does Delangin recognize chromatin regions that have a particular chromatin organization that allows Delangin and subsequent Cohesin binding? Alternetively, Delangin triggers changes in chromatin structure that allow recruitment of Cohesin to specific sites. It would be interesting to investigate these possibilities in mammalian cells.

In addition to its role in histone deacetylation and a strong correlation with metastatic potential (Nicolson, 1998), MTA1, a ubiquitously expressed protein, is also involved in the regulation of other cellular pathways including hormonal action, protein stability and development by modifying the acetylation status of crucial target genes (Manavathi and Kumar, 2007). For example, it was shown that MTA1 drastically represses transcription of the BRCA1 gene by associating with the previously identified ERE in the BRCA1 gene promoter via ER α (Molli et al., 2008). It would be interesting to identify additional targets of MTA1 and correlate them with Delangin targets in the genome since the direct interaction of Delangin and MTA1 indicates that they might act together to regulate gene expression. The availability of complete genome sequences and high-throughput analysis techniques have broadened the focus from

single interactions to a more complex picture. ChIP-sequencing, which combines chromatin immunoprecipitation (ChIP) with massively parallel DNA sequencing, can be used to precisely map global binding sites for any protein of interest. This technique gives numerous opportunities for identification of binding sites of many factors that bind directly or indirectly to DNA. In the case of indirect binding, the use of protein-protein cross-linkers, for example disuccinimidyl glutarate (DSG), is required before DNA-protein fixation. Mapping of binding sites of diverse factors will provide a better understanding of functional protein complexes and their formation in the crowded cellular environment.

DNA Topoisomerase II α is another protein that was identified by mass-spectrometry and confirmed to interact with endogenous Delangin. Interestingly, coimmunoprecipitation experiments indicate that HDAC1 and HDAC2 are also associated with Topo II α *in vivo* under normal physiological condition. Complexes containing Topo II α possess HDAC activity, and complexes containing HDAC1 and HDAC2 possess Topo II activity (Tsai et al., 2000). This molecule can be another possible link of the Delangin with its targets at the genomic level.

Cohesinopathies and Delangin

Recently, a growing number of Cohesins and its regulatory proteins have been associated with human developmental disorders that are therefore also referred to as "cohesinopathies". The most frequent of these diseases is Down syndrome (trisomy 21) which, in the majority of cases, is caused by missegregation of chromosome 21 during meiosis I in oocytes (Gilliland and Hawley, 2005). Maternal age correlates with an increased frequency of unpaired chromosomes and precociously separated sister chromatids suggesting that the defects in sister chromatid cohesion may be one of the causes of nondisjunction (Pellestor et al., 2003). Experimental support comes from the observation that mice in which the meiosis-specific Smc1β Cohesin subunit was mutated show age related defects in oocytes (Hodges et al., 2005). It is therefore possible that defects in Cohesin, or Cohesin regulators including Delangin, are a major cause of Down syndrome. Why these effects are age related and how they occur is still unclear. Since, in females, cohesion is established during prenatal development, albeit has to persist for several decades. During this time, Cohesin is probably reloaded many times since it is hard to believe that Cohesin is stably bound for such a long time. In this case, Delangin loads Cohesin de novo during this period. If so, it would be interesting to address if and how Delangin involved is in this process.

Cornelia de Lange Syndrome (CdLS) is associated with mutations in the Delangin gene in humans. Chapter 3 focuses on a description of the Delangin KO (Knockout) mouse that shows a complex phenotype similar to that seen in Cornelia de Lange patients. Surprisingly, cell proliferation and sister chromatid cohesion appear to be normal in cells derived from Delangin-deficient mice. We therefore suggest that the

abnormalities present in Delangin-deficient mice and CdLS patients are caused by Delangin functions other than sister chromatid cohesion.

A reduced dosage of Delangin could result in increased apoptosis during early development, with a loss of key tissue progenitor cells as a result. Alternatively, the partial loss of Delangin function could result in significant transcriptional deregulation of many genes. Hence, it is important to identify the genes whose transcription is altered when Delangin function is impaired or the dose is reduced. It is hard to predict which genes might be the targets of Delangin loss-of-function since role of the Delangin protein is probably as a non-sequence-specific chromosomal protein which regulates DNA structure to facilitate transcription of many genes. Therefore, it is reasonable to expect that genes with an altered level are not related in a functional way but rather in structural way to Delangin. For example, by having regulatory enhancers near to Cohesin binding sites.

Whole-genome mapping of Delangin binding sites in combination with transcriptional profiling should reveal target genes that are up- or down-regulated in Delangin-deficient mice and hopefully provide insight into the defects in human CdLS patients. The choice of tissue for RNA extraction is important since the transcriptional changes should be monitored in the organs before they are affected or in organs that are not affected since it is necessary to avoid the results of ongoing pathological changes. The importance of the identification of the molecular basis for CdLS will permit accurate DNA-based diagnostics. Some understanding of the genetic basis of CdLS has been obtained from the patients, but for the development of a genetic test and therapeutic treatments studies in transgenic animals are very important.

