Fingers in Action!
Chromatin organization and transcriptional regulation
by
CTCF and CTCFL

Widia Soochit

ISBN: 978-90-8891-753-0

The studies presented in this thesis were performed at the Department of Cell Biology of the Erasmus Medical Center, Rotterdam, The Netherlands.

Printing by Proefschriftmaken.nl || Uitgeverij BOXPress

Cover designed by Lena Kourkouta; CTCF and CTCFL on the DNA

Fingers in Action!
Chromatin organization and transcriptional regulation by CTCF and CTCFL

Vingers in actie!
Chromatine organisatie en transcriptionele regulatie door CTCF en CTCFL

Proefschrift

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

Prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
dinsdag 19 november 2013 om 15.30 uur

door

Widia Sabrina Wanita Soochit

geboren te Rotterdam

Promotiecommissie:

Promotor:	Prof.dr. F.G. Grosveld
Overige leden:	Prof.dr. J.H. Gribnau Prof.dr. B. van Steensel
	Dr.ir. W.M. Baarends
Copromotor:	Dr.ir. N.J. Galjart

Om Namah Shivaya

- We should consider every day lost on which we have not danced at least once -
Friedrich Nietzsche

Voor papa en mama

Table of contents

Scope of this thesis **p9**

Chapter 1: **p11**
General introduction

Chapter 2: **p45**
CTCF regulates the local epigenetic state
of ribosomal DNA repeats

Chapter 3: **p83**
Reconstitution of wild type and mutant GFP-CTCF expression
in embryonic stem cells lacking endogenous CTCF

Chapter 4: **p117**
The male germ cell gene regulator CTCFL is functionally
different from CTCF and binds CTCF-like consensus sites
in a nucleosome composition-dependent manner

Chapter 5: **p141**
CTCFL activates transcription of a limited number
of genes in the testis

Chapter 6: **p165**
General discussion

Chapter 7: **p175**

Summary/Samenvatting **p176**

List of Abbreviations **p180**

Curriculum Vitae **p181**

PhD Portfolio **p183**

Acknowledgements **p185**

Scope of this thesis

Chromatin is hierarchically folded and wrapped in order to compact DNA. It is accessible to specific proteins to allow regulation of various cellular processes. Although chromatin is organized into higher-order structures it is highly dynamic and it can influence genome configuration and transcription via interactions with various subnuclear compartments. CTCF is the most important factor involved in chromatin structure regulation, in particular the spatial organization of higher-order chromatin configurations. CTCF-like (CTCFL) is a testis specific paralogue of CTCF, whose function has been characterized to a lesser extent.

The aim of this thesis is to obtain more insight in the biological roles of CTCF and CTCFL.

A general introduction to the field of nuclear organization and transcription regulation is provided in **chapter 1**. This chapter also provides an overview of the process of spermatogenesis during which both CTCF and CTCFL are thought to perform important functions. Finally this chapter also summarizes already known aspects of CTCF, CTCFL and their functional interaction partners.

The role of CTCF and CTCFL in the regulation of ribosomal repeat DNA is the main focus of **chapter 2.** CTCF and CTCFL interact with the key regulator of RNA polymerase I, UBF. Furthermore, CTCF regulates the spacer promoter by recruiting RNA polymerase I, H2A.Z and UBF to rDNA.

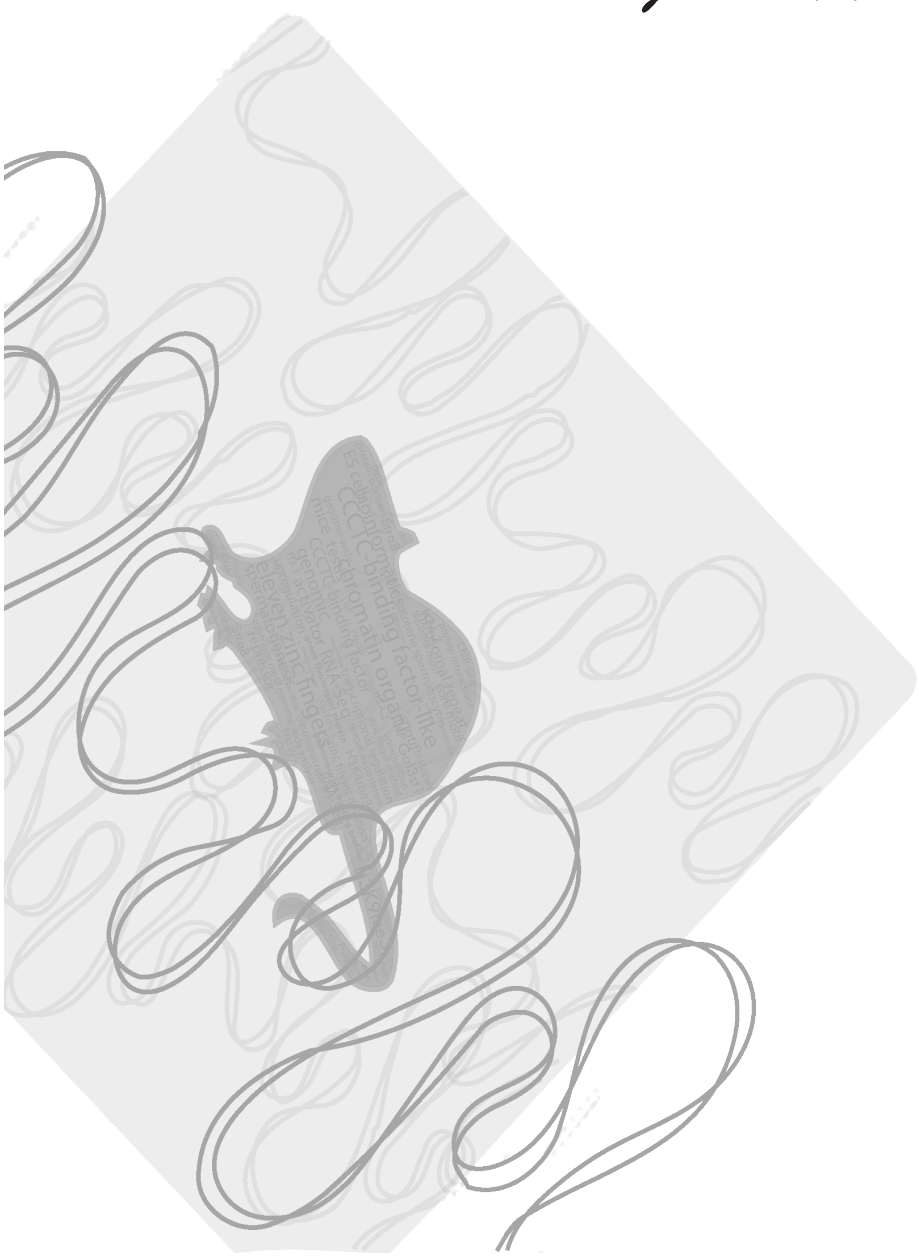
In **chapter 3** the focus shifts towards the study of CTCF binding motifs and the binding of CTCF zinc fingers to DNA. Using a genome-wide binding analysis on CTCF zinc finger mutants we propose a model for DNA binding by CTCF.

Chapter 4 and **5** examine the functional relationship between CTCF and CTCFL extensively in mouse embryonic stem cells and testis by examining genome-wide binding and transcription profiles.

Finally, **chapter 6** provides a general discussion elaborating on the findings in this thesis. Furthermore, the findings are positioned in perspective with current literature and provide recommendations for further experiments on this topic.

Chapter 1

General Introduction



§ 1-1 Chromatin organization and transcription regulation

DNA compaction in the nucleus

The hereditary information needed for the functioning of an organism is stored in DNA. DNA is a two-stranded helix in which each strand consists of four different nucleotides called adenine (A), guanine (G), cytosine (C) and thymine (T). Like letters in a very long phrase, the nucleotides are placed in a non-random order so as to convey a message: the genetic code. A small part of that code contains the information for our ~25,000 genes. Humans have 46 chromosomes that concatenated together would stretch up to 2 meters. DNA is stored in the nucleus, which on average has a diameter of 10 μm. This obviously requires an enormous compaction of the DNA, which is achieved by folding and wrapping the DNA in a hierarchical manner, using specific associated proteins. DNA together with its associated proteins is called chromatin. DNA folding requires many proteins, including the histones, which provide the first level of folding of the double helix and can be considered the core chromatin proteins.

DNA folding is hierarchically organized into several levels **(Figure 1)**. First, 147 base pairs (bp) of DNA is tightly wrapped in almost two helical turns around a histone octamer consisting of two dimers of H2A-H2B and H3-H4. This protein-DNA complex is called the nucleosome *(Luger et al., 1997; Noll and Kornberg, 1977; Richmond and Davey, 2003)*. Nucleosomes are separated by ~20-50 bp of linker DNA, which is associated with histone H1 *(Oudet et al., 1975)*. In electron microscopic (EM) images nucleosomes appear to be arrayed as "beads on a string", forming a 10 nm chromatin fiber *(Oudet et al., 1975)*. This fiber has been shown *in vitro* to be organized into a secondary structure, the 30 nm chromatin fiber *(Finch and Klug, 1976)*. However, the in vivo existence of the 30 nm fiber remains controversial *(Tremethick, 2007)*. Chromatin is further compacted into higher order structures in interphase, and even further into very tightly folded structures, called chromonema, in mitotic chromosomes *(Belmont and Bruce, 1994; Rattner and Lin, 1985; Widom and Klug, 1985)*.

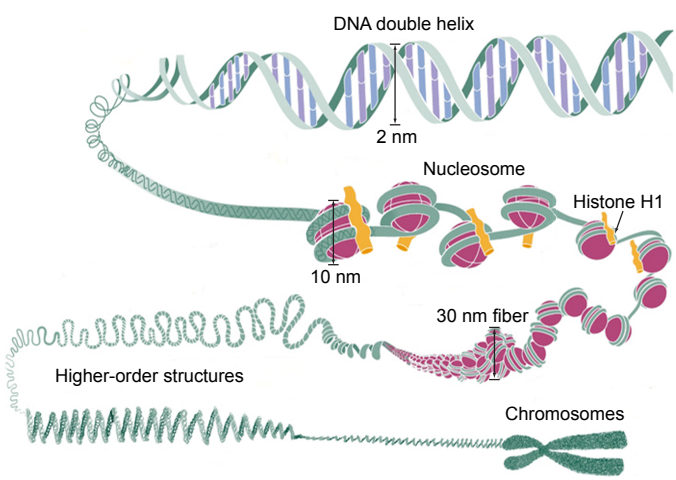


Figure 1: Chromatin organization
DNA is a double-stranded helical structure with a diameter of 2 nm. DNA and its associated proteins are called chromatin. The DNA is tightly wrapped around a histone octamer consisting of two dimers of H2A-H2B and H3-H4. This protein-DNA structure is called the nucleosome. These nucleosomes are separated by 20-50 bp of linker DNA bound by histone H1 and are further compacted in a 30 nm chromatin fiber. Chromatin is further compacted into higher-order structures in interphase and even further folded in mitotic chromosomes.
Image adapted from W. H. Freeman Pierce, Benjamin. Genetics: A Conceptual Approach, 2nd ed.

Based on EM images, chromatin was originally divided into 2 types, heterochromatin and euchromatin. Heterochromatin shows up as dark and condensed matter with often a granular composition. By contrast, euchromatin appears more lightly colored and less condensed *(Oudet et al., 1975)*. This division also has structural and functional relevance. Heterochromatin is indeed highly condensed, generally gene poor and transcriptionally inactive. It comes in two varieties: permanently silenced chromatin or constitutive heterochromatin, which is often found at centromeres, telomeres and inactive repetitive elements, and facultative heterochromatin,

which is mostly inactive but which can be activated, e.g. during development or differentiation *(Dillon and Festenstein, 2002; Le et al., 2004)*. Euchromatin is less condensed, generally gene rich, trancriptionally active, and evenly distributed along the genome *(Dillon and Festenstein, 2002)*. Despite this general classification, euchromatic regions may also contain inactive genes whereas active genes can also be located in heterochromatic regions *(Gilbert et al., 2004)*. In addition, although DNA is highly compacted, it has to be accessible to all kinds of proteins in order to allow processes like transcription, replication, recombination and repair to occur efficiently and in a regulated manner.

Epigenetic modifications

The genetic four-letter code of DNA is quite reliably passed on to the next generation, although mutations can occur. Depending on their position these changes may lead to altered gene function and/or expression. There are also other phenomena that can lead to changes in gene expression. These cannot be explained by alterations in DNA sequence, but they can nevertheless be stable, at least for one generation. Such changes are due to an “epigenetic” code, which is laid down “on top” of the genetic one. The epigenetic code does not meddle with the nucleotide sequence but instead it acts on the proteins that wrap and protect the DNA, or it modifies the DNA itself. The epigenetic code represents the chromatin state, i.e. the properties of DNA and its associated proteins, and this in turn can affect gene expression.

Epigenetic modifications on the four core histones, H2A, H2B, H3 and H4, occur post-translationally. Many epigenetic modifications have been described, including acetylation, methylation, ubiquitination and phosphorylation *(Kouzarides, 2007; Rivera and Ren, 2013)*. Together they are important to facilitate proper chromatin organization and to allow the efficient execution of biological tasks in the nucleus, e.g. gene expression or silencing, cell cycle progression and DNA repair *(Dillon and Festenstein, 2002; Kouzarides, 2007; Rivera and Ren, 2013)*. Epigenetic modifications are mainly restricted to the histone tails and are carried out by chromatin remodeling complexes, e.g. histone acetyl transferases (HATs) and histone deacetylases (HDACs). These chromatin remodeling complexes modify the histones and affect the local chromatin state *(Brownell and Allis, 1996; Brownell et al., 1996; Mizzen et al., 1996)*. Besides the canonical histones, there are many histone variants, that differ in a small subset of amino acids from the canonical proteins, and that change the chromatin state upon replacement of the core histones. These replacements can be stable too and are therefore also considered to be “epigenetic events”.

In addition to histone modifications and replacements, DNA methylation is the third major epigenetic modification. DNA methylation occurs on cytosines of CpG dinucleotides, and can be divided into *de novo* and maintenance events. *De novo* DNA methylation occurs on non-methylated CpGs and is established by DNA methyltransferases 3A and 3B (DNMT3A and DNMT3B) *(Bestor, 1992)*. Maintenance of DNA methylation is carried out by DNMT1 and occurs after DNA replication on the non-methylated cytosine of the new CpG *(Holliday and Pugh, 1975; Pradhan et al., 1999)*.

Epigenetic modifications can mark specific regions of the DNA **(Table 1)**. For example, H2A.Z and H3.3, which replace canonical H3, are often found on actively transcribed regions *(Barski et al., 2007; Jin and Felsenfeld, 2007; Jin et al., 2009).* Histone 3 lysine 4 mono-, di- and tri-methylation (H3K4me1/2/3), histone 3 lysine 36 mono- and tri-methylation (H3K36me1/3), H2A.Z, H3.3, histone 3 lysine 27 acetylation (H3K27ac) and histone 3 lysine 9 acetylation (H3K9ac) are located in actively transcribed regions and promoters and the levels of these modifications correlate with gene activation *(Barski et al., 2007; Ernst et al., 2011; Heintzman et al., 2009; Heintzman et al., 2007)*. Heterochromatic regions on the other hand are often marked by histone 3 lysine 9 (H3K9), histone 3 lysine 27 (H3K27) and histone 4 lysine 20 methylation (H4K20me) and hardly show histone acetylation. Differential modification of the same amino acid residue can be associated with different chromatin domains; for example histone 3 lysine

27 mono-methylation (H3K27me1) associates with active promoters, whereas histone 3 lysine 27 di- and tri-methylation (H3K27me2/3) are more associated with silent promoters *(Barski et al., 2007)*. Furthermore H3K27me3 is present together with H3K4me1/2/3 in "poised" promoters in embryonic stem cells (ES cells), i.e. promoters that may become expressed (and lose these marks) depending on the differentiation path of the ES cell *(Ernst et al., 2011)*.

A

Histone Modification	Residues Modified	Functions Regulated
Acetylation	K-ac	Transcription, Repair, Replication, Condensation
Methylation (lysines)	K-me1 Kme2 Kme3	Transcription, Repair
Methylation (arginines)	Rme1 Rme2a Rme2	Transcription
Phosphorylation	S-ph T-ph	Transcription, Repair, Condensation
Ubiquitylation	K-ub	Transcription, Repair
Sumoylation	K-su	Transcription
ADP ribosylation	E-ar	Transcription
Deimination	R > Cit	Transcription
Proline Isomerization	P-cis > P-trans	Transcription

B

Histone modification	Transcriptional role
H3K4me1	Activation
H3K4me2	Activation
H3K4me3	Activation
H3K27ac	Activation
H3K27me1	Activation
H3K27me3	Repression
H3K36me1/2	Activation
H3K9ac	Activation
H3K9me2/3	Repression
H4K20	Repression

C

Histone variant	Transcriptional role
H3.3	Activation
H2A.Z	Activation

Table 1: Histone modifications and their functions
(A) Possible modifications on histones and their associated functions. Modifications on specific amino acids are depicted in the center column. **(B)** Functional role on transcription of histone modifications and **(C)** histone variants. *Tables based on Kouzarides et al, 2007 and Rivera and Ren, 2013.*

DNA methylation marks chromatin domains during differentiation and in general differentiated cells have more DNA methylation than pluripotent cells *(Gifford et al., 2013; Xie et al., 2013)*. Genes that are active during early stages of differentiation are often CpG rich and are silenced due to DNA methylation in later stages. Genes active at later stages during differentiation are often CpG poor and therefore unmethylated *(Xie et al., 2013)*. In fact, the genome can be segmented into three categories regarding methylation: fully methylated regions containing most of the genome, unmethylated regions encompassing promoters and unmethylated CpG islands and low-methylated regions often identified at distal regulatory regions *(Stadler et al., 2011)*.

Transcription and regulatory elements

Mammalian genomes have ~ 25,000 genes whose transcription is tightly regulated in order to carry out the spatio-temporally controlled, lineage-specific gene expression program that is required for the development of an organism. Transcription is a dynamic and complicated process that involves multiple proteins and consists of three phases: initiation, elongation and termination *(Roeder, 2005; Venters and Pugh, 2009; Weake and Workman, 2010)*. The mega dalton, multisubunit proteins, RNA polymerase I, and III, are responsible for transcription of ribosomal RNA (rRNA), occurring in the nucleolus, and of transfer RNA (tRNA), respectively. RNA polymerase II is responsible for the transcription of protein coding genes and of non-coding RNAs (ncRNAs) other than rRNA and tRNA.

In order to start transcription a gene needs to be(come) accessible to the basic transcription machinery, which involves chromatin remodeling *(Muller et al., 2001; Tumbar et al., 1999)*. Epigenetic modifications, that change local chromatin structure to increase or decrease accessibility, can obviously impact on transcription. Like initiation, transcription elongation and termination, which are linked to RNA splicing, are also highly regulated, by factors acting on the processes themselves as well as by epigenetic modifications.

At the DNA level relatively small domains called regulatory elements regulate transcription. These regulatory elements include promoters, enhancers (or silencers), locus control regions, and insulator sequences **(Figure 2)**. The promoter is a small DNA sequence upstream of the transcription start site on to which the basal transcription machinery assembles and from where transcription is initiated *(Butler and Kadonaga, 2002; Roeder, 2005)*. Enhancers, silencers, and locus control regions are distal regulatory elements that can be located inside or outside the transcription unit regulating transcription in cis, often over a long distance. Transcription requires contact between promoters and regulatory elements, bringing together the pre-initiation complex (basal transcription factors and the RNA polymerase) and transcription regulators (e.g. activators) and/or co-regulators (e.g. chromatin remodelers). This requires folding of the chromatin into loops. This type of looping, which one could call "regulatory looping", differs from the type of looping (or folding) described above, that is required to compact chromatin and that one might call "architectural looping".

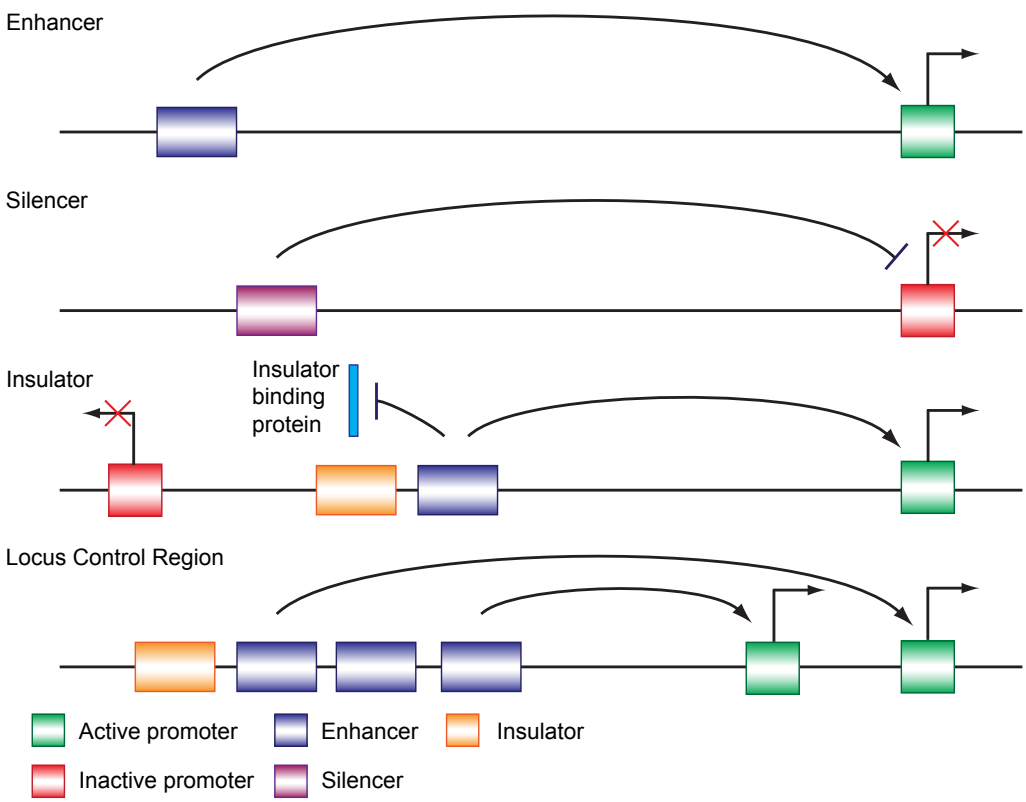


Figure 2. Distal transcriptional regulatory elements
Enhancers and silencers are elements that can activate or repress transcription, respectively. Insulators can act as enhancer-blocker by preventing interactions between promoter and enhancer when placed in between these two elements. Locus control regions are cis-regulatory elements that regulate transcription of clusters of genes. *Image adapted from Maston et al 2006*

Enhancers are operationally defined as DNA sequences that can activate transcription when brought in the vicinity to the promoter *(Banerji et al., 1981; Nolis et al., 2009; Vilar and Saiz, 2005)*. These elements are marked by specific histone modifications e.g. H3K4me1, and chromatin remodelers (e.g. p300) and are bound by cell type specific transcription factors *(Visel et al., 2009)*. Silencing/repressor elements function to repress transcription of a gene, and are also regulated by different histone modifications, chromatin remodelers and cell type specific factors *(Maston et al., 2006)*. Insulators (also termed boundary elements), which are often bound by the protein CTCF (see below), are elements that prevent the spreading of modified histones from one chromatin domain into another, thereby for example blocking the expansion of heterochromatin. Insulators can also act as enhancer-blockers, preventing enhancer-promoter interactions when placed between these two elements *(Maston et al., 2006)*. While some insulators exert both functions, others only act as a boundary element or as enhancer-blocker *(Maston et al., 2006; Parkhurst et al., 1988; Recillas-Targa et al., 2002; Scott et al., 1999; Spana et al., 1988)*. Locus control regions (LCRs) were defined as cis-regulatory elements that regulate transcription of an entire gene locus or cluster. Like enhancers they act over a distance and require DNA looping to exert their effect *(Dean, 2011; Grosveld et al., 1987; Maston et al., 2006)*. The best studied LCR, which will be described later in this introduction, is the one regulating the β-globin locus.

Links between the spatial organization of chromatin and transcription

Microscopic experiments using fluorescent in situ hybridization (FISH) to detect chromosomes, loci, and genes, and genome-wide experiments examining intra- and interchromosomal interactions, have revealed that chromatin is organized in more intricate manners and into smaller domains or sub-compartments than the folding and compaction, and heterochromatic and euchromatic divisions mentioned earlier. The chromosomes during interphase each occupy a distinct part of the nuclear space, called chromosome territory (CT). Epigenetic modifications have a big impact on this refined spatial organization *(Cremer and Cremer, 2010; Lichter et al., 1988; Lieberman-Aiden et al., 2009; Pinkel et al., 1988; Zhang et al., 2012)*.

Consistent with the CT view, interactions between DNA elements within a locus or between loci on the same chromosome occur much more often than interactions between chromosomes *(Dixon et al., 2012; Lieberman-Aiden et al., 2009)*. The interaction frequency of intrachromosomal loci revealed that CTs consist of smaller chromosomal areas, topologically associated domains (TADs). Frequent long-range interactions occur within loci in a TAD and much less frequent between loci in different TADs *(Dixon et al., 2012; Nora et al., 2012)*.

Several other sub-compartments have been detected within nuclei, for example, areas combining gene-rich regions within a CT, or in multiple CTs, and excluding gene-poor regions *(Dixon et al., 2012; Lieberman-Aiden et al., 2009)*. Actively transcribed genes have also been show to come together in so-called transcription factories, which suggest that these genes might be coordinately regulated *(Ghamari et al., 2013; Iborra et al., 1996; Osborne et al., 2004)*. Inactive regions often associate with the nuclear lamina, and are called lamina associated domains (LADs), while active regions often localize to the inner part of the nucleus and the nuclear pore complexes *(Brown et al., 2008; Capelson et al., 2010; Guelen et al., 2008)*. These regions are dynamic and change during differentiation or upon reception of environmental signals, e.g. nucleolus associated domains (NADs) can switch to LADs and vice versa *(Nemeth et al., 2010; Peric-Hupkes et al., 2010; van Koningsbruggen et al., 2010)*.

Long-range interactions between DNA elements cause DNA looping, which include both “regulatory looping” and “architectural looping”. Interactions are mediated by protein-protein interactions and create a complex three-dimensional structure. Long-range interactions involve the structural proteins CTCF and cohesin. These proteins are described in more detail later in this introductory chapter. Long-range interactions also involve other proteins, such as transcription factors and chromatin remodelers. Together these proteins establish cell

type specific TADs and other domains, which are thought to regulate gene expression and epigenetic events *(Gibcus and Dekker, 2013)*.

In 2002 a paralogue of CTCF, called CTCFL or BORIS (Brother of the Regulator of Imprinted Sites) was discovered *(Loukinov et al., 2002)*. A short overview of the published literature regarding CTCFL will be given at the end of this chapter. Since CTCFL is specifically expressed in the male germ line, the following paragraph will provide a description of mouse spermatogenesis and an overview on what is known about chromatin organization during this process.

§ 1-2 Chromatin organization during mouse spermatogenesis

Mammalian spermatogenesis

Spermatogenesis in mammals is a complex process in which male haploid cells are generated from diploid precursors. Spermatogenesis involves multiple rounds of cell division (mitotic phase) but without cytokinesis in order to form a syncytium, a group of nuclei sharing a common cytoplasm and without the usual complete plasma membrane division. This syncytium is maintained throughout spermatogensis. The mitotic phase is followed by the generation of haploid spermatocytes (meiotic phase). The final stage is the differentiation into spermatids and spermatozoa. The latter process, which is called spermiogenesis, involves the compaction of the nucleus that requires replacement of histones with, first, transition proteins and then protamines. During spermiogenesis the cytoplasm is extruded and a flagellum is built for sperm motility. Spermatogenesis takes place in the seminiferous tubules of the testis in an outside-in fashion, i.e. diploid cells are located towards the outside of the tubule whereas more differentiated cells (e.g. spermatids and spermatozoa) are located more and more on the inside. Within the seminiferous tubules basal and luminal compartments are present, which form a blood-testis barrier, and which are maintained by Sertoli cells, that connect to each other via junctional complexes *(Griswold, 1998; Yoshida et al., 2007; Zhou and Griswold, 2008)*. Shortly after cells have entered meiotic prophase, they move from the basal to the luminal compartment through an intricate regulatory system that allows the blood-testis barrier to remain intact during the passage of these cells. It is interesting to realize that luminal testicular cells are separated from the immune system by the blood-testis barrier for proper development and maturation of the testis *(Mital et al., 2011)*. Germ cell specific genes that are aberrantly expressed in cancer cells are termed cancer germ cell genes (CG genes) *(Cheng et al., 2011)*. The proteins of these genes can be used for immunotherapy to treat cancer.

In mice diploid primordial germ cells (PGC) give rise to prospermatogonia, the spermatogonial stem cell, which is located at the basal side of the seminiferous tubule and which undergoes both self-renewal and differentiation **(Figure 3)** *(Eddy, 2002; Zhou and Griswold, 2008)*. Three types of spermatogonia are observed: A, B and intermediate. Type A spermatogonia consist of seven sub-types, A (single) A (pair), A (aligned), which are the spermatogonial stem cells, and A1-A4 spermatogonia, which are committed to differentiation and which expand via a couple of mitotic divisions. A1-A4 spermatogonia also start to migrate laterally along the tubule until they differentiate further, first to intermediate spermatogonia and then to type B spermatogonia *(Chiarini-Garcia et al., 2001; de Rooij, 1998; Dettin et al., 2003; Yoshida et al., 2007)*. It has been proposed that retinoic acid plays an important role in spermatogonial differentiation. Retinoic acid can induce differentiation *in vitro* *(Haneji et al., 1983)*, and positively regulates stem cell factor KIT, STRA8 (stimulated by retinoic acid), and DAZL (deleted in azoospermia-like) *(Schrans-Stassen et al., 1999; Zhou and Griswold, 2008)*, proteins that are essential for spermatogonial differentiation. Mutation or knock out experiments of these genes result in a pre-meiotic block, increased apoptosis and infertility *(Baltus et al., 2006; Brannan et al., 1992; de Kretser, 1997; Koubova et al., 2006; Mark et al., 2008; Schrans-Stassen et al., 2001)*.

Type B spermatogonia give rise to primary spermatocytes that enter meiosis starting with the pre-leptotene phase. In these spermatocytes DNA is replicated in a prolonged S-phase and the sister chromatids are tightly bound by cohesin complexes, a multi-subunit protein complex, which function will be further discussed in section 1-3. Subsequently to S-phase a prolonged G2 phase is initiated, meiotic prophase I, in which DAZL and STRA8 are required *(Anderson et al., 2008; Jan et al., 2012; Lin et al., 2008)*.

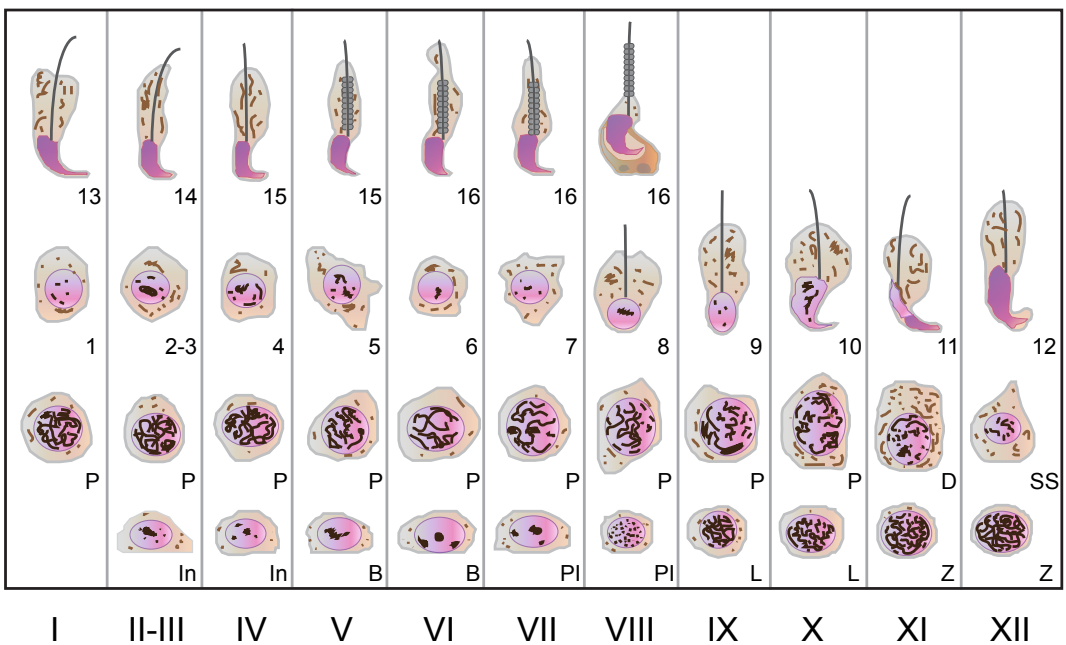


Figure 3. Schematic overview of the seminiferous tubule of mouse during spermatogenesis.
In mouse there are 12 stages for the production of spermatozoa in the seminiferous epithelium. During differentiation certain cell types are always aligned from the basal to luminal compartment. In each diagram, the lower and upper rows of cells are located closer to basal and luminal compartments, respectively. Undifferentiated and type A spermatogonia are not depicted. In: intermediate-type spermatogonia; B: B-type spermatogonia; Pl: primary spermatocytes of the preleptotene stages; L: leptotene stage; Z: zygotene stage; P: pachytene stage; SS: secondary spermatocytes. Arabic numbers represent the step of spermiogenesis. *Modified from Russel et al 1990.*

There are four stages during meiotic prophase I: leptonema, zygonema, pachynema and diplonema *(Handel and Schimenti, 2010; Jan et al., 2012)*. First, during the beginning of leptonema chromosomes begin to align but are not yet paired. Chromatin condensation occurs and DNA double-strand breaks (DSB) are induced by the enzyme SPO11. DSB repair and checkpoint proteins like e.g. the checkpoint kinases ATM and ATR, the recombinases RAD51 and its meiosis-specific paralogue DMC1 are recruited to the chromosomes *(Handel and Schimenti, 2010)*. In addition, the assembly of the axial elements composed of synaptonemal complex (SC)-specific proteins, SYCP2 and SYCP3, and cohesin is initiated *(Handel and Schimenti, 2010; Jan et al., 2012; Page and Hawley, 2004)*. The SC with its axial elements is required for the synapsis of the chromosomes *(Handel and Schimenti, 2010; Jan et al., 2012; Page and Hawley, 2004)*.

By zygonema, the homologous chromosomes have paired and synapsis is initiated. Prior to synapsis, the axial elements become the lateral elements of the SC and interact with the central element formed of e.g. SYCP1 *(Handel and Schimenti, 2010; Jan et al., 2012; Page and Hawley, 2004)*. In the third stage, pachynema, the synapses are established and chromosome crossover occurs. In this process non-sister chromatids of the homologous chromosomes exchange segments over homologous regions and chiasmatas are formed were this exchange happens

(Page and Hawley, 2004). Crossover formation is established in late pachynema *(Handel and Schimenti, 2010; Jan et al., 2012; Page and Hawley, 2004)*. During the fourth and last stage, diplonema, the SC is disassembled, but the homologous chromosomes are still tightly bound to each other at the region of the chiasmata *(Handel and Schimenti, 2010; Page and Hawley, 2004)*. Prophase I ends with diakinesis, the stage of transition to metaphase I. During metaphase I the microtubles are attached to the kinetochore and cohesin complexes are dissociated along the sister chromatid arms, which resolves the crossovers and allows the duplicated homologs to separate at anaphase I *(Handel and Schimenti, 2010; Jan et al., 2012)*. The sister chromatids are still together by the residual cohesin. During anaphase II the remaining cohesin complex is dissociated, which allows the sister chromatids to separate resulting in four haploid round spermatids *(Handel and Schimenti, 2010; Jan et al., 2012)*.

During spermiogenesis the round spermatids undergo morphological changes and develop a flagellum, which is composed of a microtubular structure, the axoneme *(Jan et al., 2012; Kierszenbaum and Tres, 2004)*. Nuclear elongation is established by another microtubular structure, the manchette, which is transient in contrast to the axoneme. During nuclear elongation chromatin remodeling takes place. A major event is the replacement of histones, first by transition proteins and then by protamines *(Jan et al., 2012; Meistrich et al., 2003)*. This results in an enormous chromatin compaction and is accompanied by a general silencing of transcription *(Gaucher et al., 2010)*. Prior to global histone replacement open chromatin domains are created by incorporation of histone variants and global H4 hyperacetylation, which results in unstable nucleosomes *(Boussouar et al., 2008; Jin et al., 2009)*. It has been proposed that these chromatin domains with unstable nucleosome are targets for nucleosome disassembly and histone displacement *(Gaucher et al., 2010)*. Histone hyperacetylation gradually disappear with the appearance of transition proteins, TNP1 and TNP1. Transition proteins are not essential for histone removal and protamines loading as these processes still occur in *Tnp1* and *Tnp2* knock out mice *(Yu et al., 2000; Zhao et al., 2001)*. However, they are important for the proper regulation of chromatin structure, since the lack of both TNP1 and TNP1 results in irregular chromatin condensation. Transition proteins are subsequently replaced by protamines, PRM1 and PRM2 *(Gaucher et al., 2010)*. These proteins are essential for male fertility and the incorporation of protamines in the DNA results in further compaction of chromatin.

During spermiogenesis the cytoplasm is extruded from cells, the acrosome and the mitochondrial sheet around the midpiece of the sperm cells are established *(Kierszenbaum and Tres, 2004)*. The acrosome is a granular vesicle that is required for the penetration of the zona pellucida during fertilization. The zona pellucida is a layer of fibrous glycoproteins secreted by the oocyte, which surrounds the plasma membrane of the oocyte. It acts as physical barrier and is required for the initiation of the acrosome reaction, the release of enzymes contained in the acrosome that breaks down the zona pellucida. Before this is possible, spermatogenesis has to be completed by the release of spermatozoa into the lumen of the seminiferous tubule *(Jan et al., 2012)*.

Regulation of DNA imprints during spermatogenesis

In addition to histone modifications in spermatids, major epigenetic reprogramming events need to take place during spermatogenesis, in particular to erase somatic imprints and establish sex-specific (i.e. paternal) ones. In mice the genomes of PGCs of both females and males become demethylated during early development *(Mayer et al., 2000; Monk et al., 1987; Reik et al., 2001)*. After embryonic day (E) 13-14 imprinted genes and single-copy genes become demethylated *(Reik et al., 2001; Tada et al., 1998)*. PGCs in males and females enter mitosis or arrest in meiosis, respectively, after complete erasure of DNA methylation. Gain of methylation in the female germ line is initiated after birth during oocyte growth. In the male germ line remethylation of the genome is initiated at the prospermatogonia stage (E15-16 in mice) *(Reik et al., 2001)*.

Reprogramming takes place once more after fertilization, in the early embryo. Reprogramming occurs on both maternal and paternal genomes *(Kafri et al., 1992; Reik et al., 2001)*. However, not all marks are erased, and most sex-specific imprinted genes are actually protected from global DNA methylation erasure *(Brandeis et al., 1993; Kafri et al., 1992; Tada et al., 1998)*. DNA methylation is again established during blastocyst formation in the inner cell mass *(Reik et al., 2001)*. Both CTCF and its paralogue CTCFL, which will be discussed in the following section, have been suggested to play a role in reprogramming of imprinted genes.

§ 1-3 Key players in chromatin organization: CTCF, cohesin and CTCFL

Maintenance of genetic and epigenetic integrity of the genome is essential to control various cellular processes. Several studies on chromatin organization have shown that CTCF and cohesin play an important role in the regulation of chromatin structure and spatial organization in order to regulate transcription and epigenetic events. The function of testis specific CTCF paralogue CTCFL in relation to transcriptional regulation and chromatin organization is less clear.

CTCF - Identification and basal function

CTCF (CCCTC-binding factor) was first identified as a transcriptional repressor that binds a highly divergent 50-60 bp DNA sequence on human/mouse and chicken c-myc promoters *(Filippova et al., 1996; Klenova et al., 1993; Lobanenkov et al., 1990)*. Additionally, CTCF was discovered in an independent study as negative protein 1 (NeP1) that binds a silencer element and negatively regulates the chicken lysozyme gene *(Baniahmad et al., 1990; Burcin et al., 1997)*. Subsequently, two other independent studies identified factors that bound to the APPβ promoter of the human amyloid precursor protein (APP) *(Quitschke et al., 1996; Vostrov and Quitschke, 1997)*, or to the HS4 insulator sequence upstream of the chicken beta globin locus *(Walters et al., 1999)*. These factors were also identified as CTCF.

CTCF is a ubiquitously expressed nuclear protein that consists of a centrally located eleven zinc finger (ZF) domain, which is surrounded by the N- and C-terminal domains. The first ten ZFs of CTCF are composed of approximately 30 amino acids and belong to the C2H2 ZF class while the eleventh ZF belongs to the C2HC class *(Ohlsson et al., 2001)*. The ZF region binds DNA, RNA and other proteins. An AT-hook domain in the C-terminus might also have a role in DNA binding and/or protein interactions, although additional research is needed to confirm this *(Ohlsson et al., 2001)*. The C-terminal domain is subjected to posttranslational modifications (e.g. phosphorylation of four serine residues), which was proposed to influence CTCF binding to its target sites on DNA *(Klenova et al., 2001)*.

CTCF is an essential protein, as *Ctcf* knock out mice die early during embryogenesis *(Heath et al., 2008)*. Furthermore, transgenic females with RNAi constructs against CTCF produce less offspring than wild type and show increased zygotic lethality and defects in pre-implantation development *(Fedoriw et al., 2004)*. *Ctcf* knock down mice develop normally until 16-31 cell stage. Hereafter, increased apoptosis occurs and development of the inner cell mass and trophoectoderm are inhibited *(Moore et al., 2012)*.

The importance of CTCF is further emphasized by studies focusing on the conservation of the protein. CTCF is highly conserved in higher order eukaryotes as it has been identified in basal nematodes (e.g. T. spiralis), zebrafish, drosophila, frogs, birds, rodents and humans *(Burke et al., 2002; Heger et al., 2009; Moon et al., 2005; Ohlsson et al., 2001; Pugacheva et al., 2006)*. CTCF has ~90% amino acid identity in the mammals and its eleven zinc finger shows an almost 100% conservation between mouse, human and chicken *(Ohlsson et al., 2001)*. Zebrafish CTCF shows an overall 70% amino acid identity and the zinc finger region shows 98% amino acid identity compared to aves (birds) and mammals *(Pugacheva et al., 2006)*.

CTCF - DNA binding specificity

Early *in vitro* approaches suggested that CTCF uses different combinations of ZFs to recognize its binding sites, ZF 2-7 being able to bind the chicken c-myc site while ZF 3-11 were required for binding to the human site *(Filippova et al., 1996)*. By contrast, CTCF ZF 6-11 were required for the F1 lysozyme silencer element *(Burcin et al., 1997)*, whereas ZF 5-7 facilitated binding to the APP promoter and deletion of the flanking ZFs or N-terminus decreased but did not abolish CTCF binding to this site *(Quitschke et al., 2000; Vostrov et al., 2002)*. Together, these data suggested that the peripheral ZFs are needed to stabilize CTCF binding.

A more precise *in vitro* study demonstrated, by deleting individual ZFs, that ZF 4-7 were absolutely required for DNA binding *(Renda et al., 2007)*. Interestingly, ZF 8 becomes essential when ZF 4 is deleted and vice versa, suggesting that at least four ZFs are required for CTCF binding. Deletion of the peripheral ZFs reduced CTCF affinity for DNA *(Renda et al., 2007)*, supporting the hypothesis that the flanking ZFs stabilize protein binding and increase binding affinity but are not involved in specific DNA recognition. Furthermore, four ZFs were able to bind at least 23 base pairs (bp), and the N-terminal ZFs were able to bind the 3'end of the CTCF target site while the C-terminal ZFs were bound to the 5'end. DNA methylation inhibited the binding of ZF 7 indicating that this finger acts as a methylation sensor that prevents binding to methylated sites *(Renda et al., 2007)*.

The first attempt to map CTCF binding sites on a genome-wide level used chromatin immunoprecipitation (ChIP) followed by detection of the precipitated DNA on chip arrays (ChIP-on-chip). Using this technique in primary human fibroblasts (IMR90) ~14000 CTCF binding sites were found dispersed throughout the genome (Kim et al., 2007). Subsequent ChIP experiments followed by high throughput sequencing (ChIP-Seq) using $CD4^+$ T cells identified ~20,000 CTCF sites (Barski et al., 2007). Other studies in mouse embryonic stem cells then revealed ~40,000 binding sites (Chen et al., 2008). A recently developed method, ChIP-exo, in which ChIP is first followed by lambda exonuclease treatment, which trims the DNA sequence up until the formaldehyde crosslink, and then by high throughput sequencing, redefined CTCF binding sites with a near single base pare resolution. This method detected ~35,000 CTCF sites in HeLa cells (Rhee and Pugh, 2011). A very recent report discovered ~48,000 binding sites in primary B cells *(Nakahashi et al., 2013)*, whereas a bio-informatics approach in which all CTCF binding sites derived from 56 human cell lines were compared identified ~450,000 unique CTCF binding sites from which ~24,000 sites are conserved in >90 % of the cell lines *(Li et al., 2013)*. This shows how detection method and bioinformatic analysis can influence the estimation of the amount of CTCF sites throughout the genome.

Initial genome-wide study in IMR90 human fibroblasts using ChIP-on-chip revealed that 46% of the CTCF binding sites are located in intergenic regions, 20% within 2.5 kb of transcription start sites, 22% in introns and 12% in exons *(Kim et al., 2007)*. An independent study in human $CD4^+$ T cells using ChIP-Seq which has been reanalyzed identified a similar distribution in which CTCF binds in 45% in intergenic region, 7% in 5' UTR, 3% in exons, 29% in introns, 2% in 3'UTR, and 13% within 5 kb of the transcription start site *(Barski et al., 2007; Xie et al., 2007)*. 10% of all CTCF binding sites identified by ChIP-exo are located in annotated genes from which 6% of these CTCF sites are located at core promoters that are located ~85 bp upstream of the transcription start site *(Rhee and Pugh, 2011)*.

The distribution of CTCF binding sites seemed to be correlated with gene density *(Barski et al., 2007; Kim et al., 2007; Xie et al., 2007)*, although many CTCF sites are also detected in intergenic regions. While there is a high correlation between CTCF binding and gene density there are regions in the genome that deviate from this trend. Two areas can be distinguished 1) regions that contain clusters of related gene families that are CTCF depleted (less than 2 CTCF sites within 2 Mb) 2) regions that contain high CTCF binding density within clusters of related genes. Here, ~81% of these genes have two or more alternative promoters *(Kim et al.,*

2007). Euchromatic regions are marked by DNase I hypersensitivity sites and CTCF (Wen et al., 2012). Furthermore, CTCF was located at boundaries of different chromatin domains where it acts as a barrier to prevent spreading of heterochromatic regions to euchromatic regions (Barski et al., 2007; Cuddapah et al., 2009). It has been proposed that the amount of CTCF binding sites expanded during evolution and are also located in retrotransposon elements, where they function as chromatin and transcription insulators in a cell-specific fashion (Bourque et al., 2008; Kunarso et al., 2010; Schmidt et al., 2012). Approximately 5000 highly stable CTCF sites are conserved in eutherian mammals (Schmidt et al., 2012). It has also been proposed that cell-specific CTCF sites flank developmental genes associated with disease (Martin et al., 2011). Tissue-specific CTCF sites are preferentially associated with enhancers while ubiquitous CTCF sites are more frequently linked to promoters (Shen et al., 2012).