Cohesin and Delangin binding sites in eukaryotic genomes

The data from yeast, *Drosophila* and mammalian cells reveal a striking diversity in the targeting mechanisms of Cohesin and Delangin. The distribution of Cohesin appears drastically different in these organisms, and the positions do not coincide with the position of Delangin that loads it onto the chromosomes.

Cohesin-binding sites were first identified in the genome of the budding yeast. ChIP-on-chip experiments revealed that Cohesin binds to discrete Cohesin attachment regions (CARs) on chromosome arms and to larger domains in the centromeric regions (Tanaka et al., 1999). The arm sites are on average 10 kb apart from each other (Laloraya et al., 2000). Although CARs do not have specific sequence elements they are enriched for A/T (Blat and Kleckner, 1999) and are positioned in a particular manner in intergenic regions at sites of convergent transcription in budding yeast (Glynn et al., 2004; Lengronne et al., 2004). One would expect that Delangin required for Cohesin loading onto DNA, would share a binding motif with Cohesin. ChIP-on-Chip experiments, in combination with microscopic observations, revealed that the distribution of Cohesin differs from Delangin's binding pattern in yeast (Lengronne et al., 2004) and probably also in mammalian cells.

An interesting explanation of these results is that Delangin and/or Cohesin are relocated from the initial binding site once the loading reaction takes place. One possibility is that Cohesin is removed from DNA and subsequently reloaded. This scenario is less probable since Delangin would be needed for reloading of Cohesin and that is not the case since Delangin is not found at the final destination. Another possibility is that Cohesin is relocated from a Delangin site by, for example RNA polymerase II transcriptional machinery (Bausch et al., 2007; Lengronne et al., 2004) to the sites of convergent transcription. This is possible only if Cohesin rings can slide along the chromatin fiber. To reveal a mechanism that can explain Cohesin relocation from one site to another, it is necessary to understand how Cohesin interacts with a DNA molecule.

Initial models suggested that the ATPase heads of SMC1 and SMC3 may directly contact DNA (Nasmyth et al., 2000). Consistent with this possibility, recombinant fragments of SMC3 and purified Cohesin core complexes can bind to DNA *in vitro* (Losada and Hirano, 2001) with a low affinity and in an ATP independent manner unlike Cohesin loading *in vivo*. Therefore, it is unknown whether these interactions reflect how Cohesin associates with DNA *in vivo*. According to the "ring" model, Cohesin mediates sister chromatid cohesion by embracing sister chromatids. It has been suggested that a tripartite ring formed by SMC1, SMC3 and Scc1/Rad21 could encircle two 10-nm chromatin fibers not contacting DNA directly (Haering et al., 2002). This hypothesis is attractive since it can explain why proteolytic cleavage of SMC3 or Scc1/Rad21 causes dissociation of Cohesin from DNA (Gruber et al., 2003) and why Cohesin binds to DNA in a sequence independent manner in yeast cells. Furthermore, if Cohesin subunits are covalently linked to each other, Cohesin cannot associate with DNA meaning that a Cohesin ring has to be opened to be loaded onto DNA (Gruber et al., 2006).

If Cohesin encircles chromatin fibers by embracing it without having preference for particular sequences, it is possible to explain the relocation of Cohesin from the initial loading sites to its final destination without opening the ring. Furthermore, it would be interesting to find out why Cohesin is relocated. Is it because it presents an obstacle for transcription or it is relocated to facilitate transcription? There are two possible explanations and in both cases Cohesin relocation is a consequence of the transcription machinery passing through. Transcriptional elongation may be directly responsible for displacing Cohesin, that is initially loaded randomly but transcription (and possible other metabolic processes) may push Cohesin to regions that are not actively transcribed.

In *Drosophila*, Cohesin preferentially binds to transcribed regions, opposite to the situation in yeast. Also, in contrast to the situation in budding yeast, the *Drosophila* Delangin ortholog Nipped-B has been reported to colocalize with Cohesin on chromosomes (Misulovin et al., 2008).

In mammalian cells, Delangin binding sites are yet to be identified, but it would be interesting to see whether there is an overlap with the Cohesin binding pattern. Chip-sequencing, a great tool for mapping global binding sites of any protein of interest could be used to profile Delangin binding sites. For Cohesin, yet a different binding

pattern has been observed when compared to that in *Drosophila* (Parelho et al., 2008; Wendt et al., 2008). In human cells, Cohesin sites are present in introns (35%), in intergenic regions (50%) and in the regions that are directly up-stream or down-stream from the genes (13%). Cohesin sites are enriched in the regions adjacent to the genes (Wendt et al., 2008), favoring the possibility that these sites have a role in gene regulation. In addition, there is no evidence that transcription relocates Cohesin like in yeast (Parelho et al., 2008; Wendt et al., 2008).

A series of recent studies gave exciting new insights into the targeting mechanisms and gene regulatory functions of Cohesin (Parelho et al., 2008; Rubio et al., 2008; Stedman et al., 2008; Wendt et al., 2008). The recent mapping of CTCF and Cohesin binding sites in several systems, including the erythroid lineage (unpublished data), throughout mammalian genomes can be used to address the question of the importance of CTCF and Cohesin cooperation to regulate gene transcription through organizing chromatin structure. However, a convincing experiment that shows a direct interaction between these two factors is still missing (van de Nobelen unpubl.). It has been reported that CTCF interacts with Cohesin directly (Stedman et al., 2008) but such an interaction could not be detected in several other studies including ours. It is possible that the interaction is weak or Cohesin interacts with CTCF transiently. Alternatively, indirect interactions via other proteins or DNA might be necessary to mediate their communication. More sensitive methods in combination with Chip-sequencing should be used to answer this question.

The knockdown of CTCF substantially reduced the association of the Cohesin complex with its genomic target loci but not the overall binding on the chromatin (Wendt et al., 2008). This observation suggests that CTCF binding sites are “meeting points” for those two factors. Depletion of CTCF binding motifs abolishes the association (Stedman et al., 2008) while Cohesin depletion had no (Parelho et al., 2008) or a moderate (Wendt et al., 2008) effect on CTCF binding.

The next step is of course to determine the functional relevance of CTCF and Cohesin shared binding sites with respect to gene expression. CTCF was first described for its role in regulation of the c-myc gene (Lobanenkov et al., 1990), an its role in insulation was later demonstrated for multiple vertebrate insulators, including the β-globin and *H19/Igf2* locus. Using reporter assays, Wendt et al. showed that Cohesin regulates the endogenous *H19/Igf2* insulator in HeLa cells.

One question is how Cohesin contributes to the insulator function? One of the possibilities is that Cohesin stabilizes long-range chromosomal interactions mediated by CTCF. A generally accepted idea is that CTCF forms chromosome loops that facilitate some enhancer-promoter interactions and prevent others (Wallace and Felsenfeld, 2007; Williams and Flavell, 2008). The Chromosome Conformation Capture (3C) assays can reveal if two distant DNA fragments are held in close proximity by crosslinked protein. Using this technique, several laboratories demonstrated CTCF-dependent loops in different loci including the mouse β-globin locus (Splinter et al., 2006). Since Cohesin accumulates at CTCF-binding sites, it is tempting to speculate that Cohesin stabilizes the CTCF-mediated interaction probably by encircling CTCF formed loops. It would be interesting to investigate the

possibility that some proportion of the Cohesin molecules is diverted to support or regulate long-range interchromosomal loops. Performing 3C assays in the set up where Cohesin and/or CTCF are depleted should answer this question. Even though the mechanism remains to be clarified, it seems that Cohesin and Delangin have important and broad functions in gene regulation and development.

REFERENCE

Arumugam, P., Gruber, S., Tanaka, K., Haering, C. H., Mechtler, K. and Nasmyth, K. (2003). ATP hydrolysis is required for cohesin's association with chromosomes. Curr Biol 13, 1941-53.

Bausch, C., Noone, S., Henry, J. M., Gaudenz, K., Sanderson, B., Seidel, C. and Gerton, J. L. (2007). Transcription alters chromosomal locations of cohesin in Saccharomyces cerevisiae. Mol Cell Biol 27, 8522-32.

Blat, Y. and Kleckner, N. (1999). Cohesins bind to preferential sites along yeast chromosome III, with differential regulation along arms versus the centric region. Cell 98, 249-59.

Ciosk, R., Shirayama, M., Shevchenko, A., Tanaka, T., Toth, A., Shevchenko, A. and Nasmyth, K. (2000). Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. Mol Cell 5, 243-54.

de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., Grosveld, F. and Strouboulis, J. (2003). Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. Proc Natl Acad Sci U S A 100, 7480-5.

Gillespie, P. J. and Hirano, T. (2004). Scc2 couples replication licensing to sister chromatid cohesion in Xenopus egg extracts. Curr Biol 14, 1598-603.

Gilliland, W. D. and Hawley, R. S. (2005). Cohesin and the maternal age effect. Cell 123, 371-3.

Glynn, E. F., Megee, P. C., Yu, H. G., Mistrot, C., Unal, E., Koshland, D. E., DeRisi, J. L. and Gerton, J. L. (2004). Genome-wide mapping of the cohesin complex in the yeast Saccharomyces cerevisiae. PLoS Biol 2, E259.

Gruber, S., Arumugam, P., Katou, Y., Kuglitsch, D., Helmhart, W., Shirahige, K. and Nasmyth, K. (2006). Evidence that loading of cohesin onto chromosomes involves opening of its SMC hinge. Cell 127, 523-37.