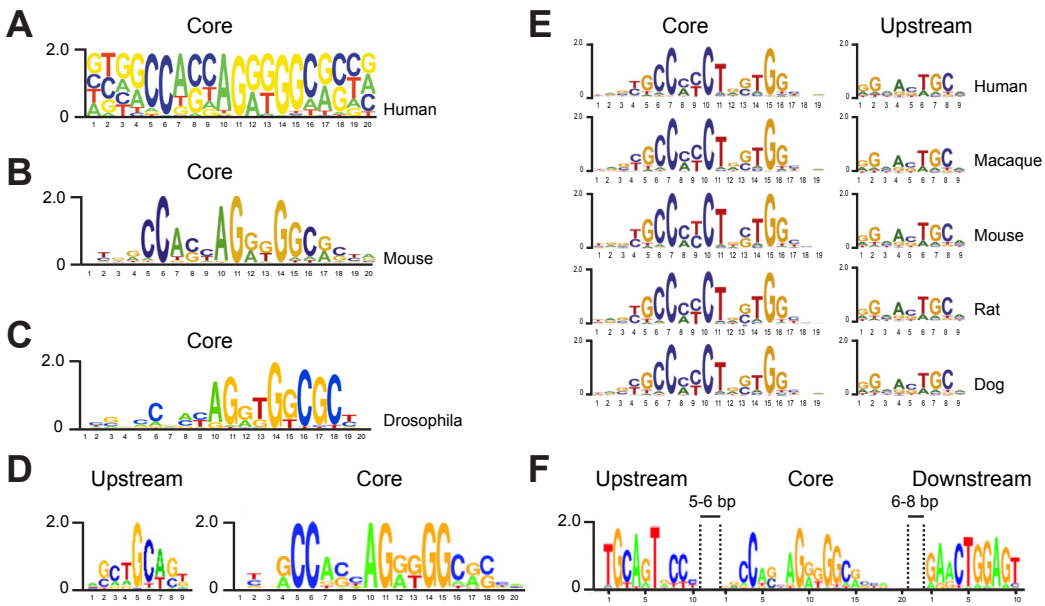


Figure 4. CTCF consensus sequence
(A) CTCF 20 bp core consensus sequence identified in human, **(B)** mouse and **(C)** drosophila. **(D)** CTCF's upstream motif in human is located ~10 bp from the core motif and is identified by DNase I hypersensitivity assays combined with ChIP-Seq. **(E)** This upstream motif and position relative to the core motif is conserved in eutherian mammals. The sequence in this panel is inverted compared to the sequences in the other panels. **(F)** An additional downstream motif 6-8 bp from the core motif was recently identified in mouse. Nucleotide position is shown underneath each sequence. The height of each letter represents the relative frequency of each nucleotide. *Images adapted from Kim et al 2007, Chen et al 2008, Holohan 2007, Boyle 2011, Schmidt et al 2012 and Nakahashi et al 2013.*

Next to the attempts to determine the amount and distribution of CTCF binding sites a search for the CTCF consensus sequence has also been performed. This revealed that ~75% of the CTCF binding sites harbor a 20 bp consensus sequence, termed the "core" motif (C), which is highly conserved in all eutherian mammals, opossum, chicken and pufferfish Tetraodon **(Figure 4A and B)** (Chen et al., 2008; Jothi et al., 2008; Kim et al., 2007; Xie et al., 2007). In addition a similar CTCF motif was discovered in drosophila **(Figure 4C)** (Holohan et al., 2007; Smith et al., 2009b). Interestingly, a computational analysis discovered 233 motifs in 60019 conserved noncoding elements (CNE) in the human genome, of which one fourth was determined to be a CTCF motif (Xie et al., 2007). An additional 10 bp motif, found ~21-22 bp upstream of the core motif, was discovered using DNase I hypersensitivity assays combined with CTCF ChIP-Seq (Boyle et al., 2011), ChIP-exo (Rhee and Pugh, 2011) and regular ChIP-Seq experiments (Schmidt et al., 2012) **(Figure 4D)**. Fifteen percent of all CTCF sites harbor this upstream sequence, which was

termed the “M2” or upstream (U) motif, and which is highly conserved in mammals **(Figure 4E)** *(Schmidt et al., 2012)*. In addition a downstream motif (D) was discovered, which is located 6-8 bp 3’ to the core **(Figure 4F)** *(Nakahashi et al., 2013)*. This motif was proposed to bind a competitor of CTCF. However, the proportion of CTCF sites containing this downstream motif is low, also compared to the prevalence of two other downstream motifs *(Li et al., 2013)*.

In line with *in vitro* studies *(Renda et al., 2007)*, recent genome-wide *in vivo* studies in primary B lymphocytes, using CTCF ZF mutants to determine how individual ZFs contribute to CTCF binding, showed that ZFs 4-7 are essential for recognition of its core sequence motif *(Nakahashi et al., 2013)*. In this approach histidine residues in the zinc binding domain of each ZF were mutated. The orientation of CTCF binding to its target sites *in vivo* was also consistent with *in vitro* results *(Nakahashi et al., 2013; Renda et al., 2007)*. It was proposed that CTCF binding is modular, i.e. groups of adjacent zinc fingers are involved in binding certain sequences, with ZFs 1-2, 3-7, 4-7, 8-11 and 9-11 each comprising a specific module. It was shown that ZFs 8-11 bind to an upstream (5’end) domain and ZFs 4-7 bind the core part of the CTCF motif. ZFs 1-3 would not recognize a specific sequence but stabilize binding. It was hypothesized that ZFs 3 and 8 act as ‘spacer ZFs’ that are required for binding simultaneously to the individual CTCF modules in the consensus *(Nakahashi et al., 2013*).

CTCF - Transcription regulation

A wealth of studies has tried to link the binding of CTCF to the transcriptional activation or repression of specific genes. However, although the CTCF binding pattern follows gene distribution none of the genome-wide studies in which global transcription was investigated has revealed a specific or overlapping set of genes activated or repressed by CTCF, despite the fact that many CTCF sites are located near a TSS *(Barski et al., 2007; Kim et al., 2007; Phillips and Corces, 2009; Ribeiro de Almeida et al., 2011; Soshnikova et al., 2010; Xie et al., 2007)*. Consistently, in mouse embryonic stem cells CTCF does not exhibit a high overlap with transcriptional regulatory networks at the genome-wide level *(Chen et al., 2008)*. Nonetheless, being a chromatin organizer, or DNA looper, CTCF does affect transcription. Two specific examples, which have lead to different concepts, are described below.

The mouse beta-globin-locus contains four functional genes, εy, βh1, βmaj and βmin whereas the human locus has five functional genes, ε, Gγ, Aγ, δ, and β **(Figure 5A)**. They are expressed in sequential order during different stages of erythroid development and their expression is regulated by the LCR *(Epner et al., 1998; Grosveld et al., 1987)*. In mouse and man the LCR contains DNAse hypersensitive sites, termed HS. CTCF binds the LCR (5’HS5), as well as the 3’end (3’HS1) of the locus, and sites upstream of the LCR (HS-62.5 and HS-85) *(Bulger et al., 2003; Farrell et al., 2002; Saitoh et al., 2000)*. Chromatin conformation capture (3C) experiments revealed higher order structure in the active and inactive beta-globin locus. These structures were established by CTCF binding and CTCF-mediated interactions which places actively transcribed genes in a so-called active chromatin hub (ACH) *(Hou et al., 2008; Splinter et al., 2006; Tolhuis et al., 2002)*. The ACH is a dynamic structure in which active genes enter the ACH before RNA polymerase II-mediated transcription *(Palstra et al., 2008; Palstra et al., 2003)*. Surprisingly, CTCF binding to 5’HS and 3’HS sites is not essential for proper beta-globin gene expression but influences the local chromatin state *(Bender et al., 2006; Splinter et al., 2006)*. This result can be explained by assuming that in the absence of 5’HS or 3’HS CTCF binding sites neighboring sites could take part in the organization of the ACH. Indeed, genome-wide analysis confirmed that both CTCF and cohesin facilitate long-range interactions throughout the beta-globin locus and influence beta-globin gene transcription and chromatin state *(Chien et al., 2011; Hou et al., 2010)*.

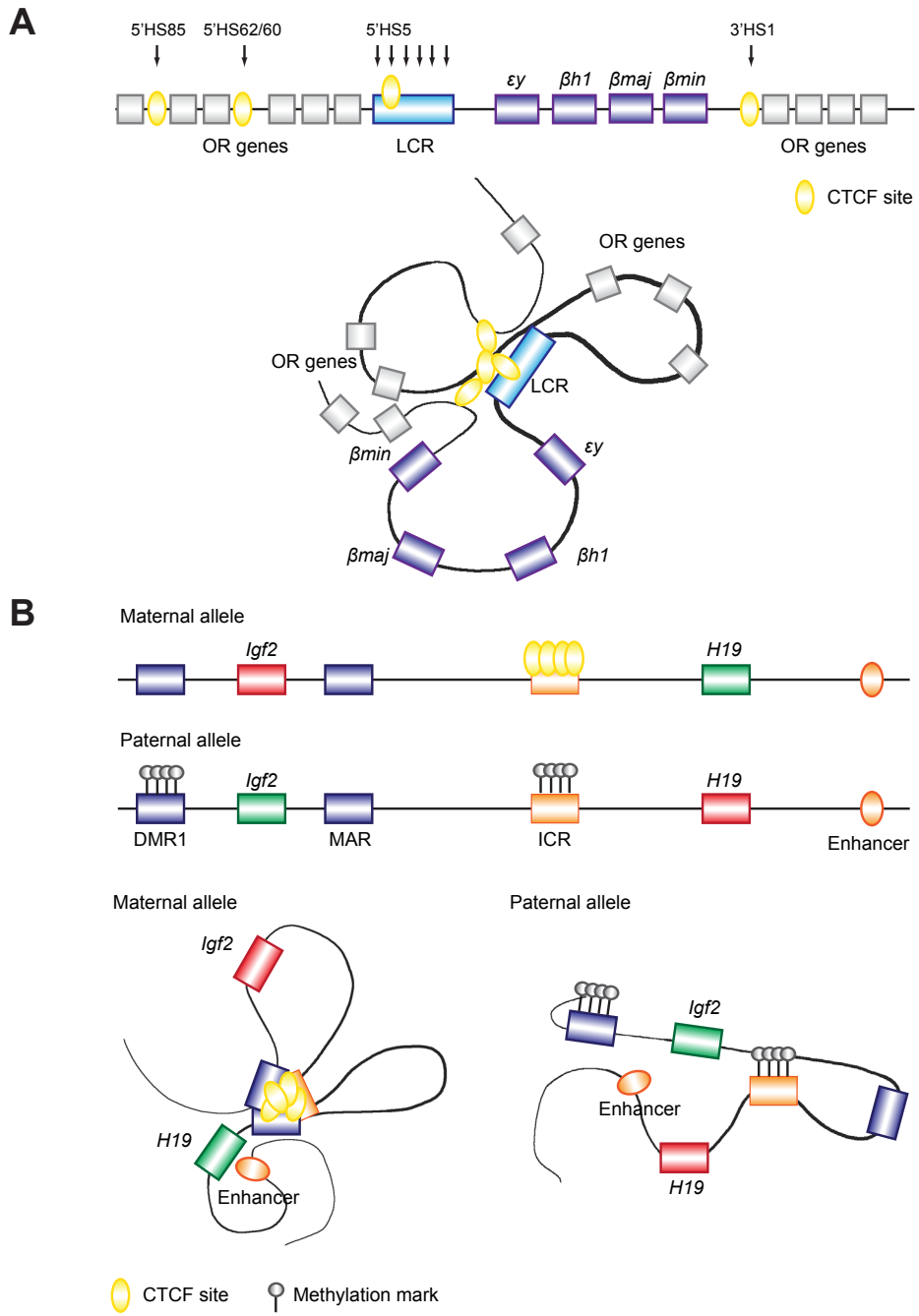


Figure 5. Long-range interactions in beta-globin and Igf2-H19 locus.
(A) Schematic representation of the mouse beta globin locus. DNase I hypersensitivity sites (HS) are depicted with arrows. The beta globin locus are looped out from the CTCF-mediated chromatin hub in erythroid progenitor cells. **(B)** Schematic representation of the Igf2-H19 locus on the maternal and paternal alleles. On the maternal allele CTCF mediated loops are established via CTCF interactions with the unmethylated ICR, DMR and MAR resulting into the expression of H19. On the paternal allele DNA methylation prevents CTCF binding, which leads to Igf2 expression. OR: olfactory genes; LCR: locus control region; DMR: differentially methylated region; MAR: matrix attachment region; ICR: imprinting control region. *Modified from Ribeiro de Almeida et al 2012.*

The second case in which the role of CTCF in transcription was intensively studied is the imprinted *Igf2-H19* locus **(Figure 5B)**. Transcription of the *Igf2* and *H19* genes is regulated via genomic imprinting of the parental alleles. The Insulin-like growth factor 2 (*Igf2*) gene is located 100 kb from the *H19* gene *(Zemel et al., 1992)*. In between these genes a 2 kb imprinting control region (ICR) is located, which is 2kb upstream of the *H19* promoter/enhancer *(Tremblay et al., 1997)*. On the paternal allele the ICR is fully methylated and only *Igf2* is expressed. By contrast, the maternal ICR is unmethylated allowing *H19* expression *(Tremblay et al., 1997)*. CTCF binds the ICR in a methylation-sensitive fashion: it only binds to the maternal unmethylated ICR and not to the fully methylated paternal ICR *(Bell and Felsenfeld, 2000; Hark et al., 2000)*. The parental-specific regulation of *Igf2-H19* expression by selective binding of CTCF to the ICR has underscored to the notion of CTCF acting as an enhancer blocker by preventing the formation of specific loops. On the maternal allele, when CTCF is bound to the ICR, the enhancer located distally of *H19* cannot contact the *Igf2* promoter and hence *Igf2* is not transcribed. By contrast, on the paternal allele, which carries a fully methylated ICR, CTCF does not bind. Consequently, the *H19* downstream enhancer and *Igf2* promoter now can contact each other and the *Igf2* gene is transcribed.

Chromatin conformation capture techniques revealed the higher-order structure of the *Igf2/H19* locus and provided evidence for the model described above. It was shown that on the paternal allele enhancers interacted with the *Igf2* promoter. This interaction was prevented on the maternal allele, where CTCF was bound to the ICR and regulated interactions with matrix attachment region (MAR) and differentially methylated region 1 (DMR1) at the *Igf2* gene. *Igf2* was thereby trapped in a separate loop, and remained inactive and its promoter was not available for the enhancer. All interactions were shown to be CTCF-dependent *(Kurukuti et al., 2006; Murrell et al., 2004)*. Later studies revealed that cohesin was also bound to CTCF sites and facilitated higher order chromatin conformation *(Nativio et al., 2009)*.

It has been proposed that CTCF protects the ICR against *de novo* methylation and maintains it in a methylation-free state *(Pant et al., 2003)*. However, other studies suggested that CTCF is not necessary to keep the maternal allele unmethylated *(Schoenherr et al., 2003; Szabo et al., 2004)*. A more recent study reported that CTCF binding to the maternal ICR is essential to maintain monoallelic *Igf2* expression *(Engel et al., 2006)*.

Although many studies have been performed regarding the regulatory mechanism of *Igf2/H19* imprinting not a lot is known about the chromatin landscape of this locus. The ICR is occupied by nucleosomes but the four CTCF sites are in an open chromatin conformation. Disruption of the nucleosomes interfered with CTCF binding and its insulator function. Interestingly, it has been suggested that CTCF is not a nucleosome positioning factor *(Kanduri et al., 2002)*. H3K9ac, H3K4me were enriched to *H19* while H3K27me3 marked the maternal ICR, DMR's, promoter and gene of *Igf2*. On the other hand the paternal allele exhibited H3K27me3 and macroH2A1 at the H19 promoter and H3K9ac and H3K4me on the *Igf2* DMR. These marks were established by the allelic specific binding of CTCF supporting the role of CTCF in establishing chromatin composition *(Guibert et al., 2012; Han et al., 2008)*.

CTCF - Chromatin organization and higher order structure

As described above, CTCF mediates the special organization of chromatin via long-range interactions, thereby regulating transcriptional activity *(Phillips and Corces, 2009)*. CTCF not only regulates the Igf2/H19 *(Kurukuti et al., 2006; Murrell et al., 2004; Nativio et al., 2009)* and beta-globin loci *(Hou et al., 2010; Palstra et al., 2003; Splinter et al., 2006; Tolhuis et al., 2002)*, but CTCF-mediated chromatin looping also occurs in the Hox cluster *(Ferraiuolo et al., 2010)*, the interferon-γ locus *(Sekimata et al., 2009)*, the Myb locus *(Stadhouders et al., 2012)*, the major histocompatibility complex class II (MHC-II) *(Majumder et al., 2006; Majumder et al., 2008)* and Immunoglobin locus (Ig) *(Ribeiro de Almeida et al., 2011)*. However, CTCF also regulates chromatin beyond specific loci.

Within the human genome LADs are present. These LADs are 0.1-10 Mb in size and are specifically associated with the nuclear lamina at one point during the cell cycle. They are marked by H3K9me2 and H3K27me3 and generally encompass genes that are expressed at a low level *(Guelen et al., 2008)*. CTCF mediated loops are enriched at the boundaries of LADs, suggesting that CTCF acts as a boundary element that maintains the LAD borders and prevents silencing of neighboring regions *(Handoko et al., 2011)*.

A "CTCF chromatin interactome" was established using the novel paired-end tag (ChIA-PET) technique followed by deep next generation sequencing. This approach revealed that CTCF mediates both inter- and intra-chromosomal interactions *(Handoko et al., 2011)*. CTCF-mediated interactions could be categorized into five clusters: 1) CTCF-mediated loops with active chromatin marks (H3K4me1, H3K4me2 and H3K36me3), RNA polymerase II and p300 binding, and depleted of repressive chromatin marks; 2) chromatin interactions encompassing repressive marks (H3K9, H3K20 and H3K27me) and lacking active marks; 3) hubs containing enhancers and promoters with H3K4me1, H3K4me2 within the loop, H3K4me3 at the boundaries, whereas other repressive or active marks are outside the hub; 4) areas demarcating active and repressive chromatin outside the loop and 5) loops with no particular chromatin marks present inside *(Handoko et al., 2011)*. Interestingly, genome-wide studies suggest that classical CTCF insulator function is rare and context-dependent *(Handoko et al., 2011; Phillips-Cremins and Corces, 2013; Phillips-Cremins et al., 2013; Sanyal et al., 2012)*.

CTCF is also enriched at the borders of conserved topologically associated domains (TADs) together with histone modifications H3K4me3, H3K36me3 but depleted for H3K9me3, chromatin binding proteins and transcription factors *(Dixon et al., 2012)*. These TADs contain subdomains, which can be distinguished in constitutive and cell type specific domains. Hundred kb - 1 Mb interactions within and between constitutive subdomains are anchored by CTCF and cohesin, whereas cohesin and mediator has been suggested to mediate small interactions (<100 kb) in cell type specific subdomains *(Phillips-Cremins et al., 2013)*. Knock down of CTCF shows that the number of interactions genome-wide decrease, but that the number of interactions between TADs increases, probably caused by the delocalization of cohesin *(Zuin et al., Submitted)*.

CTCF - Interaction partners

Besides mediating its function via interaction with DNA, CTCF also regulates certain processes interacting protein partners **(Table 2)**. Interaction partners linked to transcription regulation are, for example, the largest subunit of RNA polymerase II *(Chernukhin et al., 2007)*, and YB-1 (as a repressor of c-myc transcription) *(Chernukhin et al., 2000)*. Partners linked to the insulator function of CTCF are the methylation insensitive transcription factor KAISO *(Defossez and Gilson, 2002)*. CTCF also interacts with the class II transactivator (CIITA) protein, to mediate long-range interactions in the MHC-II locus *(Majumder et al., 2008)*.

Interaction partners related to chromatin modification were also discovered. CTCF interacts for example with SNF2-like chromodomain helicase CHD8 to regulate epigenetic changes *(Ishihara et al., 2006)*. CTCF-SIN3A complexes repress transcription via recruitment of histone deacetylase (HDAC) activity *(Lutz et al., 2000)*. The Polycomb repressive complex 2 (PRC2) is recruited via CTCF-SUZ12 interactions at CTCF target sites resulting in H3K27me and suppression of transcription *(Li et al., 2008)*. DNMT1 and PARP1 form a complex with CTCF to maintain the unmethylated status of CTCF binding sites *(Guastafierro et al., 2008; Zampieri et al., 2012)*. The loss of PARP and CTCF binding result in *de novo* methylation of CpG. Next to CTCF-DNA interactions CTCF-protein interactions are also highly important to mediate its function in chromatin and spatial organization.

CTCF Interaction partner	Function	CTCF interaction domain	References
Large subunit Pol II	Transcription factor	C-terminus	Chernukhin et al 2007
YBI	Transcription factor	Zinc finger domain	Chernukhin et al 2000
KAISO	Transcription factor	C-terminus	Defossez and Gilson 2002
YYI	Transcription factor	N-terminus	Donohoe et al 2007
CIITA	Transcriptional co-activator	Unknown	Majumder et al 2008
CHD8	Chromatin modification	Zinc finger domain	Ishihara et al 2006
SIN3A	Chromatin modification	Zinc finger domain	Lutz et al 2000
PRC2	Chromatin remodeler	Zinc finger domain	Li et al 2008
SUZ12	Transcription factor	Unknown	Li et al 2008
DNMT1	DNA methylation	Unknown	Guastafierro et al 2008, Zampieri et al 2012
PARP1	Protein modification	Unknown	Guastafierro et al 2008, Zampieri et al 2012
Cohesin	Chromatin organization	C-terminus	Xiao et al 2011
Nucleophosmin	Nucleolar protein	Unknown	Yusufzai et al 2004

Table 2. CTCF interacting partners
Known CTCF interacting partners, their main function and CTCF interaction domain.

Cohesin and CTCF

Cohesin is a highly conserved multisubunit protein complex. It consists of four subunits: two coiled-coil ATPases, SMC1 and SMC3, that are rod-shaped subunits forming a 45 nm long ring-like structure with RAD21, that connects SMC1 and SMC3, and SA1 (or SA2), that binds to RAD21 *(Anderson et al., 2002; Haering et al., 2002)*. Cohesin is classically known as an essential factor for the establishment of sister chromatid cohesion, prior to chromosome segregation. Cohesin is loaded on the DNA in G1 via loading factors NIPBL and MAU-2 *(Ciosk et al., 2000; Misulovin et al., 2008)*. During S-phase cohesion is established and maintained by SORORIN *(Lafont et al., 2010; Schmitz et al., 2007)*. During prophase AURORA B, PLK1, and WAPL remove the majority of cohesin, except centromeric-bound cohesin, which is protected by SGO1 *(Kitajima et al., 2006; Lipp et al., 2007; Salic et al., 2004; Sumara et al., 2002)*. Sister chromatid cohesion at the centromere persists until the chromatids are properly attached to the opposite poles of the mitotic spindle. At this point SEPARASE cleaves RAD21 subunit, which results in removal of the remaining cohesin. Removal of cohesin allows chromosomes to segregate *(Salic et al., 2004; Uhlmann et al., 2000; Waizenegger et al., 2000)*.

In addition to this classical function cohesin was more recently shown to also be involved in gene regulation and chromatin organization. In fact, genome-wide studies showed that the vast majority of cohesin-bound sites co-localize with CTCF sites *(Parelho et al., 2008; Wendt et al., 2008)*. A bio-informatics approach in which all CTCF binding sites derived from 56 human cell lines and all cohesin binding sites derived from 8 human cell lines were compared identified ~12,000 CTCF-cohesin sites *(Li et al., 2013)*. It has been proposed that the direct interaction of CTCF and cohesin is established between the cohesin subunit SA2 and the C-terminal tail of CTCF *(Xiao et al., 2011)*. Regardless of whether a direct interaction exists, CTCF is responsible for cohesin localization along the genome and therefore mediates long-range interactions together with cohesin *(Hadjur et al., 2009; Mishiro et al., 2009; Nativio et al., 2009)*.

Besides shared cohesin-CTCF sites, cohesin-specific binding sites (not bound by CTCF) have also been described; they coincide with enhancer marker p300, and tissue specific transcription factors. Moreover, these sites tend to be more tissue-specific when compared to CTCF-cohesin sites *(Faure et al., 2012; Nitzsche et al., 2011; Schmidt et al., 2010)*. Cohesin-specific binding sites are able to mediate long-range interactions in a tissue specific fashion *(Demare et al., 2013; Phillips-Cremins et al., 2013; Schmidt et al., 2010)*. Cohesin mediates together with CTCF long-range interactions e.g. in the apolipoprotein locus, interferon gamma locus and Igf2/H19

locus. Both proteins control proper expression of these genes, since depletion of either cohesin or CTCF results in aberrant gene expression and disruption of chromatin architecture *(Hadjur et al., 2009; Mishiro et al., 2009; Nativio et al., 2009)*. Thus, CTCF mediates long-range interactions in collaboration with cohesin and other factors.

CTCFL - Identification and gene conservation

In 2002 a CTCF paralogue, termed CTCFL or BORIS (Brother Of the Regulator of Imprinted Sites), was identified that is specifically expressed in testis *(Loukinov et al., 2002)*. Just like CTCF, CTCFL has N- and C-terminal domains flanking an eleven ZF domain that shows 71% identity to the DNA binding domain of CTCF. In contrast, CTCFL's N- and C-termini are distinct compared to CTCF *(Loukinov et al., 2002)*. It has been hypothesized that a *Ctcf* gene duplication event during vertebrate evolution gave rise to *Ctcfl* since the gene is present in mammals, amphibians, reptiles, and monotremes, but not in avians (birds) *(Hore et al., 2008; Loukinov et al., 2002)*. CTCFL expression gradually constricted during evolution to the gonads in marsupials and cattle and eventually to the testis in mice and man *(Hore et al., 2008)*. This tissue-specific pattern of expression is in stark contrast to that of CTCF. No cell so far has been detected lacking CTCF.

The *CTCFL* gene is located on chromosome 20q13.2 in humans and on chromosome 2 in mice. In humans this region is frequently amplified in human cancer *(Klenova et al., 2002)*. The human *CTCFL* gene has 16 (10 coding) exons and 3 alternative promoters (A, B and C) and encodes at least 23 *CTCFL* isoforms, only 7 contain the full-length 11 ZF domain *(Pugacheva et al., 2010; Renaud et al., 2007)*. All three *CTCFL* promoters in human have multiple CTCF sites and, at least in cell culture, their activities appear to be partly controlled by CTCF *(Renaud et al., 2007)*. Remarkably, as shown in chapter 4, murine *Ctcfl* is much simpler in terms of isoform expression.

CTCFL - Intracellular distribution and function

As mentioned above CTCFL is mainly expressed in the testis. The exact cellular localization of CTCFL is under debate. The first immunostainings of adult mouse testis showed a mutually exclusive expression of CTCFL and CTCF *(Loukinov et al., 2002)*. Here, CTCFL was restricted to primary spermatocytes in meiotic prophase. Remarkably, CTCFL localization was cytoplasmic rather than nuclear. In contrast, CTCF showed a completely nuclear localization pattern and could only be detected during later stages of spermatogenesis, in round spermatids. It was proposed that although both CTCF and CTCFL could bind to the same target site they could not compete with each other due to the mutually exclusive expression pattern during male germ cell differentiation. This also led to the hypothesis that CTCFL substitutes for the absence of CTCF and that this switch is associated with reprogramming of DNA methylation during spermatogenesis *(Loukinov et al., 2002)*.

Another report described CTCFL in gonocytes during embryonic development, at 14.5 d.p.c. and revealed a nuclear localization of the protein in spermatogonia after birth *(Jelinic et al., 2006)*. This finding was linked to ongoing methylation and appearance of de novo DNA methyltransferases. In addition, CTCFL interaction with PRMT7, a protein arginine methyltransferase, was associated with the establishment of DNA methylation of imprinted genes by marking histones H2A and H4 with methylated arginine residues at sites for de novo DNA methylation by DNMT3A *(Jelinic et al., 2006)*. These data suggested an active role for CTCFL in DNA methylation reprogramming. However, *Ctcfl* knock out mice did not show any embryonic phenotype that could be associated with imprinting defects *(Suzuki et al., 2010)*. Instead, male knock out mice displayed a subfertility phenotype accompanied by small testis and increased apoptosis *(Suzuki et al., 2010)*. Intriguingly, in this study CTCFL was localized in round spermatids

and not in spermatocytes.

To gain further insights into the function of CTCFL, knock out mice were generated *(Suzuki et al., 2010)*. In the *Ctcfl* knock out mice, two testis-related genes, *Prss50* (Protease serine 50, also known as testis-specific protease 50 (*Tsp50*)) and the testis specific isoform of *Gal3st1* (*Cst*), were shown to be aberrantly expressed *(Suzuki et al., 2010)* and to be activated by CTCFL *(Kosaka-Suzuki et al., 2011; Suzuki et al., 2010)*. CTCFL positively regulated *Gal3st1* and *Prss50*, suggesting a role for CTCFL in the regulation of these genes during spermatogenesis. CTCF also binds the testis-specific promoter of *Gal3st1*, tested *in vitro*, but does not appear to affect transcription of this gene *(Suzuki et al., 2010)*.

Several groups have found CTCFL to be expressed outside the testis. An interesting observation is the detection of CTCFL in human embryonic ovary *(Monk et al., 2008; Pugacheva et al., 2010)*. Besides the reproductive organs, CTCFL was also detected in human embryonic stem cells *(Pugacheva et al., 2010)* and in human skin *(Rosa-Garrido et al., 2012)*. In human skin CTCFL expression is restricted to keratinocytes in the differentiating layers of the epidermis. CTCFL accumulates in nuclear and peri-nuclear spots identified as centrosomes in keratinocytes. Within the nucleus CTCFL localizes to the nucleolus. CTCFL overexpression resulted in the accumulation of S and G2/M phased cells and in polyploidy, suggesting a function for CTCFL during the keratinocyte cell cycle and in the maintenance of genomic stability *(Rosa-Garrido et al., 2012)*. Furthermore, CTCFL was also localized to DNA sites with high histone occupancy indicating that the protein localizes to less condensed and euchromatic regions *(Rosa-Garrido et al., 2012)*.

In vitro analysis revealed that CTCF and CTCFL bind similar DNA sequences *in vitro* *(Loukinov et al., 2002)*. A number of cell culture experiments have subsequently been carried out to further investigate the DNA binding specificity of CTCFL. In one study, the unmethylated promoter of the telomerase reverse transcriptase (*hTERT*) gene was found to be bound by CTCFL in testicular (NCCIT) and ovarian (OVCAR-3) tumor cell lines *(Renaud et al., 2011)*. In addition, both CTCFL and CTCF were shown to bind a site within the first exon, here CTCFL competes with CTCF to initiate *hTERT* expression *(Renaud et al., 2011; Renaud et al., 2005)*. In contrast, while CTCF binds DNA in a methylation-sensitive manner, a methylation-independent binding of CTCFL to the human *Igf2/H19 ICR* has been reported *(Nguyen et al., 2008; Pugacheva et al., 2010)*. Together these and other studies raise a confusing picture about the mode of binding, intracellular localization and function of CTCFL.

CTCFL - A role in human cancer?

During germ cell development many germ cell specific factors are essential to establish proper spermatogenesis and oogenesis. However, a subset of these genes, most of them part of multigene families, are also expressed in various cancers and are called cancer germ line genes (CG genes). These CG genes are usually activated in cancer due to promoter demethylation *(Cheng et al., 2011)*. Approximately 50% of CG genes can be mapped to the X-chromosome and, as explained above, the majority shows immunogenicity of their protein products (giving rise to the term cancer testis antigen (CTA)), which makes them an interesting target for cancer immunotherapy *(Cheng et al., 2011; Simpson et al., 2005)*. So far 70 gene families and more than 140 proteins have been identified as CTA, including CTCFL itself *(Cheng et al., 2011)*. In addition, CTCFL is also expressed in cancer cell lines derived from melanoma, neuroblastoma, breast cancer, prostate cancer and colon cancer, and in primary tumors including breast cancer, prostate cancer and colon cancer *(Vatolin et al., 2005)*. The fact that CTCFL is normally only found in testis but is upregulated in cancer was the basis to suggest that CTCFL could act as tumor-promoting protein. Higher frequency of *Ctcfl* expression and other CG genes in metastatic tumors could be due to selection of these genes in this phase of tumorigenesis. It has been proposed that aberrant activation of CG genes in somatic cells could result in

abnormal re-programming events that would lead to fully transformed cancer cells *(Wang et al., 2011)*. In addition, the expression of a subset of CG genes in cancer could lead to the activation of a significant number of germline genes that can initiate the germ cell genetic program *(Wang et al., 2011)*. This occurs often in malignant cells, which suggests that the activation of the CG genes results in a selective advantage in the process of oncogenic transformation *(Rousseaux et al., 2013; Wang et al., 2011)*.

A role for CTCFL during carcinogenesis has been investigated in cultured cell lines and in cancer tissues but has not yet been clearly demonstrated. In one of the first genome-wide expression studies it was shown that in spermatocytic seminomas, tumors originating from primary spermatocytes, *CTCFL*, *PRSS50* and *SYCO1* expression was elevated *(Looijenga et al., 2006)*. In addition, *in vitro* transient *CTCFL* expression in fibroblasts leads to the reactivation of several X-linked cancer testis genes, *MAGE* genes, *NY-ESO-1* and pluripotency factor *POU5F1 (Bhan et al., 2011)*. Here, CTCFL expression coincides with partial demethylation of the *MAGE-A1* promoter suggesting that CTCFL disrupts the regulation of these cancer-testis genes in somatic cells. Remarkably, CTCF binds in a methylation-independent manner to the *MAGE-A1* promoter in NHDF cells and is competed from its binding site by CTCFL *(Vatolin et al., 2005)*. Other members of the *MAGE-A* family were also positively regulated by CTCFL in NHDF cells *(Bhan et al., 2011),* and were expressed in primary head and neck squamous cell carcinoma *(Smith et al., 2009a)*. CTCFL binding correlated with an active chromatin state, containing H3K8ac, H3K14ac and H3K4me3, on *MAGEA* promoters *(Bhan et al., 2011)*. Interestingly, expression of these genes and *CTCFL* promote cell growth *(Smith et al., 2009a)*. In general, in cultured cells CG expression coincides with *CTCFL* expression and DNA demethylation.

In breast cancer patients, elevated *CTCFL* expression levels were detected in neutrophil polymorphonuclear granulocytes (PMNs) and in all types of breast cancer (non-malignant and malignant) but not in primary breast cells *(D'Arcy et al., 2006)*. *CTCFL* expression in PMNs could therefore be used as a diagnostic tool to detect breast cancer *(D'Arcy et al., 2006)*. Taxane docetaxel is a cytotoxic antitumor agent used in patients with breast cancer and has an anti-proliferative effect. Histone deacetylase inhibitors, such as apicidin, are another type of anticancer agents used in breast cancer patients, which suppress growth of human breast cancer cells by modulating the cell cycle and inducing apoptosis. Treatment with a combination of apicidin and docetaxel in breast cancer cell lines induced *CTCFL* expression in highly metastatic breast cancer cells *(Buoncervello et al., 2012)*. Together with *CTCFL* the expression of presumed target genes, *NY-ESO-1 ad MAGE-A1*, were up-regulated in MDA-MB-435 cells *(Buoncervello et al., 2012)*. This data suggests that treatment with these antitumor agents enhances expression of these particular CG (onco)genes. How the expression of CG genes affects the efficacy of antitumor drugs has to be determined.

In ovarian cancer *CTCFL* is expressed due to promoter hypomethylation *(Link et al., 2013; Woloszynska-Read et al., 2007)*. *CTCFL* expression can be induced via DAC (DNMT inhibitor 5-aza-2'-deoxycytidine) that demethylates the *CTCFL* promoter *(Woloszynska-Read et al., 2007)*. However, *CTCFL* over-expression in ovarian cells did not induce expression of CG genes (*MAGEA-1, NY-ESO-1 and XAGE-1*) and global DNA hypomethylation *(Woloszynska-Read et al., 2010)*. Another report by the same group showed DNA hypomethylation in epithelial ovarian cancer and on CG promoters *(Woloszynska-Read et al., 2011)*.

As shown in the examples above, virtually all studies relating CTCFL to cancer, either examine CTCFL expression and function in cell culture focusing on a restricted set of CTCFL binding sites and presumed target genes, or document (lack of) correlation of *CTCFL* expression with that of other genes, in particular the CTAs. In a very recent study, it was found that 50% of the esophageal squamous cell cancers (ESCCs) stage pT1N+ express *CTCFL (Okabayashi et al., 2012)*. Importantly, it was shown that patients with CTCFL-positive ESCC show a poor 5-year survival rate. Furthermore, a CTCFL knock down resulted in decreased cell proliferation and invasion ability of ESCC cell lines *(Okabayashi et al., 2012)*. This is the first time

that CTCFL expression was linked to the survival of cancer patients. Together, studies on CTCFL (and CTCF) in cancer suggest that CTCFL plays a role during carcinogenesis, however no mechanistic pathway has been revealed so far.

References

Anderson, D.E., Losada, A., Erickson, H.P., and Hirano, T. (2002). Condensin and cohesin display different arm conformations with characteristic hinge angles. J Cell Biol 156, 419-424.

Anderson, E.L., Baltus, A.E., Roepers-Gajadien, H.L., Hassold, T.J., de Rooij, D.G., van Pelt, A.M., and Page, D.C. (2008). Stra8 and its inducer, retinoic acid, regulate meiotic initiation in both spermatogenesis and oogenesis in mice. Proc Natl Acad Sci U S A 105, 14976-14980.

Baltus, A.E., Menke, D.B., Hu, Y.C., Goodheart, M.L., Carpenter, A.E., de Rooij, D.G., and Page, D.C. (2006). In germ cells of mouse embryonic ovaries, the decision to enter meiosis precedes premeiotic DNA replication. Nat Genet 38, 1430-1434.

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.

Baniahmad, A., Steiner, C., Kohne, A.C., and Renkawitz, R. (1990). Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid hormone receptor binding site. Cell 61, 505-514.

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, 823-837.

Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature 405, 482-485.

Belmont, A.S., and Bruce, K. (1994). Visualization of G1 chromosomes: a folded, twisted, supercoiled chromonema model of interphase chromatid structure. J Cell Biol 127, 287-302.

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, 1395-1401.

Bestor, T.H. (1992). Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. Embo J 11, 2611-2617.

Bhan, S., Negi, S.S., Shao, C., Glazer, C.A., Chuang, A., Gaykalova, D.A., Sun, W., Sidransky, D., Ha, P.K., and Califano, J.A. (2011). BORIS binding to the promoters of cancer testis antigens, MAGEA2, MAGEA3, and MAGEA4, is associated with their transcriptional activation in lung cancer. Clin Cancer Res 17, 4267-4276.

Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.L., Ruan, Y., Wei, C.L., Ng, H.H., et al. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res 18, 1752-1762.

Boussouar, F., Rousseaux, S., and Khochbin, S. (2008). A new insight into male genome reprogramming by histone variants and histone code. Cell Cycle 7, 3499-3502.

Boyle, A.P., Song, L., Lee, B.K., London, D., Keefe, D., Birney, E., Iyer, V.R., Crawford, G.E., and Furey, T.S. (2011). High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. Genome Res 21, 456-464.

Brandeis, M., Kafri, T., Ariel, M., Chaillet, J.R., McCarrey, J., Razin, A., and Cedar, H. (1993). The ontogeny of allele-specific methylation associated with imprinted genes in the mouse. Embo J 12, 3669-3677.

Brannan, C.I., Bedell, M.A., Resnick, J.L., Eppig, J.J., Handel, M.A., Williams, D.E., Lyman, S.D., Donovan, P.J., Jenkins, N.A., and Copeland, N.G. (1992). Developmental abnormalities in Steel17H mice result from a splicing defect in the steel factor cytoplasmic tail. Genes Dev 6, 1832-1842.

Brown, C.R., Kennedy, C.J., Delmar, V.A., Forbes, D.J., and Silver, P.A. (2008). Global histone acetylation induces functional genomic reorganization at mammalian nuclear pore complexes. Genes Dev 22, 627-639.

Brownell, J.E., and Allis, C.D. (1996). Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. Curr Opin Genet Dev 6, 176-184.

Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., and Allis, C.D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell 84, 843-851.

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, 5234-5244.

Buoncervello, M., Borghi, P., Romagnoli, G., Spadaro, F., Belardelli, F., Toschi, E., and Gabriele, L. (2012). Apicidin and docetaxel combination treatment drives CTCFL expression and HMGB1 release acting as potential antitumor immune response inducers in metastatic breast cancer cells. Neoplasia 14, 855-867.

Burcin, M., Arnold, R., Lutz, M., Kaiser, B., Runge, D., Lottspeich, F., Filippova, G.N., Lobanenkov, V.V., and Renkawitz, R. (1997). Negative protein 1, which is required for function of the chicken lysozyme gene silencer in conjunction with hormone receptors, is identical to the multivalent zinc finger repressor CTCF. Mol Cell Biol 17, 1281-1288.

Burke, L.J., Hollemann, T., Pieler, T., and Renkawitz, R. (2002). Molecular cloning and expression of the chromatin insulator protein CTCF in Xenopus laevis. Mech Dev 113, 95-98.

Butler, J.E., and Kadonaga, J.T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes Dev 16, 2583-2592.

Capelson, M., Liang, Y., Schulte, R., Mair, W., Wagner, U., and Hetzer, M.W. (2010). Chromatin-bound nuclear pore components regulate gene expression in higher eukaryotes. Cell 140, 372-383.

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, 1106-1117.

Cheng, Y.H., Wong, E.W., and Cheng, C.Y. (2011). Cancer/testis (CT) antigens, carcinogenesis and spermatogenesis. Spermatogenesis 1, 209-220.

Chernukhin, I., Shamsuddin, S., Kang, S.Y., Bergstrom, R., Kwon, Y.W., Yu, W., Whitehead, J., Mukhopadhyay, R., Docquier, F., Farrar, D., 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.

Chernukhin, I.V., Shamsuddin, S., Robinson, A.F., Carne, A.F., Paul, A., El-Kady, A.I., Lobanenkov, V.V., and Klenova, E.M. (2000). Physical and functional interaction between two pluripotent proteins, the Y-box DNA/RNA-binding factor, YB-1, and the multivalent zinc finger factor, CTCF. J Biol Chem 275, 29915-29921.

Chiarini-Garcia, H., Hornick, J.R., Griswold, M.D., and Russell, L.D. (2001). Distribution of type A spermatogonia in the mouse is not random. Biol Reprod 65, 1179-1185.

Chien, R., Zeng, W., Kawauchi, S., Bender, M.A., Santos, R., Gregson, H.C., Schmiesing, J.A., Newkirk, D.A., Kong, X., Ball, A.R., Jr., et al. (2011). Cohesin mediates chromatin interactions that regulate mammalian beta-globin expression. J Biol Chem 286, 17870-17878.

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

Cremer, T., and Cremer, M. (2010). Chromosome territories. Cold Spring Harb Perspect Biol 2, a003889.

Cuddapah, S., Jothi, R., Schones, D.E., Roh, T.Y., Cui, K., and Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. Genome Res 19, 24-32.

D'Arcy, V., Abdullaev, Z.K., Pore, N., Docquier, F., Torrano, V., Chernukhin, I., Smart, M., Farrar, D., Metodiev, M., Fernandez, N., et al. (2006). The potential of BORIS detected in the leukocytes of breast cancer patients as an early marker of tumorigenesis. Clin Cancer Res 12, 5978-5986.

de Kretser, D.M. (1997). Male infertility. Lancet 349, 787-790.

de Rooij, D.G. (1998). Stem cells in the testis. Int J Exp Pathol 79, 67-80.

Dean, A. (2011). In the loop: long range chromatin interactions and gene regulation. Brief Funct Genomics 10, 3-10.

Defossez, P.A., and Gilson, E. (2002). The vertebrate protein CTCF functions as an insulator in Saccharomyces cerevisiae. Nucleic Acids Res 30, 5136-5141.

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

Dettin, L., Ravindranath, N., Hofmann, M.C., and Dym, M. (2003). Morphological characterization of the spermatogonial subtypes in the neonatal mouse testis. Biol Reprod 69, 1565-1571.

Dillon, N., and Festenstein, R. (2002). Unravelling heterochromatin: competition between positive and negative factors regulates accessibility. Trends Genet 18, 252-258.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376-380.

Eddy, E.M. (2002). Male germ cell gene expression. Recent Prog Horm Res 57, 103-128.

Engel, N., Thorvaldsen, J.L., and Bartolomei, M.S. (2006). CTCF binding sites promote transcription initiation and prevent DNA methylation on the maternal allele at the imprinted H19/Igf2 locus. Hum Mol Genet 15, 2945-2954.

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, 447-455.

Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473, 43-49.

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, 3820-3831.

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

Fedoriw, A.M., Stein, P., Svoboda, P., Schultz, R.M., and Bartolomei, M.S. (2004). Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting. Science 303, 238-240.

Ferraiuolo, M.A., Rousseau, M., Miyamoto, C., Shenker, S., Wang, X.Q., Nadler, M., Blanchette, M., and Dostie, J. (2010). The three-dimensional architecture of Hox cluster silencing. Nucleic Acids Res 38, 7472-7484.

Filippova, G.N., Fagerlie, S., Klenova, E.M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P.E., Collins, S.J., and Lobanenkov, V.V. (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol Cell Biol 16, 2802-2813.

Finch, J.T., and Klug, A. (1976). Solenoidal model for superstructure in chromatin. Proc Natl Acad Sci U S A 73, 1897-1901.

Gaucher, J., Reynoird, N., Montellier, E., Boussouar, F., Rousseaux, S., and Khochbin, S. (2010). From meiosis to postmeiotic events: the secrets of histone disappearance. FEBS J 277, 599-604.

Ghamari, A., van de Corput, M.P., Thongjuea, S., van Cappellen, W.A., van Ijcken, W., van Haren, J., Soler, E., Eick, D., Lenhard, B., and Grosveld, F.G. (2013). In vivo live imaging of RNA polymerase II transcription factories in primary cells. Genes Dev 27, 767-777.

Gibcus, J.H., and Dekker, J. (2013). The hierarchy of the 3D genome. Mol Cell 49, 773-782.

Gifford, C.A., Ziller, M.J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A.K., Kelley, D.R., Shishkin, A.A., Issner, R., et al. (2013). Transcriptional and Epigenetic Dynamics during Specification of Human Embryonic Stem Cells. Cell 153, 1149-1163.

Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N.P., and Bickmore, W.A. (2004). Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. Cell 118, 555-566.

Griswold, M.D. (1998). The central role of Sertoli cells in spermatogenesis. Semin Cell Dev Biol 9, 411-416.

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, 975-985.

Guastafierro, T., Cecchinelli, B., Zampieri, M., Reale, A., Riggio, G., Sthandier, O., Zupi, G., Calabrese, L., and Caiafa, P. (2008). CCCTC-binding factor activates PARP-1 affecting DNA methylation machinery. J Biol Chem 283, 21873-21880.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., et al. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature.

Guibert, S., Zhao, Z., Sjolinder, M., Gondor, A., Fernandez, A., Pant, V., and Ohlsson, R. (2012). CTCF-binding sites within the H19 ICR differentially regulate local chromatin structures and cis-acting functions. Epigenetics 7, 361-369.