Gruber, S., Haering, C. H. and Nasmyth, K. (2003). Chromosomal cohesin forms a ring. Cell 112, 765-77.

Haering, C. H., Lowe, J., Hochwagen, A. and Nasmyth, K. (2002). Molecular architecture of SMC proteins and the yeast cohesin complex. Mol Cell 9, 773-88.

Hakimi, M. A., Bochar, D. A., Schmiesing, J. A., Dong, Y., Barak, O. G., Speicher, D. W., Yokomori, K. and Shiekhattar, R. (2002). A chromatin remodelling complex that loads cohesin onto human chromosomes. Nature 418, 994-8.

Hodges, C. A., Revenkova, E., Jessberger, R., Hassold, T. J. and Hunt, P. A. (2005). SMC1beta-deficient female mice provide evidence that cohesins are a missing link in age-related nondisjunction. Nat Genet 37, 1351-5.

Jahnke, P., Xu, W., Wulling, M., Albrecht, M., Gabriel, H., Gillessen-Kaesbach, G. and Kaiser, F. J. (2008). The Cohesin loading factor NIPBL recruits histone deacetylases to mediate local chromatin modifications. Nucleic Acids Res 36, 6450-8.

Kumar, R., Wang, R. A. and Bagheri-Yarmand, R. (2003). Emerging roles of MTA family members in human cancers. Semin Oncol 30, 30-7.

Laloraya, S., Guacci, V. and Koshland, D. (2000). Chromosomal addresses of the cohesin component Mcd1p. J Cell Biol 151, 1047-56.

Lechner, M. S., Schultz, D. C., Negorev, D., Maul, G. G. and Rauscher, F. J., 3rd. (2005). The mammalian heterochromatin protein 1 binds diverse nuclear proteins through a common motif that targets the chromoshadow domain. Biochem Biophys Res Commun 331, 929-37.

Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G. P., Itoh, T., Watanabe, Y., Shirahige, K. and Uhlmann, F. (2004). Cohesin relocation from sites of chromosomal loading to places of convergent transcription. Nature 430, 573-8.

Lobanenkov, V. V., Nicolas, R. H., Adler, V. V., Paterson, H., Klenova, E. M., Polotskaja, A. V. and Goodwin, G. H. (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, 1743-53.

Losada, A. and Hirano, T. (2001). Intermolecular DNA interactions stimulated by the cohesin complex in vitro: implications for sister chromatid cohesion. Curr Biol 11, 268-72.

Manavathi, B. and Kumar, R. (2007). Metastasis tumor antigens, an emerging family of multifaceted master coregulators. J Biol Chem 282, 1529-33.

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

Molli, P. R., Singh, R. R., Lee, S. W. and Kumar, R. (2008). MTA1-mediated transcriptional repression of BRCA1 tumor suppressor gene. Oncogene 27, 1971-80.

Nasmyth, K., Peters, J. M. and Uhlmann, F. (2000). Splitting the chromosome: cutting the ties that bind sister chromatids. Science 288, 1379-85.

Nicolson, G. L. (1998). Breast cancer metastasis-associated genes: role in tumour progression to the metastatic state. Biochem Soc Symp 63, 231-43.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H. C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T. et al. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell 132, 422-33.

Pellestor, F., Andreo, B., Arnal, F., Humeau, C. and Demaille, J. (2003). Maternal aging and chromosomal abnormalities: new data drawn from in vitro unfertilized human oocytes. Hum Genet 112, 195-203.

Rodriguez, P., Bonte, E., Krijgsveld, J., Kolodziej, K. E., Guyot, B., Heck, A. J., Vyas, P., de Boer, E., Grosveld, F. and Strouboulis, J. (2005). GATA-1 forms distinct activating and repressive complexes in erythroid cells. Embo J 24, 2354-66.

Rollins, R. A., Korom, M., Aulner, N., Martens, A. and Dorsett, D. (2004). Drosophila nipped-B protein supports sister chromatid cohesion and opposes the stromalin/Scc3 cohesion factor to facilitate long-range activation of the cut gene. Mol Cell Biol 24, 3100-11.

Rollins, R. A., Morcillo, P. and Dorsett, D. (1999). Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. Genetics 152, 577-93.

Rubio, E. D., Reiss, D. J., Welcsh, P. L., Disteche, C. M., Filippova, G. N., Baliga, N. S., Aebersold, R., Ranish, J. A. and Krumm, A. (2008). CTCF physically links cohesin to chromatin. Proc Natl Acad Sci U S A 105, 8309-14.

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

Splinter, E., Heath, H., Kooren, J., Palstra, R. J., Klous, P., Grosveld, F., Galjart, N. and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev 20, 2349-54.

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

Embo J 27, 654-66.

Tanaka, T., Cosma, M. P., Wirth, K. and Nasmyth, K. (1999). Identification of cohesin association sites at centromeres and along chromosome arms. Cell 98, 847-58.

Toh, Y., Pencil, S. D. and Nicolson, G. L. (1994). A novel candidate metastasis-associated gene, mta1, differentially expressed in highly metastatic mammary adenocarcinoma cell lines. cDNA cloning, expression, and protein analyses. J Biol Chem 269, 22958-63.

Tsai, S. C., Valkov, N., Yang, W. M., Gump, J., Sullivan, D. and Seto, E. (2000). Histone deacetylase interacts directly with DNA topoisomerase II. Nat Genet 26, 349-53.

Wallace, J. A. and Felsenfeld, G. (2007). We gather together: insulators and genome organization. Curr Opin Genet Dev 17, 400-7.

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

Williams, A. and Flavell, R. A. (2008). The role of CTCF in regulating nuclear organization. J Exp Med 205, 747-50.

SUMMARY

Genetic information required for the development and functioning of an organism is stored in billions of base pairs of DNA. In eukaryotes, DNA is organized in large units called chromosomes. These are the essential units for cellular division and must be replicated, divided, and passed successfully to the daughter cells. They are packaged by proteins into a condensed structure, called chromatin, which enables the very long DNA molecules to fit into the cell nucleus. Despite the compacted structure of the DNA, the right set of genes is transcribed. Different levels of control are involved in the regulation of proper gene expression throughout the process of cellular differentiation. Numerous general, cell-type and stage specific proteins that bind to chromatin and DNA coordinate these processes. In the DNA of metazoans, long-range interactions between distant regulatory elements, and the genes, are responsible for the specific expression of many developmentally regulated genes. Often these interactions occur over considerable distances as in the some cases control elements are located hundreds or even thousands of kilobases away from the genes. Over the last few years, there is a growing interest in distant regulatory elements and the factors that are involved in the regulation of the interactions. The main focus, of research described in this thesis, are several factors that have a role in long-range interactions and chromatin organization.

The first chapter gives an introduction of several mechanisms that have been proposed for long distance interactions. The roles of vertebrate insulator protein CTCF (CCCTC binding factor) and Cohesin in long-range interactions are also addressed. In addition the Cohesion loading factor Delangin and the Ldb1 transcription factor complex are introduced in Chapter 1.

In Chapter 2 novel binding partners of Ldb1 are described. Ldb1, a ubiquitously expressed LIM domain binding protein, is known to be essential in several tissues during development. By using systems biology approach we have identified known and many new binding partners of Ldb1. We have also determined their importance for hematopoietic development, especially for the earliest stages.

To better understand the role of Delangin in mammals, we have generated mice lacking this gene that are described in Chapter 3. Delangin is a protein involved in loading of Cohesin onto the chromatin and hence thought to be important for sister chromatid cohesion and chromatin dynamics. Homozygous *Delangin* mouse mutants are lethal early in development while alive, born heterozygous mice exhibit a complex phenotype. This suggests that reduced level of Delangin can cause chromatin defects early in development. The affected mice have a severely reduced body size and abnormalities in many organs. Interestingly, the phenotype resembles to those seen in patients with Cornelia de Lange Syndrome (CdLS) that have mutations in the same gene. We suggest that our mouse model is a valuable model for the understanding of the molecular mechanisms behind Cornelia de Lange Syndrome.

A co-relation of functions of Cohesin, and CTCF protein is presented in the Chapter 4. We show that subset of CTCF binding sites are shared with Cohesin. We have specifically focused on shared binding sites in the mouse β-globin locus in erythroid

cells. We have shown that Cohesin binding is required in a CTCF dependent manner. Since Cohesin and CTCF share several binding sites in this locus, we investigated whether Cohesin is required for the long-range interactions.

An interesting role that CTCF has in the Th2 cytokine expression in T-cells is addressed in the Chapter 5. Chromatin immunoprecipitation (Chip) assays revealed the presence of multiple CTCF binding sites in the Th2 cytokine locus in the T-cells (thymocites). The conditional deletion of *Ctcf* in thymus, which allows this gene to be deleted in spatio-temporally manner, caused Th2 cytokine expression defect.

In Chapter 6, the specific functional interactions between Delangin and Cohesin and the functional relevance of shared Cohesin and CTCF binding sites are discussed in more detail. In this chapter, I suggest that different regulatory mechanisms are involved in their functional interactions. The last chapter is concluded with a number of suggestions for future experiments to further elucidate this area of research.