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

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

Han, L., Lee, D.H., and Szabo, P.E. (2008). CTCF is the master organizer of domain-wide allele-specific chromatin at the H19/Igf2 imprinted region. Mol Cell Biol 28, 1124-1135.

Handel, M.A., and Schimenti, J.C. (2010). Genetics of mammalian meiosis: regulation, dynamics and impact on fertility. Nat Rev Genet 11, 124-136.

Handoko, L., Xu, H., Li, G., Ngan, C.Y., Chew, E., Schnapp, M., Lee, C.W., Ye, C., Ping, J.L., Mulawadi, F., et al. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. Nat Genet 43, 630-638.

Haneji, T., Maekawa, M., and Nishimune, Y. (1983). Retinoids induce differentiation of type A spermatogonia in vitro: organ culture of mouse cryptorchid testes. J Nutr 113, 1119-1123.

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, 486-489.

Heath, H., Ribeiro de Almeida, C., Sleutels, F., Dingjan, G., van de Nobelen, S., Jonkers, I., Ling, K.W., Gribnau, J., Renkawitz, R., Grosveld, F., et al. (2008). CTCF regulates cell cycle progression of alphabeta T cells in the thymus. Embo J 27, 2839-2850.

Heger, P., Marin, B., and Schierenberg, E. (2009). Loss of the insulator protein CTCF during nematode evolution. BMC Mol Biol 10, 84.

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

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet 39, 311-318.

Holliday, R., and Pugh, J.E. (1975). DNA modification mechanisms and gene activity during development. Science 187, 226-232.

Holohan, E.E., Kwong, C., Adryan, B., Bartkuhn, M., Herold, M., Renkawitz, R., Russell, S., and White, R. (2007). CTCF genomic binding sites in Drosophila and the organisation of the bithorax complex. PLoS Genet 3, e112.

Hore, T.A., Deakin, J.E., and Marshall Graves, J.A. (2008). The evolution of epigenetic regulators CTCF and BORIS/CTCFL in amniotes. PLoS Genet 4, e1000169.

Hou, C., Dale, R., and Dean, A. (2010). Cell type specificity of chromatin organization mediated by CTCF and cohesin. Proc Natl Acad Sci U S A 107, 3651-3656.

Hou, C., Zhao, H., Tanimoto, K., and Dean, A. (2008). CTCF-dependent enhancer-blocking by alternative chromatin loop formation. Proc Natl Acad Sci U S A 105, 20398-20403.

Iborra, F.J., Pombo, A., Jackson, D.A., and Cook, P.R. (1996). Active RNA polymerases are localized within discrete transcription “factories’ in human nuclei. J Cell Sci 109 (Pt 6), 1427-1436.

Ishihara, K., Oshimura, M., and Nakao, M. (2006). CTCF-dependent chromatin insulator is linked to epigenetic remodeling. Mol Cell 23, 733-742.

Jan, S.Z., Hamer, G., Repping, S., de Rooij, D.G., van Pelt, A.M., and Vormer, T.L. (2012). Molecular control of rodent spermatogenesis. Biochim Biophys Acta 1822, 1838-1850.

Jelinic, P., Stehle, J.C., and Shaw, P. (2006). The testis-specific factor CTCFL cooperates with the protein methyltransferase PRMT7 in H19 imprinting control region methylation. PLoS Biol 4, e355.

Jin, C., and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes Dev 21, 1519-1529.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark ‘nucleosome-free regions’ of active promoters and other regulatory regions. Nat Genet 41, 941-945.

Jothi, R., Cuddapah, S., Barski, A., Cui, K., and Zhao, K. (2008). Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. Nucleic Acids Res 36, 5221-5231.

Kafri, T., Ariel, M., Brandeis, M., Shemer, R., Urven, L., McCarrey, J., Cedar, H., and Razin, A. (1992). Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. Genes Dev 6, 705-714.

Kanduri, M., Kanduri, C., Mariano, P., Vostrov, A.A., Quitschke, W., Lobanenkov, V., and Ohlsson, R. (2002). Multiple nucleosome positioning sites regulate the CTCF-mediated insulator function of the H19 imprinting control region. Mol Cell Biol 22, 3339-3344.

Kierszenbaum, A.L., and Tres, L.L. (2004). The acrosome-acroplaxome-manchette complex and the shaping of the spermatid head. Arch Histol Cytol 67, 271-284.

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, 1231-1245.

Kitajima, T.S., Sakuno, T., Ishiguro, K., Iemura, S., Natsume, T., Kawashima, S.A., and Watanabe, Y. (2006). Shugoshin collaborates with protein phosphatase 2A to protect cohesin. Nature 441, 46-52.

Klenova, E.M., Chernukhin, I.V., El-Kady, A., Lee, R.E., Pugacheva, E.M., Loukinov, D.I., Goodwin, G.H., Delgado, D., Filippova, G.N., Leon, J., et al. (2001). Functional phosphorylation sites in the C-terminal region of the multivalent multifunctional transcriptional factor CTCF. Mol Cell Biol 21, 2221-2234.

Klenova, E.M., Morse, H.C., 3rd, Ohlsson, R., and Lobanenkov, V.V. (2002). The novel BORIS + CTCF gene family is uniquely involved in the epigenetics of normal biology and cancer. Semin Cancer Biol 12, 399-414.

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, 7612-7624.

Kosaka-Suzuki, N., Suzuki, T., Pugacheva, E.M., Vostrov, A.A., Morse, H.C., 3rd, Loukinov, D., and Lobanenkov, V. (2011). Transcription factor BORIS (Brother of the Regulator of Imprinted Sites) directly induces expression of a cancer-testis antigen, TSP50, through regulated binding of BORIS to the promoter. J Biol Chem 286, 27378-27388.

Koubova, J., Menke, D.B., Zhou, Q., Capel, B., Griswold, M.D., and Page, D.C. (2006). Retinoic acid regulates sex-specific timing of meiotic initiation in mice. Proc Natl Acad Sci U S A 103, 2474-2479.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell 128, 693-705.

Kunarso, G., Chia, N.Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.S., Ng, H.H., and Bourque, G. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat Genet 42, 631-634.

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, 10684-10689.

Lafont, A.L., Song, J., and Rankin, S. (2010). Sororin cooperates with the acetyltransferase Eco2 to ensure DNA replication-dependent sister chromatid cohesion. Proc Natl Acad Sci U S A 107, 20364-20369.

Le, H.D., Donaldson, K.M., Cook, K.R., and Karpen, G.H. (2004). A high proportion of genes involved in position effect variegation also affect chromosome inheritance. Chromosoma 112, 269-276.

Li, T., Hu, J.F., Qiu, X., Ling, J., Chen, H., Wang, S., Hou, A., Vu, T.H., and Hoffman, A.R. (2008). CTCF regulates allelic expression of Igf2 by orchestrating a promoter-polycomb repressive complex 2 intrachromosomal loop. Mol Cell Biol 28, 6473-6482.

Li, Y., Huang, W., Niu, L., Umbach, D.M., Covo, S., and Li, L. (2013). Characterization of constitutive CTCF/cohesin loci: a possible role in establishing topological domains in mammalian genomes. BMC Genomics 14, 553.

Lichter, P., Cremer, T., Borden, J., Manuelidis, L., and Ward, D.C. (1988). Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. Hum Genet 80, 224-234.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326, 289-293.

Lin, Y., Gill, M.E., Koubova, J., and Page, D.C. (2008). Germ cell-intrinsic and -extrinsic factors govern meiotic initiation in mouse embryos. Science 322, 1685-1687.

Link, P.A., Zhang, W., Odunsi, K., and Karpf, A.R. (2013). BORIS/CTCFL mRNA isoform expression and epigenetic regulation in epithelial ovarian cancer. Cancer Immun 13, 6.

Lipp, J.J., Hirota, T., Poser, I., and Peters, J.M. (2007). Aurora B controls the association of condensin I but not condensin II with mitotic chromosomes. J Cell Sci 120, 1245-1255.

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, 1743-1753.

Looijenga, L.H., Hersmus, R., Gillis, A.J., Pfundt, R., Stoop, H.J., van Gurp, R.J., Veltman, J., Beverloo, H.B., van Drunen, E., van Kessel, A.G., et al. (2006). Genomic and expression profiling of human spermatocytic seminomas: primary spermatocyte as tumorigenic precursor and DMRT1 as candidate chromosome 9 gene. Cancer Res 66, 290-302.

Loukinov, D.I., Pugacheva, E., Vatolin, S., Pack, S.D., Moon, H., Chernukhin, I., Mannan, P., Larsson, E., Kanduri, C., Vostrov, A.A., et al. (2002). BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. Proc Natl Acad Sci U S A 99, 6806-6811.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389, 251-260.

Lutz, M., Burke, L.J., Barreto, G., Goeman, F., Greb, H., Arnold, R., Schultheiss, H., Brehm, A., Kouzarides, T., Lobanenkov, V., et al. (2000). Transcriptional repression by the insulator protein CTCF involves histone deacetylases. Nucleic Acids Res 28, 1707-1713.

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

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

Mark, M., Jacobs, H., Oulad-Abdelghani, M., Dennefeld, C., Feret, B., Vernet, N., Codreanu, C.A., Chambon, P., and Ghyselinck, N.B. (2008). STRA8-deficient spermatocytes initiate, but fail to complete, meiosis and undergo premature chromosome condensation. J Cell Sci 121, 3233-3242.

Martin, D., Pantoja, C., Fernandez Minan, A., Valdes-Quezada, C., Molto, E., Matesanz, F., Bogdanovic, O., de la Calle-Mustienes, E., Dominguez, O., Taher, L., et al. (2011). Genome-wide CTCF distribution in vertebrates defines equivalent sites that aid the identification of disease-associated genes. Nat Struct Mol Biol 18, 708-714.

Maston, G.A., Evans, S.K., and Green, M.R. (2006). Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet 7, 29-59.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. Nature 403, 501-502.

Meistrich, M.L., Mohapatra, B., Shirley, C.R., and Zhao, M. (2003). Roles of transition nuclear proteins in spermiogenesis. Chromosoma 111, 483-488.

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

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

Mital, P., Hinton, B.T., and Dufour, J.M. (2011). The blood-testis and blood-epididymis barriers are more than just their tight junctions. Biol Reprod 84, 851-858.

Mizzen, C.A., Yang, X.J., Kokubo, T., Brownell, J.E., Bannister, A.J., Owen-Hughes, T., Workman, J., Wang, L., Berger, S.L., Kouzarides, T., et al. (1996). The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. Cell 87, 1261-1270.

Monk, M., Boubelik, M., and Lehnert, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. Development 99, 371-382.

Monk, M., Hitchins, M., and Hawes, S. (2008). Differential expression of the embryo/cancer gene ECSA(DPPA2), the cancer/testis gene BORIS and the pluripotency structural gene OCT4, in human preimplantation development. Mol Hum Reprod 14, 347-355.

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, 165-170.

Moore, J.M., Rabaia, N.A., Smith, L.E., Fagerlie, S., Gurley, K., Loukinov, D., Disteche, C.M., Collins, S.J., Kemp, C.J., Lobanenkov, V.V., et al. (2012). Loss of maternal CTCF is associated with peri-implantation lethality of Ctcf null embryos. PLoS One 7, e34915.

Muller, W.G., Walker, D., Hager, G.L., and McNally, J.G. (2001). Large-scale chromatin decondensation and recondensation regulated by transcription from a natural promoter. J Cell Biol 154, 33-48.

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

Nakahashi, H., Kwon, K.R., Resch, W., Vian, L., Dose, M., Stavreva, D., Hakim, O., Pruett, N., Nelson, S., Yamane, A., et al. (2013). A Genome-wide Map of CTCF Multivalency Redefines the CTCF Code. Cell Rep 3, 1678-1689.

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

Nemeth, A., Conesa, A., Santoyo-Lopez, J., Medina, I., Montaner, D., Peterfia, B., Solovei, I., Cremer, T., Dopazo, J., and Langst, G. (2010). Initial genomics of the human nucleolus. PLoS Genet 6, e1000889.

Nguyen, P., Cui, H., Bisht, K.S., Sun, L., Patel, K., Lee, R.S., Kugoh, H., Oshimura, M., Feinberg, A.P., and Gius, D. (2008). CTCFL/BORIS is a methylation-independent DNA-binding protein that preferentially binds to the paternal H19 differentially methylated region. Cancer Res 68, 5546-5551.

Nitzsche, A., Paszkowski-Rogacz, M., Matarese, F., Janssen-Megens, E.M., Hubner, N.C., Schulz, H., de Vries, I., Ding, L., Huebner, N., Mann, M., et al. (2011). RAD21 cooperates with pluripotency transcription factors in the maintenance of embryonic stem cell identity. PLoS One 6, e19470.

Nolis, I.K., McKay, D.J., Mantouvalou, E., Lomvardas, S., Merika, M., and Thanos, D. (2009). Transcription factors mediate long-range enhancer-promoter interactions. Proc Natl Acad Sci U S A 106, 20222-20227.

Noll, M., and Kornberg, R.D. (1977). Action of micrococcal nuclease on chromatin and the location of histone H1. J Mol Biol 109, 393-404.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485, 381-385.

Ohlsson, R., Renkawitz, R., and Lobanenkov, V. (2001). CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. Trends Genet 17, 520-527.

Okabayashi, K., Fujita, T., Miyazaki, J., Okada, T., Iwata, T., Hirao, N., Noji, S., Tsukamoto, N., Goshima, N., Hasegawa, H., et al. (2012). Cancer-testis antigen BORIS is a novel prognostic marker for patients with esophageal cancer. Cancer Sci 103, 1617-1624.

Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W., et al. (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. Nat Genet 36, 1065-1071.

Oudet, P., Gross-Bellard, M., and Chambon, P. (1975). Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. Cell 4, 281-300.

Page, S.L., and Hawley, R.S. (2004). The genetics and molecular biology of the synaptonemal complex. Annu Rev Cell Dev Biol 20, 525-558.

Palstra, R.J., Simonis, M., Klous, P., Brasset, E., Eijkelkamp, B., and de Laat, W. (2008). Maintenance of long-range DNA interactions after inhibition of ongoing RNA polymerase II transcription. PLoS One 3, e1661.

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.

Pant, V., Mariano, P., Kanduri, C., Mattsson, A., Lobanenkov, V., Heuchel, R., and Ohlsson, R. (2003). The nucleotides responsible for the direct physical contact between the chromatin insulator protein CTCF and the H19 imprinting control region manifest parent of origin-specific long-distance insulation and methylation-free domains. Genes Dev 17, 586-590.

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, 422-433.

Parkhurst, S.M., Harrison, D.A., Remington, M.P., Spana, C., Kelley, R.L., Coyne, R.S., and Corces, V.G. (1988). The Drosophila su(Hw) gene, which controls the phenotypic effect of the gypsy transposable element, encodes a putative DNA-binding protein. Genes Dev 2, 1205-1215.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W., Solovei, I., Brugman, W., Graf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M., et al. (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol Cell 38, 603-613.

Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. Cell 137, 1194-1211.

Phillips-Cremins, J.E., and Corces, V.G. (2013). Chromatin insulators: linking genome organization to cellular function. Mol Cell 50, 461-474.

Phillips-Cremins, J.E., Sauria, M.E., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. Cell 153, 1281-1295.

Pinkel, D., Landegent, J., Collins, C., Fuscoe, J., Segraves, R., Lucas, J., and Gray, J. (1988). Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. Proc Natl Acad Sci U S A 85, 9138-9142.

Pradhan, S., Bacolla, A., Wells, R.D., and Roberts, R.J. (1999). Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. J Biol Chem 274, 33002-33010.

Pugacheva, E.M., Kwon, Y.W., Hukriede, N.A., Pack, S., Flanagan, P.T., Ahn, J.C., Park, J.A., Choi, K.S., Kim, K.W., Loukinov, D., et al. (2006). Cloning and characterization of zebrafish CTCF: Developmental expression patterns, regulation of the promoter region, and evolutionary aspects of gene organization. Gene 375, 26-36.

Pugacheva, E.M., Suzuki, T., Pack, S.D., Kosaka-Suzuki, N., Yoon, J., Vostrov, A.A., Barsov, E., Strunnikov, A.V., Morse, H.C., 3rd, Loukinov, D., et al. (2010). The structural complexity of the human BORIS gene in gametogenesis and cancer. PLoS One 5, e13872.

Quitschke, W.W., Matthews, J.P., Kraus, R.J., and Vostrov, A.A. (1996). The initiator element and proximal upstream sequences affect transcriptional activity and start site selection in the amyloid beta-protein precursor promoter. J Biol Chem 271, 22231-22239.

Quitschke, W.W., Taheny, M.J., Fochtmann, L.J., and Vostrov, A.A. (2000). Differential effect of zinc finger deletions on the binding of CTCF to the promoter of the amyloid precursor protein gene. Nucleic Acids Res 28, 3370-3378.

Rattner, J.B., and Lin, C.C. (1985). Radial loops and helical coils coexist in metaphase chromosomes. Cell 42, 291-296.

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, 6883-6888.

Reik, W., Dean, W., and Walter, J. (2001). Epigenetic reprogramming in mammalian development. Science 293, 1089-1093.

Renaud, S., Loukinov, D., Alberti, L., Vostrov, A., Kwon, Y.W., Bosman, F.T., Lobanenkov, V., and Benhattar, J. (2011). BORIS/CTCFL-mediated transcriptional regulation of the hTERT telomerase gene in testicular and ovarian tumor cells. Nucleic Acids Res 39, 862-873.

Renaud, S., Loukinov, D., Bosman, F.T., Lobanenkov, V., and Benhattar, J. (2005). CTCF binds the proximal exonic region of hTERT and inhibits its transcription. Nucleic Acids Res 33, 6850-6860.

Renaud, S., Pugacheva, E.M., Delgado, M.D., Braunschweig, R., Abdullaev, Z., Loukinov, D., Benhattar, J., and Lobanenkov, V. (2007). Expression of the CTCF-paralogous cancer-testis gene, brother of the regulator of imprinted sites (BORIS), is regulated by three alternative promoters modulated by CpG methylation and by CTCF and p53 transcription factors. Nucleic Acids Res 35, 7372-7388.

Renda, M., Baglivo, I., Burgess-Beusse, B., Esposito, S., Fattorusso, R., Felsenfeld, G., and Pedone, P.V. (2007). Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. J Biol Chem 282, 33336-33345.
Rhee, H.S., and Pugh, B.F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell 147, 1408-1419.

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

Richmond, T.J., and Davey, C.A. (2003). The structure of DNA in the nucleosome core. Nature 423, 145-150.
Rivera, C.M., and Ren, B. (2013). Mapping human epigenomes. Cell 155, 39-55.

Roeder, R.G. (2005). Transcriptional regulation and the role of diverse coactivators in animal cells. FEBS Lett 579, 909-915.

Rosa-Garrido, M., Ceballos, L., Alonso-Lecue, P., Abraira, C., Delgado, M.D., and Gandarillas, A. (2012). A cell cycle role for the epigenetic factor CTCF-L/BORIS. PLoS One 7, e39371.

Rousseaux, S., Wang, J., and Khochbin, S. (2013). Cancer hallmarks sustained by ectopic activations of placenta/male germline genes. Cell Cycle 12, 2331-2332.

Saitoh, N., Bell, A.C., Recillas-Targa, F., West, A.G., Simpson, M., Pikaart, M., and Felsenfeld, G. (2000). Structural and functional conservation at the boundaries of the chicken beta-globin domain. Embo J 19, 2315-2322.

Salic, A., Waters, J.C., and Mitchison, T.J. (2004). Vertebrate shugoshin links sister centromere cohesion and kinetochore microtubule stability in mitosis. Cell 118, 567-578.

Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature 489, 109-113.

Schmidt, D., Schwalie, P.C., Ross-Innes, C.S., Hurtado, A., Brown, G.D., Carroll, J.S., Flicek, P., and Odom, D.T. (2010). A CTCF-independent role for cohesin in tissue-specific transcription. Genome Res 20, 578-588.

Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Goncalves, A., Kutter, C., Brown, G.D., Marshall, A., Flicek, P., and Odom, D.T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell 148, 335-348.

Schmitz, J., Watrin, E., Lenart, P., Mechtler, K., and Peters, J.M. (2007). Sororin is required for stable binding of cohesin to chromatin and for sister chromatid cohesion in interphase. Curr Biol 17, 630-636.

Schoenherr, C.J., Levorse, J.M., and Tilghman, S.M. (2003). CTCF maintains differential methylation at the Igf2/H19 locus. Nat Genet 33, 66-69.

Schrans-Stassen, B.H., Saunders, P.T., Cooke, H.J., and de Rooij, D.G. (2001). Nature of the spermatogenic arrest in Dazl -/- mice. Biol Reprod 65, 771-776.

Schrans-Stassen, B.H., van de Kant, H.J., de Rooij, D.G., and van Pelt, A.M. (1999). Differential expression of c-kit in mouse undifferentiated and differentiating type A spermatogonia. Endocrinology 140, 5894-5900.

Scott, K.C., Taubman, A.D., and Geyer, P.K. (1999). Enhancer blocking by the Drosophila gypsy insulator depends upon insulator anatomy and enhancer strength. Genetics 153, 787-798.

Sekimata, M., Perez-Melgosa, M., Miller, S.A., Weinmann, A.S., Sabo, P.J., Sandstrom, R., Dorschner, M.O., Stamatoyannopoulos, J.A., and Wilson, C.B. (2009). CCCTC-binding factor and the transcription factor T-bet orchestrate T helper 1 cell-specific structure and function at the interferon-gamma locus. Immunity 31, 551-564.

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

Simpson, A.J., Caballero, O.L., Jungbluth, A., Chen, Y.T., and Old, L.J. (2005). Cancer/testis antigens, gametogenesis and cancer. Nat Rev Cancer 5, 615-625.

Smith, I.M., Glazer, C.A., Mithani, S.K., Ochs, M.F., Sun, W., Bhan, S., Vostrov, A., Abdullaev, Z., Lobanenkov, V., Gray, A., et al. (2009a). Coordinated activation of candidate proto-oncogenes and cancer testes antigens via promoter demethylation in head and neck cancer and lung cancer. PLoS One 4, e4961.

Smith, S.T., Wickramasinghe, P., Olson, A., Loukinov, D., Lin, L., Deng, J., Xiong, Y., Rux, J., Sachidanandam, R., Sun, H., et al. (2009b). Genome wide ChIP-chip analyses reveal important roles for CTCF in Drosophila genome organization. Dev Biol 328, 518-528.

Soshnikova, N., Montavon, T., Leleu, M., Galjart, N., and Duboule, D. (2010). Functional analysis of CTCF during mammalian limb development. Dev Cell 19, 819-830.

Spana, C., Harrison, D.A., and Corces, V.G. (1988). The Drosophila melanogaster suppressor of Hairy-wing protein binds to specific sequences of the gypsy retrotransposon. Genes Dev 2, 1414-1423.

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

Stadhouders, R., Thongjuea, S., Andrieu-Soler, C., Palstra, R.J., Bryne, J.C., van den Heuvel, A., Stevens, M., de Boer, E., Kockx, C., van der Sloot, A., et al. (2012). Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development. Embo J 31, 986-999.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., et al. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature 480, 490-495.

Sumara, I., Vorlaufer, E., Stukenberg, P.T., Kelm, O., Redemann, N., Nigg, E.A., and Peters, J.M. (2002). The dissociation of cohesin from chromosomes in prophase is regulated by Polo-like kinase. Mol Cell 9, 515-525.

Suzuki, T., Kosaka-Suzuki, N., Pack, S., Shin, D.M., Yoon, J., Abdullaev, Z., Pugacheva, E., Morse, H.C., 3rd, Loukinov, D., and Lobanenkov, V. (2010). Expression of a testis-specific form of Gal3st1 (CST), a gene essential for spermatogenesis, is regulated by the CTCF paralogous gene BORIS. Mol Cell Biol 30, 2473-2484.

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, 4791-4800.

Tada, T., Tada, M., Hilton, K., Barton, S.C., Sado, T., Takagi, N., and Surani, M.A. (1998). Epigenotype switching of imprintable loci in embryonic germ cells. Dev Genes Evol 207, 551-561.

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, 1453-1465.

Tremblay, K.D., Duran, K.L., and Bartolomei, M.S. (1997). A 5' 2-kilobase-pair region of the imprinted mouse H19 gene exhibits exclusive paternal methylation throughout development. Mol Cell Biol 17, 4322-4329.

Tremethick, D.J. (2007). Higher-order structures of chromatin: the elusive 30 nm fiber. Cell 128, 651-654.

Tumbar, T., Sudlow, G., and Belmont, A.S. (1999). Large-scale chromatin unfolding and remodeling induced by VP16 acidic activation domain. J Cell Biol 145, 1341-1354.

Uhlmann, F., Wernic, D., Poupart, M.A., Koonin, E.V., and Nasmyth, K. (2000). Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast. Cell 103, 375-386.

van Koningsbruggen, S., Gierlinski, M., Schofield, P., Martin, D., Barton, G.J., Ariyurek, Y., den Dunnen, J.T., and Lamond, A.I. (2010). High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. Mol Biol Cell 21, 3735-3748.

Vatolin, S., Abdullaev, Z., Pack, S.D., Flanagan, P.T., Custer, M., Loukinov, D.I., Pugacheva, E., Hong, J.A., Morse, H., 3rd, Schrump, D.S., et al. (2005). Conditional expression of the CTCF-paralogous transcriptional factor BORIS in normal cells results in demethylation and derepression of MAGE-A1 and reactivation of other cancer-testis genes. Cancer Res 65, 7751-7762.

Venters, B.J., and Pugh, B.F. (2009). How eukaryotic genes are transcribed. Crit Rev Biochem Mol Biol 44, 117-141.

Vilar, J.M., and Saiz, L. (2005). DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. Curr Opin Genet Dev 15, 136-144.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457, 854-858.

Vostrov, A.A., and Quitschke, W.W. (1997). The zinc finger protein CTCF binds to the APBbeta domain of the amyloid beta-protein precursor promoter. Evidence for a role in transcriptional activation. J Biol Chem 272, 33353-33359.

Vostrov, A.A., Taheny, M.J., and Quitschke, W.W. (2002). A region to the N-terminal side of the CTCF zinc finger domain is essential for activating transcription from the amyloid precursor protein promoter. J Biol Chem 277, 1619-1627.

Waizenegger, I.C., Hauf, S., Meinke, A., and Peters, J.M. (2000). Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase. Cell 103, 399-410.

Walters, M.C., Fiering, S., Bouhassira, E.E., Scalzo, D., Goeke, S., Magis, W., Garrick, D., Whitelaw, E., and Martin, D.I. (1999). The chicken beta-globin 5'HS4 boundary element blocks enhancer-mediated suppression of silencing. Mol Cell Biol 19, 3714-3726.

Wang, J., Emadali, A., Le Bescont, A., Callanan, M., Rousseaux, S., and Khochbin, S. (2011). Induced malignant genome reprogramming in somatic cells by testis-specific factors. Biochim Biophys Acta 1809, 221-225.

Weake, V.M., and Workman, J.L. (2010). Inducible gene expression: diverse regulatory mechanisms. Nat Rev Genet 11, 426-437.

Wen, B., Wu, H., Loh, Y.H., Briem, E., Daley, G.Q., and Feinberg, A.P. (2012). Euchromatin islands in large heterochromatin domains are enriched for CTCF binding and differentially DNA-methylated regions. BMC Genomics 13, 566.

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.
Widom, J., and Klug, A. (1985). Structure of the 300A chromatin filament: X-ray diffraction from oriented samples. Cell 43, 207-213.

Woloszynska-Read, A., James, S.R., Link, P.A., Yu, J., Odunsi, K., and Karpf, A.R. (2007). DNA methylation-dependent regulation of BORIS/CTCFL expression in ovarian cancer. Cancer Immun 7, 21.

Woloszynska-Read, A., James, S.R., Song, C., Jin, B., Odunsi, K., and Karpf, A.R. (2010). BORIS/CTCFL expression is insufficient for cancer-germline antigen gene expression and DNA hypomethylation in ovarian cell lines. Cancer Immun 10, 6.

Woloszynska-Read, A., Zhang, W., Yu, J., Link, P.A., Mhawech-Fauceglia, P., Collamat, G., Akers, S.N., Ostler, K.R., Godley, L.A., Odunsi, K., et al. (2011). Coordinated cancer germline antigen promoter and global DNA hypomethylation in ovarian cancer: association with the BORIS/CTCF expression ratio and advanced stage. Clin Cancer Res 17, 2170-2180.

Xiao, T., Wallace, J., and Felsenfeld, G. (2011). Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. Mol Cell Biol 31, 2174-2183.

Xie, W., Schultz, M.D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J.W., Tian, S., Hawkins, R.D., Leung, D., et al. (2013). Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell 153, 1134-1148.

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, 7145-7150.

Yoshida, S., Sukeno, M., and Nabeshima, Y. (2007). A vasculature-associated niche for undifferentiated spermatogonia in the mouse testis. Science 317, 1722-1726.

Yu, Y.E., Zhang, Y., Unni, E., Shirley, C.R., Deng, J.M., Russell, L.D., Weil, M.M., Behringer, R.R., and Meistrich, M.L. (2000). Abnormal spermatogenesis and reduced fertility in transition nuclear protein 1-deficient mice. Proc Natl Acad Sci U S A 97, 4683-4688.

Zampieri, M., Guastafierro, T., Calabrese, R., Ciccarone, F., Bacalini, M.G., Reale, A., Perilli, M., Passananti, C., and Caiafa, P. (2012). ADP-ribose polymers localized on Ctcf-Parp1-Dnmt1 complex prevent methylation of Ctcf target sites. Biochem J 441, 645-652.

Zemel, S., Bartolomei, M.S., and Tilghman, S.M. (1992). Physical linkage of two mammalian imprinted genes, H19 and insulin-like growth factor 2. Nat Genet 2, 61-65.

Zhang, Y., McCord, R.P., Ho, Y.J., Lajoie, B.R., Hildebrand, D.G., Simon, A.C., Becker, M.S., Alt, F.W., and Dekker, J. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. Cell 148, 908-921.

Zhao, M., Shirley, C.R., Yu, Y.E., Mohapatra, B., Zhang, Y., Unni, E., Deng, J.M., Arango, N.A., Terry, N.H., Weil, M.M., et al. (2001). Targeted disruption of the transition protein 2 gene affects sperm chromatin structure and reduces fertility in mice. Mol Cell Biol 21, 7243-7255.

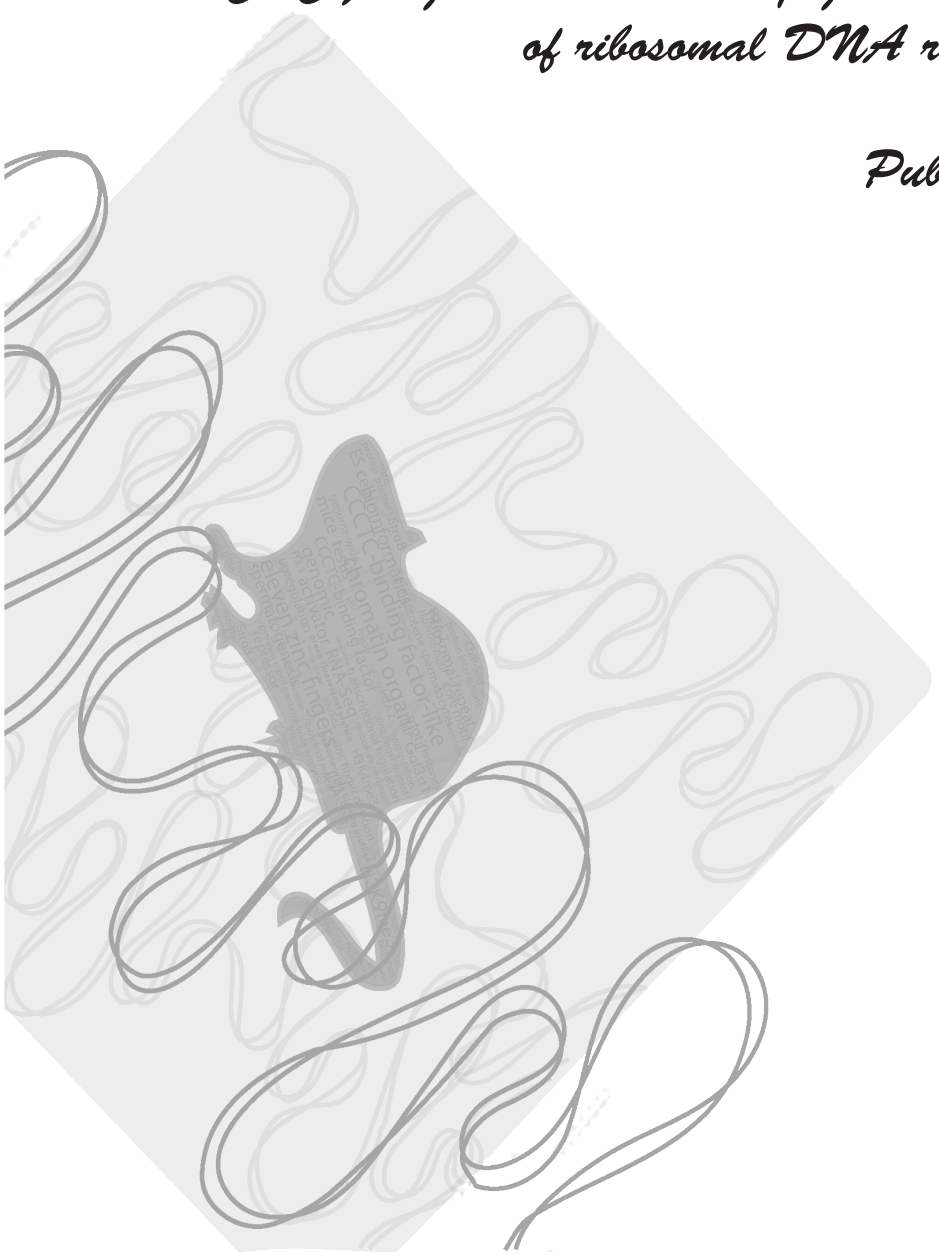
Zhou, Q., and Griswold, M.D. (2008). Regulation of spermatogonia.

Zuin J, Dixon.J.R., van der Reijden M.I.J.A., Ye Z., Kolovos P., Brouwer R.W.W., van de Corput M.P.C., van de Werken H.J.G., Knoch T.A., van IJcken W.F.J., Grosveld F.G., Ren B. and Wendt K.S. (Submitted). Cohesin and CTCF Differentially Affect Chromatin Architecture and Gene Expression in Human Cells

Chapter 2

CTCF regulates the local epigenetic state of ribosomal DNA repeats

Published



van de Nobelen *et al. Epigenetics & Chromatin* 2010, **3**:19
http://www.epigeneticsandchromatin.com/content/3/1/19

Epigenetics & Chromatin

RESEARCH Open Access

CTCF regulates the local epigenetic state of ribosomal DNA repeats

Suzanne van de Nobelen[1,6], Manuel Rosa-Garrido[2], Joerg Leers[3], Helen Heath[1,7], Widia Soochit[1], Linda Joosen[1], Iris Jonkers[4], Jeroen Demmers[5], Michael van der Reijden[1], Verónica Torrano[2], Frank Grosveld[1], M Dolores Delgado[2], Rainer Renkawitz[3], Niels Galjart[1*], Frank Sleutels[1*]

Abstract

Background: CCCTC binding factor (CTCF) is a highly conserved zinc finger protein, which is involved in chromatin organization, local histone modifications, and RNA polymerase II-mediated gene transcription. CTCF may act by binding tightly to DNA and recruiting other proteins to mediate its various functions in the nucleus. To further explore the role of this essential factor, we used a mass spectrometry-based approach to screen for novel CTCF-interacting partners.

Results: Using biotinylated CTCF as bait, we identified upstream binding factor (UBF) and multiple other components of the RNA polymerase I complex as potential CTCF-interacting partners. Interestingly, CTCFL, the testis-specific paralog of CTCF, also binds UBF. The interaction between CTCF(L) and UBF is direct, and requires the zinc finger domain of CTCF(L) and the high mobility group (HMG)-box 1 and dimerization domain of UBF. Because UBF is involved in RNA polymerase I-mediated ribosomal (r)RNA transcription, we analyzed CTCF binding to the rDNA repeat. We found that CTCF bound to a site upstream of the rDNA spacer promoter and preferred non-methylated over methylated rDNA. DNA binding by CTCF in turn stimulated binding of UBF. Absence of CTCF in cultured cells resulted in decreased association of UBF with rDNA and in nucleolar fusion. Furthermore, lack of CTCF led to reduced binding of RNA polymerase I and variant histone H2A.Z near the rDNA spacer promoter, a loss of specific histone modifications, and diminished transcription of non-coding RNA from the spacer promoter.

Conclusions: UBF is the first common interaction partner of CTCF and CTCFL, suggesting a role for these proteins in chromatin organization of the rDNA repeats. We propose that CTCF affects RNA polymerase I-mediated events globally by controlling nucleolar number, and locally by regulating chromatin at the rDNA spacer promoter, similar to RNA polymerase II promoters. CTCF may load UBF onto rDNA, thereby forming part of a network that maintains rDNA genes poised for transcription.

Background

CTCF is a conserved and ubiquitously expressed protein, which binds DNA through an 11-zinc finger (ZF) domain and organizes chromatin into loops [1]. CTCF may act as an insulator, mainly by inhibiting inappropriate interactions between regulatory elements on adjacent or distal chromatin domains. In many instances, CTCF binds cognate sites in a methylation-sensitive manner, allowing the regulation of imprinted loci, such as the *H19/Igf2* locus. A testis-specific paralog of CTCF has been characterized, called CTCFL or BORIS (brother of the regulator of imprinted sites), which has strong similarity to CTCF in the ZF domain and has overlapping DNA-binding specificity [2]. CTCF and CTCFL share little similarity outside their ZF region. To date, no common interaction partners of CTCF and CTCFL have been reported.

Genomewide studies have revealed a multitude of CTCF binding sites, whose distribution over chromosomes correlates with gene density [3]. The cohesin complex, which mediates sister chromatid cohesion in dividing cells, was shown to colocalize with CTCF on CTCF binding sites [4-6]. Recent data suggest that CTCF/cohesin are together involved in the organization

* Correspondence: n.galjart@erasmusmc.nl; f.sleutels@erasmusmc.nl
[1]Department of Cell Biology and Genetics, Erasmus MC, The Netherlands
Full list of author information is available at the end of the article

of chromatin loops, with CTCF recruiting cohesin to specific sites, and cohesin in turn mediating chromosomal interactions [7]. CTCF may also colocalize with the variant histone H2A.Z [8]. When CTCF is bound near an RNA polymerase II-regulated transcription start site (TSS), it is mostly located upstream of a DNAse I hypersensitive site (HS) which in turn precedes the TSS [9]. These data suggest a global role played by CTCF as an organizer of RNA polymerase II-mediated transcription. By contrast, we have shown that loss of a CTCF-binding site affects chromatin looping and local histone modifications in the mouse β-globin locus, without significantly perturbing transcription [10]. Collectively, these data indicate that CTCF is able to regulate the balance between active and repressive chromatin modifications near its binding sites, with different outcomes in terms of transcription. CTCF may control epigenetic modifications by binding to the chromatin remodeling factor CHD8 [11].

The nucleolus is a nuclear subcompartment in which the 18S, 5.8S and 28S ribosomal (r)RNAs are synthesized by RNA polymerase I, processed and, together with 5S rRNA, assembled into ribosomes [12]. Ribosome biogenesis is tightly coordinated with cellular metabolism and cell proliferation. In all organisms, ribosomal genes are repeated many times, so that enough rRNA can be produced when demand for ribosomes is high. However, even in metabolically active cells, a significant number of repeats are not transcribed. In human and mouse, there are approximately 200 rDNA repeats per haploid genome (that is, ~400 per interphase nucleus). These are clustered in five nucleolar organizer regions (NORs), located on different chromosomes. Two promoters have been identified within the mouse rDNA repeat: the spacer promoter and the gene promoter. The spacer promoter is located upstream of the gene promoter within the intergenic spacer (IGS). Transcription from this promoter is thought to serve a regulatory function and gives rise to non-coding RNAs (ncRNAs or nc-rRNAs). Transcription from the gene promoter yields a ~13 kb (or 47S) ribosomal precursor RNA (pre-rRNA), which is processed in a complex manner into the mature 18S, 5.8S and 28S rRNAs.

Efficient transcription from the ribosomal gene promoter requires a multiprotein complex including selectivity factor (SL)1, RNA polymerase I, and upstream binding factor (UBF) [13]. UBF is an abundant nucleolar protein that contains several HMG domains involved in DNA binding [14]. UBF binds dynamically throughout the rDNA repeat [15], and not only plays a role as a transcriptional activator of RNA polymerase I, but also in transcription elongation [16] and in the maintenance of the specific chromatin structure of NORs [17]. More recent data suggest that UBF is involved in determining the number of active rDNA genes [18].

To better understand the function of CTCF, we performed a screen for CTCF-interacting proteins. We found that both CTCF and CTCFL interact directly with UBF. CTCF binds immediately upstream of the ribosomal spacer promoter in a methylation-sensitive manner, and activates spacer promoter transcription. CTCF binding controls the loading of UBF onto rDNA, and the binding of RNA polymerase I and H2A.Z near the spacer promoter. Our data show that CTCF regulates the local epigenetic state of the rDNA repeat. CTCF may organize RNA polymerase I and II promoters in a similar manner. We propose that CTCF binding maintains rDNA repeats in a state poised for activation.

Results

Characterization of biotinylated CTCF

To identify CTCF-binding partners, we used a biotinylation tagging and proteomics approach (Figure 1A) [19]. As CTCF levels are crucial for cell proliferation, we did not generate cell lines overexpressing biotinylated CTCF. Instead, we used homologous recombination in embryonic stem (ES) cells to generate a novel *Ctcf* knock-in allele. DNA encoding a small peptide tag of 23 amino acids was inserted in the last exon of the *Ctcf* gene, before the stop codon of CTCF (Figure 1B). This tag is biotinylated upon addition of the bacterial biotin ligase enzyme, BirA. Southern blot and PCR analysis identified homologous recombination events (Figure 1C). The resulting allele was termed $Ctcf^{bio\text{-}neo}$, as it contains both the biotinylation sequence and the neomycin resistance gene.

$Ctcf^{bio\text{-}neo/+}$ ES cells were transfected with a plasmid expressing Cre recombinase to remove the neomycin resistance gene and generate the $Ctcf^{bio}$ allele (Figure 1B). Then, using homologous recombination, the BirA biotin ligase was placed into the *Rosa26* locus (data not shown). Genotyping and verification of these targeting events was performed by PCR (Figure 1D). This method yielded an ES cell line expressing normal CTCF (from the wild type allele) and biotinylated CTCF (from the $Ctcf^{bio}$ allele). The biotin tag is placed at the C-terminus of CTCF, hence the fusion protein was called CTCF-bio. $Ctcf^{bio\text{-}neo}$ ES cells were also injected into blastocysts to generate knock-in mice. These mice were subsequently crossed with a mouse line expressing BirA from the *Rosa26* locus [20]. From these mice, CTCF-interacting proteins could be identified in a developmental and tissue specific manner.

CTCF-bio cannot be distinguished from untagged CTCF with anti-CTCF antibodies because the biotin tag does not cause a major difference in migration behavior in SDS-PAGE gels (Figure 1E, upper panel). However, CTCF-bio is detected using streptavidin-based methods (Figure 1E, middle panel). Our results

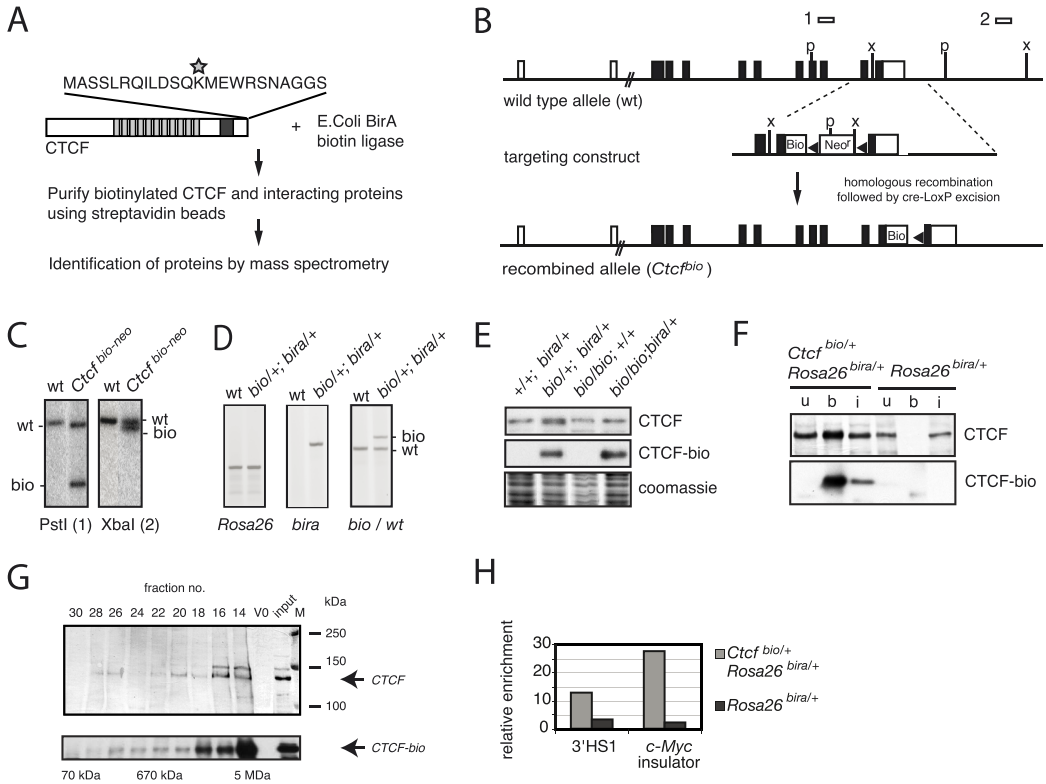


Figure 1 Characterization of biotinylated CCCTC binding factor (CTCF). **(A)** Identification of CTCF-binding partners. CTCF (light grey boxes) was tagged with biotin at its C-terminus (biotinylated lysine indicated by a star) and pulled down, and interacting proteins were identified by mass spectrometry. Zinc fingers are indicated by dark grey box; proline-rich region with AT hook. **(B)** Targeting of mouse *Ctcf*. The *Ctcf* gene is shown on the top line. Exons (black boxes are coding, open boxes are non-coding) and probes (1, 2) are indicated. X = *Xba*I, P = *Pst*I. Targeting construct is shown in the middle. Bio = biotin-tag, black triangles = loxP sites, Neo^r = PMC1-neomycin cassette. After targeting, $Ctcf^{bio\text{-}neo}$ was obtained. Cre-mediated excision yielded the $Ctcf^{bio}$ knock-in allele (bottom line). **(C)** Southern blot analysis. Digested embryonic stem (ES) cell DNA was analyzed by Southern blotting using probes 1 or 2. Alleles are indicated. **(D)** PCR analysis of mouse tail DNA. **(E)** Expression of biotin tagged CTCF (CTCF-bio). Nuclear thymic extracts were analyzed by western blotting. Coomassie staining shows total protein. **(F)** CTCF-bio pull-down assay with streptavidin beads. Nuclear ES cell extracts were incubated with streptavidin-coupled magnetic beads. Input (i) = 5% of nuclear extract, (u) = 6% of unbound fraction, b = material bound to beads. **(G)** Size fractionation of CTCF and CTCF-bio. Size fractionated nuclear extracts were analyzed by western blotting. Molecular mass markers are indicated. V0 = void volume, input = nuclear extract (5%). **(H)** Chromatin immunoprecipitations. CTCF-bio precipitated with streptavidin-coupled magnetic beads from formaldehyde fixed nuclei bound known CTCF sites (β-globin 3" HS1 and c-Myc insulator).

indicate that CTCF-bio and CTCF are expressed at similar levels (Figure 1E). Pull-down assays using ES cell extracts with streptavidin-coupled magnetic beads results in efficient and specific binding of CTCF-bio to the beads (Figure 1F). Size fractionation experiments suggest that CTCF and CTCF-bio are present in high molecular weight complexes in ES cells (Figure 1G). Furthermore, CTCF-bio binds known CTCF target sites such as the *c-Myc* insulator and the 3" HS1 of *β-globin* (Figure 1H). Importantly, mice expressing CTCF-bio are viable and fertile (data not shown). Combined, these results indicate that CTCF-bio is a functional protein.

CTCF and CTCFL interact with UBF

CTCF-bio was purified from ES cell nuclear extracts under mild conditions using streptavidin-coupled magnetic beads (Figure 2A). Known CTCF-interacting partners, including Yin Yang (YY)-1, poly(ADP-ribose) polymerase (Parp)1 and nucleophosmin, co-precipitated with CTCF-bio (see Additional file 1, Figure S1A), further confirming that CTCF-bio is a functional fusion

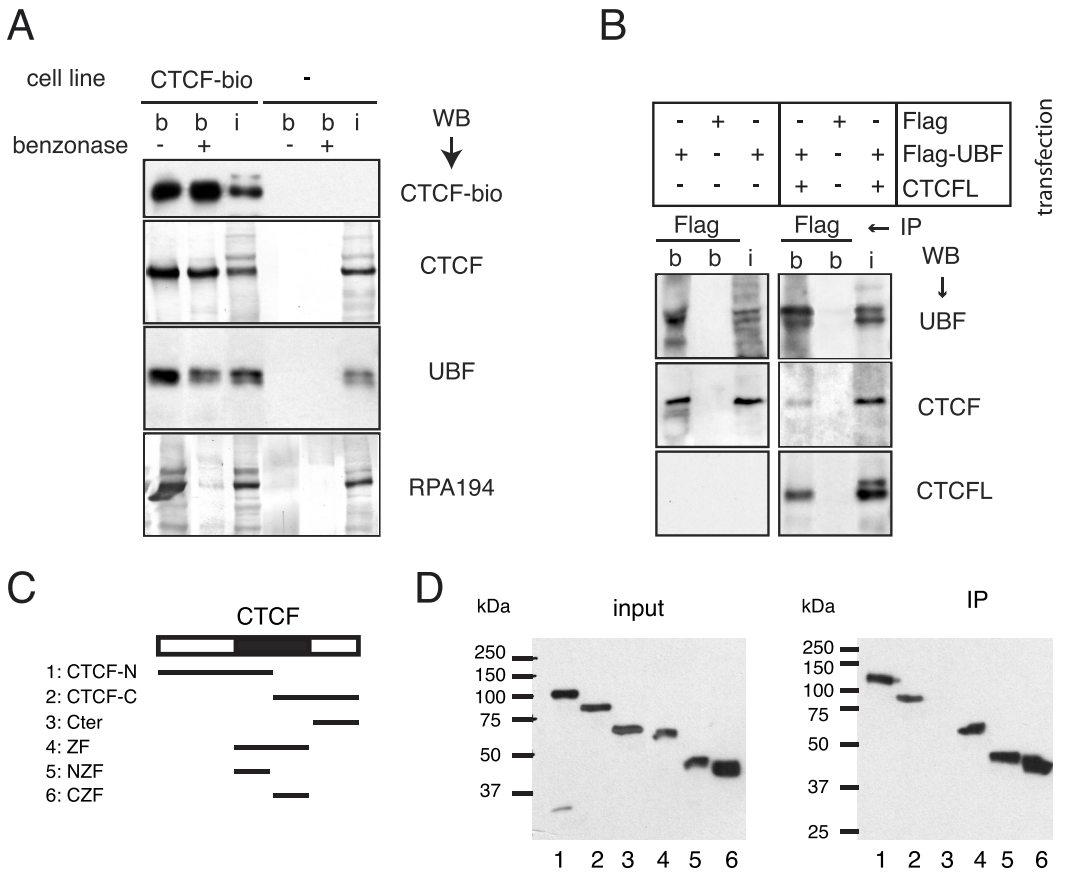


Figure 2 Interaction of CCCTC binding factor (CTCF) and CTCFL with upstream binding factor (UBF). **(A)** Biotin tagged CTCF (CTCF-bio) interacted with UBF and RNA polymerase I in embryonic stem (ES) cells. Nuclear extracts from ES cells expressing CTCF-bio and control ES cells (-) were incubated with streptavidin beads, and CTCF-bio was purified with interacting proteins. Extracts were treated with (+) or without (-) benzonase for 2 hours at 4°C. Western blots were incubated with the indicated antibodies (CTCF-bio detected with streptavidin-coupled horseradish peroxidase). UBF was detected as a doublet consisting of UBF1 and UBF2; RPA194 = the large subunit of RNA polymerase I. B = bound fraction; i = input (5%). **(B)** Both CTCF and CTCFL interacted with UBF. Cells were transfected with different cDNAs (indicated below the lanes). CTCFL was not tagged. Immunoprecipitations (IPs) were carried out using anti-Flag antibodies. Western blots were incubated with antibodies against the indicated proteins. **(C)** Green fluorescent protein (GFP) tagged CTCF deletion mutants. The regions of CTCF used for making the different fusion proteins are indicated by lines. **(D)** UBF interacts with the zinc-finger domain of CTCF. GFP tagged CTCF deletion mutants (in (C)) were co-expressed with Flag tagged UBF in HEK293T cells. All fusion proteins are expressed at similar levels (input). After a Flag pull-down (IP), co-precipitating proteins were detected with an antibody against GFP. Lane numbers correspond to mutant numbers in (C).

protein, and suggesting that the conditions used to isolate CTCF-bio were sufficiently mild to allow identification of novel interaction partners. Proteins co-purifying with CTCF-bio were detected by mass spectrometry and classified by BLAST searches; 58 of these co-purified specifically with CTCF in more than one pull-down experiment (data not shown). We noted that several CTCF-interacting proteins are involved in RNA polymerase I-mediated transcription (see Additional File 2 Table S1), including UBF and proteins that form a complex with UBF, such as the large subunit of RNA polymerase I (RPA194) and its associated factor PAF53 [21]. Moreover, the 40 kDa and 135 kDa subunits of RNA polymerase I (RPA40, RPA135) and polymerase associated factor (PAF)49 were pulled down by CTCF-bio (data not shown). These data suggest that CTCF interacts with essential components of the machinery that regulates the synthesis of rRNA. We therefore decided to further analyze the function of CTCF in rRNA transcription.

Streptavidin pull-down assays followed by western blot analysis confirmed the CTCF-bio interaction with UBF and the large subunit of RNA polymerase I (Figure 2A). We also detected interaction of CTCF-bio and UBF in lung (see Additional file 1, Figure S1B) and thymus (not shown), indicating that the association between these two proteins is not confined to ES cells. When ES cell nuclear extracts were treated with benzonase, the CTCF-bio interaction with UBF remained detectable, indicating that the interaction is not mediated by DNA (Figure 2A). Co-immunoprecipitation (co-IP) with anti-CTCF antiserum revealed an interaction between untagged CTCF and UBF (see Additional file 1, Figure S1C).

As CTCF and CTCFL are very similar in their ZF domains, we tested the possibility that CTCFL also interacts with UBF. We overexpressed a Flag-tagged form of UBF in 293T cells, either alone or with CTCFL, and performed a Flag co-IP on extracts from these cells. Flag-UBF brings down endogenous CTCF and overexpressed CTCFL (Figure 2B). Interestingly, diminished interaction between CTCF and UBF was detected in cells expressing CTCFL. These results identify UBF as the first common interaction partner of CTCF and CTCFL, and also indicate that CTCF and CTCFL compete for binding to UBF.

Experiments with bacterially purified proteins revealed a direct interaction between the CTCF and CTCFL ZF domains and the UBF dimerization domain plus HMG-box 1 (see Additional file 3, Figure S2). Using CTCF deletion mutants [22], we observed that both the N- and C-terminal ZFs of CTCF interacted with UBF (Figure 2C, D). Taken together, our data show that CTCF and CTCFL bind UBF directly via their ZF domains.

Identification of a CTCF binding site upstream of the rDNA spacer promoter

To provide a functional explanation for the CTCF-UBF interaction, we tested binding of both proteins using chromatin immunoprecipitation (ChIP) in mouse embryonic fibroblasts (MEFs). Consistent with published experiments [15,23], UBF bound throughout the enhancer/promoter regions and transcribed portion of the mouse rDNA repeat, with hardly any enrichment in the IGS (Figure 3B, blue line). By contrast, ChIP of CTCF revealed a highly specific accumulation immediately upstream of the rDNA spacer promoter (Figure 3B, black line). We also detected CTCF binding to the rDNA spacer promoter region in extracts of adult thymus from wild type and CTCF-bio-expressing mice (see Additional file 4, Figure S3). The CTCF binding coincided with (and was adjacent to) RNA polymerase I enrichment (Figure 3B, red line). Strong RNA polymerase I association to the spacer promoter relative to the gene promoter has also been shown by others [24-26].

The ChIP experiments suggest the presence of a CTCF binding site near the spacer promoter of the mouse rDNA locus. An algorithm was devised to search for potential binding sites within this locus. One site (R30), which conforms to the CTCF consensus sequence [3], is present in the spacer promoter area (Figure 3C). A probe (also called R30) was designed and tested in band-shift analysis, using nuclear extracts of non-transfected cells and of cells overexpressing CTCF. The known chicken lysozyme F1 site was used as control. We detected binding of endogenous CTCF and bacterially purified glutathione-S-transferase (GST)-CTCF-ZF to the R30 probe (Figure 3D, lanes 6 to 10 and 11 to 13, respectively). Competition experiments indicated that CTCF bound the FI probe less efficiently than it did R30, (Figure 3E, lanes 3 and 4). These data demonstrate that CTCF binds R30 through its ZF domain.

Previous studies have shown that mouse, rat and hamster rDNA repeats share significant sequence similarity in the spacer promoter region of the IGS [27]. Rat and hamster rDNA also contain the CTCF binding site (Figure S4, Additional file 5). Based on alignment information, we mutated three residues within R30, and performed band-shifts with normal and mutant R30 probes. As shown in Figure 3E, CTCF bound less efficiently to mutant R30 (lanes 5 and 7). Combined, these results identify a novel CTCF binding site in the mouse rDNA repeat that is conserved in rat and hamster.

The IGS of the human rDNA repeat is completely divergent in sequence from the mouse IGS (see Additional file 6, Figure S5A to C) and the presence of a spacer promoter has not been accurately described. Nevertheless, we identified two potential CTCF binding sites in the rDNA repeat, which were 0.9 kb and 5.1 kb upstream of the ribosomal gene promoter (called H42.1 and H37.9, respectively, for their respective positions) (see Additional file 6, Figure S5B). ChIP analysis revealed occupancy of CTCF at both sites, although binding was more prominent in the region near H42.1 than near H37.9 (see Additional file 7, Figure S6A). As K562 cells express both CTCF and CTCFL, we also tested whether CTCFL could bind the human rDNA repeat. Protein-DNA complexes were immunoprecipitated with two different CTCFL antibodies. CTCFL bound both H37.9 and H42.1, with a preference for site H42.1 (see Additional file 7, Figure S6A). We also detected binding of UBF to these rDNA regions (see Additional file 7, Figure S6A) and using sequential (ChIP-reChIP) analysis, found that CTCF and UBF were present on the same rDNA repeats (see Additional file 7, Figure S6B).

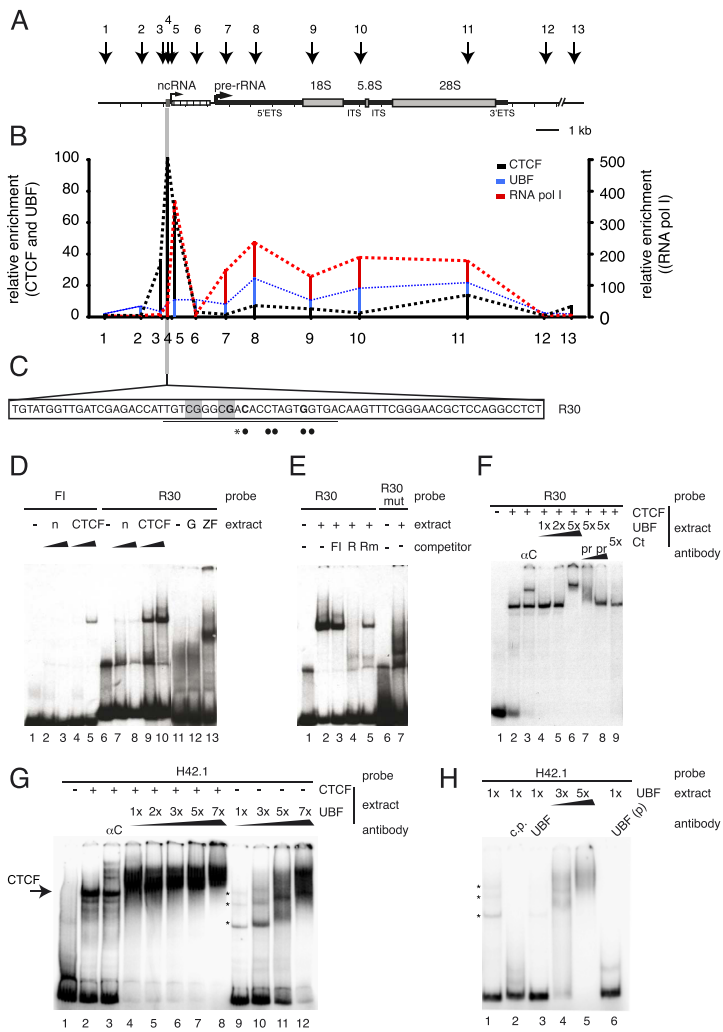


Figure 3 Identification of a novel CCCTC binding factor (CTCF) binding site near the ribosomal (r)DNA spacer promoter. **(A)** Outline of mouse rDNA repeat. Transcription (spacer promoter = ncRNA, gene promoter = pre-rRNA, right-pointing arrows), chromatin immunoprecipitation (ChIP) primers (downward-pointing arrows), organization of 47S pre-rRNA (external transcribed spacers (ETS), internal transcribed spacers (ITS), 18S, 5.8S and 28S rRNA, CTCF consensus site (grey box) and enhancer repeats (white boxes)), are indicated. **(B)** Binding of CTCF, upstream binding factor (UBF) and RNA polymerase I to mouse rDNA was analyzed by ChIP on formaldehyde-fixed mouse embryonic fibroblast (MEF) nuclei. **(C)** CTCF binding site in mouse rDNA. The R30 sequence is shown, with CTCF consensus site (underlined), highly conserved CTCF site residues (dots), deviation from consensus (asterisk) and CpG dinucleotides (gray). Nucleotides in bold were mutated (R30mut). **(D, E)** Band-shift analysis. Increasing amounts of HEK293T extracts (non-transfected (n) or transfected with CTCF), were incubated with the indicated probes (a known CTCF binding site (F1, chicken lysozyme, R30 and R30mut). Competitor was added in 300-fold excess. Band shifts were also performed with purified bacterial proteins (glutathione-S-transferase (GST) only (G) and zinc fingers (ZF) of CTCF tagged with GST. **(F)** CTCF and UBF bound R30 DNA together. HEK293T extracts (increasing amounts indicated, transfected with GFP-CTCF (+) Flag-UBF (1 to 5×) or the C-terminal domain of CTCF tagged with GFP (Ct, see Figure 2C) were incubated with R30. αC = addition of anti-CTCF; pr = addition of anti-UBF and anti-FLAG (preclear). **(G)** CTCF and UBF bound human rDNA together. HEK293T extracts (increasing amounts indicated, transfected with GFP-CTCF (+) or Flag-UBF (1 to 7×) were incubated with H42.1. Asterisks indicate UBF binding to H42.1. αC = addition of anti-CTCF. **(H)** UBF weakly bound human rDNA. HEK293T extracts (increasing amounts indicated), transfected with Flag-UBF (1 to 5×) were incubated with H42.1. Asterisks indicate UBF binding to H42.1 Cp = cold probe; UBF = addition of anti-UBF; UBF(p) = addition of anti-UBF and anti-FLAG before incubation with probe (preclear).

The CTCF ChIP results were confirmed *in vitro* by electrophoretic mobility shift assay (EMSA) analysis (see Additional file 7, Figure S6C). Nuclear extracts from cells transfected with CTCF showed stronger binding to H42.1 and H37.9 rDNA probes compared with extract from mock-transfected cells. The specificity of the binding was shown by competition with unlabeled probes and by supershift assays using anti-CTCF antibody. Incubation with an anti-actin antibody, used as a negative control, did not produce supershifts (data not shown). Together, these results demonstrate that CTCF associates upstream of the gene promoter in human rDNA and suggest that CTCF and UBF are bound together to the rDNA.

We next tested whether the *in vitro* binding of CTCF to DNA influences binding of UBF. Extracts of cells transfected with GFP-CTCF or Flag-UBF were incubated separately or together with the H42.1 probe, and binding of CTCF and UBF was examined by EMSA (Figure 3G, H). Binding of UBF alone to the H42.1 probe resulted in a relatively weak signal (Figure 3G, lane 9; Figure 3H, lane 1; asterisks), that was specific for UBF (Figure 3H, lanes 2, 3, 6). Increasing the amount of UBF in the reaction eventually led to enhanced and cooperative binding of UBF (Figure 3G, lanes 10 to 12; Figure 3H, lanes 4, 5). Interestingly, binding of CTCF to H42.1 resulted in enhanced binding of UBF at much lower levels of this protein (Figure 3G, compare lanes 4 to 8 with 9 to 12). These data suggest that CTCF helps to load UBF onto rDNA.

CTCF binds rDNA in a methylation-sensitive manner

The CTCF R30 binding site in the mouse rDNA repeat includes two CpG residues (Figure 3C), which might be methylated *in vivo*. The CpG residues are conserved in rat and hamster (see Additional file 5, Figure S4). As CTCF often binds DNA in a methylation-sensitive manner, we tested whether the *in vitro* methylation of these two sites in R30 affected CTCF binding. We found that this was the case to some extent, as CTCF bound the non-methylated R30 probe slightly more efficiently, and this probe was a better competitor than methylated R30 (see Additional file 8, Figure S7A).

Human 37.9 and 42.1 CTCF binding sites contain three CpG residues instead of two (not shown). One of these overlaps with the highly conserved GG dinucleotide that is part of the 'core' CTCF binding site (Figure 3C). The second CpG is conserved between human and mouse rDNA sites (it is the 5"end CpG in R30) (Figure 3C), whereas the third CpG in the human rDNA sites is not conserved between mouse and human, nor between 37.9 and 42.1 (not shown). We used SssI methyltransferase to completely methylate the human H37.9 and H42.1 probes (see Additional file 8, Figure S7B). Interestingly, CTCF binding to these fully methylated probes was severely reduced (Figure 4A). These data indicate that CTCF binds rDNA in a methylation-sensitive manner *in vitro*. Both the position and number of methylated CpG residues appear to influence CTCF binding to cognate sites.

To test whether DNA methylation might interfere with CTCF binding to the rDNA *in vivo*, we performed ChIP-chop experiments. Before quantitative PCR, the input and CTCF-enriched DNA samples were subjected to digestion with the methylation-sensitive enzyme *Hpa*II or the methylation-insensitive enzyme *Msp*I. CTCF did not bind *Hpa*II-resistant (that is, methylated) H42.1 rDNA (Figure 4B). Similar data were obtained for H37.9 rDNA (not shown). These results indicate that CTCF prefers non-methylated over methylated ribosomal DNA *in vivo*. A ChIP-chop assay performed on mouse ES cell DNA also showed CTCF binding to non-methylated rDNA (data not shown).

Fully methylated rDNA repeats are thought to be inactive [28]. To investigate an *in vivo* correlation between CTCF binding in the spacer promoter and methylation status of the rDNA repeats, we used 3T3L1 cells. These cells can be differentiated into adipocytes, which results in the repression of rRNA transcription by more than 50% [29]. Increased heterochromatin features at the rDNA promoter accompany this repression [29,30]. ChIP analysis revealed binding of CTCF, UBF and RNA polymerase I at the spacer promoter in undifferentiated 3T3L1 cells (Figure 4C, left panel). As reported previously [29], UBF and RNA polymerase I binding to the rDNA repeat was reduced in differentiated 3T3L1 cells (Figure 4C, right panel). CTCF binding was also significantly reduced (Figure 4C, right panel). These data suggest that increasing heterochromatinization *in vivo* significantly affects binding of CTCF, UBF and RNA polymerase I. We propose that *in vivo* CTCF binds rDNA repeats in a methylation-sensitive manner.

CTCF regulates nucleolar number, and is required for UBF and RNA pol I binding near the spacer promoter

To examine the physiological significance of a CTCF-UBF interaction and of CTCF binding to the rDNA spacer promoter, we generated a system to efficiently deplete CTCF *in vitro*. MEFs were isolated from mice homozygous for a *Ctcf* conditional knockout allele [31], and CTCF was deleted by infecting confluent MEFs with a replication-deficient lentivirus expressing Cre recombinase [10]. After 4 days of culture, only very low levels of CTCF protein were detected on western blot (Figure 5A). Immunofluorescence analysis revealed that a small proportion of MEFs still expressed CTCF (data not shown), suggesting that these were not infected by

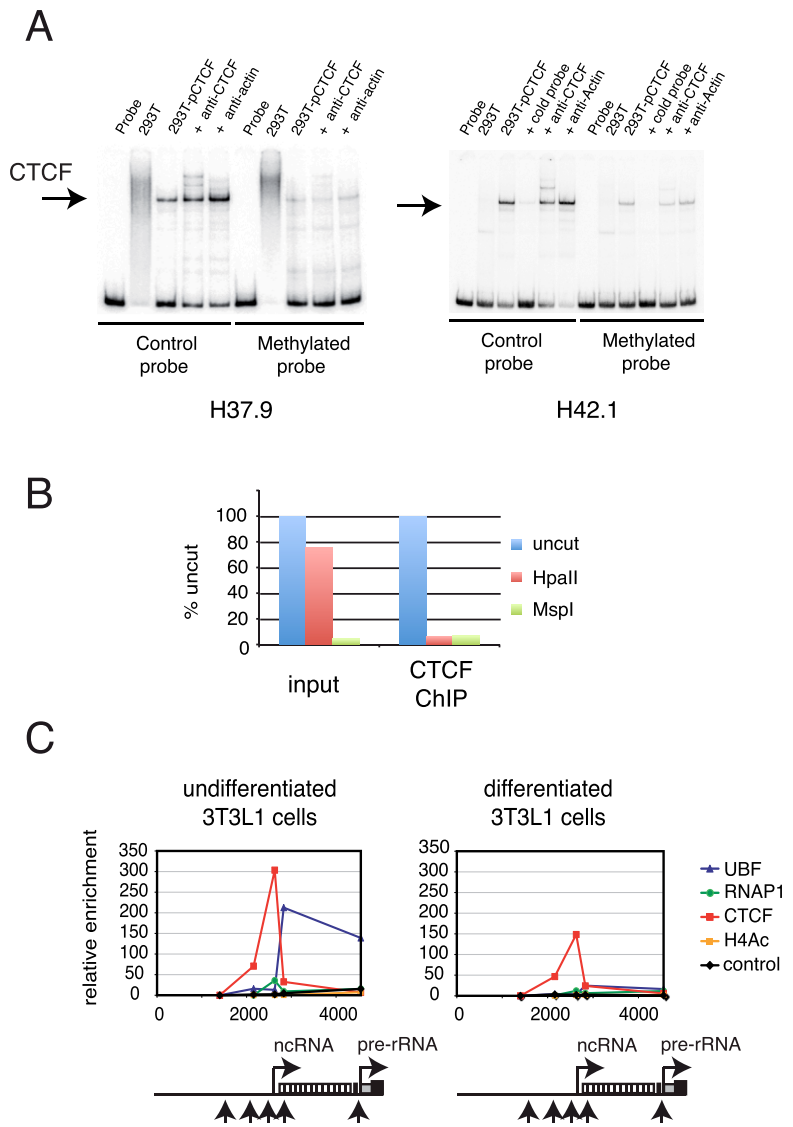


Figure 4 CCCTC binding factor (CTCF) binding to ribosomal (r)DNA is methylation-sensitive. **(A)** Influence of methylation on the binding of CTCF to rDNA. Band-shift assays using human H37.9 and H42.1 rDNA probes, either completely methylated with SssI methyltransferase (methylated probe) or non-methylated (control probe) on HEK293T extracts, transfected or not with CTCF. Competition was assessed by adding increasing amounts of non-labeled probe. In some cases, extracts were incubated with the indicated antibodies. **(B)** CTCF prefers non-methylated rDNA *in vivo*. Chromatin from K562 cells was immunoprecipitated with anti-CTCF (CTCF ChIP). Purified DNA was left uncut (mock digestion), or digested with *Hpa*II or *Msp*I. Quantitative PCR was then performed with H42.1 primers, both on non-precipitated K562 DNA (input) and on DNA enriched for CTCF binding sites (CTCF ChIP). Note the high content of *Hpa*II-resistant H42.1 rDNA in K562 cells (input), which represents methylated rDNA. In the CTCF-enriched sample, the *Hpa*II-resistant rDNA was not present, suggesting that CTCF does not bind well to methylated rDNA. **(C)** ChIP analysis on (left panel) undifferentiated and (right panel) differentiated 3T3L1 cells. Nuclei were fixed with 1% formaldehyde, and protein-DNA complexes were immunoprecipitated with antibodies against the indicated proteins (the large subunit of RNA polymerase I (RPA194), CTCF, upstream binding factor (UBF) and acetylated histone H4). The position of the primer sets (upward arrows, see also ChiP2 in Figure 6A), spacer promoter (right-pointing arrow with ncRNA (part A), enhancer repeats (white rectangles) and gene promoter (right-pointing arrow with pre-rRNA)) is indicated on the rDNA. The horizontal axis of the panels is co-aligned with the rDNA underneath and shows distance in base pairs.

the virus. MEFs lacking CTCF could be maintained as a confluent layer for several days (data not shown), but they could not be passaged, because they are severely impaired in division. These results are in line with *in vivo* data showing that CTCF is essential for the proliferation and growth of β-selected T cells [31].

Next, we investigated the intracellular distribution of UBF in MEFs. Interestingly, deletion of CTCF reduced the number of UBF-positive spots, and thus the number of nucleoli, in MEFs (Figure 5B, C). However, the average size of a UBF-positive area, and thus that of a nucleolus, was larger in CTCF-deleted MEFs. As a result the total fluorescence intensity (and hence the level) of UBF was similar in CTCF-deleted and normal MEFs, a result supported by western blot analysis (not shown) and data in T cells [31]. We conclude that deletion of CTCF in MEFs results in fusion of nucleoli but does not affect UBF levels.

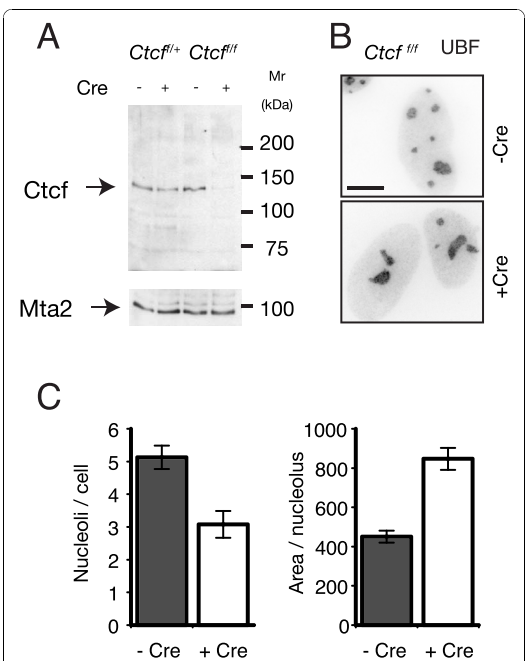


Figure 5 CCCTC binding factor (CTCF) regulates nucleolar number. **(A)** Efficient deletion of CTCF in mouse embryonic fibroblasts (MEFs). MEFs carrying the conditional *Ctcf*^{f} knockout allele were infected (+) or not (-) with a lentivirus expressing Cre recombinase. After 4 days, cell extracts were analyzed for residual CTCF. Mta2 was used as loading control. In MEFs, *Ctcf* deletion with lentiviral Cre was very efficient, with > 90% of cells infected. **(B, C)** Distribution of upstream binding factor (UBF) in MEFs. Primary MEFs carrying the conditional *Ctcf*^{f} knockout allele were either infected (+CrE) or not (-CrE) with a lentivirus expressing Cre recombinase. Cells were fixed, and incubated with antibodies against UBF (because both the CTCF and UBF antibodies used for immunofluorescence analysis were mouse monoclonals, we could not perform a combined CTCF/UBF stain. **(B)** Single image; **(C)** immunofluorescence results from several images were quantified.

CTCF binding to the rDNA spacer promoter was virtually undetectable in MEFs treated with Cre virus (Figure 6B, right-hand panels, red line) compared with non-treated MEFs (Figure 6B, left-hand panels, red line). In the absence of CTCF, binding of UBF and RNA polymerase I was severely reduced (Figure 6B, right hand panels, blue and green line, respectively). Remarkably, the absence of CTCF did not significantly perturb RNA polymerase I binding to the gene promoter. Thus, CTCF exerts a local influence.

In mouse ES cells, distribution of CTCF, UBF and RNA Pol I over the rDNA repeat, as analyzed by ChIP, was similar to that in MEFs (see Additional file 9, Figure S8). We used an RNA interference (RNAi)-based approach to knock down *Ctcf* mRNA in ES cells. Real time PCR and immunofluorescence analysis suggested knock down of CTCF of > 70% after 4 days of culture. Although the depletion of CTCF in ES cells was less effective than Cre-treatment of *Ctcf*$^{f/f}$ MEFs, this reduction did lead to a loss in UBF and RNA pol I binding (see Additional file 9, Figure S8B). These results confirm the role of CTCF in UBF and RNA polymerase I localization.

CTCF maintains specific histone marks at the spacer promoter

Given the role of CTCF in epigenetic chromatin remodeling near its binding sites, we examined the distribution of specific histone marks across the rDNA regulatory region in the presence and absence of CTCF. ChIP analysis in normal MEFs revealed peaks of histone H3 acetylation, H3K4 dimethylation and H2A.Z just upstream of the CTCF binding site (Figure 6C, left panel). In the absence of CTCF H2A.Z, H3K4 dimethylation and H3 acetylation (that is, markers of 'active' chromatin and of insulator sites) were clearly downregulated (Figure 6C, right panel). A control ChIP experiment revealed similar amounts of histone H3 in the presence or absence of CTCF (see Figure S9A, Additional file 10), showing that the reduction in H2A.Z, H3K4me2 and H3ac levels is specific. Furthermore, ChIP analysis in the human rDNA repeat revealed specific accumulation of H2A.Z, H3K4me2 and H3ac at both CTCF binding sites in K562 cells (see Figure S9C, Additional file 10). Combined, our data suggest that CTCF is required for local histone modifications and the accumulation of a histone variant at the spacer promoter.

Because we found that CTCF is required for H2A.Z accumulation at the rDNA spacer promoter, we tested whether this also occurs with H2A.Z sites near RNA polymerase II-dependent genes. In the absence of CTCF, H2A.Z was indeed lost from the *c-Myc* promoter

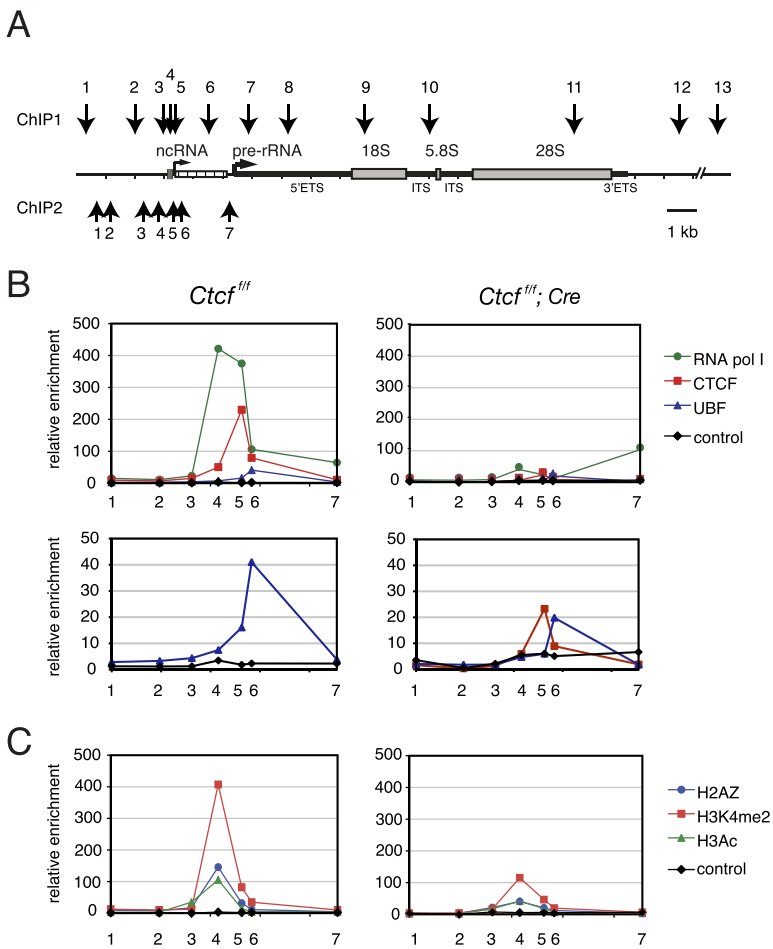


Figure 6 CCCTC binding factor (CTCF) organizes chromatin on ribosomal (r)DNA. **(A)** Outline of mouse rDNA repeat. The position of the primer pairs (1 to 7) used in the second chromatin immunoprecipitation (ChIP2; panel B) is indicated by upward-pointing arrows. For comparison, the primers used in the first ChIP (ChIP1; Figure 3) are also shown. For other explanations, see Figure 3A. **(B)** Binding of CTCF, upstream binding factor (UBF) and RNA polymerase I to the mouse spacer promoter. ChIP analysis was performed on mouse embryonic fibroblasts (MEFs) homozygous for the conditional *Ctcf* knockout allele ($Ctcf^{f/(F)}$). MEFs were either (right panels) infected or not (left panels) with a lentivirus expressing Cre recombinase. Nuclei were fixed with 1% formaldehyde, and protein-DNA complexes were immunoprecipitated with antibodies against the indicated proteins (control = rabbit IgG). The upper and lower panels show the same results, but with a different vertical axis. Numbers on the horizontal axis refer to primer pairs. **(C)** Binding of modified and variant histones to the mouse spacer promoter. ChIP analysis was performed as described above. Protein-DNA complexes were immunoprecipitated with antibodies against the indicated proteins (control = rabbit IgG). Numbers on the horizontal axis refer to primer pairs.

(see Additional file 10, Figure S9B), implying that CTCF can mediate the deposition of this histone variant close to RNA polymerase I and II promoters. The relatively constant levels of histone H3 in the rDNA locus in normal and Cre-treated MEFs (see Additional file 10, Figure S9A) indicate that the observed loss of histone modifications and H2A.Z are not caused by a reduction in nucleosomes in the absence of CTCF. Furthermore, the changes in DNA binding by specific proteins (for example, UBF and RNA polymerase I) seen in the absence of CTCF, are not the result of changes in the total amount of these proteins (data not shown).

CTCF activates transcription from the spacer promoter

We next examined the effect of a CTCF deletion on steady state RNA levels using total RNA isolated from

Ctcf$^{f/f}$ MEFs that were either treated or not treated with Cre virus. Using northern blot analysis, we found similar ratios of pre-rRNA (47S) to *Gapdh* mRNA in CTCF depleted MEFs (Figure 7B). Furthermore, the ratio of mature 18S rRNA to *Gapdh* mRNA was comparable in normal and CTCF-depleted cells. These results indicate that a deletion of CTCF does not affect steady state rRNA amounts in confluent non-dividing fibroblasts. Using nuclear run-on analysis, we investigated transcription from spacer and gene promoters in the presence and absence of CTCF. Deletion of CTCF significantly reduced transcription from the spacer promoter but did not affect transcription from the ribosomal gene promoter (Figure 7C). These results show that CTCF can activate transcription from the spacer promoter independently of the gene promoter.

We next examined the influence of CTCF on rRNA biogenesis in a cell type other than MEF. For these experiments, we used mouse ES cells, because RNAi-mediated knockdown of CTCF is effective in these cells. We investigated spacer and gene-promoter derived transcription in the presence or absence of CTCF using real-time PCR and fluorescent *in situ* hybridization (FISH). For the latter, we used probes against spacer promoter-derived transcripts ('ncRNA' probe), together with a previously described probe ('pre-rRNA' probe [32]) that covers the unstable 5" end of the external transcribed spacer (see Figure 7A for position of the probes on the rDNA). RNA FISH experiments showed that both the ncRNA probe (green) and the pre-rRNA probe (red) were located inside ES cell nucleoli (see Additional file 11, Figure S10C-E). Whereas the pre-rRNA signal was detected at a relatively constant level in each ES cell nucleolus, the ncRNA signal varied with respect to intensity and number of spots (on average ~4 per cell). These FISH experiments indicate that only a small subset of the rDNA repeats express ncRNA, consistent with recently published data [26]. Remarkably, ncRNA spots were often located at the periphery of the nucleolus, whereas pre-rRNA was detected throughout (see Figure S10C, D, Additional file 11). Our results suggest that ncRNA and pre-rRNA transcription can be independently regulated in space and time inside the nucleoli of ES cells.

Real-time PCR analysis suggests that pre-rRNA levels are reduced in ES cells lacking CTCF (Figure 7D). We also quantified pre-rRNA levels by measuring the fluorescence intensity in FISH experiments in non-treated, control RNA-treated, and CTCF RNAi-treated ES cells. Consistent with the PCR data, the pre-rRNA transcript was mildly affected in cells knocked down for CTCF (Figure 7E). Thus, in ES cells, lack of CTCF resulted in a very mild reduction of pre-rRNA levels, to ~80% of control. By contrast, both real-time PCR experiments using two different primers sets (Figure 7D) and RNA FISH data (see Additional file 11, Figure S10F) showed that ncRNA levels were significantly decreased in ES cells lacking CTCF. Combined with the run-on analysis in MEFs (Figure 7C) these results strongly suggest that CTCF activates transcription from the spacer promoter, an activity that is independent of cell type.

Discussion

We have identified UBF as the first common interaction partner of CTCF and CTCFL, emphasizing a role for these proteins in the organization of rDNA chromatin. It will be interesting to determine how CTCFL influences rRNA transcription *in vivo*, as in normal tissues this protein is expressed in a very restricted manner [2], whereas its expression is upregulated in various types of cancers [33]. The ZF domain of CTCF and CTCFL mediates the interaction with UBF. In CTCF, this domain has also been shown to be responsible for interaction with other proteins, including CHD8 [11]. Interestingly, a ZF-dependent, pan-nucleolar localization of CTCF was described in K562 cells, which correlated with poly (ADP-ribosyl)ation and growth arrest of cells [34]. The pan-nucleolar distribution of CTCF indicates that the protein must be bound to rRNA and/or nucleolar proteins in addition to rDNA. It is therefore not surprising that CTCF function is different in K562 cells compared with MEFs or ES cells. Post-translational modifications may alter the function, localization and interactions of CTCF in a cell type-specific manner. We conclude that the ZF-domain of CTCF is a versatile nucleic acid and protein-protein interaction surface, explaining why it is so conserved.

Previously, the *Xenopus laevis* rDNA repeat was reported to contain multiple weak CTCF binding sites near its spacer promoter [35]. Although the physiological significance for rDNA transcription was not investigated in that study, the result is consistent with our data. The importance of CTCF binding near the spacer promoter is emphasized by the observation that the mouse binding site is conserved in rat and hamster. Furthermore, we identified two CTCF sites in the human rDNA (H37.9 and H42.1, respectively) upstream of the gene promoter. We found a specific accumulation of H2A.Z, H3K4me2 and H3ac at CTCF binding sites in the human and mouse rDNA repeats. Interestingly, enrichment of the acetylated histones H3 and H4 and of TATA binding protein (TBP) was observed 100 bp upstream of site H42.1, whereas UBF accumulates 3' to this site [36]. Thus, despite the fact that the IGS regions of mouse and human are not conserved (see Additional file 6, Figure S5A to C), critical factors and chromatin modifications are similarly organized around CTCF binding sites in rDNA (see Additional file 6, Figure S5D). In fact, our

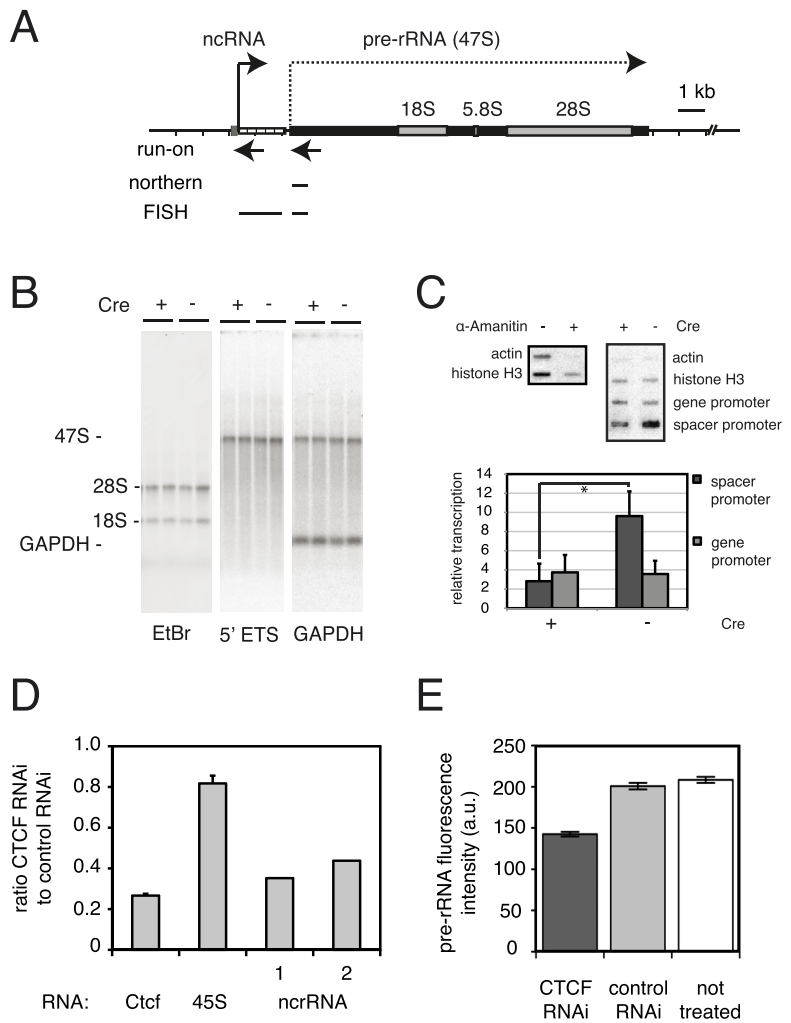


Figure 7 CCCTC binding factor (CTCF) activates spacer promoter transcription. **(A)** Transcribed portion of mouse ribosomal (r)DNA. Transcription initiation (spacer promoter = ncRNA, gene promoter = pre-rRNA, right-pointing arrows) and positions of run-on (left-pointing arrows), northern blot and fluorescent *in situ* hybridization (lines) probes are indicated. **(B)** Northern blot analysis. Total RNA from *Ctcf*$^{f/f}$ primary mouse embryonic fibroblasts (MEFs) (either treated (+) or not (-) with Cre) was analyzed sequentially with probes against the 47S pre-rRNA [32] and *Gapdh* (two samples per genotype shown, > 10 independent samples per genotype analyzed. **(C)** Spacer promoter transcription in the absence of CTCF. Nuclear run-on of spacer and gene promoter in *Ctcf*$^{f/f}$ MEFs (Cre-treated (+) or not (-)). The suppression effect of α-amanitin on RNA polymerase II-mediated transcription was measured. Left panel shows *Actin* and *Histone H3* genes; a typical run-on is shown on the right. The graph shows the relative transcription activity of spacer and gene promoter (three independent experiments; **P* = 0.005, Student t-test). **(D, E)** CTCF knockdown in embryonic stem (ES) cells. Cells were transfected with a pSUPER plasmid to knock down CTCF (controls were either no transfection, or transfection with a control vector). After 4 days, cells were harvested for **(D)** RNA analysis by PCR, or fixed and analyzed by fluorescent *in situ* hybridization **(E)**. ncRNA was determined twice with independent primers (1, 2), *Ctcf* and pre-rRNA were determined four times (SD indicated). Taking the average of the four ncRNA experiments showed a reduction of spacer transcript in CTCF-depleted cells to 39 ± 22% (mean ± SD). **(E)** Quantification was performed with > 300 cells per treatment (SEM indicated).

data suggest that the local organization of chromatin at CTCF sites near the RNA polymerase I-regulated spacer promoter and near RNA polymerase II promoters [9] is also similar. First, CTCF binds ~200 bp upstream of the TSS in both types of promoters, and an HS is present between the TSS and CTCF binding site. Furthermore, H2A.Z and H3K4me2 accumulate ~200 to 300 bp upstream of the CTCF binding site. Enrichment of H2A.Z at CTCF binding sites appears to be a general phenomenon [8,9,37]. In RNA polymerase II promoters, H2A.Z and H3K4me2 marks are associated with active or 'poised' promoters. We propose that binding of CTCF to the spacer promoter also maintains rDNA repeats 'poised' for transcription.

With one high affinity binding site per mouse rDNA, and with CTCF preferring non-methylated (and thus active) rDNA repeats, it is expected that only a small number of DNA-bound nucleolar CTCF molecules would be present. By contrast, UBF is abundantly present in the nucleolus, where it binds rDNA with low specificity [23] and is highly dynamic [15,38]. Thus, a UBF-CTCF interaction is expected to be transient. However, the interaction is crucial, as CTCF binding enhances UBF binding both *in vitro* and *in vivo*. Nucleolar UBF in turn ensures that rDNA repeats remain accessible to RNA polymerase I [18]. UBF, as part of the architectural HMG-box protein family, could change the topology of the rDNA, thereby facilitating binding of other factors [39], and allowing formation of small ~175 bp DNA loop structures called enhanceosomes [40]. In addition, CTCF and UBF may together recruit RNA polymerase I to the spacer promoter. Binding by CTCF to components of the RNA polymerase I complex would aid in this recruitment.

The biological function of the spacer promoter and the ncRNA transcript that is generated from it are still not completely understood. Early experiments suggested that the spacer promoter and the enhancer region act together to stimulate pre-rRNA transcription [27,41]. More recent data have shown that ncRNAs generated from the spacer promoter are unstable; transcripts are rapidly processed and degraded, and only the 3' end (~150 nucleotides) of the transcript, which matches the rDNA gene promoter, is bound to the nucleolar remodeling complex (NoRC) and is required for the establishment and maintenance of inactive rDNA repeats [24,30]. In this context, the spacer promoter transcript functions in rDNA silencing instead of activation.