Samenvatting

De genetische informatie die nodig is voor de ontwikkeling en het functioneren van een organisme is opgeslagen in miljoenen DNA basenparen. In eukaryoten is DNA georganiseerd in grote structuren, genaamd chromosomen. Deze zijn de essentiële eenheden voor celdeling en moeten verdubbelen, delen en op de juiste manier verdeeld worden over de nieuwe dochter cellen. De chromosomen worden door middel van eiwitten verpakt in een compacte structuur, genaamd chromatine. Dit zorgt er voor dat de lange DNA strengen in de celkern passen. Ondanks de compacte structuur van het chromatine, kunnen de juiste genen worden afgelezen. Er zijn verschillende niveaus van controle nodig voor de regulatie van de juiste expressie van genen tijdens het proces van differentiatie. Deze processen zijn gecoördineerd door verscheidene algemene cel en stadium specifieke eiwitten die binden aan chromatine en DNA. In het DNA van metazoans zijn interacties tussen verscheidene regulerende elementen en hun genen verantwoordelijk voor de specifieke expressie van vele ontwikkelingsgereguleerde genen. Vaak vinden deze interacties plaats over grote afstanden, in sommige gevallen liggen de regulerende elementen zelfs honderden en soms wel duizenden kilobasen van de desbetreffende genen verwijderd. Recentelijk is er een groeiende interesse in deze regulerende elementen ontstaan en in de factoren die betrokken zijn bij deze interacties. In dit proefschrift worden een aantal factoren besproken die een rol spelen in interacties over lange afstand en de organisatie van chromatine.

In hoofdstuk 1 worden een aantal mechanismen besproken die van invloed zijn op interacties over lange afstand. De rol van het vertebraat insulator eiwit CTCF (CCCTC binding factor) en Cohesin worden besproken tesamen met Delangin, en transcriptie factor complex Ldb1. Beiden zijn ook betrokken bij interacties over lange afstand.

Hoofdstuk 2 beschrijft de nieuwe bindingspartners van Ldb1. Ldb1 is een overvloedig tot expressie komend LIM domein eiwit, dat betrokken is bij de ontwikkeling van verschillende weefsels. Door gebruik te maken van een systeem biologische benadering hebben wij reeds bekende, maar ook nieuwe bindingpartners van LDB1 geïdentificeerd. Daarnaast is ook de rol van deze nieuwe partners tijdens de hematopoetische ontwikkeling bepaald, en dan met namen in de vroege stadia van de hematopoese.

Om de rol van Delangin in zoogdieren beter te begrijpen, hebben we knock-out muizen gemaakt, die beschreven staan in hoofdstuk 3. Delangin is een eiwit dat betrokken is bij het laden van cohesin op het chromatine en speelt daardoor dus een belangrijke rol in zuster chromatide cohesie en in overall chromatine dynamiek. Muizen met een homozygote Delangin mutatie zijn vroeg embryonaal lethaal, terwijl muizen met een heterozygote mutatie een complex fenotype hebben. Dit suggereert dat een gereduceerd niveau van Delangin al voor chromatine defecten kan zorgen tijdens de vroege ontwikkeling van de muis. De heterozygote muizen zijn een stuk kleiner dan wild-type muizen en hebben afwijkingen in een groot aantal organen. Het fenotype van de Delangin muizen lijkt sterk op dat van Cornelia de Lange Syndroom

(CdLS) patiënten. Deze patiënten hebben een mutatie in hetzelfde gen. Wij denken daarom dat onze muis model een goed model is om het moleculaire mechanismen achter het Cornelia de Lange Syndroom te bestuderen.

In hoofdstuk 4 presenteer ik de co-relatie tussen de functie van Cohesin, bekent om zijn rol in zuster chromatide cohesie, en CTCF, het meest bekende insulator eiwit. We laten zien dat een subpopulatie van CTCF bindingsplaatsen gedeeld wordt met Cohesin. We hebben ons met namen geconcentreerd op de gedeelde bindingsplaatsen binnen het β-globine locus in rode bloedcellen. In deze cellen laten we zien dat cohesin binding afhankelijk is van CTCF . Omdat CTCF en Cohesin een aantal bindingsplaatsen gemeen hebben in de β-globine locus, hebben we gekeken of Cohesin noodzakelijk is voor lange afstand interacties.

In hoofdstuk 5 bespreken we de interessante rol die CTCF speelt in de expressie van Th2 cytokine in T-cellen. Chromatine immunoprecipitaties (Chip) in T-cellen laten de aanwezigheid van meerdere CTCF bindingsplaatsen binnen het Th2 cytokine locus zien. De conditionele deletie van CTCF in de thymus, wat ervoor zorgt dat dit gen op een specifieke plek en tijd kan worden gedeleteerd, zorgt voor misregulatie van Th2 expressie.

In hoofdstuk 6 wordt de functionele interactie tussen Delangin en Cohesin besproken. Ook wordt de functionele relevantie van de gedeelde Cohesin en CTCF bindingsplaatsen bediscussieerd. In dit hoofdstuk suggereer ik dat verschillende regulerende mechanismen betrokken zijn bij hun functionele interacties. Het hoofdstuk wordt beeindigd met een aantal suggesties voor toekomstige experimenten om meer inzicht in dit onderzoeksgebied te krijgen.