Recent data implicate UBF in the balance between active and inactive rRNA genes, via a 'pseudosilencing' mechanism that is reversible and does not involve DNA methylation [18]. Thus, there appear to be two different epigenetic mechanisms that regulate the number of active rRNA genes. An attractive hypothesis is that CTCF, by binding to the spacer promoter of non-methylated (and thus active) rDNA repeats, and by interacting with UBF and 'loading' it onto these repeats, is involved in the 'pseudo-silencing' mechanism and maintains rDNA repeats 'poised' for transcription. At the same time, by generating spacer promoter transcripts, CTCF is 'feeding' NoRC with its 3' end degradation product, allowing this protein complex to function in a second epigenetic rRNA gene silencing mechanism. Consistent with this notion, ncRNA transcription appears to take place on a small subset of hypomethylated mouse rDNA repeats [26].

CTCF not only acts locally, but also regulates nucleolar number. Results in MEFs are consistent with data in T cells, where we found that the number of rDNA-positive signals decreases when CTCF is deleted *in vivo* [31]. Interestingly, B23 (or nucleophosmin) is a protein partner of CTCF, and B23-enriched insulator sequences are tethered to the nucleolar rim in a CTCF-dependent manner [42]. B23 is important for nucleolar structure [43]. Moreover, the borders of lamina-associated domains, detected by lamin B1, are demarcated by CTCF binding sites [44]. Lamin B1 interacts with B23, and is also involved in maintaining nucleolar structure [45]. We hypothesize that control of nucleolar number by CTCF is linked to its global function as an architectural factor, in association with proteins such as B23 and lamin B1.

Ribosome biogenesis controls cell growth and proliferation, as it determines the protein synthesis capacity of a cell. Recently, we showed that CTCF positively regulates cell growth in rapidly dividing thymocytes [31]. In the present study we detected multiple components of the RNA polymerase I complex in the mass spectrometry analysis of CTCF-bio-interacting partners. Knockdown of CTCF in ES cells resulted in slightly lowered levels of pre-rRNA. Conversely, under conditions of repressed pre-rRNA transcription, as in differentiated 3T3L1 cells [29], CTCF binding to the spacer promoter is reduced. Combined, these data suggest a link between CTCF, rRNA synthesis and cell growth control, whereby CTCF appears to act at a local and a global level.

Conclusions

We show that CTCF and CTCFL bind UBF directly. CTCF organizes the local epigenetic state of rDNA repeats by regulating the binding UBF and RNA polymerase I and of other crucial components, and by altering chromatin modifications near its binding site. By tightly binding the rDNA spacer promoter, CTCF may enhance UBF binding and ensure that rDNA repeats are accessible to RNA polymerase I. CTCF binding at the spacer promoter stimulates transcription of non-coding RNA from the spacer promoter. The local organization

of chromatin at CTCF sites near the RNA polymerase I-regulated spacer promoter and near RNA polymerase II promoters is remarkably similar. The CTCF-dependent enrichment of H2A.Z and H3K4me2 near the spacer promoter indicates that CTCF binding maintains rDNA repeats 'poised' for transcription.

Methods

Accession numbers and primers

We used mouse (accession number BK000964), human (U13369), rat (X04084) and hamster (DQ235090) rDNA sequences for alignments, to design primers for PCR and ChIP experiments and for probe generation. Primers used in all the different experiments are shown in Tables S2 tp S7 (Additional file 12, 13, 14, 15, 16, 17 respectively).

Antibodies and cDNAs

CTCF mouse monoclonal antibodies were from BD Biosciences (Breda, NL), and CTCFL polyclonal rabbit antibodies (18337) were from Abcam. CTCFL (#4) polyclonal rabbit antibodies are described elsewhere (Sleutels *et al*, manuscript in preparation). The anti-CTCF (N3) and anti-RPA194 rabbit polyclonal antisera have been described previously [31,46]. Anti-histone H2A.Z (ab4174), anti-dimethyl-histone H3 (Lys4) (ab7766) and anti-histone H3 (ab1791) antibodies were from Abcam. Anti-acetyl histone H3 (06-599) and anti-acetyl Histone H4 (06-866) antibodies were from Upstate (Millipore, Amsterdam, NL). Anti-UBF (sc-13125) and anti-actin (sc-8432) antibodies were from Santa Cruz Biotechnologies (Santa Cruz, CA, USA). Streptavidin-HRP (RPN1231VS) and secondary HRP-labeled anti-mouse (NA931VS) and anti-rabbit antibodies (NA934V) were from Amersham (GE Healthcare, Uppsala, Sweden). Anti-His antibody was from Qiagen (Valencia, CA, USA), and anti-Flag M2 antibody was from Sigma Chemical Co. (St Louis, MO, USA).

His-tagged UBF fusion proteins were generated by PCR using mouse UBF cDNA from a Flag-tagged UBF construct as template (kind gift of Dr I. Grummt). Primers contained *Nhe*I and *Bam*HI sites for subcloning into the pET28a vector. GST-tagged fusions of mouse CTCF and CTCFL were amplified using mouse CTCF (IMAGE 6825952) and CTCFL (Sleutels *et al*, manuscript in preparation) cDNAs as templates. cDNAs were cloned into plasmid pGEX-3X and purified (glutathione-Sepharose 4B; Amersham Biosciences). GST-tagged fusion proteins derived from chicken CTCF have been described previously [47].

Mass spectrometry

For mass spectrometry samples were treated and analyzed as described [48]. Data analysis and protein identification was performed as reported [49]. The Mascot search algorithm (version 2.0; MatrixScience) was used for searching against the NCBI database (taxonomy: *Mus musculus*). The Mascot score cut-off value for a positive protein hit was set to 60. Individual peptide tandem mass spectrometry (MS/MS) spectra with scores of < 40 were checked manually, and either interpreted as valid identifications or discarded. A number of CTCF-bio interacting proteins are listed in Table S1 (see Additional file 2). It should be noted that CTCF is difficult to purify under the mild conditions that are required to isolate associating proteins, although CTCF binds DNA tightly, the majority of its protein-protein interactions are of a transient nature.

Affinity chromatography and size fractionation

Nuclear extracts were prepared as described previously [50]. Salt concentration in the extract was adjusted to 100 mmol/l NaCl. Unless stated differently, all IP and pull-down reactions were performed in IP buffer (100 mmol/l NaCl, 0.3% NP40, 20 mmol/l Hepes pH8, 0.2 mmol/l EDTA, 10 mmol/l $MgCl_2$, with protease inhibitors) (Complete; Roche). Benzonase (Novagen) was added where indicated to remove DNA and RNA.

Streptavidin pull-down assays were performed as described previously [19], with the exception that the wash buffer and binding buffer were the same as the IP buffer described above. For IPs, nuclear extracts were pre-cleared at 4°C (Protein A sepharose beads, Sigma). Washes were performed at 4°C in wash buffer (100 mmol/l NaCl, 20 mmol/l Tris pH7.5, 0.3% NP40 and protease inhibitors). IPs were performed by adding antibodies to the samples and incubating for 1 hour at 4°C. Subsequently, protein-A sepharose beads were added, and incubation was continued for another hour at 4°C while rotating. Beads were washed six times with wash buffer.

Flag-IPs were performed using the same protocol as for IPs, except that anti-Flag M2 agarose (Sigma) incubation was performed for 3h at 4°C.

His-tagged proteins were bound to nickel-nitrilotriacetic (Ni-NTA) beads (Qiagen) in low salt buffer (20 mmol/l Hepes pH 7.5, 100mmol/l KCl, 10 mmol/l β-mercaptoethanol and 10% glycerol v/v). Proteins were purified by extensive washing of the beads, first in low-salt buffer, followed by washing in buffer with 1 mol/l KCl, and washing again in low-salt buffer. Proteins were eluted from the beads with 200 mmol/l imidazole in low-salt buffer, then the imidazole removed by dialysis. GST-tagged proteins were purified on glutathione-Sepharose 4B columns (Amersham Biosciences), using low and high salt buffers as above. To remove contaminating nucleic acids, benzonase was first added to bacterial extracts and again during

washing of the (Ni-NTA) and glutathione beads. GST-based pull-downs were performed in binding buffer (20 mmol/l Tris-HCl pH 8, 100 mmol/l NaCl, 0.05% Triton X-100) containing benzonase, for 2 hours at 4° C. Washes were performed in binding buffer, followed by washes in high salt wash buffer (20 mmol/l Tris-HCl pH8, 400 mol/l NaCl, 0.05% Triton X-100). GST pull-downs on ES cell nuclear extracts were performed using the binding and washing conditions as described in the IP section.

Size fractionation of protein complexes was performed on a fast protein liquid chromatography apparatus (AKTA FPLC; Amersham Biosciences) with a Superose 6 10/30 column (Amersham Biosciences). Fractions were precipitated with 100% trichloroacetic acid and analyzed by western blotting as described previously [51]. Molecular size standards were thyroglobulin (670 kDa) and albumin (66 kDa) (Amersham Biosciences).

SDS-PAGE, western blotting and EMSA

Bound proteins were eluted from beads by boiling in sample buffer (1 × Laemmli buffer). For western blot analysis, samples were separated by electrophoresis in SDS polyacrylamide gels and blotted onto poly(vinylidene fluoride membranes), (MilliPore) using a semi-dry blotting apparatus (BioRad). Signal detection was performed using enhanced chemiluminescence (Amersham).

For EMSA or band-shift analysis, protein extracts were preincubated with bandshift buffer (10% glycerol, 20 mmol/l Hepes pH7.4, 20 mmol/l KCl, 1 mmol/l $MgCl_2$, 5 mmol/l dithiothreitol (DTT), 10 μmol/l $ZnCl_2$, 100 μg/ml bovine serum albumin (BSA), 0.02% NP40) and 2 to 4 μg of salmon sperm DNA as a non-specific competitor. The reaction was incubated for 20 minutes at room temperature. Upon addition of probe the binding reaction was performed for another 20 minutes. Complexes were analyzed by electrophoresis through a 5% acrylamide (37,5:1) 0.5 × Tris/borate/EDTA non-denaturing gel at 8V/cm^2 at 4°C. Where specified, 300-fold excess of unlabeled probe or specific competitor was added at the same time as the probe.

Mouse probes for EMSA were end-labeled with ^{32}P, whereas human probes (MYC-N, H42.1 rDNA and H37.9 rDNA) were ^{32}P-labeled PCR fragments. For EMSA with *in vitro* methylated probes, purified H37.9 and H42.1 rDNA fragments (5 μl) were methylated *in vitro* using 12 U *Sss*I methyltransferase (New England Biolabs) and 1 μl S-adenosyl-L-methionine (32 mmol/l) in a final volume of 30 μl. Reactions were performed twice for 4 hours at 37°C, after which probes were purified. For supershift experiments, 1 μl of anti-CTCF mouse monoclonal or anti-actin (used as non-specific antibody) was added to the binding reaction before the radiolabeled probe.

ChIP

Preparation of cross-linked chromatin (2 × 10^7 cells treated with 1% formaldehyde for 10 minutes at room temperature), sonication of chromatin to yield fragments of 300 to 800 bp, and immunoprecipitation were performed as described in the Upstate protocol http://www.upstate.com. At least two independent ChIPs were carried out per experiment. For streptavidin ChIPs, minor modifications were used: streptavidin beads were blocked for 1 hour at room temperature in 0.2 mg/ml sonicated salmon sperm DNA, elution was performed for 16 hours at 65°C in elution buffer (0.1% $NaHCO_3$, 1% SDS, 0.2 mol/l NaCl). Quantitative real-time PCR (Opticon I, MJ Research and MyiQ, BioRad) was performed using SYBR Green (Sigma), Platinum *Taq* DNA polymerase (Invitrogen) and 100 ng of each primer under the following cycling conditions: 95°C for 3 minutes, followed by 40 cycles of 10 seconds at 95°C, 30 seconds at 60°C and 15 seconds at 72°C (during which measurements were taken). Values were normalized to input measurements, and enrichment was calculated relative to the *Amylase* gene using the comparative Ct method. PCR products were all < 150 bp.

For ChIP analysis with nuclei derived from human cell lines, 5 × 10^7 cells were fixed in 1% formaldehyde, lysed and sonicated. ChIP was performed using Dynabeads-protein G (Dynal Biotech) coupled to anti-CTCF, anti-CTCFL or anti-UBF antibodies. Dynabeads were incubated with lysates for 4 h at 4°C, and washed consecutively with commercial buffers (Low Salt, High Salt and LiCl Immune Complex Wash Buffers; Upstate). Chromatin was eluted with 200 μl of elution buffer (Upstate), de-crosslinked for 8 hours at 65°C, and purified (Qiaquick columns; Qiagen). Real-time PCR of immunoprecipitated DNA was performed with primers shown in Table S7 (see Additional file 17). The MYC-N and NY-ESO1 amplicons were used as positive controls for CTCF and CTCFL, respectively, and the MYC-H.1 amplicon as negative control. Enrichment for a specific DNA sequence was calculated as above.

Methylation-sensitive ChIP assay (ChIP-chop)

To analyze the methylation density of rDNA precipitated with CTCF antibodies, post-ChIP hydrolysis ('chopping') of DNA was performed using the methylation sensitive enzyme *Hpa*II and its isoschizomer *Msp*I. Input sample (60 ng) and DNA from the CTCF ChIP reaction were divided into three equal aliquots, which were digested with *Hpa*II or *Msp*I, or left undigested (mock digested control). Digestions were carried out in a final volume of 20 μl for 3 hours at 37°C. Enzymes were inactivated for 30 minutes at 65°C. From each digestion, 10 μl was subjected to quantitative PCR with H42.1 rDNA primers, as described above. The uncut

rDNA was set at 100%. The percentage of *Hpa*II and *Msp*I resistance was calculated as a percentage of mock digested DNA, by measuring the difference in Ct values in the qPCR (mock-*Msp*I or mock-*Hpa*II), taking the inverse of the fold difference in expression level, and multiplying this value by 100.

Cell lines, transfections and lentiviral transduction

To generate the $Ctcf^{bio\text{-}neo}$ knock-in allele, a CTCF-TEV-bio in-frame fusion DNA was generated by PCR. In this construct, the biotinylation sequence [19] is preceded by a tobacco etch virus (TEV) protease cleavage site of seven amino acids. The neomycin-resistant LoxP-Neo-loxP vector and targeting procedures have been described previously [51]. IB10 129 ES cell DNA was analyzed by Southern blotting using radiolabeled probes outside of the region of homology (Figure 1A). For confirmation of homologous recombination, we used different 5' end and 3' end probes, and a PCR-based genotyping assay.

$Ctcf^{bio\text{-}neo}$ ES cells were transfected with CMV-Cre to remove the neomycin resistance cassette. A second round of homologous recombination was performed to target the Rosa26 locus with hemagglutinin (HA)-tagged BirA [20]. Verification of homologous recombined clones was performed by PCR. Control BirA-positive ES cell lines have been described previously [20].

3T3L1 cells (CL-173; ATCC) [29] and 293T cells [10] were cultured as described previously. The $Ctcf^{f/f}$ primary MEFs were isolated as described previously [51] at embryonic day 13.5 from embryos derived from conditional $Ctcf^{f/f}$ knockout mice [31].

Transient transfections in 293T cells with Flag-UBF and pcDNA3-CTCFL were performed using a transfection reagent (Lipofectamine™2000; Invitrogen) in reduced serum media (Optimem; GibcoBRL). Cells were analyzed 24 hours after transfection. Cre-lentivirus production and transduction of confluent primary MEFs was performed as described [10], with the exception that cells were split and diluted two-fold at 24 hours after transduction. Virus titers and Cre functionality were tested using serial dilutions. Recombination was tested after 4 days of infection by quantitative RT-PCR.

KCTCFD11 is a sub-line derived from K562 myeloid leukemia cells, which is stably transfected with a constitutive CTCF expression vector that moderately overexpresses CTCF (two to three-fold) compared with cells transfected with the empty vector (KpCDNA subline) [52]. For EMSA experiments, 293T cells or K562 cells were transfected with pcDNA3-CTCF expression vector (Lipofectamine™2000; Invitrogen).

Northern blot analysis

Total RNA was isolated using an isolation solvent (RNA-Bee RNA Isolation Solvent; Tel-Test Inc.), size separated by gel electrophoresis (~6 μg per lane) and blotted onto membrane (Hybond N+; Amersham). Probes were radioactively labeled by PCR. Blots were exposed to screens (PhophorImager; Molecular Dynamics) to quantify results.

Nuclear run-on

Cells were collected and washed twice with cold phosphate-buffered saline (PBS). The cells were lysed in nuclear isolation buffer (10 mmol/l Tris pH7.5; 10 mmol/l NaCl, 10 mmol/l $MgCl_2$, 0.5% NP40). The nuclei were spun at 1000 g and resuspended in storage buffer (50 mmol/l Tris pH8.5, 0.1 mmol/l EDTA, 5 mmol/l $MgCl_2$, 40% glycerol). Approximately 10^6 nuclei (50 μl) were pre-incubated for 20 minutes on ice with 100 μg/ml α-amanitin. Nuclei were then mixed with 50 μl 2 × reaction buffer (300 mmol/l KCl, 5 mmol/l $MgCl_2$, 10 mmol/l Tris pH 7.5, 5 mmol/l DTT, 20 U RNA Guard, 0.5 mmol/l of each ATP, UTP and GTP, and 100 μCi of α^{32}P CTP (800 Ci/mmol, 10 mCi/ml); Amersham). The labeling reaction was performed for 30 minutes at 30°C. The reaction was stopped on ice by adding 1 ml of isolation solvent (RNA-Bee) and total RNA was extracted as indicated above. Using a slot blot hybridization system with nylon membranes (Hybond-N+), 5 μg of DNA PCR fragments were hybridized with2×10^5 cpm of labeled RNA. Hybridization and detection was performed as described above. Incubation was performed in 2 ml of Church hybridization mix (0.5 mol/l Na_2HPO_4 pH 7.2, 7% SDS, 1 mmol/l EDTA) in a rotating hybridizer at 65°C for 24 h. Membranes were washed extensively at 65°C with Church wash buffer (40 mmol/l Na_2HPO_4 pH 7.2, 1% SDS). Hybridization signals were quantified with an imager (Phosphor Imager; Typhoon Amersham) using Imagequant software. The signal was corrected for the amount of CTG in the probe.

Real-time PCR on ES cell RNA

Total ES cell RNA was isolated using Trizol (Invitrogen), treated with DNAseI, and converted into cDNA using random hexamers and reverse transcriptase (Superscript II; Invitrogen). Real-time PCR was performed using specific rRNA-covering primers and Sybr Green mix (Quantitect; Qiagen) on a performed on an automated PCR system (7500 Fast RT-PCR; Applied Biosystems). The negative control was as above with omission of the reverse transcriptase. The obtained Ct values were normalized to the Ct value of *Hprt*.

FISH and immunofluorescence analysis

For FISH in ES cells, the cells were grown on coverslips. RNAi treatment of the cells was performed using a pSUPER vector-based system (CTCF RNAi sequence:

5"-GCAGAGAAAGTAGTTGGTAAT-3"). After transfection, cells were treated with puromycin to positively select for infected cells, thereby increasing the number of cells in which CTCF was knocked down. After 4 days of RNAi treatment, cells were fixed for 10 minutes with 4% paraformaldehyde (PFA) in PBS. Slides were stored in 70% ethanol until further use. For RNA FISH, cells were pretreated by two PBS washing steps, followed by a permeabilization step of 5 minutes in a solution of 25 μg/ml proteinase K in PBS. Slides were washed once in PBS, dehydrated and hybridized as described previously [53]. For DNA FISH, slides were pretreated by two PBS washing steps followed by a permeabilization step (4 minutes incubation in 0.1% pepsin in 0.01 mol/l HCl at 37°C). Slides were washed once in PBS on ice and fixed again for 5 minutes in 4% PFA in PBS. Slides were washed twice in PBS and dehydrated. Denaturation was performed for 2 minutes at 80°C in denaturing solution (70% formamide, 2 × saline sodium citrate, 10 mmol/l phosphate buffer pH 7), after which the slides were cooled in 70% ethanol, dehydrated, and hybridized as described previously [53].

The unstable 5" external transcribed spacers (ETS) probe has been described previously [32]. The enhancer probe used for DNA and RNA FISH (ncRNA; see Figure 7A for its position) was isolated as a 1.7 kb *Sal*I fragment from a cosmid covering a large part of the mouse rDNA repeat [32]. Probes were labeled by nick translation (Roche) using digoxygenin or biotin. Control DNA FISH experiments in ES cells showed that the ncRNA probe specifically localized to the nucleolus, as on prometaphase chromosomes the probe localized in discrete spots adjacent to centromeric DNA, indicative of NORs (see Additional file 11, Figure S10A), whereas in interphase cells the ncRNA probe localized within the nucleolus (see Additional file 11, Figure S10B). These data strongly suggest that the ncRNA probe specifically recognizes rDNA. When ES cells were treated with α-amanitin to inhibit RNA polymerase II transcription, both ncRNA and pre-rRNA signals remained visible (data not shown), confirming that RNA polymerase I is responsible for transcription of spacer and gene promoters.

For immunofluorescence staining, cells were fixed in 4% PFA in PBS for 15 minutes at room temperature, permeabilized in 0.15% Triton X-100 in PBS, blocked in 1% BSA in PBS and incubated with antibodies as described previously [32,51]. Images of cells were collected with a microscope (DMRBE; Leica) equipped with a camera (ORCA ER; Hamamatsu) or with a confocal lens (LSM510; Zeiss), as described previously [51].

For quantification of pre-rRNA signals, images of ES cells were collected with a microscope (DMRBE; Leica), using the same exposure time for all images. Five images each were collected of non-treated ES cells, control RNAi-treated ES cells and CTCF RNAi-treated ES cells. Collectively, more than 300 cells were present in the images, which were imported into Image J software (Rasband W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA; 1997 to 2008; http://rsb.info.nih.gov/ij/,). Regions of interest (ROIs) were placed around individual pre-rRNA signals, using the freehand tool of Image J. ROIs were saved with the ROI manager. Both background fluorescence and mean fluorescent intensities of ROIs were calculated in each image. After deduction of the background fluorescence, mean fluorescence intensity data were collected into a spreadsheet (Excel; Microsoft), pooled and analyzed (Aabel software; GigaWiz). Quantification was performed in two independent experiments using different ES cell cultures, different probes and different RNAi treatments. Both experiments yielded similar results; that is, knock-down of CTCF leads to mildly reduced pre-rRNA levels.

Additional material

Additional file 1: Figure S1: Characterization of CCCTC binding factor (CTCF) and biotin tagged CTCF (CTCF-bio) interactions. **(A)** CTCF-bio interacts with known CTCF binding partners. To identify CTCF-interacting proteins, CTCF-bio was purified from embryonic stem (ES) cell nuclear extracts under mild conditions. We validated our approach by showing that known interaction partners of CTCF, such as Yin Yang (YY)-1 [54], poly(ADP-ribose) polymerase (Parp)1 and nucleophosmin [42] co-precipitate with CTCF-bio. **(B)** CTCF-bio interacts with upstream binding factor (UBF) *in vivo*. Streptavidin pull-downs were performed using lung nuclear extracts isolated from mice expressing biotinylated (CTCF-bio) or normal (-) CTCF. Western blot analysis (b = bound fraction, i = input (5%)) revealed that CTCF-bio interacts with UBF. **(C)** Immunoprecipitation (IP) analysis of CTCF and UBF. IP was carried out on extracts of ES cells expressing both CTCF and CTCF-bio. We used specific antibodies against CTCF and UBF to precipitate endogenous proteins (IgG = control rabbit IgG). CTCF-bio was detected with horseradish peroxidase-coupled streptavidin. B = bound fraction, i = input (5%).

Additional file 2: Table S1: Mass spectrometry results for biotin tagged CTCF (CTCF-bio).

Additional file 3: Figure S2: Direct interaction of CCCTC binding factor (CTCF) and CTCFL with UBF. **(A)** Schematic representation of the glutathione-S-transferase (GST)- and histidin (His)-tagged fusion proteins used. **(B)** Expression of GST- and His-tagged fusion proteins. Proteins were expressed in bacteria and affinity purified. Fusion proteins are indicated by asterisks. **(C)** Interaction between bacterially produced proteins. Purified GST- and His-tagged fusion proteins were incubated together, followed by GST pull-down. Western blots were incubated with an anti-His antibody. The experiments revealed a direct interaction between the CTCF and CTCFL zinc finger (ZF) domains and the upstream binding factor (UBF) dimerization domain plus high mobility group (HMG)-box 1. His-tagged proteins containing either the dimerization domain of UBF or HMG-box 1 only weakly bound CTCF and CTCFL, indicating that both regions are necessary for efficient interaction. **(D)** Bacterially produced CTCF and CTCFL interacted with UBF derived from embryonic stem (ES). GST pull-down assays of bacterially produced CTCF and CTCFL mutants were performed with nuclear protein extracts from ES cells. Equal amounts of GST fusion proteins were incubated with nuclear extracts from ES cells. Binding was performed under low-salt conditions, and washing was performed under more stringent conditions. Western blots were incubated with an antibody against UBF

to detect ES cell-derived UBF. GST-tagged CTCF and CTCFL were both able to pull down specifically UBF. The ZF domains of CTCF (1) and CTCFL (5) displayed prominent interaction with ES cell-derived UBF. **(E)** Bacterially produced ZFP37 did not interact with histidine (His)-tagged UBF. The ZF domain of murine ZFP37 a protein that is enriched in the nucleolus [55] was tagged with GST. To provide further evidence for the specificity of the CTCF-UBF interaction, we examined whether this ZF domain interacts with UBF. Purified GST-tagged ZFP37 was incubated with His-tagged UBF (constructs 9, 10, 13). The interaction between CTCF (construct 1) and UBF was clear, but we could not detect any binding between UBF and ZFP37.

Additional file 4: Figure S3: Both CCCTC binding factor (CTCF) and biotin-tagged CTCF (CTCF-bio) bind the ribosomal (r)DNA spacer promoter *in vivo*. Extracts of adult thymus from wild type and $Ctcf^{bio/+}$; $Rosa26^{bira/+}$ mice were analyzed for CTCF and CTCF-bio binding to the rDNA spacer promoter using anti-CTCF antibodies or a control serum (-).

Additional file 5: Figure S4: Comparison of mouse, rat and hamster ribosomal (r)DNA repeat regions. Comparison of nucleotide sequences of the mouse, rat and hamster rDNA repeats [27]. Only the regions around the spacer promoter are indicated. Numbers to the left indicate distance (in base pairs) from the transcription start site of the gene promoter. The CCCTC binding factor (CTCF) consensus site [3] is underlined. Highly conserved CTCF consensus site residues are indicated by a dot (the asterisk indicates deviation between consensus site prediction and real residue). Conserved CpG dinucleotides are boxed. The transcription start site of the spacer promoter is indicated by a right-pointing arrow.

Additional file 6: Figure S5: Comparison of mouse and human rDNA repeat regions. **(A-C)** Matrix plot comparisons of nucleotide sequences of **(A)** mouse versus mouse, (**(B)** human versus human and **(C)** human versus mouse rDNA repeats in the region upstream of the gene promoter. CCCTC binding factor (CTCF) binding sites are indicated (mCTCF BS for mouse, H37.9 and H42.1 for human). A highly repetitive Alu sequence is present ~2.5 kb upstream of the gene promoter of the human rDNA. Mouse rDNA does not have this repeat, but instead contains the well known 'enhancer repeat' region. Why CTCF binds twice in human and only once in mouse rDNA is unclear. One possibility is that CTCF has additional regulatory functions in the human rDNA repeat. For example, the H37.9 site is conserved in the rDNA of the great apes, as is the highly repetitive Alu repeat [56]. We speculate that H37.9 might be linked to the presence of this repetitive region in human and great ape rDNAs. **(D)** Similar chromatin organization of mouse and human rDNA repeat regions upstream of the gene promoter. The upper line represents the mouse rDNA (enhancer repeats are indicated by the open rectangles), and the lower line represents the human repeat. Only regions upstream of the gene promoter are shown. Right-pointing arrows indicate transcription from the gene promoter, giving rise to pre-rRNA. The spacer promoter has been clearly identified for the mouse but its location has not yet been mapped accurately for the human RNA. The chromatin organization surrounding the CTCF binding site (indicated by a lollipop) that is most proximal to the gene promoter, is strikingly similar in both mouse and human. In both organisms, the CTCF binding sites are embedded within a CpG island (as predicted with EMBOSS-CpG Plot [57]; the length of the CpG domains is indicated below the respective rDNAs). Immediately upstream of the CTCF binding site, mouse rDNA chromatin is enriched in 'active' histone modifications. A surprisingly similar result was previously obtained in the human locus (see Figure 5, site H42 in the paper by Grandori *et al.* [36]). Furthermore, TATA binding protein (TBP) has been shown to accumulate near the CTCF binding site, both in human [36] and mouse [25] rDNA repeats. We therefore propose that the spacer promoter in the human rDNA is located immediately downstream of the H42.1 CTCF binding site.

Additional file 7: Figure S6: CCCTC binding factor (CTCF) and CTCFL interact with human ribosomal (r)DNA *in vivo*. **(A)** Chromatin immunoprecipitation (ChIP) analysis on human rDNA. ChIP analysis with CTCF, CTCFL (two independent antibodies (Abs) were used) and UBF antisera, showing binding to the IGS of the rDNA repeat (sites H4, H37.9 and H42.1). Chromatin was prepared from non-transfected K562 cells or from cells stably transfected with CTCF (KCTCFD11) or the empty vector (KpCDNA). Relative enrichment was quantified by real-time PCR with the indicated primer sets. Known CTCF (MYC-N) and CTCFL (NY-ESO1) target sites were used as positive control for ChIP. Data were normalized against the enrichment for the negative control MYC-H.1. The value for the amount of PCR product present from the ChIP assay without antibody was set as 1 (white bars). Error bars represent the SEM of five to seven independent experiments for CTCF, eight to 10 for upstream binding factor (UBF), and four for CTCFL. **(B)** Sequential ChIP (ChIP-reChiP) analysis on human rDNA. Primary ChIP was performed as above, and CTCF or UBF ChIP products were subjected to a second immunoprecipitation (reChIP) with anti-UBF or anti-CTCF antisera, respectively. Relative enrichment was quantified by real-time PCR with primers for H37.9 or H42.1 rDNA, and data were normalized as in part (A). Error bars represent SEM of three independent experiments. Results show *in vivo* binding of CTCF and UBF simultaneously at rDNA sites. **(C)** CTCF interacts with human rDNA *in vitro*. Electrophoretic mobility shift assay (EMSA) analysis with nuclear extracts from 293T cells or K562 cells transfected with CTCF or mock transfected. ^{32}P-labeled PCR fragments of MYC-N (positive control), H42.1 rDNA and H37.9 rDNA were used as probes. Unlabeled (cold) probes were used as competitors (Myc-N = 90% competition (compare lanes 1 and 3); H42.1 = 95% competition (compare lanes 5 and 7); H37.9 = 85% competition (compare lanes 11 and 13). Arrowheads indicate binding of CTCF; asterisks indicate supershift bands that appear after incubation with anti-CTCF antibody.

Additional file 8: Figure S7: CCCTC binding factor (CTCF) binds human ribosomal (r)DNA in a methylation-sensitive manner. **(A)** Influence of methylation on the binding of CTCF to mouse rDNA. Band-shifts were performed using the mouse rDNA probe (R30) and the same probe methylated on two cytosine residues (R30-CmE). Competition was assessed by adding increasing amounts of unlabeled probe. **(B)** Control and SssI methyltransferase-treated H37.9 and H42.1 rDNA probes were digested with the methylation-sensitive enzyme *Hpa*II to assess the level of *in vitro* methylation. Fragments were separated in 8% polyacrylamide gels.

Additional file 9: Figure S8: Chromatin immunoprecipitation (ChIP) analysis in embryonic stem (ES) cells. **(A)** Outline of mouse rDNA repeat. The position of the primer pairs used in the ChIP in panel B is indicated by downward-pointing arrows. Transcription initiation from the spacer promoter (yielding ncRNA) and the gene promoter (yielding pre-rRNA) is indicated by right-pointing arrows. The 47S pre-rRNA is divided into 5' and 3' external transcribed spacer (ETS), internal transcribed spacers (ITS), and 18S, 5.8S and 28S rRNA genes. The approximate positions of the CCCTC binding factor (CTCF) consensus site (gray box) and enhancer repeats (white boxes) are indicated. **(B)** ChIP assay on mouse rDNA. Binding of CTCF (black), upstream binding factor (UBF) (purple/blue) and RNA polymerase I (red) to mouse rDNA was analyzed using the primer pairs indicated in part (A). Embryonic stem (ES) cells were treated with control (straight lines) or *Ctcf* (stippled lines) RNAi constructs. ES cell nuclei were fixed with 1% formaldehyde, and protein-DNA complexes were immunoprecipitated with antibodies against the indicated proteins. Upon depletion of CTCF, binding of both RNA Pol I and UBF was diminished. Strikingly, for both proteins, loss in binding was greatest near the CTCF binding site, strongly suggesting an important role for CTCF in the binding of these proteins at or near the spacer promoter (RNA Pol I, UBF) and on the enhancer repeat (UBF). The fact that RNA Pol I binding was not affected at or downstream of the gene promoter is consistent with previous data.

Additional file 10: Figure S9: CCCTC binding factor (CTCF) regulates histone deposition. **(A)** Binding of histone H3 to mouse ribosomal (r) DNA. Chromatin immunoprecipitation (ChIP) analysis is the same as shown in Figure 6, but a lower y-axis scale is used to demonstrate the histone H3 binding pattern. Enrichment was normalized to input and is shown relative to the *Amylase* gene (note that in this case the *Amylase* gene is not a negative control, because histone H3 will also bind this gene, hence the 'low' relative enrichment). Histone H3 was distributed in a similar manner in Cre-treated $Ctcf^{f/f}$ mouse embryonic fibroblasts (MEFs) compared with non-treated cells. Interestingly, binding appeared to diminish as the ribosomal gene promoter area is approached. This might be due to the fact that active ribosomal genes contain fewer nucleosomes [12]. **(B)** Binding of CTCF and H2A.Z to the *c-Myc* gene.

ChIP analysis was performed as in Figure 6C, using the regulatory region upstream of the *c-Myc* transcriptional start site (SD of three independent experiments indicated). The position of the primer sets is indicated with arrows. **(C)** Binding of modified and variant histones to human rDNA. ChIP analysis was performed as in Figure S6A (see Additional file 7). Chromatin was prepared from K562 cells. Protein-DNA complexes were immunoprecipitated with antibodies against the indicated proteins. ChIP analysis showed specific binding of H2A.Z, H3ac and H3K4me2 to sites H37.9 and H42.1 of the rDNA.

Additional file 11: Figure S10: Spatial segregation of non-coding (nc)RNA and pre-rRNA transcription. A, **(B)** DNA fluorescent *in situ* hybridization (FISH) analysis. The ncRNA probe (biotin-labeled, green) was hybridized to fixed and denatured ES cells. **(A)** Cell in prometaphase, with chromosomes condensed but not yet aligned. The ncRNA probe localized in distinct spots (arrows) adjacent to the strongly 4',6-diamidino-2-phenylindole (DAPI)-stained centromeric DNA. **(B)** An interphase cell, with the ncRNA probe localized to the nucleolus (visualized as weakly staining DAPI regions). Scale bars = (A) 2 μm, (B) 3 μm. **(C-F)** RNA FISH analysis. The ncRNA probe (biotin-labelled, green) and pre-rRNA probe (digoxygenin-labeled, red) were hybridized to fixed non-denatured ES cells. **(C-E)** Embryonic stem (ES) cells contain normal levels of CCCTC binding factor (CTCF), whereas **(F)** ES cells transfected with a pSUPER plasmid have CTCF knockdown. **(C)** Low resolution image of a small ES cell colony (cells in the middle are less well visualized because these cells grow in clumps). Multiple nuclei (one is outlined), particularly on the edge of the colony, had readily detectable ncRNA and pre-rRNA signals. Scale bar = 10 μm. **(D)** High resolution confocal image of a single DAPI-stained ES cell nucleus. Both ncRNA and pre-rRNA signals were localized exclusively to the three nucleoli present within this cell. Five ncRNA spots are visible (arrows), localized at the periphery of the nucleoli. Scale bar = 1 μm. **(E, F)** Confocal images taken with similar settings. **(E)** non-treated ES cells; **(F)** CTCF RNAi-treated ES cells. The ncRNA signal is indicated by arrows. Depletion of CTCF led to a reduction in ncRNA. ncRNA was lacking in many cells throughout a 3D confocal stack. In cells lacking ncRNA, pre-rRNA levels also seemed to be affected (see asterisks). Scale bars (E, F) = 8 μm. **(G)** Knock-down of CTCF in ES cells. ES cells were transfected with a pSUPER plasmid to knock down CTCF. After 4 days, < 50% of the cells expressed detectable levels of CTCF (red), as detected by immunofluorescent staining with anti-CTCF antibodies. Nuclei were stained with 4',6-diamidino-2-phenylindole (DAPI). By contrast, cells treated with a control RNAi vector all expressed CTCF (not shown). Scale bar = 8 μm.

Additional file 12: Table S2: Primers used for band-shift assays [58].

Additional file 13: Table S3: Primers used for genotyping.

Additional file 14: Table S4: Primers used for mouse chromatin immunoprecipitation (ChIP).

Additional file 15: Table S5: Primers used for northern blotting and nuclear run-on assays.

Additional file 16: Table S6: Primers used for real-time PCR on embryonic stem (ES) cell ribosomal (r)RNA.

Additional file 17: Table S7: Primers used for human chromatin immunoprecipitation (ChIP) and band-shift assays [59-61].

Acknowledgements
We thank Dr I. Grummt for kindly providing us with Flag-tagged UBF, Dr D. N. Meijer for the Rosa26-BirA construct and K. Bezstarosti for helping with the mass spectrometry analysis. This research was supported by the Dutch Cancer Society (KWF), a Dutch Horizon Breakthrough grant, the ESF EUROCORES Programme EuroDYNA (ERAS-CT-2003-980409) and the Spanish Instituto de Salud Carlos III (FIS08/29 and RTICC-RD06/0020/017).

Author details
[1]Department of Cell Biology and Genetics, Erasmus MC, The Netherlands. [2]Department of Molecular Biology, Instituto de Biomedicina y Biotecnología de Cantabria, IBBTEC, Universidad de Cantabria-CSIC-IDICAN, Santander, Spain. [3]Institute for Genetics, Justus-Liebig-Universitaet Giessen, Heinrich-Buff-Ring 58-62, D-35392 Giessen, Germany. [4]Department of Reproduction and Development, Erasmus MC, The Netherlands. [5]Proteomics Center, Erasmus MC, The Netherlands. [6]Department of Epigenetics, Max-Planck Institute of Immunobiology, Freiburg, Germany. [7]Institute of Cell and Molecular Science, Centre for Gastroenterology, London, UK.

Authors' contributions
SvdN, MR-G, JL, FS and NG carried out specific experiments, participated in the design of the studies and helped to write the manuscript, HH, WS, IJ, JD, MvdR, VT and LJ carried out specific experiments, and participated in the design of the studies, FG, MDD and RR participated in the design of the studies, and helped to write the manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Received: 5 July 2010 Accepted: 8 November 2010
Published: 8 November 2010

References

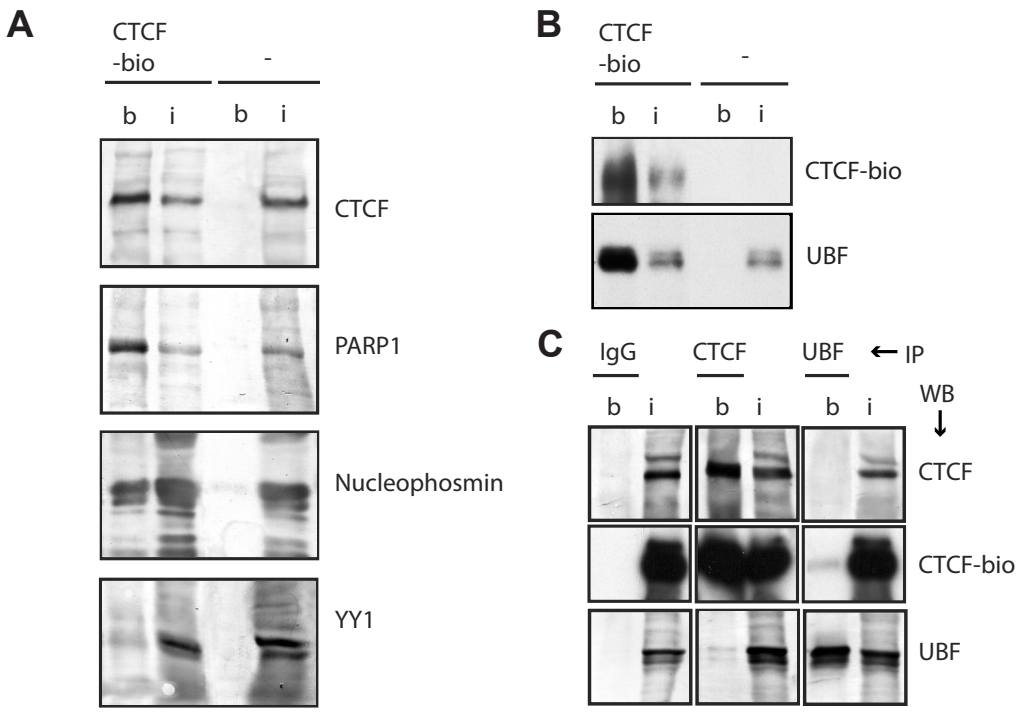
1. Phillips JE, Corces VG: **CTCF: master weaver of the genome.** *Cell* 2009, **137**:1194-1211.
2. Loukinov DI, Pugacheva E, Vatolin S, Pack SD, Moon H, Chernukhin I, Mannan P, Larsson E, Kanduri C, Vostrov AA, *et al*: **BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma.** *Proc Natl Acad Sci USA* 2002, **99**:6806-6811.
3. Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenkov VV, Ren B: **Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome.** *Cell* 2007, **128**:1231-1245.
4. Parelho V, Hadjur S, Spivakov M, Leleu M, Sauer S, Gregson HC, Jarmuz A, Canzonetta C, Webster Z, Nesterova T, *et al*: **Cohesins functionally associate with CTCF on mammalian chromosome arms.** *Cell* 2008.
5. Stedman W, Kang H, Lin S, Kissil JL, Bartolomei MS, Lieberman PM: **Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators.** *Embo J* 2008.
6. Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T, *et al*: **Cohesin mediates transcriptional insulation by CCCTC binding factor.** *Nature* 2008, **451**:796-801.
7. Hadjur S, Williams LM, Ryan NK, Cobb BS, Sexton T, Fraser P, Fisher AG, Merkenschlager M: **Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus.** *Nature* 2009, **460**:410-413.
8. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K: **High-resolution profiling of histone methylations in the human genome.** *Cell* 2007, **129**:823-837.
9. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE: **High-resolution mapping and characterization of open chromatin across the genome.** *Cell* 2008, **132**:311-322.
10. Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, Galjart N, de Laat W: **CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus.** *Genes Dev* 2006, **20**:2349-2354.
11. Ishihara K, Oshimura M, Nakao M: **CTCF-dependent chromatin insulator is linked to epigenetic remodeling.** *Mol Cell* 2006, **23**:733-742.
12. McStay B, Grummt I: **The epigenetics of rRNA genes: from molecular to chromosome biology.** *Annu Rev Cell Dev Biol* 2008, **24**:131-157.
13. Bell SP, Learned RM, Jantzen HM, Tjian R: **Functional cooperativity between transcription factors UBF1 and SL1 mediates human ribosomal RNA synthesis.** *Science* 1988, **241**:1192-1197.
14. Jantzen HM, Admon A, Bell SP, Tjian R: **Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins.** *Nature* 1990, **344**:830-836.
15. O'Sullivan AC, Sullivan GJ, McStay B: **UBF binding *in vivo* is not restricted to regulatory sequences within the vertebrate ribosomal DNA repeat.** *Mol Cell Biol* 2002, **22**:657-668.
16. Stefanovsky V, Langlois F, Gagnon-Kugler T, Rothblum LI, Moss T: **Growth factor signaling regulates elongation of RNA polymerase I transcription in mammals via UBF phosphorylation and r-chromatin remodeling.** *Mol Cell* 2006, **21**:629-639.

17. Mais C, Wright JE, Prieto JL, Raggett SL, McStay B: **UBF-binding site arrays form pseudo-NORs and sequester the RNA polymerase I transcription machinery.** *Genes Dev* 2005, **19**:50-64.
18. Sanij E, Poortinga G, Sharkey K, Hung S, Holloway TP, Quin J, Robb E, Wong LH, Thomas WG, Stefanovsky V, *et al*: **UBF levels determine the number of active ribosomal RNA genes in mammals.** *J Cell Biol* 2008, **183**:1259-1274.
19. Rodriguez P, Bonte E, Krijgsveld J, Kolodziej KE, Guyot B, Heck AJ, Vyas P, de Boer E, Grosveld F, Strouboulis J: **GATA-1 forms distinct activating and repressive complexes in erythroid cells.** *Embo J* 2005, **24**:2354-2366.
20. Driegen S, Ferreira R, van Zon A, Strouboulis J, Jaegle M, Grosveld F, Philipsen S, Meijer D: **A generic tool for biotinylation of tagged proteins in transgenic mice.** *Transgenic Res* 2005, **14**:477-482.
21. Hanada K, Song CZ, Yamamoto K, Yano K, Maeda Y, Yamaguchi K, Muramatsu M: **RNA polymerase I associated factor 53 binds to the nucleolar transcription factor UBF and functions in specific rDNA transcription.** *Embo J* 1996, **15**:2217-2226.
22. Burke LJ, Zhang R, Bartkuhn M, Tiwari VK, Tavoosidana G, Kurukuti S, Weth C, Leers J, Galjart N, Ohlsson R, Renkawitz R: **CTCF binding and higher order chromatin structure of the H19 locus are maintained in mitotic chromatin.** *Embo J* 2005, **24**:3291-3300.
23. Sanij E, Hannan RD: **The role of UBF in regulating the structure and dynamics of transcriptionally active rDNA chromatin.** *Epigenetics* 2009, **4**:374-382.
24. Mayer C, Schmitz KM, Li J, Grummt I, Santoro R: **Intergenic transcripts regulate the epigenetic state of rRNA genes.** *Mol Cell* 2006, **22**:351-361.
25. Nemeth A, Guibert S, Tiwari VK, Ohlsson R, Langst G: **Epigenetic regulation of TTF-I-mediated promoter-terminator interactions of rRNA genes.** *Embo J* 2008.
26. Santoro R, Schmitz KM, Sandoval J, Grummt I: **Intergenic transcripts originating from a subclass of ribosomal DNA repeats silence ribosomal RNA genes in *trans*.** *EMBO Rep* 2010, **11**:52-58.
27. Tower J, Henderson SL, Dougherty KM, Wejksnora PJ, Sollner-Webb B: **An RNA polymerase I promoter located in the CHO and mouse ribosomal DNA spacers: functional analysis and factor and sequence requirements.** *Mol Cell Biol* 1989, **9**:1513-1525.
28. Grummt I: **Different epigenetic layers engage in complex crosstalk to define the epigenetic state of mammalian rRNA genes.** *Hum Mol Genet* 2007, **16(Spec No 1)**:R21-27.
29. Li J, Langst G, Grummt I: **NoRC-dependent nucleosome positioning silences rRNA genes.** *Embo J* 2006, **25**:5735-5741.
30. Santoro R, Li J, Grummt I: **The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription.** *Nat Genet* 2002, **32**:393-396.
31. Heath H, de Almeida CR, Sleutels F, Dingjan G, van de Nobelen S, Jonkers I, Ling KW, Gribnau J, Renkawitz R, Grosveld F, *et al*: **CTCF regulates cell cycle progression of alphabeta T cells in the thymus.** *Embo J* 2008, **27**:2839-2850.
32. Akhmanova A, Verkerk T, Langeveld A, Grosveld F, Galjart N: **Characterisation of transcriptionally active and inactive chromatin domains in neurons.** *J Cell Sci* 2000, **113(Pt 24)**:4463-4474.
33. Vatolin S, Abdullaev Z, Pack SD, Flanagan PT, Custer M, Loukinov DI, Pugacheva E, Hong JA, Morse H, Schrump DS, *et al*: **Conditional expression of the CTCF-paralogous transcriptional factor BORIS in normal cells results in demethylation and derepression of MAGE-A1 and reactivation of other cancer-testis genes.** *Cancer Res* 2005, **65**:7751-7762.
34. Torrano V, Navascues J, Docquier F, Zhang R, Burke LJ, Chernukhin I, Farrar D, Leon J, Berciano MT, Renkawitz R, *et al*: **Targeting of CTCF to the nucleolus inhibits nucleolar transcription through a poly(ADP-ribosyl)ation-dependent mechanism.** *J Cell Sci* 2006, **119**:1746-1759.
35. Bell AC, West AG, Felsenfeld G: **The protein CTCF is required for the enhancer blocking activity of vertebrate insulators.** *Cell* 1999, **98**:387-396.
36. Grandori C, Gomez-Roman N, Felton-Edkins ZA, Ngouenet C, Galloway DA, Eisenman RN, White RJ: **c-Myc binds to human ribosomal DNA and stimulates transcription of rRNA genes by RNA polymerase I.** *Nat Cell Biol* 2005, **7**:311-318.
37. Jin C, Zang C, Wei G, Cui K, Peng W, Zhao K, Felsenfeld G: **H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions.** *Nat Genet* 2009, **41**:941-945.
38. Chen D, Huang S: **Nucleolar components involved in ribosome biogenesis cycle between the nucleolus and nucleoplasm in interphase cells.** *J Cell Biol* 2001, **153**:169-176.
39. Chen D, Belmont AS, Huang S: **Upstream binding factor association induces large-scale chromatin decondensation.** *Proc Natl Acad Sci USA* 2004, **101**:15106-15111.
40. Stefanovsky VY, Pelletier G, Bazett-Jones DP, Crane-Robinson C, Moss T: **DNA looping in the RNA polymerase I enhancesome is the result of non-cooperative in-phase bending by two UBF molecules.** *Nucleic Acids Res* 2001, **29**:3241-3247.
41. Paalman MH, Henderson SL, Sollner-Webb B: **Stimulation of the mouse rRNA gene promoter by a distal spacer promoter.** *Mol Cell Biol* 1995, **15**:4648-4656.
42. Yusufzai TM, Tagami H, Nakatani Y, Felsenfeld G: **CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species.** *Mol Cell* 2004, **13**:291-298.
43. Amin MA, Matsunaga S, Uchiyama S, Fukui K: **Depletion of nucleophosmin leads to distortion of nucleolar and nuclear structures in HeLa cells.** *Biochem J* 2008, **415**:345-351.
44. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, van Steensel B: **Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions.** *Nature* 2008, **453**:948-951.
45. Martin C, Chen S, Maya-Mendoza A, Lovric J, Sims PF, Jackson DA: **Lamin B1 maintains the functional plasticity of nucleoli.** *J Cell Sci* 2009, **122**:1551-1562.
46. Hoogstraten D, Nigg AL, Heath H, Mullenders LH, van Driel R, Hoeijmakers JH, Vermeulen W, Houtsmuller AB: **Rapid switching of TFIIH between RNA polymerase I and II transcription and DNA repair *in vivo*.** *Mol Cell* 2002, **10**:1163-1174.
47. Lutz M, Burke LJ, Barreto G, Goeman F, Greb H, Arnold R, Schultheiss H, Brehm A, Kouzarides T, Lobanenkov V, Renkawitz R: **Transcriptional repression by the insulator protein CTCF involves histone deacetylases.** *Nucleic Acids Res* 2000, **28**:1707-1713.
48. Wilm M, Shevchenko A, Houthaeve T, Breit S, Schweigerer L, Fotsis T, Mann M: **Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry.** *Nature* 1996, **379**:466-469.
49. Lansbergen G, Grigoriev I, Mimori-Kiyosue Y, Ohtsuka T, Higa S, Kitajima I, Demmers J, Galjart N, Houtsmuller AB, Grosveld F, Akhmanova A: **CLASPs attach microtubule plus ends to the cell cortex through a complex with LL5beta.** *Dev Cell* 2006, **11**:21-32.
50. Andrews NC, Faller DV: **A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells.** *Nucleic Acids Res* 1991, **19**:2499.
51. Akhmanova A, Mausset-Bonnefont AL, van Cappellen W, Keijzer N, Hoogenraad CC, Stepanova T, Drabek K, van der Wees J, Mommaas M, Onderwater J, *et al*: **The microtubule plus-end-tracking protein CLIP-170 associates with the spermatid manchette and is essential for spermatogenesis.** *Genes Dev* 2005, **19**:2501-2515.
52. Torrano V, Chernukhin I, Docquier F, D'Arcy V, Leon J, Klenova E, Delgado MD: **CTCF regulates growth and erythroid differentiation of human myeloid leukemia cells.** *J Biol Chem* 2005, **280**:28152-28161.
53. Gribnau J, Luikenhuis S, Hochedlinger K, Monkhorst K, Jaenisch R: **X chromosome choice occurs independently of asynchronous replication timing.** *J Cell Biol* 2005, **168**:365-373.
54. Donohoe ME, Zhang LF, Xu N, Shi Y, Lee JT: **Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch.** *Mol Cell* 2007, **25**:43-56.
55. Payen E, Verkerk T, Michalovich D, Dreyer SD, Winterpacht A, Lee B, De Zeeuw CI, Grosveld F, Galjart N: **The centromeric/nucleolar chromatin protein ZFP-37 may function to specify neuronal nuclear domains.** *J Biol Chem* 1998, **273**:9099-9109.
56. Netchvolodov KK, Boiko AV, Ryskov AP, Kupriyanova NS: **Evolutionary divergence of the pre-promotor region of ribosomal DNA in the great apes.** *DNA Seq* 2006, **17**:378-391.
57. Larsen F, Gundersen G, Lopez R, Prydz H: **CpG islands as gene markers in the human genome.** *Genomics* 1992, **13**:1095-1107.
58. Arnold R, Burcin M, Kaiser B, Muller M, Renkawitz R: **DNA bending by the silencer protein NeP1 is modulated by TR and RXR.** *Nucleic Acids Res* 1996, **24**:2640-2647.

59. Ohlsson R, Renkawitz R, Lobanenkov V: **CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease.** *Trends Genet* 2001, **17**:520-527.
60. Gombert WM, Farris SD, Rubio ED, Morey-Rosler KM, Schubach WH, Krumm A: **The c-myc insulator element and matrix attachment regions define the c-myc chromosomal domain.** *Mol Cell Biol* 2003, **23**:9338-9348.
61. Hong JA, Kang Y, Abdullaev Z, Flanagan PT, Pack SD, Fischette MR, Adnani MT, Loukinov DI, Vatolin S, Risinger JI, *et al*: **Reciprocal binding of CTCF and BORIS to the NY-ESO-1 promoter coincides with derepression of this cancer-testis gene in lung cancer cells.** *Cancer Res* 2005, **65**:7763-7774.

doi:10.1186/1756-8935-3-19
Cite this article as: van de Nobelen *et al.*: CTCF regulates the local epigenetic state of ribosomal DNA repeats. *Epigenetics & Chromatin* 2010 **3**:19.

Supplemental Figure S1

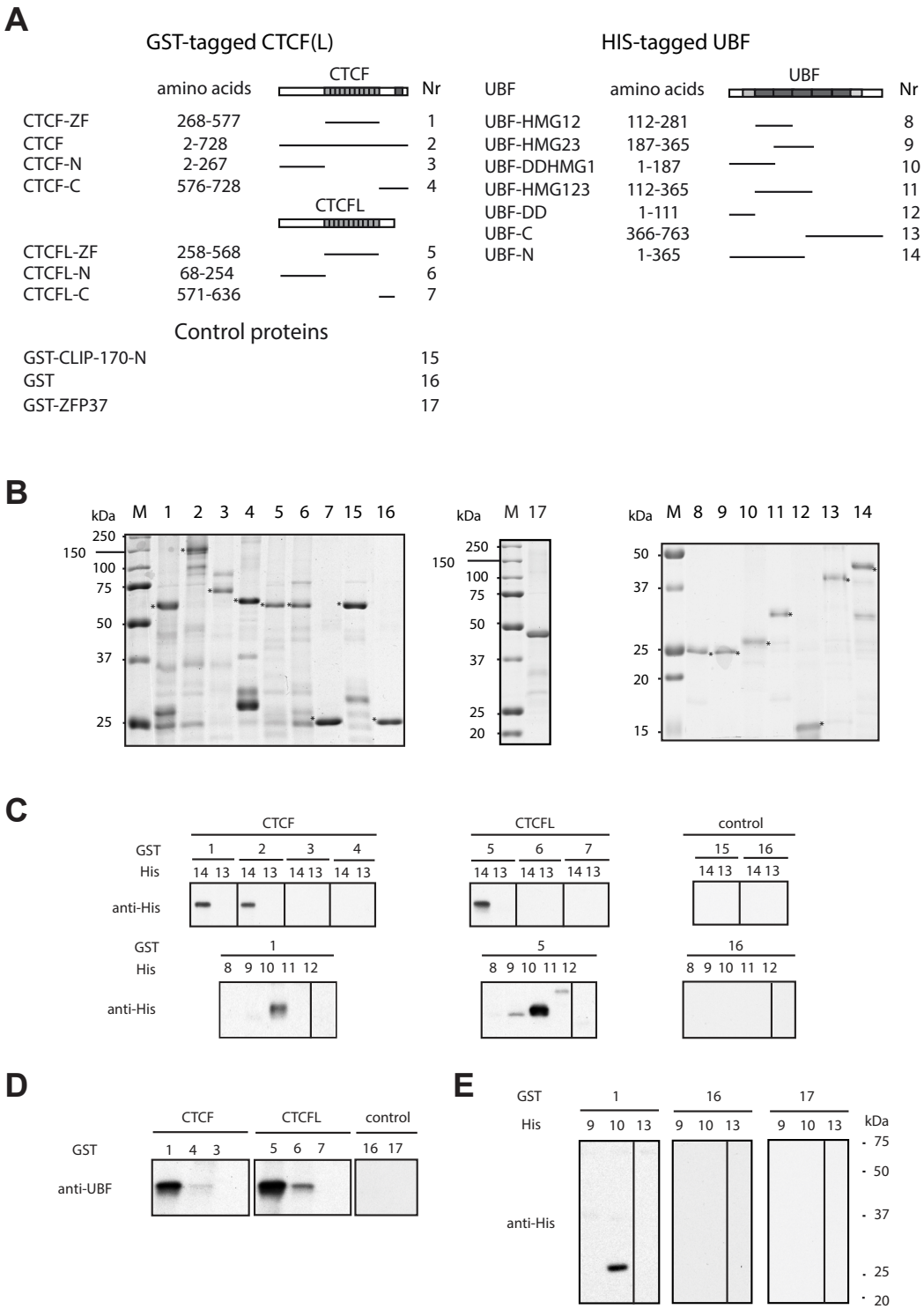


Supplemental Table S1.
Mass spectrometry results for CTCF-bio

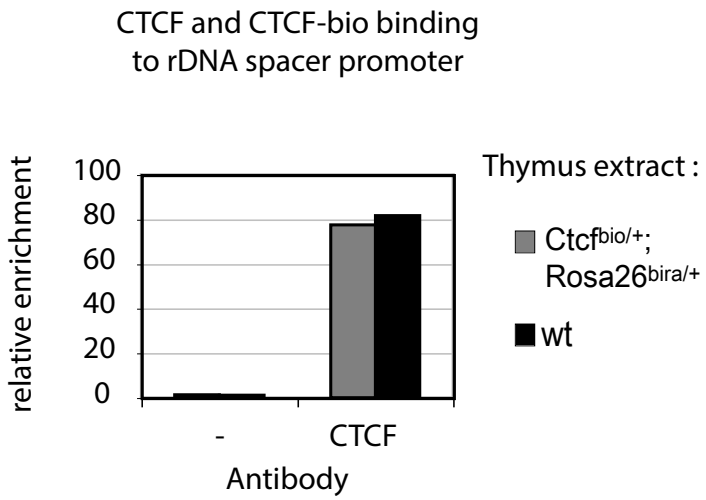
Identified Protein	Molecular w eight (kDa)	Acc. Number	Mascot Score*	Identified Peptides*
UBF	97	gi 136653	110, 85	2, 2
RNA pol I (RPA116)	116	gi 38614338	38	1
RNA pol I (RPA194)	194	gi 2330007	210, 294	6, 8
RNA pol I (RPA40)	40	gi 120538451	124	1
PAF49	49	gi 38602694	165, 213	4, 6
PAF53	53	gi 12328816	115, 108	2, 2
WDR5	36	gi 14250247	47	1
CTCF	80	gi 6681073	69, 102	2, 2

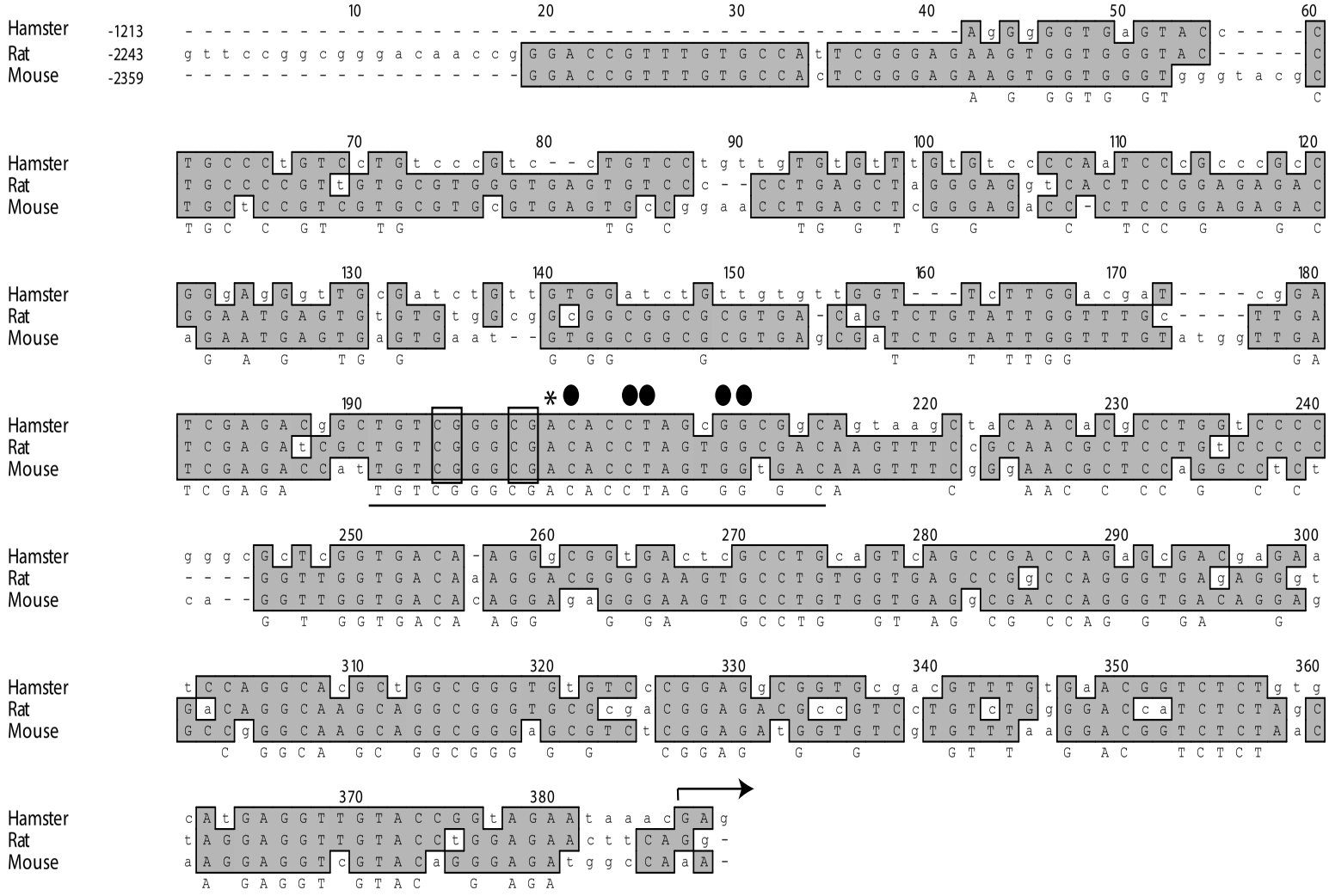
*When two numbers are listed the data are from two independent mass spectrometry experiments. Proteins listed above were not detected in the control samples (cells expressing BirA only)

Supplemental Figure S2

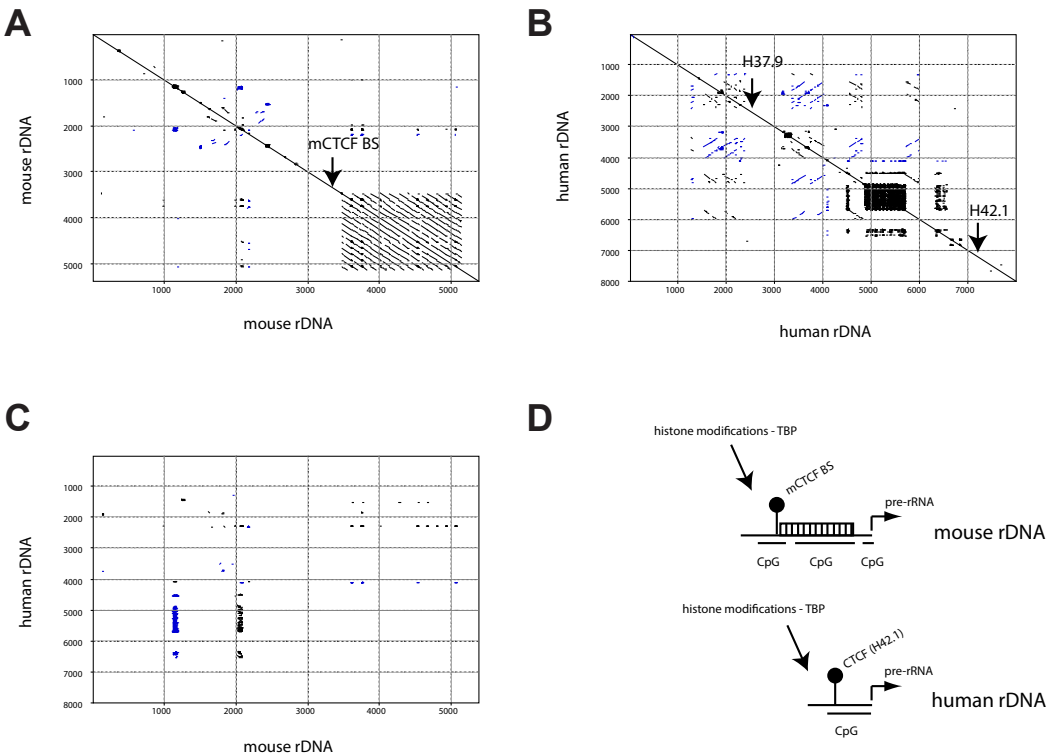


Supplemental Figure S3

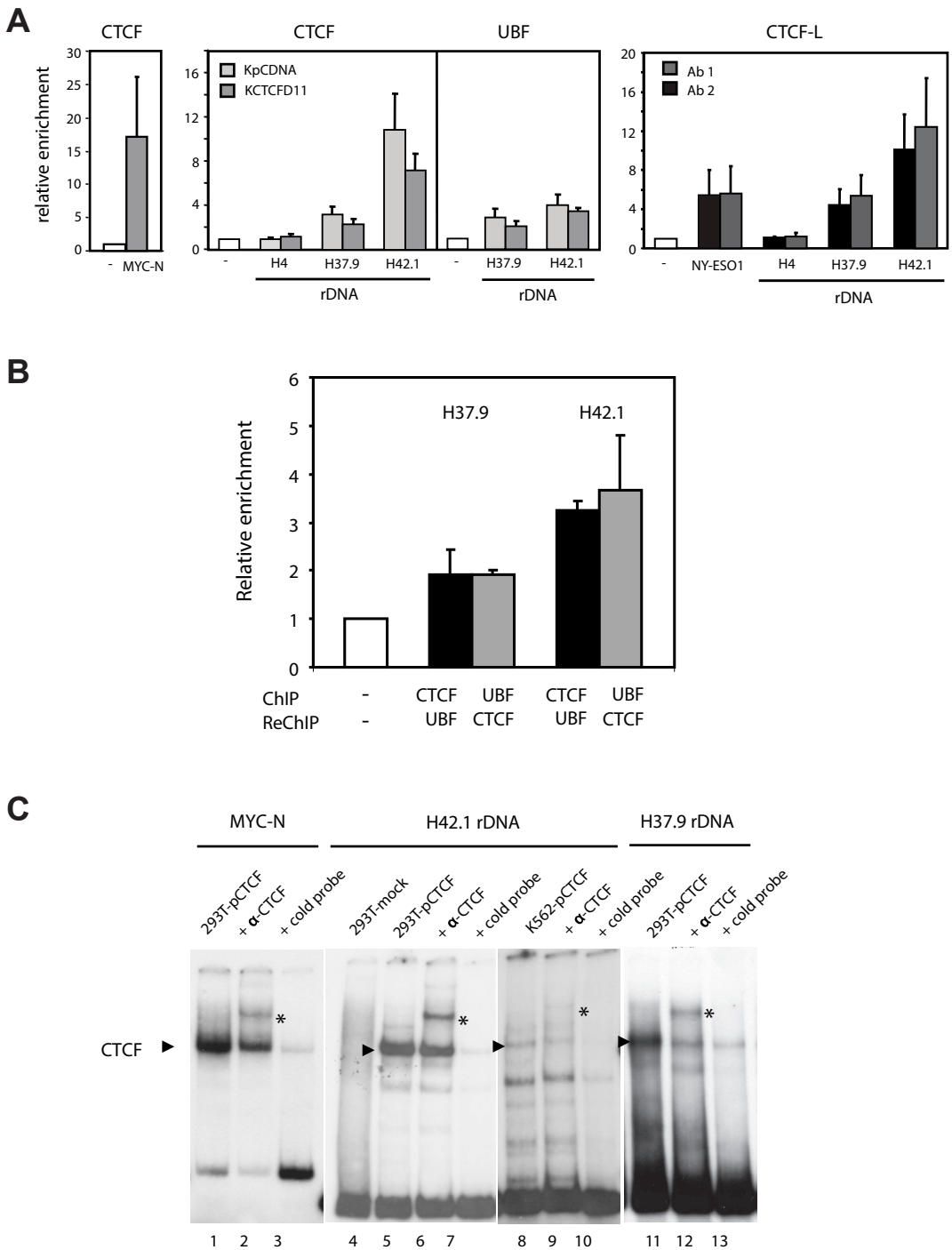



Hamster
Rat
Mouse
-1213
-2243
-2359
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
220
230
240
250
260
270
280
290
300
310
320
330
340
350
360
370
380

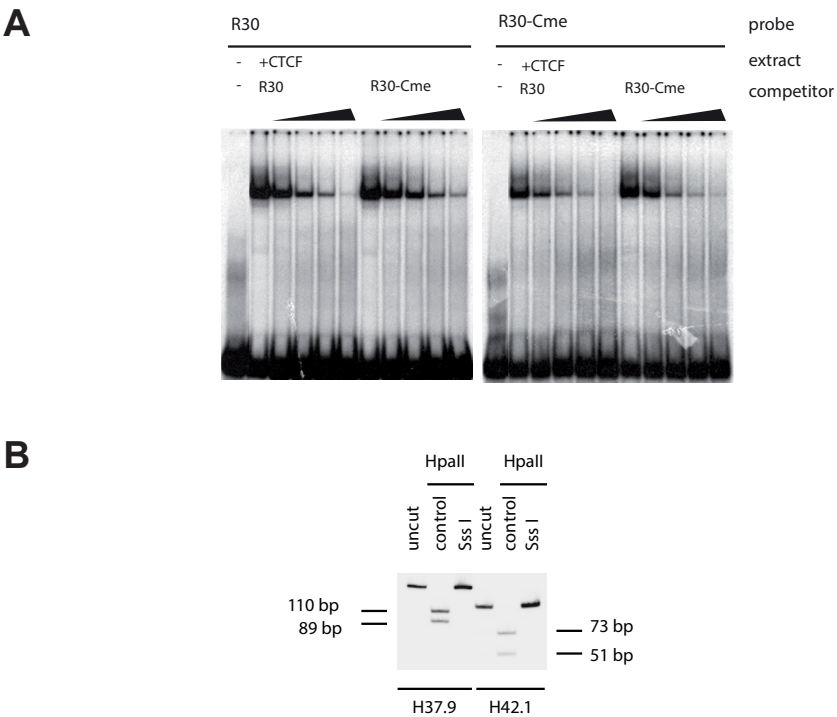
Supplemental Figure S5



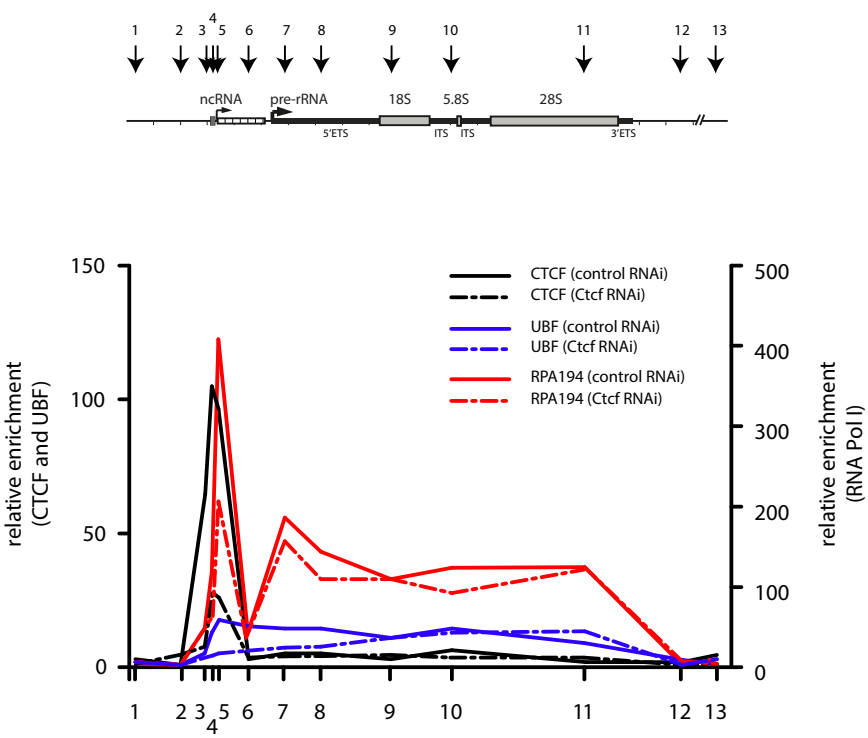
Supplemental Figure S6



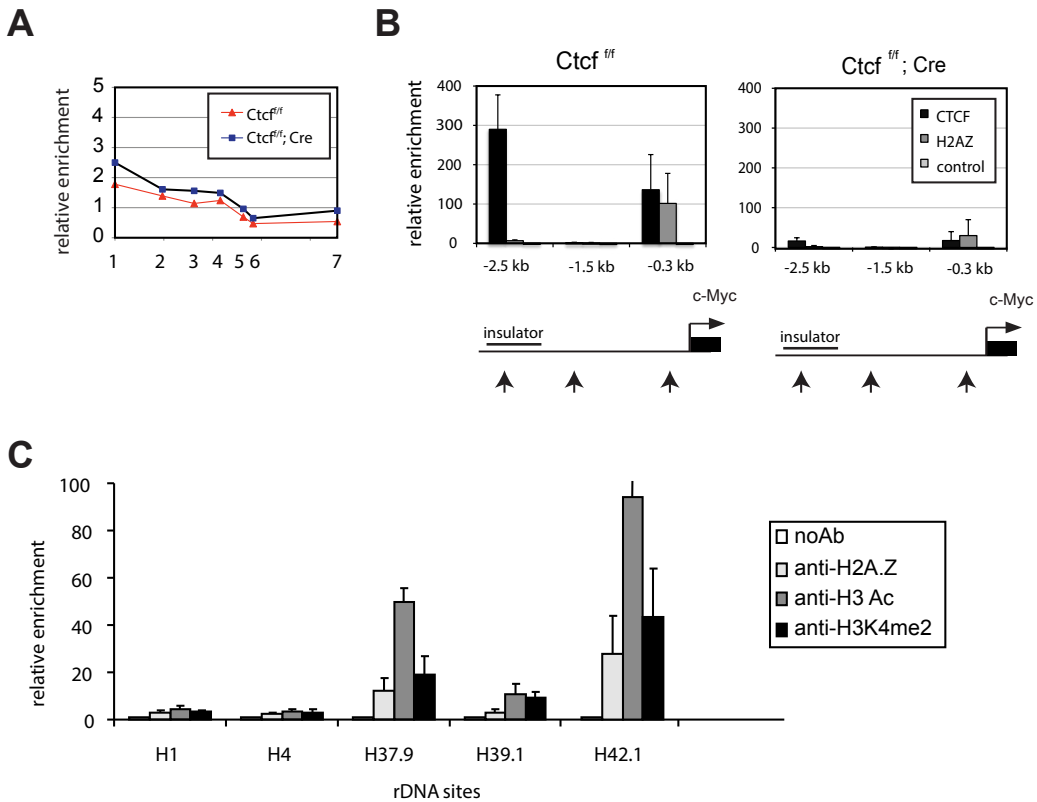
Supplemental Figure S7



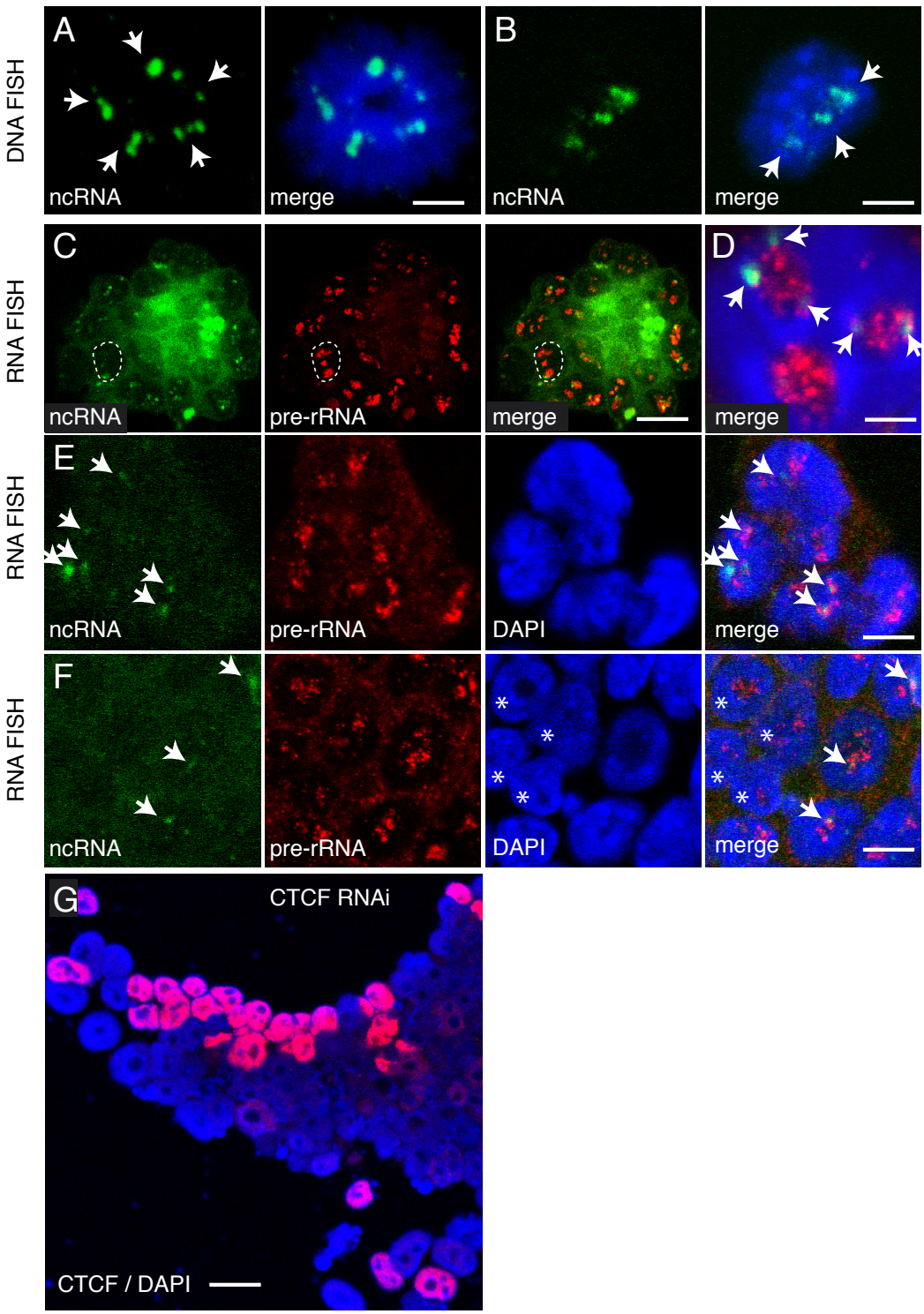
Supplemental Figure S8



Supplemental Figure S9



Supplemental Figure S10



Supplemental Table S2. Primers used for band-shifts

name	sequence (5' to 3')
R30s	TGTATGGTTGATCGAGACCATTGTCGGGCGACACCTAG TGGTGACAAGTTTCGGGAACGCTCCAGGCCTCT
R30as	AGAGGCCTGGAGCGTTCCCGAAACTTGTCACCACTAGG TGTCGCCCGACAATGGTCTCGATCAACCATACA
R30mut-s	TGTATGGTTGATCGAGACCATTGTCGGGCAATACCTAG TAGTGACAAGTTTCGGGAACGCTCCAGGCCTCT-3'
R30metss[1)]	TGTATGGTTGATCGAGACCATTGT[5Me-dC]GGG[5Me-C]GACACCTAGTGGTGACAAGTTTCGGGAACGCTCCAG GCCTCT
R30 metas[1)]	AGAGGCCTGGAGCGTTCCCGAAACTTGTCACCACTAGG TGT[5Me-dC]GCC[5Me-dC]GACAATGGTCTCGATCAACCATACA
F1 _F[2)]	CTAGATGAAGAAATTGAGACCTCTACTGGATAGCTATG GTATTTACGTGTCTA
F1_B[2)]	AGCTTAGACACGTAAATACCATAGCTATCCAGTAGAGG TCTCAATTTCTTCAT

1) Ordered at Operon Biotechnologied, Germany
2) From the chicken lysozyme gene [58]

Supplemental Table S3. Primers used for genotyping

name	sequence (5' to 3')
Probe 1F	TCCTGCCTCTGTCCAGTCAGAGA
Probe 1B	GCAGATCACTGTGTGTTCAAGGC
Probe 2F	CGAATGCCACCTTTGACTCTACC
Probe 2B	AAGCCTCGTCCTTCCGAGCCT
Rosa26 F (265)	GTGTAACTGTGGACAGAGGAG
Rosa26 F (266)	GAACTTGATGTGTAGACCAGG
BirA_F (91)	TTCAGACACTGCGTGACT
BirA_B (92)	GGCTCCAATGACTATTTGC
CTCFGB1	AGCAAAAGCAAAACCAGGTTA
CTCFGF14	AGGAGCCAGATGCCGAGCCTG

Genotyping yields fragments of 350 bp (Rosa26), 514 bp (BirA). 599 bp (*Ctcf*^{bio}), and 549 bp (wild type allele).

Supplemental Table S4. Primers used for mouse ChIP

name	sequence (5' to 3')
Enh4_F (APS1) -4736	GTCACCATTCTGCACTTGCAA
Enh4_B (APS1) -4584	ACATGTGCATGGCAGCCATCTTG
Enh5_F -3736	GTGTGTTTGTGCTCTATCTGCTG
Enh5_B -3641	CACTTATTCTCAGGAGCTGCATG
Enh6_F -3088	GTGAGTTCCAGGACTTACCAGAG
Enh6_B -2988	CTGTGTAGCCCTATCGGACTTG
Enh3_F -2561	CACTGCTTAGATGCTCCCTTCC
Enh3_B -2446	ATCGTTCTTGAAGTCAAAGTACGTC
Enh2_F (spacer prom.) -2087	AGGAGGCCGGGCAAGCA
Enh2_B (spacer prom.) -1975	CGTACAGCAACTCGGTCTGCT
Enh_F (enhancer repeat) -1882	CCTCCAGAAGCCCTCTCTTGTCCC
Enh_B (enhancer repeat) – 1779	CAGCTGGCCGAGCCACACCGG
Prom_UCE_F -162	AGTTGTTCCTTTGAGGTCCGGT
Prom_UCE_B -52	GAGACAGGGAGGAAAGTGACAG
ChIP1s	ACCTCACTATGACCGGCTGAGATTC
ChIP1a	CCACCCGTAATGAGATCTGATGTCC
ChIP2s	ATGTGCCACCACTGCCCGGACTGA
ChIP2a	CACCTTTAACCTTAGGCAAATTTTA
ChIP3s	CACTGCTTAGATGCTCCCTTCC
ChIP3a	CACTGCTTAGATGCTCCCTTCC
ChIP4s	TGTTCGGGCGGGACGATG
ChIP4a	AGGTGTCGCCCGACAATG
ChIP5s	TGACAGGAGGCCGGGCAAGCA
ChIP5a	GCGTACAGCAACTCGGTCTGCT
ChIP7s	GTCACCCGGGGCGCTTGTACTTCTGAT
ChIP7a	TCGTGTCCTCTAGGCCTCAGATGTAA
ChIP9s	CCAAGTGTTCATGCCACGTG
ChIP10s	GCGCAGCGTTTGCTCTCT
ChIP10a	CACACAAGCCGAGCCACAT

name	sequence (5' to 3')
ChIP11s	GCTTGTCTCAAAGATTAAGCCATGC
ChIP11a	TATTAGCTCTAGAATTACCACAGTT
ChIP12s	CCGGCTTGCCCGATTTCCGCGGGT
ChIP12a	GCCAGCAGGAACGAAACG
ChIP13s	GTAACTATGACTCTCTTAAGGTAGCCA
ChIP13a	CTTCACCGTGCCAGACTAGAG
ChIP14s	TGGTTGCTGGGATTTGAACTC
ChIP14a	CAGAGAAATACTGTCTCAGAAG
ChIP15s	ACTTGCAAACCGGGCCACTAAA
ChIP15a	TTCCTTGTTCTGTCACTCGGTTGC
amylase 99	CTCCTTGTACGGGTTGGT
amylase 100	AATGATGTGCACAGCTGAA
Myc_ChIP_1F (cMyc 1,5kb downstr prom.)	GCTCCTAAACCAGAGTCTGCTG
Myc_ChIP_1B (cMyc 1,5kb downstr prom.)	CATACACCTCCACACAGTTCCAG
Myc_ChIP_2F (cMyc promoter)	TGACTCGCTGTAGTAATTCCAGC
Myc_ChIP_2B (cMyc promoter)	TCTCACTCCAGAGCTGCCTTC
Myc_5'INS_F (cMyc insulator)	CAGAACCTGGAAACCCTGCAG
Myc_5'INS_B (cMyc insulator)	GTTGTGGCTCTCGGATTTGTG
3'HS1_529_F (3'HS1)	AATCAGTGGAACACTTCTGC
3'HS1_530_B (3'HS1)	GTCTCAGGTTGTCAACTAAAGC

Supplemental Table S5. Primers used for northern blot and nuclear run-on

name	sequence (5' to 3')
Northern blot	
5'ETS_F	GTTCCTATTGGACCTGGAGA
5'ETS_B	CGGTTGGAATGGTGGAGCCA
GAPDH_F	TGAACGGGAAGCTCACTGG
GAPDH_B	TCCACCACCCTGTTGCTGTA
Run-on : promoter probe (290 bp; 129bp overlap transcript)	
rDNAprom_F -161	GTTGTCAGGGTCGACCAGTTGT
IGS_F_B +129	GACAGCTTCAGGCACCGC
Run-on: spacer promoter probe (361bp; 219bp overlap transcript)	
IGSB_F -2140	CAGGTTGGTGACACAGGAGAG
Enh_B - 1779	CAGCTGGCCGAGCCACACCGG
Run-on: actin and histone H3 probes	
Actin-exon5-F	ATCATGTTTGACACCTTCAACACC
Actin-exon5-B	GAACCGCTCGTTGCCAATAGT
Histone H3i-F	AAGCAGCTGGCCACCAAG
Histone H3i-B	CTCCTGCAGAGCCATGACG

Supplemental Table S6. Primers used for real-time PCR on ES cell rRNA

name	sequence (5' to 3')
ncrRNA2s (ChIP7s)	GTCACCCGGGGCGCTTGTACTTCTGAT
ncrRNA2a (ChIP7a)	TCGTGTCCTCTAGGCCTCAGATGTAA
ncrRNA1s (IGSB_F -2140)	CAGGTTGGTGACACAGGAGAG
ncrRNA1a (Enh_B - 1779)	CAGCTGGCCGAGCCACACCGG
45S forward	GACACGCTGTCCTTTCCCTA
45S reverse	AGGCTGGACAAGCAAAACAG

Supplemental Table S7. Primers used for human ChIP and band-shifts

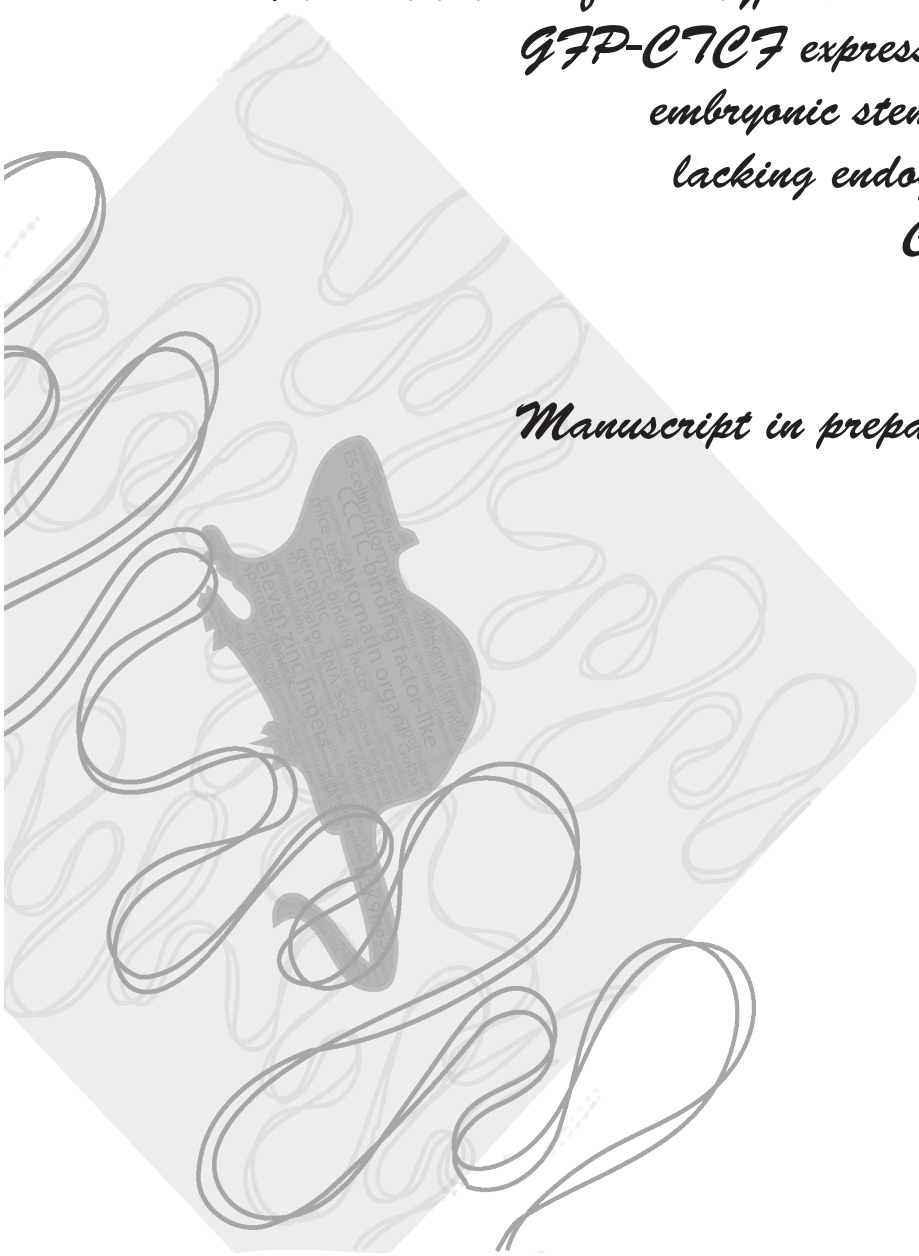
Name	Sequence	Human rDNA coordinate site
MYC-N_F[1)]	ACAAGGAGGTGGCTGGAAAC	
MYC-N_B	TTCCCCTCCTGGCTTTTAGT	
MYC-H.1_F[2)]	CAACGCAACACAGGATATGG	
MYC_H.1_B	TTCCCCTCCTGGCTTTTAGT	
NY-ESO1_F[3)]	ACCCGCAACCCACCCCACAC	
NY-ESO1_B	GGGGCAGGCCTCTAACTGGG	
H42.1 rDNA_F[4)]	GCTTCTCGACTCACGGTTTC	42012-42031
H42.1 rDNA_B	CCGAGAGCACGATCTCAAA	42117-42135
H37.9 rDNA_F[4)]	CCCTGGTCGATTAGTTGTGG	37818-37837
H37.9 rDNA_B	GTGCTCCCTTCCTCTGTGAG	37997-38016
H4 rDNA_F[5)]	CGACGACCCATTCGAACGTCT	3990-4010
H4 rDNA B	CTCTCCGGAATCGAACCCTGA	4072-4092

1) See ref [59]
2) See ref [60]
3) See ref [61]
4) For the position of the CTCF binding sites H37.9 and H42.1 in the IGS of the human rDNA repeat, see Figure S5 (additional File 6).
5) See ref [15]

Chapter 3

Reconstitution of wild type and mutant GFP-CTCF expression in embryonic stem cells lacking endogenous CTCF

Manuscript in preparation



Reconstitution of wild type and mutant GFP-CTCF expression in embryonic stem cells lacking endogenous CTCF

Widia Soochit[1]*, Frank Sleutels[1]*, Marek Bartkuhn[2]*, Michael van der Reijden[1], Wilfred van IJcken[3], Frank Grosveld[1,4], Rainer Renkawitz[2], and Niels Galjart[1,4,5]

*: These authors contributed equally to this work
1: Department of Cell Biology Erasmus Medical Center, Rotterdam, The Netherlands.
2: Institut für Genetik, Justus-Liebig-Universität, Heinrich-Buff-Ring 58-62, 35392, Giessen, Germany
3: Center for Biomics Erasmus Medical Center, Rotterdam, The Netherlands
4: Epigenetics Consortium, NGI, Erasmus Medical Center, Rotterdam, The Netherlands
5: Corresponding author: n.galjart@erasmusmc.nl

Abstract

The multifunctional and highly conserved chromatin organizer CTCF binds 25,000-50,000 sites in mouse and human genomes using its eleven zinc finger (ZF) domain. To determine how the different ZFs of CTCF contribute to binding specificity and how this relates to cellular function, we replaced the endogenous CTCF gene in mouse embryonic stem (ES) cells with wild type GFP-CTCF and with mutants in which individual ZF domains were deleted. We find that ZF1, and ZF8-11 of CTCF are not required for ES cell viability. Fluorescence-based microscopy studies reveal that more than a million immobile GFP-CTCF molecules are present in an ES cell nucleus. Compared to wild type protein GFP-CTCF-Δ1, and 8 are relatively mobile. Interestingly, ES cells expressing GFP-CTCF-Δ1 and -Δ8 grow less fast, revealing a link between defective dynamic behavior of mutant CTCF proteins and aberrant cell growth. Based on ChIP-Sequencing analysis of GFP-CTCF- and GFP-CTCF-ZF-mutant-expressing ES cells we propose that ZF1-3 bind nine contiguous nucleotides immediately downstream of the CTCF core motif bound by ZF4-7, and that ZF8-11 are required for binding a spacer sequence and upstream motif. CTCF binding sites containing the upstream motif are specifically depleted from transcription start sites and exons, and are associated with the repressive chromatin mark H3K9me3. These sites are less well bound by GFP-CTCF-Δ8-11. Genes in the vicinity of affected CTCF binding sites can show altered expression. Combined our data suggest that CTCF organizes chromatin at two distinct levels: 1) the protein binds DNA throughout the genome to regulate nuclear processes essential for ES cell growth and proliferation. 2) CTCF also binds conserved CTCF target sites near genes to regulate transcription locally, but its position to the TSS is independent on effects on transcription. Additionally, CTCF does not act as a classic transcription factor, but rather makes chromatin near genes (in)accessible i.e. by mediating long-range interactions.