CURRICULUM VITAE

Name:	Sanja Krpić
Date of birth:	13.08.1975
Place of birth:	Belgrade, Serbia

1990 –1994 Student of The Fourteenth Belgrade Gymnasium,
group: Natural sciences
Belgrade, Serbia

1994 – 2001 Student of Molecular Biology and Physiology, Department of Experimental biomedicine, University of Belgrade, Serbia
Research project for Diploma Thesis at the Institute of Molecular Genetics and Genetic Engineering, Laboratory for Human Molecular Genetics, Belgrade, Serbia
Promoter: Prof. Dr. Milena Stevanović

2002 – 2003 Researcher in the Developmental Neurogenetics Unit, Department of Neuroscience, The Scientific Institute San Raffaelle, Milan, Italy
Promoter: Dr.G. Giacomo Consalez

2004 – 2009 PhD student in The Department of Cell Biology and Genetics, Erasmus MC, Rotterdam
Promoter: Prof. Dr. Frank Grosveld

LIST OF PUBLICATIONS:

Optimal use of tandem biotin and V5 tags in ChIP assays.
Kolodziej KE, Poufarzad F, de Boer E, **Krpic S**, Grosveld F, Strouboulis J.
BMC Mol Biol. 2009 Feb 5;10(1):6.

Critical role for the transcription regulator CCCTC-binding factor in the control of Th2 cytokine expression.
Ribeiro de Almeida C*, Heath H*, **Krpic S**, Dingjan GM, van Hamburg JP, Bergen I, van de Nobelen S, Sleutels F, Grosveld F, Galjart N, Hendriks RW.
J Immunol. 2009 Jan 15;182(2):999-1010.

Novel binding partners of Ldb1 are required for haematopoietic development.
Meier N, **Krpic S**, Rodriguez P, Strouboulis J, Monti M, Krijgsveld J, Gering M, Patient R, Hostert A, Grosveld F.
Development. 2006 Dec;133(24):4913-23.

Metazoan Scc4 homologs link sister chromatid cohesion to cell and axon migration guidance.
Seitan VC, Banks P, Laval S, Majid NA, Dorsett D, Rana A, Smith J, Bateman A, **Krpic S**, Hostert A, Rollins RA, Erdjument-Bromage H, Tempst P, Benard CY, Hekimi S, Newbury SF, Strachan T.
PLoS Biol. 2006 Jul;4(8):e242.

Isolation and characterization of hematopoietic transcription factor complexes by in vivo biotinylation tagging and mass spectrometry.
Grosveld F, Rodriguez P, Meier N, **Krpic S**, Pourfarzad F, Papadopoulos P, Kolodziej K, Patrinos GP, Hostert A, Strouboulis J.
Ann N Y Acad Sci. 2005;1054:55-67. Review.

Erasmus MC
Universitair Medisch Centrum Rotterdam

PhD Portfolio Summary

Summary of PhD training and teaching activities

Name PhD student: **Sanja Krpic** Erasmus MC Department: Cell Biology Postgraduate School Molecular Medicine	PhD period: January 2004-March 2009 Promotor: Prof.Dr. F. Grosveld

1. PhD training

	Year
General academic skills	
- Laboratory animal science (Artikel 9), Utrecht University	2004
- Safe handling of radioactive materials, Belgrade, Serbia	2001
In-depth courses	
- Analysis of microarray gene expression data, Rotterdam	2005
- Cells into Organ Summer School, Basel, Switzerland	2006
- From Development to Disease, Rotterdam	2005
Presentations	
- 16th MGC-Symposium, Rotterdam	2006 2005,2006
- Winter School of the International Graduiertenkolleg (International Research Training School): Transcriptional Control In Developmental Processes, Kleinwalstertal, Austria	2005
- Presentation "Cell into Organs", Braga, Portugal	2005
- Sixth joint MGC-Cancer Research UK Graduate Student Conference, Liege, Belgium	
International conferences	
- European Science Foundation Development of a Stem Cell Tool Box (EuroSTELLS) Exploring Chromatin in Stem Cells Workshop, Montpellier, France (poster)	2007
- Seventh joint MGC-Cancer Research UK Graduate Student Conference, Oxford, UK (poster)	2007

Acknowledgments

This thesis owes its existence to the help, support, and inspiration of many people.

First, I would like to thank my Promoter Prof. Dr. Frank Grosveld, for his continuous support and encouragement. You showed me different ways to approach a research problem and the need to be persistent to accomplish any goal. Thanks for teaching me how to ask questions and express my ideas. Frank, I have learnt a lot from you!

Big thanks to my Small Committee: Sjaak, Dies and Niels who reviewed my work on a very short notice and still gave many insightful comments. Also, for always asking me the right questions and helping me to think through my scientific challenges.
I would like to thank the other members of my thesis committee: Wouter, Joost and Rudi for helping me during my PhD period with very useful advice. I appreciate it a lot.
Arnie, though it was short, it was a great and unique experience to work with you. Thanks for the good advice and friendship.