Introduction

Eukaryotic genomes are tightly organized in order to maintain proper transcriptional regulation in a cell type-specific and developmental fashion *(Misteli, 2007)*. Chromatin is arranged into dynamic higher-order structures, which can localize to various subnuclear compartments that affect genome configuration and consequently transcription *(Brown et al., 2008; Capelson et al., 2010; Guelen et al., 2008; Nemeth et al., 2010; Peric-Hupkes et al., 2010; van Koningsbruggen et al., 2010)*. Regulatory regions are marked by specific chromatin features and combined through long-range interaction to influence transcription. Long-range interactions can also separate domains to establish independent transcription regulation in each individual domain *(Dixon et al., 2012; Handoko et al., 2011; Sanyal et al., 2012)*.

One of the key players in chromatin structure and spatial organization is CTCF (for review, see *(Phillips and Corces, 2009)*). CTCF is characterized by an eleven zinc finger (ZF) domain that is surrounded by N- and C-terminal regions, which do not contain any conspicuous motif, except for an AT-hook motif *(Ohlsson et al., 2001)*. CTCF is a highly conserved ubiquitously expressed nuclear protein *(Heger et al., 2012; Heger et al., 2009; Moon et al., 2005; Pugacheva et al., 2006)*, which is essential for cell viability *(Fedoriw et al., 2004; Heath et al., 2008; Moore et al., 2012)*. CTCF has been identified in independent studies as a transcriptional repressor *(Baniahmad et al., 1990; Klenova et al., 1993; Lobanenkov et al., 1990)* and as well as transcriptional activator *(Quitschke et al., 1996; Vostrov and Quitschke, 1997)*. The protein has since been implicated in many other cellular processes, including the regulation of genomic imprinting to maintain mono-allelic gene expression. Additionally, CTCF is often located at transitions of distinct chromatin states where it appears to act as a boundary element that prevents spreading of distinct chromatin signatures across the genome *(Barski et al., 2007; Cuddapah et al., 2009)*. It also organizes chromatin by binding to the borders of topological domains and mediating long-range interactions within subdomains *(Dixon et al., 2012; Guo et al., 2011; Phillips-Cremins et al., 2013; Splinter et al., 2006; Tolhuis et al., 2002; Zuin J, Submitted)*.

CTCF mediates its function via protein-DNA, protein-protein, and protein-RNA interactions *(Ohlsson et al., 2001; Phillips and Corces, 2009; Sun et al., 2013)*. CTCF binds to DNA via its ZF domain, and genome-wide studies reveal that it can bind approximately 25,000-50,000 sites in human and mouse genomes *(Boyle et al., 2011; Kim et al., 2007; Ohlsson et al., 2001; Rhee and Pugh, 2011)*. Initial genome-wide study in IMR90 human fibroblasts using ChIP-on-chip revealed that 46% of the CTCF binding sites are located in intergenic regions, 20% within 2.5 kb of transcription start sites, 22% in introns and 12% in exons *(Kim et al., 2007)*. An independent study in human $CD4^+$ T cells using ChIP-Seq, which has been reanalyzed, identified a similar distribution in which CTCF binds in 45% in intergenic region, 7% in 5' UTR, 3% in exons, 29% in introns, 2% in 3'UTR, and 13% within 5 kb of the transcription start site *(Barski et al., 2007; Xie et al., 2007)*. About 80% of all CTCF binding sites harbor a 20 base pair (bp) motif that contains a subset of highly conserved nucleotides and are contacted by ZFs of CTCF *(Boyle et al., 2011; Chen et al., 2008; Kim et al., 2007; Nakahashi et al., 2013; Rhee and Pugh, 2011; Schmidt et al., 2012; Xie et al., 2007)*. Additionally, about 15-30% of the CTCF binding sites contain another motif of ~9 nucleotides (nt), which is upstream of the canonical CTCF motif and which is separated from the core motif by a spacer of 5-6 nt *(Boyle et al., 2011; Nakahashi et al., 2013; Rhee and Pugh, 2011; Schmidt et al., 2012)*. The upstream and core motif together are called either bipartite motif or upstream and core (UC) motif. Recently, a 10 bp motif downstream of the CTCF core motif was discovered in a small subset of all CTCF sites *(Nakahashi et al., 2013)*. This motif is not bound by CTCF itself, but by a factor that might compete with CTCF.

The binding orientation of CTCF to its motifs is 'inverted', with the N- and C-terminal ZFs binding to the 3' and 5'end of the motif, respectively *(Nakahashi et al., 2013; Renda et al., 2007)*. Initial *in vitro* experiments revealed that subsets of ZF were needed to establish DNA binding *(Burcin et al., 1997; Filippova et al., 1996; Quitschke et al., 2000)*. This is in line with *in vivo* data that indicate

that CTCF binds its different motifs using groups of adjacent ZFs *(Nakahashi et al., 2013)*. Both approaches identified ZFs 4-7 as the most essential ZFs to establish CTCF binding *(Nakahashi et al., 2013; Renda et al., 2007)*. The peripheral ZFs would contribute to binding affinity and are required for stabilizing CTCF binding to its target sites *(Quitschke et al., 2000; Renda et al., 2007)*.

It is not known how each ZF contributes to CTCF's function. To address this question we generated CTCF mutants with deletions of individual ZFs and expressed these in ES cells lacking endogenous CTCF. We find that mutants with individual deletions of ZF 1, 8, 9, 10, and 11 are able to substitute for wild type CTCF. Genome-wide binding profiles reveal the full CTCF binding motif and suggest that ZF1 binds to nucleotides downstream of the core consensus, whereas ZF8-11 are important for binding the 9 bp motif upstream. Our data also show that CTCF sites containing a bipartite motif are virtually absent from transcription start sites and more often localized to intronic and intergenic regions and are associated with repressive chromatin mark H3K9me3. We propose that CTCF regulates transcription locally, but this regulation is independent of the position of CTCF to the TSS. Additionally, CTCF does not act as a classic transcription factor, but rather makes chromatin near genes (in)accessible by mediating long-range interactions.

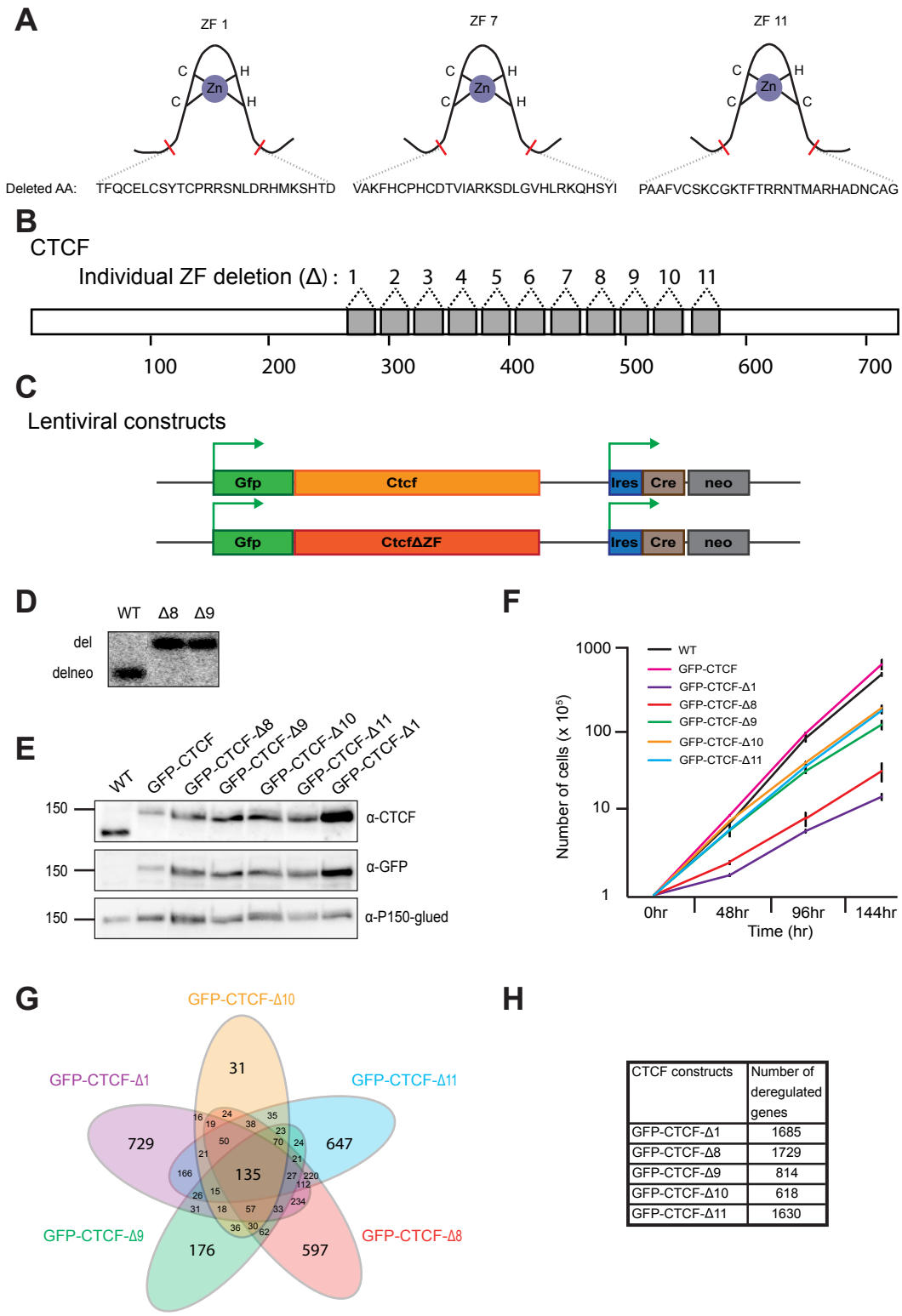
Results

Generation of ES cells expressing GFP-tagged CTCF mutants

In order to determine which ZFs are important for CTCF function and which specific nucleotides are bound by each ZF, we generated cDNAs encoding CTCF with mutations in individual CTCF ZFs **(Figure 1A, B and Table S1)**. To express these ZF mutants in a *Ctcf*-negative background, we isolated embryonic stem (ES) cells from *Ctcf* floxed mice ($Ctcf^{lox/lox}$, see *(Heath et al., 2008)*), and removed the neomycin resistance cassette at the 5' end of the $Ctcf^{lox}$ allele using Cre recombinase, thereby generating $Ctcf^{delneo/delneo}$ ES cells **(Figure S1A)**. This genotype will be further referred to as wild type (WT). We then substituted the endogenous *Ctcf* gene with GFP-tagged wild type CTCF (referred to as GFP-CTCF) or with CTCF mutants carrying individual ZF deletions (referred to as GFP-CTCF-ZF or GFP-CTCF-ΔZF), by infecting ES cells with lentiviral vectors expressing either wild type or mutant GFP-CTCF driven by a CAG promoter (CMV early enhancer/chicken beta actin) in combination with Ires-driven Cre recombinase **(Figure 1C**, *(Sleutels et al., 2012)***)**.

Deletion of *Ctcf* results in lethality, both in mice *(Heath et al., 2008)* and in ES cells *(Sleutels et al., 2012)*. Rescue of knock out ES cells from this fate requires exogenous expression of a functional CTCF (fusion) protein to support ES cell growth and proliferation. After lentiviral infection, clones that were both GFP-positive and neomycin-resistant were picked. Genotyping of each clone by southern blot analysis of Hind III-digested DNA **(Figure 1D and S1B)** revealed which clones had a deletion of both $Ctcf^{delneo}$ alleles, of only one allele or of neither of the endogenous alleles. One clone of each mutant ES cell line with a homozygous deletion was further analyzed by DNA-, RNA-Sequencing **(Figure S1C and S2)** and further assays. In addition to the individual ZF deletions, we found one (silent) mutation (G→A) at position 1476, counted from the A of the ATG of the CTCF cDNA sequence, in GFP-CTCF-Δ8 (data not shown). We did not obtain any ES cell clone expressing mutants with individual deletion of GFP-CTCF-Δ2-6 in a $Ctcf^{lox/lox}$ background, indicating that each of these ZFs is essential for the function of CTCF. We could, however, replace endogenous CTCF with wild type GFP-tagged CTCF (see also *(Sleutels et al., 2012)*), as well as with GFP-CTCF-Δ1, GFP-CTCF-Δ8, GFP-CTCF-Δ9, GFP-CTCF-Δ10 and GFP-CTCF-Δ11 **(Figure S1C)**. Thus, these 5 ZFs are not absolutely essential for ES cell viability.

Figure 1



CTCF constructs	Number of deregulated genes
GFP-CTCF-Δ1	1685
GFP-CTCF-Δ8	1729
GFP-CTCF-Δ9	814
GFP-CTCF-Δ10	618
GFP-CTCF-Δ11	1630

Figure 1. Generation and characterization of CTCF mutant ES cell lines
(A) Schematic representation of zinc fingers (ZF) 1, 7 and 11 and the amino acid sequence, which have been deleted. Red strips mark the points that are fused to each other after deletion. Amino acid sequences of all zinc fingers are shown in **Table S1**.
(B) Schematic representation of the CTCF protein structure and the individual ZF deletions (Δ).
(C) Lentiviral constructs used for ES cell rescue experiments. GFP is fused to the N-terminus of CTCF (GFP-CTCF) or CTCF with an individual ZF deletion (GFP-CTCF-ZF). Green arrows show where translation starts. Cre: cre recombinase, neo: neomycin.
(D) Southern blot analysis of *Ctcf*^{delneo} (WT) and *Ctcf*^{del} genotypes in GFP-CTCF-Δ8- and GFP-CTCF-Δ9-expressing ES cell lines.
(E) Western blot analysis of WT and rescued ES cell lines, expressing GFP-CTCF, GFP-CTCF-Δ8, GFP-CTCF-Δ9, GFP-CTCF-Δ10, GFP-CTCF-Δ11 and GFP-CTCF-Δ1. Blots were incubated with rabbit polyclonal CTCF (upper panel) and GFP (middle panel) antibodies to detect (GFP-tagged) CTCF, and with mouse p150^{glued} antibody (lower panel) as loading control. The 150 kDa band is represented as 150.
(F) Proliferation assay of WT and mutant ES cell lines. Cells were seeded at time 0hr with 100000 cells and were counted every 48 hours. Y-axis represents the number of cells (x 10^5), hr: hours. Data are presented ± SEM, N=3
(G-H) Venn diagram and table displaying the number of significantly deregulated genes in ES cells expressing the indicated ZF mutants. Venn diagram shows number of unique and overlapping deregulated genes in each mutant.

In order to examine the expression of GFP-CTCF-ZF mutants we performed a western blot analysis. Endogenous CTCF was only detected in WT ES cells, whereas in mutant cell lines expression of the respective GFP-CTCF(-ZF) fusion proteins was detected **(Figure 1E)**. Interestingly, mutants were expressed without major differences as endogenous CTCF, both at the protein level **(Figure 1E)**, as well as at the RNA level **(Figure S1D**, see also **Table S2)**, as determined by RNA-sequencing.

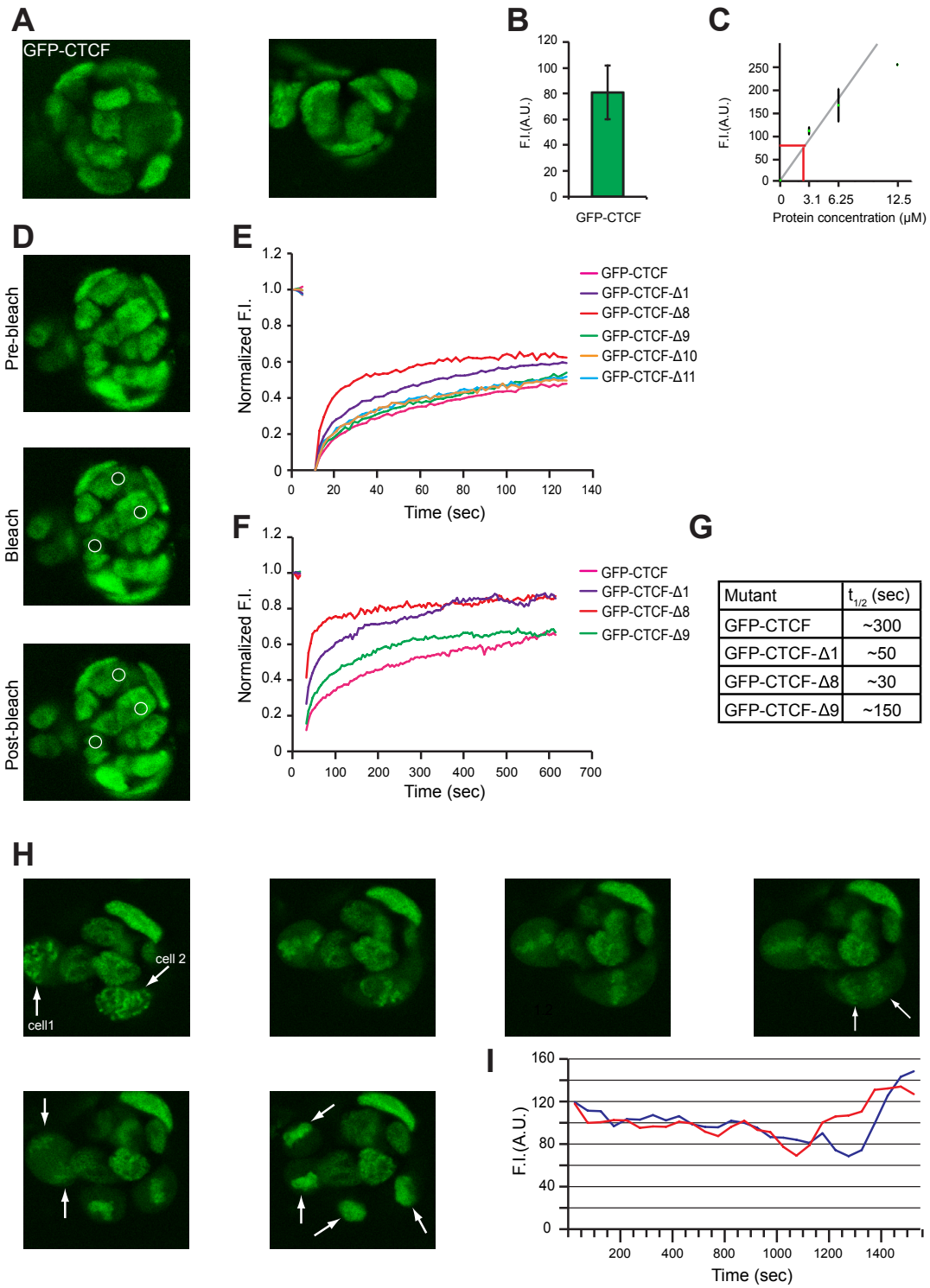
Rescued ES cells retained the expression of the pluripotency markers *Pou5f1*, *Nanog*, *Sox2* and *Alpl* (alkaline phosphatase) **(Table S2**, and data not shown**)**. We next examined their proliferation capacity. GFP-CTCF-expressing cells proliferated like WT ES cells **(Figure 1F)**, indicating that the addition of the GFP tag to the N-terminus of CTCF does not grossly affect the function of the protein. Strikingly, ES cells expressing mutant GFP-CTCF-ZF proteins proliferated less efficiently compared to WT and GFP-CTCF-expressing cells (**Figure 1F)**. Thus, while ZFs 1, and 8-11 are not essential for ES cell viability, they are required for optimal ES cell growth. Interestingly, GFP-CTCF-Δ9, GFP-CTCF-Δ10 and GFP-CTCF-Δ11 were only moderately affected in their growth capacity, while GFP-CTCF-Δ1 and GFP-CTCF-Δ8 were severely affected **(Figure 1F)**. These data suggest that individual ZFs have distinct contributions to CTCF functionality.

To link defects in ES cell growth to deregulated gene expression we performed RNA-Sequencing on each cell line. Genes with a significant change in expression ($p<0.05$) in at least one of the ZF mutant-expressing ES cells compared to GFP-CTCF were clustered and data was plotted in a Venn diagram **(Figure 1G)**. This analysis revealed 135 commonly deregulated genes and many more uniquely deregulated genes in ES cells expressing mutant CTCF proteins **(Figure 1G)**. Of all mutant ES cells, those expressing GFP-CTCF-Δ1 and GFP-CTCF-Δ8 had the highest number of deregulated genes, 1685 and 1729, respectively **(Figure 1H)**. Except for GFP-CTCF-Δ11, all other deletions showed a nice correlation between the number of affected genes and the inhibition of proliferation.

GFP-CTCF-ZF mutants with a high impact on gene regulation are highly mobile

To examine the properties and dynamic behavior of wild type and mutant GFP-CTCF in ES cells, we performed time-lapse imaging experiments, including fluorescence recovery after photobleaching (FRAP). Consistent with published studies in fixed cells using antibodies against CTCF *(Sleutels et al., 2012; van de Nobelen et al., 2010)*, GFP-CTCF was detected in a speckled pattern throughout the interphase nucleus of ES cells **(Figure 2A)**.

Figure 2



Mutant	$t_{1/2}$ (sec)
GFP-CTCF	~300
GFP-CTCF-Δ1	~50
GFP-CTCF-Δ8	~30
GFP-CTCF-Δ9	~150

Figure 2. High mobility in GFP-CTCF-Δ1 and -Δ8
(A) Images of GFP-CTCF-expressing ES cell clones showing that GFP-CTCF is located in the nucleus.
(B) Average fluorescence intensity (F.I.) of GFP-CTCF in ES cell nuclei. Data are presented ± SD, N=18 cells.
(C) Plot of the F.I. of purified GFP-EB3 versus its concentration. Data are presented ± SD, N=4. Red line represents F.I. and estimated concentration of GFP-CTCF.
(D) Images of GFP-CTCF-expressing ES cells used for FRAP experiment. Cells are shown before (Pre-Bleach, upper panel), immediately after (Bleach, middle panel) and at the end of the FRAP experiment (Post-Bleach, lower panel). White circles indicate bleached ROIs.
(E-G) FRAP analysis of GFP-CTCF- and GFP-CTCF-ZF-mutant-expressing ES cells. FRAP experiments were done for 2 minutes with image acquisition times of 2 seconds **(E)**, and for 10 minutes with image acquisition times of 5 seconds **(F)**. Fluorescence recoveries were corrected for background and normalized, and then again corrected for fluorescence intensity variations, both in non-bleached cells as well in non-bleached ROIs in the bleached cells. Table displays the times after bleaching at which 50% of the signal was observed **(G)**.
(H-I) GFP-CTCF- behavior during mitosis. In panel **(H)** still images are shown of a time-lapse movie in which two GFP-CTCF-expressing cells go through mitosis. Arrows point to segregating chromosomes after mitosis. Panel **(I)** displays FI of chromosomes over time (cell 1, blue, cell 2, red).

Although fluorescence intensity varied per cell and over time, it did not fluctuate greatly **(Figure 2B)**. We used a purified protein (i.e. GFP-EB3, a GFP tagged microtubule associated protein) to generate a standard curve of soluble GFP fluorescence versus concentration **(Figure 2C)**. Using this curve we estimated that on average the concentration of GFP-CTCF in ES cells is ~3 μM **(Figure 2C)**. Given a nuclear volume of 1 picolitre in an ellipsoid nucleus of 10x5x5 micrometers, these data suggest that ~2 million molecules of GFP-CTCF are present inside an ES cell nucleus.

We next performed FRAP experiments on GFP-CTCF- and GFP-CTCF-ZF-expressing ES cells. In order to prevent excessive bleaching of highly mobile GFP-CTCF molecules, we bleached small circular regions of interest (ROIs) in the nucleus of ES cells for a limited period of time **(Figure 2D)**. However, ES cell nuclei are mobile within the time scale of the experiment (data not shown). Nuclear movement combined with small ROIs caused inaccuracies in recovery curves at later time points. We therefore used two FRAP regimes. To obtain an accurate view of the mobility of GFP-CTCF molecules we measured recoveries for 2 minutes, using image acquisition times of 2 seconds **(Figure 2E)**. To observe the dynamic behavior of GFP-CTCF and selected mutants for longer times, we performed FRAP experiments for 10 minutes, using image acquisition times of 5 seconds **(Figure 2F)**. Both regimes gave similar results and showed that the vast majority of GFP-CTCF molecules is relatively immobile, recovering to 50% of the initial fluorescence in ~5 minutes **(Figure 2E, G)**. Taken together these results indicate that GFP-CTCF distributes over the complete genome and that at least half of the ~2 million molecules of GFP-CTCF in an ES cell nucleus are bound to DNA.

The recovery kinetics of the GFP-CTCF-ZF mutants was different from GFP-CTCF. GFP-CTCF-Δ9, GFP-CTCF-Δ10 and GFP-CTCF-Δ11 recovered slightly faster than GFP-CTCF, whereas GFP-CTCF-Δ1 and GFP-CTCF-Δ8, the two mutants with a reduced ES cell growth rate, recovered much faster **(Figure 2E, F)**. These data show that the strong increase in mobility of GFP-CTCF-Δ1 or GFP-CTCF-Δ8 correlates with a high number of deregulated genes.

Time-lapse movies revealed that GFP-CTCF remains bound to mitotic chromosomes **(Figure 2H)** although its fluorescence intensity on metaphase chromosomes is decreased **(Figure 2I)**. We estimate that ~10-fold fewer molecules are bound to metaphase chromosomes compared to interphase chromatin. Interestingly, upon completion of cytokinesis the pool of cytoplasmic CTCF is rapidly transported into the nucleus, presumably to establish proper CTCF-chromatin interactions at the beginning of G1 **(Figure 2H, I)**. Because nuclei are smaller after mitosis, they appear brighter due to a higher CTCF concentration. We conclude that CTCF remains bound to chromatin during the complete cell cycle but that the number of molecules bound to DNA can differ depending on the stage of the cell cycle.

Figure 3

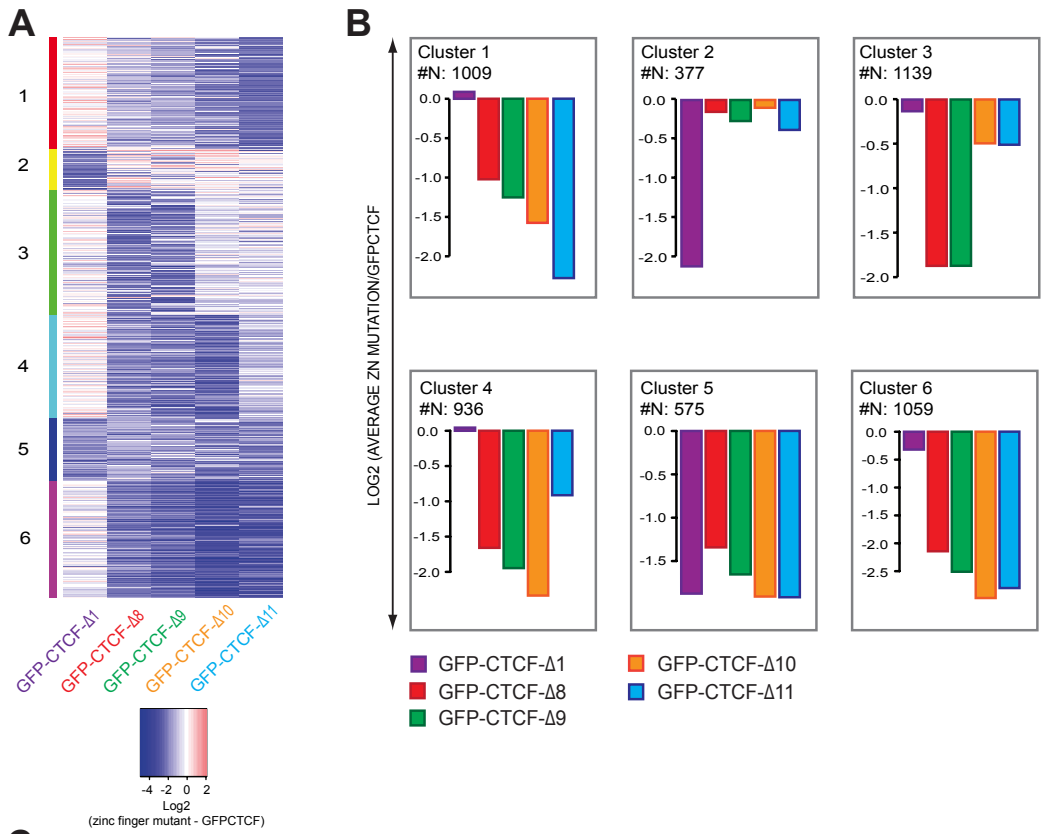

A
1
2
3
4
5
6
GFP-CTCF-Δ1
GFP-CTCF-Δ8
GFP-CTCF-Δ9
GFP-CTCF-Δ10
GFP-CTCF-Δ11
-4 -2 0 2
Log2
(zinc finger mutant - GFPCTCF)
B
LOG2 (AVERAGE ZN MUTATION/GFPCTCF
Cluster 1
#N: 1009
Cluster 2
#N: 377
Cluster 3
#N: 1139
Cluster 4
#N: 936
Cluster 5
#N: 575
Cluster 6
#N: 1059
GFP-CTCF-Δ1
GFP-CTCF-Δ8
GFP-CTCF-Δ9
GFP-CTCF-Δ10
GFP-CTCF-Δ11
C
Information content
GFP-CTCF
GFP-CTCF-Δ1
GFP-CTCF-Δ8
GFP-CTCF-Δ9
GFP-CTCF-Δ10
GFP-CTCF-Δ11

Figure 3. GFP-CTCF-Δ8-11 are specifically impaired in UC motif binding
(A) Heatmap of a k-means clustering of sites with at least 4 fold absolute change in at least one condition. Clustering reveals 6 subgroups. Scale represents log2 fold binding change between GFP-CTCF and GFP-CTCF-ZF mutant.
(B) Binding within the subgroups of GFP-CTCF-ZF mutants compared to GFP-CTCF. #N represents the number of affected binding sites within each cluster.
(C) Motif analysis of DNA binding by GFP-CTCF-ZF mutants. Sites with at least 4 fold reduction in each GFP-CTCF ZF mutant dataset and random sites in the GFP-CTCF dataset were taken along in the analysis. An upstream and core motif are visible in the GFP-CTCF-Δ8-11 mutants. Additionally, downstream of the core motif nucleotides are visible that indicate the 3'end border of the full CTCF binding motif in the GFP-CTCF-Δ1 mutant. Red, green, yellow and blue represent T, A, G, and C nucleotides, respectively.

GFP-CTCF-Δ8-11 specifically loose binding to the upstream consensus site and differ from GFP-CTCF-Δ1

To determine the DNA binding specificity of CTCF ZF mutants we performed ChIP-Sequencing (ChIP-Seq) on the different ES cell lines using antibodies against CTCF. We discovered ~22,000 CTCF target sites, defined as a peak region with a $p<0.05$ overlap with a predicted core consensus motif, in ES cells expressing GFP-CTCF. ChIP-Seq data was validated by ChIP-qPCR (data not shown). To detect specific patterns in the binding of ZF mutants we selected sites with a log2-fold binding change larger than 2 in at least one of the ZF mutants compared to GFP-CTCF and applied k-means cluster analysis. K-means clustering allowed us to distinguish 6 different groups **(Figure 3A)**. Interestingly, many more sites showed reduced binding as opposed to increased binding of mutant CTCF **(Figure 3A)**, and reduced binding was more pronounced than increased binding **(Figure 3A, B)**. These results suggest that ZF1 and 8-11 are required for optimal DNA binding by full length CTCF.

Binding of GFP-CTCF-Δ1 was clearly different from that of the other mutants, as this mutant was least affected in the total number of binding sites and clustered in a distinct manner **(Figure 3A, B)**. In contrast, GFP-CTCF-Δ8-11 clustered together, with GFP-CTCF-Δ8 and GFP-CTCF-Δ9 being affected at more sites than GFP-CTCF-Δ10 and GFP-CTCF-Δ11. Together, these data suggest that ZF 1 and ZF 8-11 contribute in distinct ways to the binding of CTCF. In addition, while nuclear mobility of the different ZF mutants is linked to proliferation of ES cells, DNA binding is not. For example, the GFP-CTCF-Δ1-expressing ES cell line is most affected in its growth but GFP-CTCF-Δ1 is least reduced in DNA binding. Furthermore, in the FRAP studies GFP-CTCF-Δ8 was the most mobile mutant protein, with ~70% of the molecules recovering within 2-3 minutes, whereas GFP-CTCF-Δ9 behaved more like GFP-CTCF. By contrast, in the ChIP-Seq analysis these two mutants are comparably poor in terms of DNA binding. This suggests that ZF 8 is involved in additional, e.g. mediation of protein interactions, functions of CTCF compared to ZF 9.

Binding defects in CTCF ZF mutants could be due to a failure to recognize specific nucleotides within the conserved CTCF binding sites. To examine this we performed a motif analysis on the 6 clusters defined above - thus, we analyzed sequences that were not bound by a particular mutant ZF protein **(Figure S3)**. Additionally, we performed a motif analysis on the GFP-CTCF dataset as control by gathering random CTCF binding sites. We found that each cluster contained the ~20 bp core CTCF motif**,** which was previously identified *(Boyle et al., 2011; Chen et al., 2008; Kim et al., 2007; Nakahashi et al., 2013; Rhee and Pugh, 2011; Schmidt et al., 2012; Xie et al., 2007)*. Binding sites affected in clusters 1, 3, 4 and 6 (i.e. sites to which ZF 8-11 bound less efficiently) contained a small motif ~8 nt upstream of the CTCF core, suggesting that ZFs 8-11 are required for binding these nucleotides. This upstream motif is smaller in cluster 5, which contained sites that were affected in all zinc finger mutants. We also noted the presence of 4 cytosines (at positions 42-45) in cluster 6 **(Figure S3)**, which contains sites that are bound less efficiently by ZFs 8-11. These data suggest that defective binding by ZFs 8-11 can affect binding downstream of the core domain. In cluster 2 (i.e. sites to which ZF 1 binds less efficiently) we did not observe the upstream signature; instead a cytosine ~10 nt downstream of to the

CTCF core motif, and a less prominent guanine showed up (positions 44 and 41, respectively) **(Figure S3)**.

To understand how individual ZFs contribute to the recognition of CTCF binding sites we generated ZF mutant-specific motifs by applying motif analysis to sites that showed at least 4-fold reduction in the binding of a particular ZF mutant. Additionally, we performed a motif analysis on the GFP-CTCF dataset as control by gathering random CTCF binding sites. Again, all affected binding sites in each ZF mutant contained the core CTCF motif **(Figure 3C)**. In addition, GFP-CTCF-Δ1-affected sites contained a prominent cytosine at position 44, as well as additional nucleotides between the core motif and position 44 (positions 37, 39, 41 and 43), indicating the binding signature of ZF 1 and possible ZF 2 and ZF 3 **(Figure 3C)**. Sites that were not bound by GFP-CTCF-Δ8-11 again contained the upstream motif. Furthermore, we noticed four cytosines at positions 42-45 in sites defective in binding of ZF mutants 9-11, most prominently in GFP-CTCF-Δ10 and 11. Together these data reveal three closely positioned regions important for CTCF binding: the core and upstream motifs, which have been described before *(Boyle et al., 2011; Chen et al., 2008; Kim et al., 2007; Nakahashi et al., 2013; Rhee and Pugh, 2011; Schmidt et al., 2012; Xie et al., 2007)*, and a novel region downstream of the core. Similar to another publication *(Nakahashi et al., 2013)*, our data suggest that ZF 9-11 are important for binding the upstream motif. In addition, we find that ZF 8 is also required for binding this region. By contrast, ZF 1 appears to mainly recognize a cytosine located 10 nt downstream of the core domain **(**at position 44 in **Figure 3C)**. ZFs 1-3 might actually bind CTCF target sites as a module, since deletion of ZF 1 coincides with the appearance of other nucleotides between position 44 and the core motif. Finally, in clusters deficient in binding of ZFs 9-11 nucleotides downstream of the core domain appear. Although these nucleotides are less prominent than the upstream motif, they coincide exactly with the domain defective in ZF 1 binding. These data indicate that binding of ZFs 9-11 affects binding downstream of the core, indicating that there is an interplay between ZFs to establish DNA binding upstream and downstream of the core motif. Combined our data reveal the complete binding motif of CTCF involving all ZFs.

We next plotted for each ZF deletion the fold change in binding to all 22,000 CTCF target sites as compared to those 2718 sites with the core and upstream motif at -18bp (UC) *(Nakahashi et al., 2013; Schmidt et al., 2012)*. An obvious relation appeared: ZF8-11 showed strongly reduced binding to these sites whereas ZF1 did not **(Figure S4)**. Thus not only did we identify the UC sites among the sites with strongly reduced binding by ZF8-11, but also virtually all UC containing CTCF target sites show decreased binding of ZF8-11.

CTCF UC sites show reduced binding for GFP-CTCF-Δ8-11, are depleted from TSS and show heterochromatic features

To explore the genomic distribution of affected sites we plotted their location with respect to transcription start sites, genes and intergenic regions, and compared this to all CTCF target sites in the GFP-CTCF dataset **(Figure 4A, B)**. We found that sites to which all mutant proteins bound less efficiently were less often located in exons, and sites affected in ES cells expressing GFP-CTCF-Δ8-11 were found less frequently at TSSs compared to the relative frequency observed in GFP-CTCF. We next examined the genomic location of CTCF binding sites containing the UC motif and compared them to all CTCF target sites. This also revealed a specific depletion of UC containing sites at TSSs and in exons **(Figure 4C, D)**. We therefore conclude that GFP-CTCF-Δ8-11 are required to bind UC CTCF target sites efficiently and that these sites are relatively rare in exons and at TSSs.

Figure 4

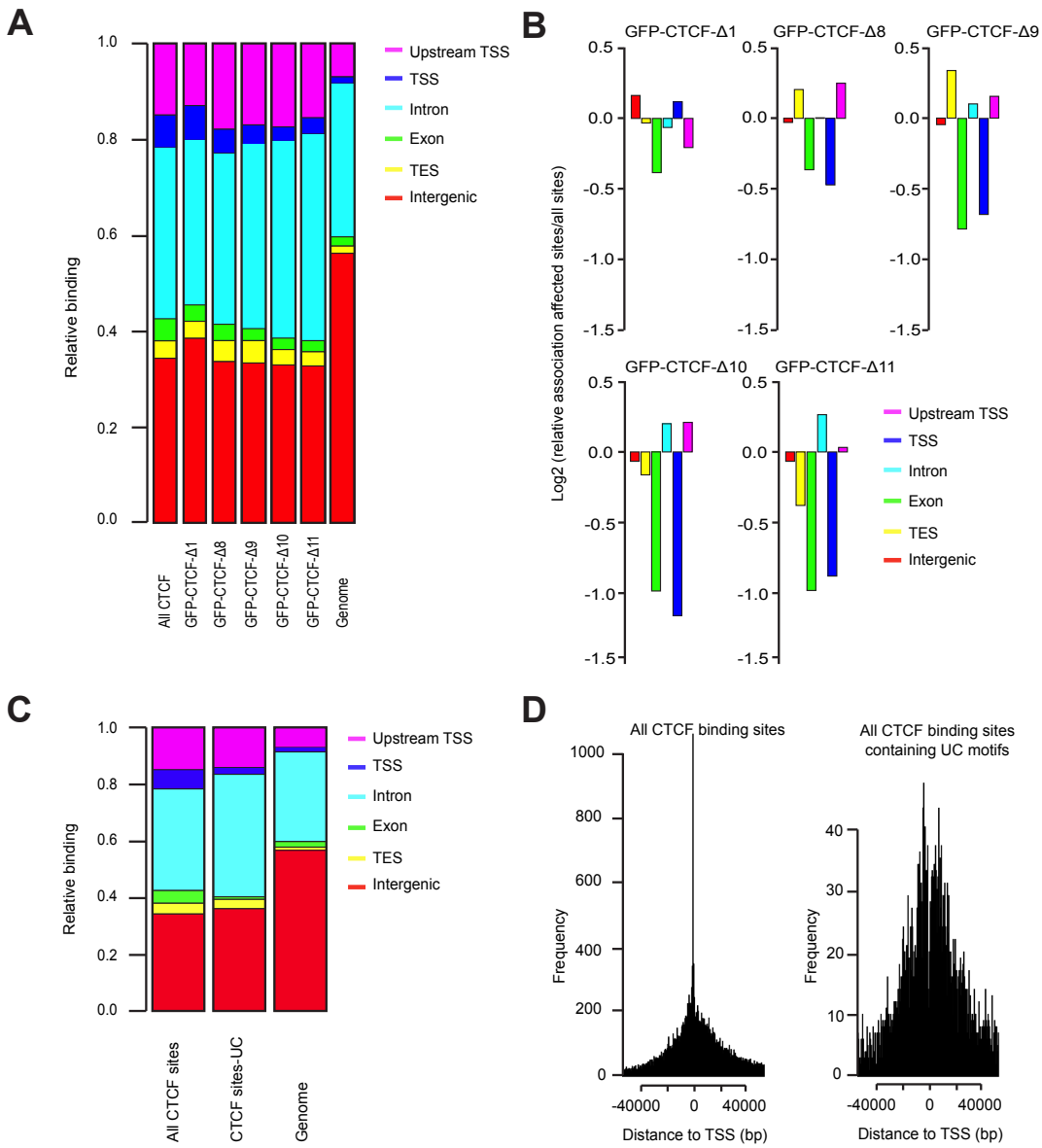


Figure 4. Reduced binding of ZF mutants 8-11 to UC motif, which is depleted from TSS
(A-B) Comparison of the genomic distribution of binding sites. Binding locations are separated into upstream of transcription start site (Upstream TSS), transcription start site (TSS), exon, intron, intergenic, and transcription end site (TES), and plotted as frequencies of total (Y-axis). The contribution of binding sites belonging to each location class to total number of binding sites is plotted in **(A)**. The distribution of all CTCF binding sites over the different location classes is plotted (left), as well as the percentage of annotated locations on the entire genome (Genome). The middle histograms in **(A)** represent the distribution of sites that show a more than 4 fold absolute binding change in the respective mutant proteins. Actual differences between binding frequencies observed in GFP-CTCF-ZF mutant compared to GFP-CTCF are plotted in **(B)**. Relative binding is plotted as log2 fold change (Y-axis). Thus, compared to all CTCF binding sites, the sites that are not bound by GFP-CTCF-Δ8-11 are less present in TSSs and exons.
(C-D) Distribution of genomic features for all CTCF binding sites and UC-containing CTCF binding sites. In **(C)** binding locations are depicted as in **(A)**. In **(D)** all CTCF binding sites and UC-containing CTCF binding sites are plotted with respect to distance to the nearest TSS. ZF mutant 8-11-sensitive binding sites show preference for genic regions.

We used published ChIP-Seq datasets (H3: *(Mullen et al., 2011)*, H3.3: *(Goldberg et al., 2010)*, H3K27ac: *(Creyghton et al., 2010)*, H3K27me3: *(Rugg-Gunn et al., 2010)*, H3K36me3: *(Mikkelsen et al., 2007)*, H3K9me3: *(Mikkelsen et al., 2007)*, OCT4: *(Chen et al., 2008)*, Pol II-ser5P: *(Kagey et al., 2010)*, SMC1: *(Kagey et al., 2010)* and CTCFL: *(Sleutels et al., 2012)*) to compare CTCF binding sites sorted for the two subsets (i.e. all CTCF sites and CTCF sites containing the UC motif) and plotted these in an area of -20 kb to +20 kb relative to the center of the peak, with specific chromatin marks and transcription factors **(Figure 5)**. This revealed that UC motif sites were reduced in H3.3 and H3K27me3 occupancy, and virtually depleted of H3K27ac and H3K4me3. Interestingly, a substantial enrichment for the chromatin mark H3K9me3, which is often associated with constitutively repressed genes, was observed in UC motif containing CTCF sites. Consistent with the relative absence of UC motif sites from TSSs, RNA polymerase II was highly reduced at these sites. Pluripotency factor OCT4, cohesin subunit SMC1 and CTCFL profiles revealed a small reduction in binding. Thus, UC motif sites, which are specifically affected in CTCF binding in ES cells expressing GFP-CTCF-Δ8-11, are less often found at TSSs and in exons and preferentially associate with a mark for repressive chromatin.