Dear Marieke, Jasperina, Corina, Melle and Arthur, you were so helpful, efficient and kind to me. Thanks for having an immediate solution to whatever problem I had.
Sjozef, Ton and Pim every computer problem seemed easy when you were around.

Thanks to everyone that I have had the opportunity to work with:
Wilfred, Christel and Erik-Jan, your efficiency and help deserves special thanks. It was a great pleasure to work with you!
Claudia and Rudi, it was an interesting and nice experience. Claudia, I wish you all the best in finishing your PhD.
Thanks to Marieke von Lindern, and her lab members, for tolerating my constant need for I/11 cells.
Thamar and Sylvia, thanks for the help.

During these 5 years, I have enjoyed an excellent working atmosphere in the 702 lab (former and present). Many people contributed to this:
Athina, it was a great experience sharing this time with you! You were a great colleague and friend. All the best with finishing your thesis and I wish you, and Chris, a beautiful future.
Erik, you were the best “desk-mate” ever! Thanks, not only for being an excellent colleague but also a very good friend. I wish you a successful scientific career and all the best in the future!
Wouter, thanks for being encouraging, helpful and friendly to me. All the best in Utrecht.
Petra, it was always pure fun being around you. You are a good, honest and true friend and I hope you will visit us soon. All the best, I will miss you!

Daan, thanks for your friendship, optimism and unselfish help whenever I needed it. Good luck in Switzerland.
Marieke, it was always a great pleasure to talk to you about all kinds of topics and to share the lab with you. All the best and I hope we will keep in touch.
Robert-Jan, thanks for all the advice and good times in the lab. Good luck with your scientific and personal future.
Jurgen, I have enjoyed working and chatting with you. I wish you all the best in your future.
Emilie, your amazing personality and kindness always impressed me. I wish you every success and happiness in France.
Despite so many changes in the lab the last couple of months, I am very lucky to have shortly shared a working space with Dorota, Maureen, Petros, Farzin, Charlotte, Eric, Boet, Ali, Ralph and Mary.
Dorota, you are a truly honest and amazing friend. Good luck to you, and your beautiful family.
Maureen, your fantastic personality made a big difference in the lab. I wish you great future.
Charlotte and Eric, I admire you for your professional and personal optimism. It was great working and being friends with you. Good luck and I hope to see you one day.
Ernie, you are such a great example of how to be a kind person, always supportive …and relaxed when it's needed. Thanks for everything.
Ali, you are a very talented person and an amazing friend. Thanks for everything and all the best in your life.
Thanks to the "ex" and new members of the 706 lab: John, Patrick, Katy, Marianne, Jennifer, Debbie, Tiago, Miyata, Eric and Umut for being excellent neighbors.
Karine, it was great to have you around and to share ups and downs with you. I wish you all the best!
Christina and Raymond, thanks for your friendship, interest and help when needed.
Susana, thanks for quality chats… and a great "Porto experience". Frank S., we should have written down all the "excellent ideas" we had over drinks. Thanks!
Laura and Esther, you were the real "dream team" for organizing the Christmas Pantomime. I have enjoyed it a lot! All the best.
Jeffrey, Suzanne, Eskew, Eva, Ekim, Noorie, Prashant, Chris D., Audrey, Filippo, Anna and Magda, I am very lucky to have met you.
Milena and Steven, you were great neighbors and even better friends. Good luck and I hope to see you soon. Divna Ana… koja me uvek oraspolozi i pomogne u svim mogucim, i nemogucim, situacijama. Sve najbolje.
Dubi, tvoj sjajan humor, pozitivna kritika i dobar savet su mi uvek znacili. Tebi i Kseniji, Srecno!
Vesna i Dejane, hvala na prijateljstvu. Vidimo se na Hvaru!

My paranimfs: Iris, Akiko and Gosia: It is great to have you around! Thank you for all the fun, laughs and great times that I have shared with you. You are truly wonderful and fantastic friends! I wish you all the fun and success in the future.

Thanks to my friends in Rotterdam, who made this time wonderful and unforgettable:
Yonatan, Dana, Vered, Sofia, Gili, Haim, Miki, Shay, Marcos...Thanks a lot. To the van der Vlist family, thanks for everything. Chen and Bart, I cannot wait to share everyday life with you in Israel.
Noa, thanks for all the help, support and amazing friendship.

Also, I would like to thank my family, for the support and encouragement to pursue my interests …all of them. Thanks for believing in me. Mama, znam da bi bila ponosna na mene. Tata i Nemanja, hvala na strpljenju, podrsci i savetima. Ljiljo, Buco, Jelena, Kajo i Djole, Hvala na svemu!
Nir, your love, support and endless patience have been truly inspirational. Without you, all of this wouldn't be possible. Thanks for everything!