Figure 5

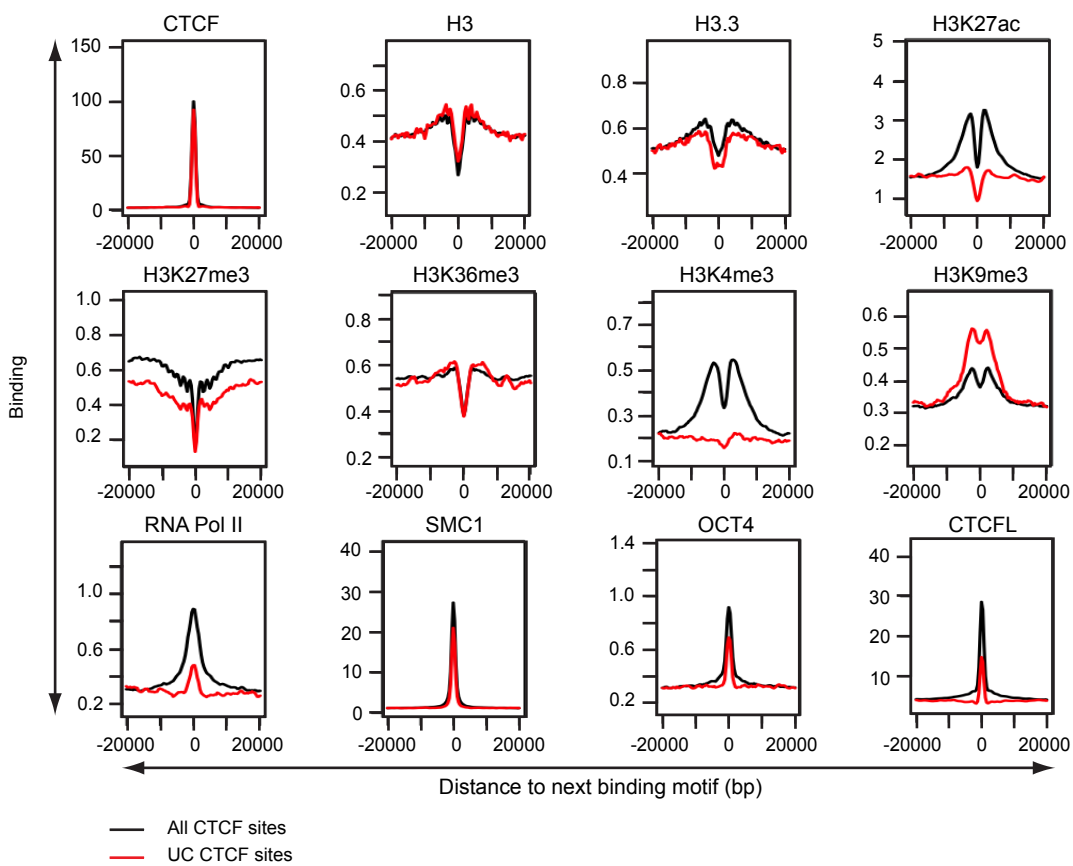


Figure 5. CTCF bound to UC sites shows heterochromatic features
Cumulative average profiles of all CTCF (black) and upstream core (UC) motif (red) containing binding sites with respect to chromatin context. Average binding was determined in 20 kb intervals around binding sites. We used published ChIP-sequencing data sets in mouse embryonic stem cells for CTCF, H3, H3.3, H3K27Ac, H3K27me3, H3K36me3, H3K4me3, H3K9me3, RNA polymerase II, SMC1, OCT4, and CTCFL.

Figure 6

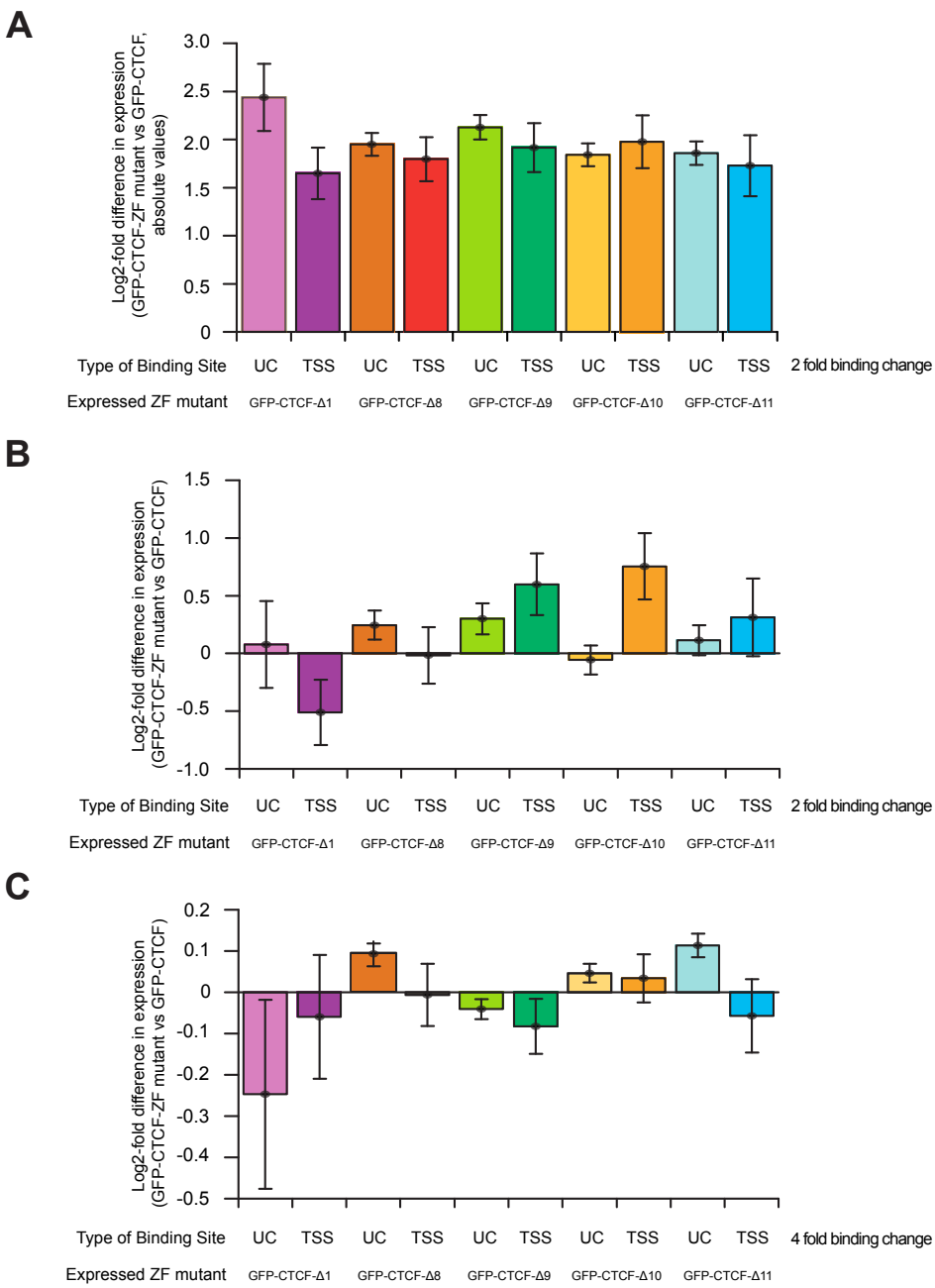


Figure 6. Correlation between loss of binding of CTCF and RNA expression of associated gene
CTCF binding sites were divided into sites near TSSs (± 2 kb) and sites with a UC motif (UC). Sites were subsequently categorized depending on whether the binding by a particular GFP-CTCF-ZF mutant was affected.
In **(A-B)** all binding sites with at least a 2-fold reduced binding were included, in **(C)** only sites were included with at least 4-fold reduced binding. We examined RNA expression of the gene nearest to the affected site. In **(A)** we looked at absolute expression levels without taking into account whether a gene is up- or down-regulated, and included outliers. In **(B)** we examined whether loss of binding resulted in increased or decreased expression levels and we also included outliers. In **(C)** we examined whether loss of binding resulted in increased or decreased expression levels, this time of the genes nearest to the most affected binding sites, excluding outliers.

Correlation between loss of binding of CTCF and RNA expression

The loss of binding of CTCF ZF mutants to a subset of CTCF binding sites in the genome provided us with the unique opportunity to ask whether there is a correlation between loss of CTCF binding and changes in gene expression. Importantly, for the first time we could address this issue in cells in which CTCF expression itself was not ablated. We generated a set of ~22,000 CTCF binding sites and linked each site to its nearest gene. We next divided sites into two categories: 1) CTCF binding sites near a TSS (± 2 kb, 2578 genes), and 2) CTCF binding sites with an upstream motif at -17 to -19 bp of the core binding sequence (i.e. UC sites, 2949 genes). We selected sites that showed a 2 fold binding change observed in a particular ZF mutant for each category. We only examined sites in which ZF binding was reduced, because reduction is observed in the majority of affected sites.

We next examined RNA levels of genes associated with affected TSS- and UC-linked CTCF binding sites, by comparing ChIP-Seq and RNA-Seq data in mutant and GFP-CTCF-expressing ES cells. To determine the overall effect of reduced CTCF binding on associated genes, we first examined absolute differences in RNA expression (i.e. without taking into account whether loss of binding resulted in increased or decreased gene expression and including outliers). This revealed that CTCF binding to target sites near genes regulates transcription **(Figure 6A)**. We observed a log2 fold change in transcription of ~2 on the TSS in each mutant. Interestingly, RNA expression level differences are similar in the different ZF mutant-expressing ES cells in the TSS category. These results indicate that it is the loss of binding that is important for transcriptional regulation by CTCF.

There is not much difference in the RNA levels of genes in which CTCF binding is affected near a TSS compared to genes where CTCF binding is reduced at a UC motif **(Figure 6A)**. These data indicate that CTCF regulates gene transcription irrespective of where it binds in a gene. To verify this conclusion we compared the expression data of all genes (this time taking into account whether loss of binding resulted in increased or decreased gene expression) in which CTCF binding was affected (two-fold or more) and in which CTCF was either near a TSS or bound to a UC motif **(Figure 6B)**. We performed a similar analysis as in **(Figure 6B)** but with CTCF binding showing a 4-fold or more reduction in each binding category (TSS and UC) and excluding outliers in the transcriptional data **(Figure 6C)**. In none of these analyses did we observe profound differences in the pattern of expression of genes when we compared loss-of-binding on TSS-linked CTCF sites versus UC-linked sites. These data suggest again that the position of CTCF binding within a gene is not essential for the regulation of transcription of that gene. Thus transcriptional regulation of CTCF is apparently independent of the position of CTCF to the TSS.

Discussion

We deleted endogenous *Ctcf* using conditional *Ctcf*$^{delneo/delneo}$ knockout ES cells and rescued ES cells by expressing GFP-tagged CTCF-ZF mutants to study the role of the different ZFs. This is a unique system since it excludes interference by wild type CTCF. It allowed us to examine effects that are specific to individual ZFs with respect to cell growth, dynamic protein behavior, DNA binding and genomic distribution, and transcription regulation. All ZFs of CTCF are highly conserved indicating that each ZF is important for proper CTCF function *(Moon et al., 2005; Pugacheva et al., 2006)*. Nevertheless, we were able to generate ES cells expressing GFP-CTCF-Δ1, GFP-CTCF-Δ8, GFP-CTCF-Δ9, GFP-CTCF-Δ10 and GFP-CTCF-Δ11 mutants replacing endogenous CTCF. It is remarkable that we were able to delete the highly conserved zinc fingers ZF1 and ZF8-11, which suggests that they contribute to a lesser extent to the functionality of CTCF compared to ZF2-7 in ES cells. It may also be that these zinc fingers act as functional unit in which deletion of one zinc finger slightly disrupt CTCF function but

deletion of more zinc fingers might result in cellular lethality. ES cells could not be rescued by GFP-CTCF-Δ2-7, which indicates that these ZFs are essential for cell survival, even if we had screened sufficient ES cell clones to find a rare GFP-CTCF-Δ2-7-expressing mutant cell line the conclusion would have remained the same.

Fluorescence-based microscopy measurements suggest that there are >1 million molecules of GFP-CTCF in an ES cell nucleus. By contrast, there are only ~25,000 conserved CTCF target sites, that consistently appear in every cell line tested *(Li et al., 2013)*. Since the majority of CTCF molecules are immobile and CTCF distributes in a speckled pattern throughout the nucleus, it is likely that the protein binds DNA at many more positions than just the ~25,000 conserved CTCF binding sites. ES cells expressing GFP-CTCF-Δ1 and -Δ8 have a growth defect and these mutant proteins display increased mobility, in particular GFP-CTCF-Δ8. These results suggest that binding of CTCF throughout the nucleus via ZF1 and -8 is important for proper ES cell growth and proliferation.

Our FRAP data are different from a recent publication, in which the recovery of fluorescent CTCF molecules in B cells was found to be in the order of seconds instead of minutes *(Nakahashi et al., 2013)*. In this study individual CTCF ZF mutants were overexpressed in primary lymphocytes in the presence of wild type CTCF, creating a situation where mutant protein competes with endogenous CTCF. Furthermore, the overexpression of CTCF-GFP proteins (both wild type and ZF mutants) by *(Nakahashi et al., 2013)* was not estimated with respect to endogenous CTCF. It is therefore unclear to what extent exogenous proteins were overexpressed. Performing FRAP under such conditions is likely to mask physiological CTCF dynamics. The rapid recovery times of the ZF mutants *(Nakahashi et al., 2013)* are not surprising if CTCF molecules are present in vast excess. By contrast, our system does not contain endogenous CTCF, hence no competition between wild type CTCF and mutant protein can occur. In addition, GFP-CTCF expression was comparable to endogenous levels in WT cells, creating a proper physiological situation to study CTCF dynamics.

Interestingly, CTCF concentration fluctuates during the cell cycle, and is most reduced during mitosis. This reduction of CTCF concentration suggests that less CTCF sites are occupied. We hypothesize that CTCF binding sites involved in the regulation of open chromatin configuration or sites located at chromatin boundary elements are likely to be reduced in CTCF binding during M-phase due to chromosome condensation. It might be that loss of CTCF binding during M-phase could be caused by the condensation of the chromosomes. On the other hand it could also be that loss of CTCF binding of these sites would allow condensation of the chromosomes.

ChIP-Seq experiments show that in ES cells expressing GFP-CTCF-Δ8-11 a subset of CTCF binding sites are affected, many of which are shared in the individual mutants. However, only GFP-CTCF-Δ8-expressing cells grow poorly and only GFP-CTCF-Δ8 shows aberrant protein dynamics. This underscores our conclusion that binding of CTCF throughout the nucleus is important for ES cell growth and proliferation and that binding defects on CTCF target sites are less important. Our data do not allow us to distinguish how exactly CTCF binds throughout the nucleus. However, it will be interesting to determine whether the *in vivo* CTCF genome-wide binding sites, which excess the ~25,000 conserved sites, documented here are as important for 3D interactions of chromatin, as are a subset of its conserved CTCF binding sites *(Handoko et al., 2011)*. Moreover, CTCF can facilitate protein-protein interactions via its ZF domain *(Chernukhin et al., 2000; Ishihara et al., 2006; Lutz et al., 2000; van de Nobelen et al., 2010)*, which could also be important for protein mobility and cell growth and could explain the different protein mobility observed between GFP-CTCF-Δ8 and GFP-CTCF-Δ9.

ZFs 4-7 were proposed to recognize the 20 bp core motif present in 80-99.5% of all CTCF binding sites *(Nakahashi et al., 2013; Renda et al., 2007; Rhee and Pugh, 2011)*. Therefore, these ZFs are essential to establish CTCF binding to virtually all conserved CTCF binding sites in the genome. The peripheral ZFs were proposed to increase CTCF binding affinity *(Renda et al., 2007)*,

with the C-terminal ZFs recognizing an upstream motif *(Nakahashi et al., 2013)* that was identified by other groups *(Boyle et al., 2011; Rhee and Pugh, 2011; Schmidt et al., 2012)*. The upstream motif is identical to the one we identify in our work. The core and upstream motifs are separated by a 10 bp 'spacer' that would allow one DNA helix turn in between the two CTCF binding motifs. We provide evidence that ZFs 8-11 together are important for recognition of the upstream motif. We hypothesize that ZFs9-11 might actually be involved in base contacts with the upstream motif, whereas ZF8 is required to bridge the 10 bp space between the motifs in the DNA. In the absence of ZF8 ZFs9-11 might not be positioned properly and fail to bind DNA altogether.

The 20 bp core motif of CTCF is highly conserved and detected in all eutherian mammals, opossum, chicken, pufferfish Tetraodon, and a similar motif has been discovered in drosophila *(Chen et al., 2008; Holohan et al., 2007; Jothi et al., 2008; Kim et al., 2007; Smith et al., 2009; Xie et al., 2007)*. By contrast, the conservation of the bipartite motif has been investigated to a lesser extent than the core motif. The bipartite motif and the 10 bp spacing between the core and upstream motif are so far only conserved in eutherian mammals *(Schmidt et al., 2012)*. It could be that the bipartite motif is less conserved than the core motif, which would indicate that these sites serve a less important role in the function of CTCF. Moreover, this would be an additional explanation why deletion of ZF8-11 did not result in cellular lethality.

Our data reveal that ZF1 recognizes nucleotides downstream of the core, in particular a C around position 44 of the CTCF consensus sequence (10 nt downstream of the core motif). Furthermore, sites deficient in binding of ZFs 9-11, contain nucleotides downstream of the core. Although these nucleotides are less prominent than the upstream motif, they coincide exactly with the domain defective in ZF 1 binding. These data indicate that binding of ZFs 9-11 affects binding downstream of the core, indicating that there is an interplay between ZFs and the DNA to establish DNA binding upstream and downstream of the core motif. Together our results identify the full CTCF consensus sequence. We propose a model in which ZFs 1-7 bind adjacent to each other, with ZF 1 binding to position 44 marking the 3'end of the CTCF binding site. ZF7 binds the 5'end of the core motif. The 10 bps in between the 2 motifs form a helical turn such that ZF9-11 continue proximal binding to the upstream motif.

Our data support previous findings *(Boyle et al., 2011; Nakahashi et al., 2013; Rhee and Pugh, 2011; Schmidt et al., 2012)* regarding sequence and position of the upstream motif relative to the core motif. However, important to realize that the majority of CTCF binding sites contain only the core motif, and that in ChIP-Seq experiments these sites show similar CTCF occupancy as bipartite motif sites. Our data show that disruption of CTCF's C-terminal ZFs results in a severe reduction of CTCF binding to bipartite motif sites. Thus, the presence of the core motif within bipartite motif sites is not sufficient to maintain binding of CTCF-ZF8-11 mutants, whereas on sites with only a core motif these mutants bind as efficiently as WT CTCF. This indicates that at UC sites a tight CTCF binding is required and that in the absence of ZF8-11 binding is reduced such that other factors or epigenetic modifications can displace mutant CTCF from the core site. On the other hand it might also suggest that the tight CTCF binding at the bipartite motif is required to maintain a proper chromatin environment.

The genomic distribution and chromatin context of CTCF's bipartite motif was unknown. Our data reveal that the bipartite motif is often located in intergenic and intronic regions and is reduced at active chromatin areas and TSSs. Interestingly, the bipartite binding sites are associated with the repressive chromatin mark H3K9me3, which suggests that CTCF is involved in establishing or maintaining repressive chromatin areas at bipartite sites. Changes in chromatin marks need to be explored in CTCF ZF mutants in order to understand the functional relevance of CTCF binding to bipartite motif sites.

In conclusion, our data indicate that CTCF organizes chromatin at two distinct levels. First, CTCF binds DNA at many sites throughout the genome and this binding is essential for proper ES cell growth and proliferation. Second, CTCF binding near genes regulates transcription locally. This transcriptional regulation is apparently independent of the position

of CTCF to the TSS. CTCF does not act as a classic transcription factor, but rather makes chromatin near genes (in)accessible via e.g. by long-range interactions.

Materials and Methods

Generation of lentivirus

Fusion PCR with fusion primers **(Table S3)** flanking each individual ZF on CTCF-encoding cDNA, amino acid sequence of each deleted zinc finger are depicted in **(Table S1)**, was used to generate individual ZF mutants. Primers contained an Xho I site before the ATG and after the stop codon. PCR products were cloned in lentiviral vectors using Xho I restriction site. Lentiviral constructs were generated as described earlier *(Sleutels et al., 2012)*. Constructs contained a CAG promoter (CMV early enhancer/chicken beta actin) driving the expression of a "bi-cistronic" cDNA encoding GFP-tagged wild type or mutant CTCF, followed by an IRES sequence to generate Cre-recombinase. A neomycin resistance cassette driven by a PGK promoter was also present. Lentivirus particles were produced as described (Addgene).

Generation of GFP-CTCF-expressing ES cells

Ctcf^{lox} mice were generated as described previously *(Heath et al., 2008)*. *Ctcf*$^{lox/lox}$ embryonic stem cells (ES cells) were isolated and transiently treated with Cre-recombinase to delete the neomycin resistance cassette generating *Ctcf*$^{delneo/delneo}$ ES cells selected on puromycin sensitivity. These ES cells were infected with lentiviral particles in suspension for 4 hours, plated and selected with 0.2 mg/ml G418 one day after infection. Neomycin-resistant, GFP-positive clones were picked and expanded as described *(Sleutels et al., 2012)*.

ES cell culture and proliferation assay

ES cells were grown on plastic dishes coated with 0.2% gelatin (Merck) in the presence of ES cell medium containing: Dulbecco's Modified Eagle Medium (DMEM) (Lonza), 15% heat-inactivated fetal calf serum (FCS), Non Essential Amino Acids (Lonza), 100 U ml^{-1} penicillin and 100 mg ml^{-1} streptamycin, 0.1 mM β-mercaptoethanol (Sigma) and 1000 U/ml leukaemia inhibitory factor (LIF). ES cells were cultured 37°C with 5% CO_2 levels. Cells were passaged by trypsinization in 1xTE (trypsin/EDTA) for 5 minutes at 37°C.

To measure ES cell growth and proliferation 100000 ES cells (N=3 for each cell line) were counted and plated on a 6-well dish. After 48 hours ES cells were harvested, counted and 100000 cells were plated back. This was repeated 3 times (144 hr).

DNA, RNA and protein isolation

Genomic DNA was isolated by incubating cell pellets with 50 mM Tris-HCl pH 8.0, 5 mM EDTA pH 8.0 and 0.5% SDS, 0.3 mg/ml proteinase K and 0.01 mg/ml RNase at 55°C on a rocking platform. 1.2 M NaCl was added to samples followed by centrifugation on 13,000 rpm for 10 minutes at room temperature (RT). Isopropanol and 70% ethanol were used for DNA precipitation and pellets were dissolved in 10 mM Tris and 1 mM EDTA.

Total RNA isolation of 2 replicates of each ES cell line was performed with Trizol-chloroform extraction. After Trizol addition, samples were incubated for 5 minutes at 30°C. Chloroform was added and aqueous phase was transferred after centrifugation on 13,000 rpm for 10 minutes 4°C. 100% ethanol was added and RNA was isolated using the RNeasy Mini Kit (Cat. no. 74104, Qiagen).

Nuclear protein extracts from 10 cm^2 cell culture dishes were generated by resuspending ES cells in 500 μl buffer A (10 mM Hepes, 1.5 mM $MgCl_2$, 10 mM KCl, and protease inhibitors (Complete, Roche)) followed by 10 minutes incubation on ice. Samples were centrifuged on

6000 rpm for 1 minute at 4°C and pellets were sub taken in 250 μl buffer C (420 mM NaCl, 1.5 mM $MgCl_2$, 200 mM EDTA, 20 mM Hepes hOH, 21.75% glycerol and protease inhibitor) followed by 20 minutes incubation on ice. After centrifugation on 10,000 rpm at 4°C supernatant was used for further assays.

Antibodies, Western blot and Southern blot analysis

Nuclear protein extracts were loaded on a SDS-PAGE gel. After electrophoresis proteins were transferred to PVDF membranes (Millipore) via semi-dry blotting (Bio-Rad, trans-blot SD, semi-dry transfer cell). Membranes were blocked either in 2% BSA, or in 5% Milk powder (Sigma) in PBS containing 0.05% Tween-20. Antibody incubations were done in the same buffer. The following antibodies were used: GFP (rabbit home made 1:1000), CTCF (rat, 1:1000, Absea) and p150 glued (mouse, 1:1000, Bd transduction laboratories). The following secondary antibodies conjugated with HRP were used for detection anti-rabbit HRP (1:10000, GE healthcare), anti-rat HRP (1:10000, GE healthcare) and anti-mouse HRP (1:10000, GE healthcare).

Genomic DNA was digested with Hind III and loaded on an agarose gel for size fractionation. Samples were blotted onto a Hybond N+ membrane (Amersham) and hybridized with radioactive probes using α^{32}P dATP as described previously *(Sleutels et al., 2012)*. Signals were detected with a Phosphor Imager (Typhoon Trio variable mode imager, GE healthcare).

Fluorescence after photobleaching (FRAP) and fluorescence intensity measurements

We bleached ES cells expressing either normal GFP-CTCF or a mutant protein with a deletion in a single ZF. FRAP experiments were performed on a Leica TSP5 confocal microscope, essentially as described *(Dragestein et al., 2008)*. We bleached ES cells expressing either GFP-CTCF or GFP-CTCF-ZF mutant. In each ES cell colony we bleached 3 nuclei, using circular ROIs with a diameter of 3 mm. We measured fluorescence recovery in the bleached region. We also measured fluorescence intensity over time in similarly sized ROIs in non-bleached regions of the same nuclei, as well as in non-bleached cells (to show there was non monitor bleaching) and in the culture medium. After background deduction fluorescence intensity values were normalized. We then compensated for fluctuations in fluorescence intensity by dividing the normalized fluorescence intensity in the bleached ROI by that of the non-bleached ROI in the same nucleus.

We performed two types of FRAP experiment. In the first we measured recovery for two minutes with image acquisition times of two seconds. In this case 12 GFP-CTCF-, 23 GFP-CTCF-Δ1, 18 GFP-CTCF-Δ8-, 12 GFP-CTCF-Δ9-, 12 GFP-CTCF-Δ10- and 19 GFP-CTCF-Δ11-expressing nuclei were bleached and measured. In the second FRAP experiment we measured the recovery of selected GFP-CTCF proteins (>8 nuclei for each protein) for 10 minutes with image acquisition times of 5 seconds.

To calculate the concentration of GFP-CTCF in ES cell nuclei we measured the fluorescence intensity of a purified fluorescent protein, (i.e. GFP-EB3, a GFP tagged microtubule associated protein) of a known concentration. Using two-sided tape and a rectangular coverslip we assembled small glass sample chambers on objective holders, with an opening on opposite sides of the coverslip to allow flow-through of buffer and proteins.To block aspecific binding of GFP-EB3 to the glass surface inside the chamber, we first added buffer containing K-CASEIN (1mg/ml) to the chamber and incubated at RT for 15 minutes. After blocking the purified GFP-EB3 was flowed through the chamber. We used separate chambers to measure GFP-EB3 at different dilutions. We measured fluorescence intensity of each sample with a Leica TSP5 confocal microscope. Average fluorescence intensities were used to generate a standard curve. Using exactly the same microscope settings we imaged GFP-CTCF in ES cell colonies.

We outlined the nucleus of 18 cells and measured total and average fluorescence intensity of GFP-CTCF in these nuclei. The average fluorescence intensity was compared to that of the purified GFP-EB3 to obtain the GFP-CTCF concentration in ES cells.

Chromatin Immunoprecipitation (ChIP)

ChIP was performed as described *(van de Nobelen et al., 2010)*. Briefly, 40-80*10^6 cells were harvested and cross-linked with 1% formaldehyde (Sigma) for 10 minutes at room temperature and quenched with Glycine (Sigma). Cell lysates were prepared with cell lysis buffer (10 mM Tris pH 8.0, 10 mM NaCl, 0.2% NP-40 (Sigma), Protease Inhibitor) followed by nuclei lysis buffer (50 mM Tris pH 8.0, 10 mM EDTA, 1% SDS, Protease Inhibitor). Sonication was performed with the bioruptor (Diagnode) to yield fragments up to 800 bp. Immunoprecipitation with CTCF antibody (N2.2, home made) or pre-immune serum rabbit (home made) was performed. ChIP-Seq was validated by three independent ChIP-qPCR experiments. Ct values from qPCR were normalized to input measurements, and enrichment was calculated relative to CTCF negative binding site. For primer sequences see **(Table S4)**.

ChIP-Sequencing

A ChIP DNA library was prepared according to the Illumina protocol (www.illumina.com). Briefly, 10 ng of end-repaired ChIPped DNA was ligated to adapters, size selected on gel (200±25 bp range), and PCR amplified using Phusion polymerase as follow: 30 sec at 98ºC, 18 cycles of (10 sec at 98ºC, 30 sec at 65ºC, 30 sec at 72ºC), 5 min at 72ºC final extension. Cluster generation was performed using the Illumina Cluster Reagents preparation. The library was sequenced on the Illumina HiSeq2000 systems to generate 36 bp reads and a 7 bp index read. Images were recorded and analyzed by the Illumina Genome Analyzer Pipeline (GAP) and processed using the IPAR (Integrated Primary Analysis Reporting Software). Samples were de-multiplexed and mapped against mouse build mm9 reference genome.

RNA-Sequencing

Purity and quality of the isolated RNA was assessed by the RNA 6000 Nano assay on a 2100 Bioanalyzer (Agilent Technologies). Two independent RNA-Seq experiments were performed for each ES cell line (n=2). 1 μg total RNA of each sample was used as starting material for Illumina Truseq sequencing. Poly-A tail containing mRNA was purified with oligo-dT attached to magnetic beads. Subsequently, mRNA was fragmented into ~ 200 bp fragments followed by first-strand cDNA synthesis using reverse transcriptase and random primers. Next, second-strand synthesis was performed using DNA polymerase I and RNaseH treatment. End-repair, phosphorylation and A-tailing were carried out followed by adapter ligation, size selection on gel and PCR amplification. PCR products were purified by Qiaquick PCR purification. Samples were sequenced on HiSeq 2000 to generate 36 bp reads and a 7 bp index read. Samples were de-multiplexed and aligned to mouse build mm9 reference genome using Tophat alignment software.

Bioinformatic analysis

Aligned reads from RNA-Seq were used to identify differentially expressed genes by CuffDiff *(Trapnell et al., 2013)* using standard settings (adjusted p-value <0.05). The resulted gene lists were imported into R and the overlap was visualized using the *venn* function from the *gplots* package.

ChIP-Seq reads were aligned to mm9 reference genome using Bowtie2 *(Langmead and Salzberg, 2012)*. Peaks were identified using PeakRanger with standard settings *(Feng et al., 2011).* Here, peaks were identified based on read dense regions defined by p-value < 0.0001. The negative (mock) control was used to subtract the background (FDR <0.05 and > 3 fold

enrichment between ChIP and mock). In order to obtain a general set of peaks we used the presence of the core CTCF motif as criteria *(Kim et al., 2007)*. We used the top 1000 sites from each dataset and constructed a position specific scoring matrix (PSSM). This PSSM was used to perform a genome-wide prediction of CTCF motifs across the complete mouse genome using the Patser tool *(Thomas-Chollier et al., 2008)* with a Patser score > 7. Peak intervals detected with the GFP-CTCF dataset that overlapped with at least one instance of a core motif were used for subsequent analysis (22,216 sites).

For these intervals we calculated the number of reads for the individual zinc finger deletion as well as for the GFP-CTCF dataset. DEseq *(Anders and Huber, 2010)* was used for proper normalization of the data as well as to determine the fold change between the individual zinc finger mutants and the GFP-CTCF dataset. Fold changes were log2-transformed and those sites with an absolute log2 (fold change) larger than 2 in at least one zinc finger mutant were further subjected to cluster analysis using k-means in order to identify co-regulated CTCF binding sites **(Figure 3A and 3B)**.

All sites within a single cluster sequence around the core consensus were collected in a 70 bp window (using the BioConductor package *BSgenome.Mmusculus.UCSC.mm9*). Sequences were summarized as a weblog using the R package *webLogo* **(Figure S3)**. Alternatively all sites with a log2 (fold change) <-2 for an individual zinc finger deletion were identified and the corresponding sequences were extracted and collected in a 70 bp window (using the BioConductor package *BSgenome.Mmusculus.UCSC.mm9*). Sequences were summarized as a weblog using the R package *webLogo* **(Figure 3C)**.

The mm9 genome was partitioned into regions based on RefSeq annotations, TSS: +/- 1kb around RefSeq TSSs, TSS upstream: -1 to -10 kb, exons, introns, TES: +/- 1kb, rest: intergenic. All intervals of a given feature class were taken together and the ratio relative to the complete genome size was calculated to obtain the genomic background distribution. The same analysis was performed for the intersection between the genomic feature intervals and the respective peaks that showed a greater than 4-fold reduction of CTCF binding in at least one zinc finger mutant compared to GFP-CTCF **(Figure 4A)**. Alternatively, reduced CTCF binding in a zinc finger mutant was expressed as a ratio relative to GFP-CTCF and data were additionally log2-transformed and plotted in a histogram **(Figure 4B)**. In addition, RefSeq annotations for mm9 were used to calculate the distance for each of the 22,216 CTCF core motifs to the next transcriptional start site. Data were plotted in a histogram for the range between +/- 40 kb around the TSS. We compared all CTCF sites to CTCF sites containing the upstream motif **(Figure 4C)**. RefSeq annotations for mm9 were used to calculate the distance for each of the ~22,000 CTCF core motifs to the next transcriptional start sites. Data is plotted as a histogram within a window of +/- 40 kb around the TSS. We compared all CTCF sites with CTCF sites containing the UC motif **(Figure 4D)**.

CTCF sites containing core and upstream motif (UC) were identified using the top 200 reduced sites identified in the GFP-CTCF-Δ9 dataset to construct a position specific scoring matrix (PSSM). This matrix was used to identify potential motif (core and upstream motifs) instances across the complete mm9 genome using Patser *(Thomas-Chollier et al., 2008)*. Under the applied setting we identified about 1 million instances of the upstream and core motifs. Next, we identified the 22,216 CTCF core motifs from the UC sites and found 2718 cases were the upstream motif was located at a position -18bp relative to the CTCF core motif. We determined the log2 (fold change) binding change for a given zinc finger mutant compared to all GFP-CTCF for those sites having an upstream motif at position -18 bp **(Figure S4)**.

Average cumulative plots show binding of histones, histone modification marks and chromatin factors in the context of all CTCF sites and UC-containing CTCF sites **(Figure 5)**. All data were published previously (H3: GSE23830 *(Mullen et al., 2011)*, H3.3: GSE16893 *(Goldberg et al., 2010)*, H3K27ac: GSE24164 *(Creyghton et al., 2010)*, H3K27me3: GSE15519 *(Rugg-Gunn et al., 2010)*, H3K36me3: GSE12241 and H3K9me3: GSE12241 *(Mikkelsen et al., 2007)*, OCT4: GSE11431

(Chen et al., 2008), Pol II-ser5P: GSE20485 and SMC1: GSE22557 *(Kagey et al., 2010)* and CTCFL GSE34094 *(Sleutels et al., 2012)*) and downloaded from the Gene Expression Omnibus in short read archive format (sra). FASTQ data were extracted using SRA tools version 2.18. Reads were aligned to the mm9 genome as described above. After read extension coverage vectors were produced. These were used to collect binding data in a +/- 20 kb window with a 200 bp step size around the CTCF binding sites.

Based on stringent selection criteria, we obtained 21,030 genes linked to CTCF sites containing the core motif of CTCF. Subsequently, genes were either sorted according to the distance from the TSS (± 2000 bp, 2578 genes), or according to the distance from the upstream motif (ranging 17 to 19 bp in distance from the core, 2949 genes) (UC motif). The TSS- and UC-linked genes were subsequently sorted according to lack of binding of each of the respective GFP-CTCF-ZF mutants. We only considered sites in GFP-CTCF-ZF mutant bound less efficiently compared to wild type GFP-CTCF. As cut-off we either took a 2-fold or a 4-fold reduction in binding. Defective binding was next coupled to the fold change in RNA expression of that gene in a particular GFP-CTCF-ZF-expressing ES cell line, as compared to GFP-CTCF-expressing ES cells. Absolute differences in expression were obtained by making all negative values positive. This allowed us to compare transcriptional effects on genes to altered binding of GFP-CTCF-ZF mutants, separating the effects of CTCF binding sites near a TSS to those of genes with CTCF binding sites near an UC motif

References

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol 11, R106.

Baniahmad, A., Steiner, C., Kohne, A.C., and Renkawitz, R. (1990). Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid hormone receptor binding site. Cell 61, 505-514.

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, 823-837.

Boyle, A.P., Song, L., Lee, B.K., London, D., Keefe, D., Birney, E., Iyer, V.R., Crawford, G.E., and Furey, T.S. (2011). High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. Genome Res 21, 456-464.

Brown, C.R., Kennedy, C.J., Delmar, V.A., Forbes, D.J., and Silver, P.A. (2008). Global histone acetylation induces functional genomic reorganization at mammalian nuclear pore complexes. Genes Dev 22, 627-639.

Burcin, M., Arnold, R., Lutz, M., Kaiser, B., Runge, D., Lottspeich, F., Filippova, G.N., Lobanenkov, V.V., and Renkawitz, R. (1997). Negative protein 1, which is required for function of the chicken lysozyme gene silencer in conjunction with hormone receptors, is identical to the multivalent zinc finger repressor CTCF. Mol Cell Biol 17, 1281-1288.

Capelson, M., Liang, Y., Schulte, R., Mair, W., Wagner, U., and Hetzer, M.W. (2010). Chromatin-bound nuclear pore components regulate gene expression in higher eukaryotes. Cell 140, 372-383.

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, 1106-1117.

Chernukhin, I.V., Shamsuddin, S., Robinson, A.F., Carne, A.F., Paul, A., El-Kady, A.I., Lobanenkov, V.V., and Klenova, E.M. (2000). Physical and functional interaction between two pluripotent proteins, the Y-box DNA/RNA-binding factor, YB-1, and the multivalent zinc finger factor, CTCF. J Biol Chem 275, 29915-29921.

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

Cuddapah, S., Jothi, R., Schones, D.E., Roh, T.Y., Cui, K., and Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. Genome Res 19, 24-32.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376-380.

Dragestein, K.A., van Cappellen, W.A., van Haren, J., Tsibidis, G.D., Akhmanova, A., Knoch, T.A., Grosveld, F., and Galjart, N. (2008). Dynamic behavior of GFP-CLIP-170 reveals fast protein turnover on microtubule plus ends. J Cell Biol 180, 729-737.

Fedoriw, A.M., Stein, P., Svoboda, P., Schultz, R.M., and Bartolomei, M.S. (2004). Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting. Science 303, 238-240.

Feng, X., Grossman, R., and Stein, L. (2011). PeakRanger: a cloud-enabled peak caller for ChIP-seq data. BMC Bioinformatics 12, 139.

Filippova, G.N., Fagerlie, S., Klenova, E.M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P.E., Collins, S.J., and Lobanenkov, V.V. (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol Cell Biol 16, 2802-2813.

Goldberg, A.D., Banaszynski, L.A., Noh, K.M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X., et al. (2010). Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell 140, 678-691.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., et al. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature.

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

Handoko, L., Xu, H., Li, G., Ngan, C.Y., Chew, E., Schnapp, M., Lee, C.W., Ye, C., Ping, J.L., Mulawadi, F., et al. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. Nat Genet 43, 630-638.

Heath, H., Ribeiro de Almeida, C., Sleutels, F., Dingjan, G., van de Nobelen, S., Jonkers, I., Ling, K.W., Gribnau, J., Renkawitz, R., Grosveld, F., et al. (2008). CTCF regulates cell cycle progression of alphabeta T cells in the thymus. Embo J 27, 2839-2850.

Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E., and Wiehe, T. (2012). The chromatin insulator CTCF and the emergence of metazoan diversity. Proc Natl Acad Sci U S A 109, 17507-17512.

Heger, P., Marin, B., and Schierenberg, E. (2009). Loss of the insulator protein CTCF during nematode evolution. BMC Mol Biol 10, 84.

Holohan, E.E., Kwong, C., Adryan, B., Bartkuhn, M., Herold, M., Renkawitz, R., Russell, S., and White, R. (2007). CTCF genomic binding sites in Drosophila and the organisation of the bithorax complex. PLoS Genet 3, e112.

Ishihara, K., Oshimura, M., and Nakao, M. (2006). CTCF-dependent chromatin insulator is linked to epigenetic remodeling. Mol Cell 23, 733-742.

Jothi, R., Cuddapah, S., Barski, A., Cui, K., and Zhao, K. (2008). Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. Nucleic Acids Res 36, 5221-5231.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature 467, 430-435.

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, 1231-1245.

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, 7612-7624.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357-359.

Li, Y., Huang, W., Niu, L., Umbach, D.M., Covo, S., and Li, L. (2013). Characterization of constitutive CTCF/cohesin loci: a possible role in establishing topological domains in mammalian genomes. BMC Genomics 14, 553.

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, 1743-1753.

Lutz, M., Burke, L.J., Barreto, G., Goeman, F., Greb, H., Arnold, R., Schultheiss, H., Brehm, A., Kouzarides, T., Lobanenkov, V., et al. (2000). Transcriptional repression by the insulator protein CTCF involves histone deacetylases. Nucleic Acids Res 28, 1707-1713.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448, 553-560.

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

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, 165-170.

Moore, J.M., Rabaia, N.A., Smith, L.E., Fagerlie, S., Gurley, K., Loukinov, D., Disteche, C.M., Collins, S.J., Kemp, C.J., Lobanenkov, V.V., et al. (2012). Loss of maternal CTCF is associated with peri-implantation lethality of Ctcf null embryos. PLoS One 7, e34915.

Mullen, A.C., Orlando, D.A., Newman, J.J., Loven, J., Kumar, R.M., Bilodeau, S., Reddy, J., Guenther, M.G., DeKoter, R.P., and Young, R.A. (2011). Master transcription factors determine cell-type-specific responses to TGF-beta signaling. Cell 147, 565-576.

Nakahashi, H., Kwon, K.R., Resch, W., Vian, L., Dose, M., Stavreva, D., Hakim, O., Pruett, N., Nelson, S., Yamane, A., et al. (2013). A Genome-wide Map of CTCF Multivalency Redefines the CTCF Code. Cell Rep 3, 1678-1689.

Nemeth, A., Conesa, A., Santoyo-Lopez, J., Medina, I., Montaner, D., Peterfia, B., Solovei, I., Cremer, T., Dopazo, J., and Langst, G. (2010). Initial genomics of the human nucleolus. PLoS Genet 6, e1000889.

Ohlsson, R., Renkawitz, R., and Lobanenkov, V. (2001). CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. Trends Genet 17, 520-527.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W., Solovei, I., Brugman, W., Graf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M., et al. (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol Cell 38, 603-613.

Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. Cell 137, 1194-1211.

Phillips-Cremins, J.E., Sauria, M.E., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. Cell 153, 1281-1295.

Pugacheva, E.M., Kwon, Y.W., Hukriede, N.A., Pack, S., Flanagan, P.T., Ahn, J.C., Park, J.A., Choi, K.S., Kim, K.W., Loukinov, D., et al. (2006). Cloning and characterization of zebrafish CTCF: Developmental expression patterns, regulation of the promoter region, and evolutionary aspects of gene organization. Gene 375, 26-36.

Quitschke, W.W., Matthews, J.P., Kraus, R.J., and Vostrov, A.A. (1996). The initiator element and proximal upstream sequences affect transcriptional activity and start site selection in the amyloid beta-protein precursor promoter. J Biol Chem 271, 22231-22239.

Quitschke, W.W., Taheny, M.J., Fochtmann, L.J., and Vostrov, A.A. (2000). Differential effect of zinc finger deletions on the binding of CTCF to the promoter of the amyloid precursor protein gene. Nucleic Acids Res 28, 3370-3378.

Renda, M., Baglivo, I., Burgess-Beusse, B., Esposito, S., Fattorusso, R., Felsenfeld, G., and Pedone, P.V. (2007). Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. J Biol Chem 282, 33336-33345.

Rhee, H.S., and Pugh, B.F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell 147, 1408-1419.

Rugg-Gunn, P.J., Cox, B.J., Ralston, A., and Rossant, J. (2010). Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo. Proc Natl Acad Sci U S A 107, 10783-10790.

Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature 489, 109-113.

Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Goncalves, A., Kutter, C., Brown, G.D., Marshall, A., Flicek, P., and Odom, D.T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell 148, 335-348.

Sleutels, F., Soochit, W., Bartkuhn, M., Heath, H., Dienstbach, S., Bergmaier, P., Franke, V., Rosa-Garrido, M., van de Nobelen, S., Caesar, L., et al. (2012). The male germ cell gene regulator CTCFL is functionally different from CTCF and binds CTCF-like consensus sites in a nucleosome composition-dependent manner. Epigenetics Chromatin 5, 8.

Smith, S.T., Wickramasinghe, P., Olson, A., Loukinov, D., Lin, L., Deng, J., Xiong, Y., Rux, J., Sachidanandam, R., Sun, H., et al. (2009). Genome wide ChIP-chip analyses reveal important roles for CTCF in Drosophila genome organization. Dev Biol 328, 518-528.

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

Sun, S., Del Rosario, B.C., Szanto, A., Ogawa, Y., Jeon, Y., and Lee, J.T. (2013). Jpx RNA Activates Xist by Evicting CTCF. Cell 153, 1537-1551.

Thomas-Chollier, M., Sand, O., Turatsinze, J.V., Janky, R., Defrance, M., Vervisch, E., Brohee, S., and van Helden, J. (2008). RSAT: regulatory sequence analysis tools. Nucleic Acids Res 36, W119-127.

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, 1453-1465.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol 31, 46-53.

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

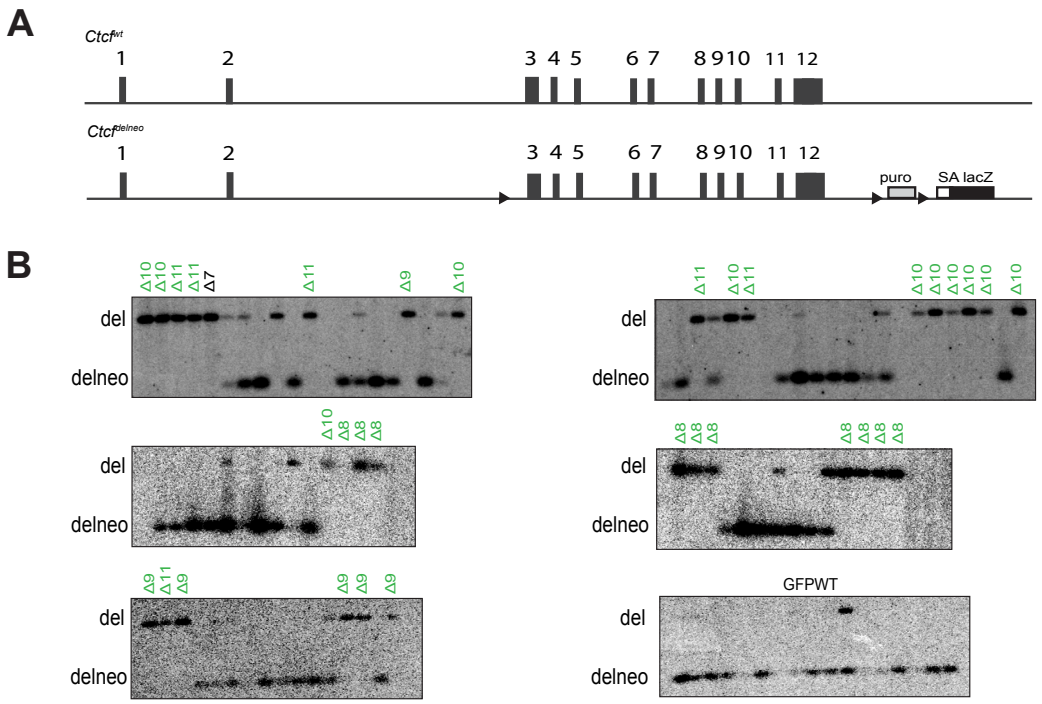
van Koningsbruggen, S., Gierlinski, M., Schofield, P., Martin, D., Barton, G.J., Ariyurek, Y., den Dunnen, J.T., and Lamond, A.I. (2010). High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. Mol Biol Cell 21, 3735-3748.

Vostrov, A.A., and Quitschke, W.W. (1997). The zinc finger protein CTCF binds to the APBbeta domain of the amyloid beta-protein precursor promoter. Evidence for a role in transcriptional activation. J Biol Chem 272, 33353-33359.

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, 7145-7150.

Zuin J, D.J.R., van der Reijden M.I.J.A., Ye Z., Kolovos P., Brouwer R.W.W., van de Corput M.P.C., van de Werken H.J.G., Knoch T.A., van IJcken W.F.J., Grosveld F.G., Ren B. and Wendt K.S. (Submitted). Cohesin and CTCF Differentially Affect Chromatin Architecture and Gene Expression in Human Cells

Supplemental Figure S1



C

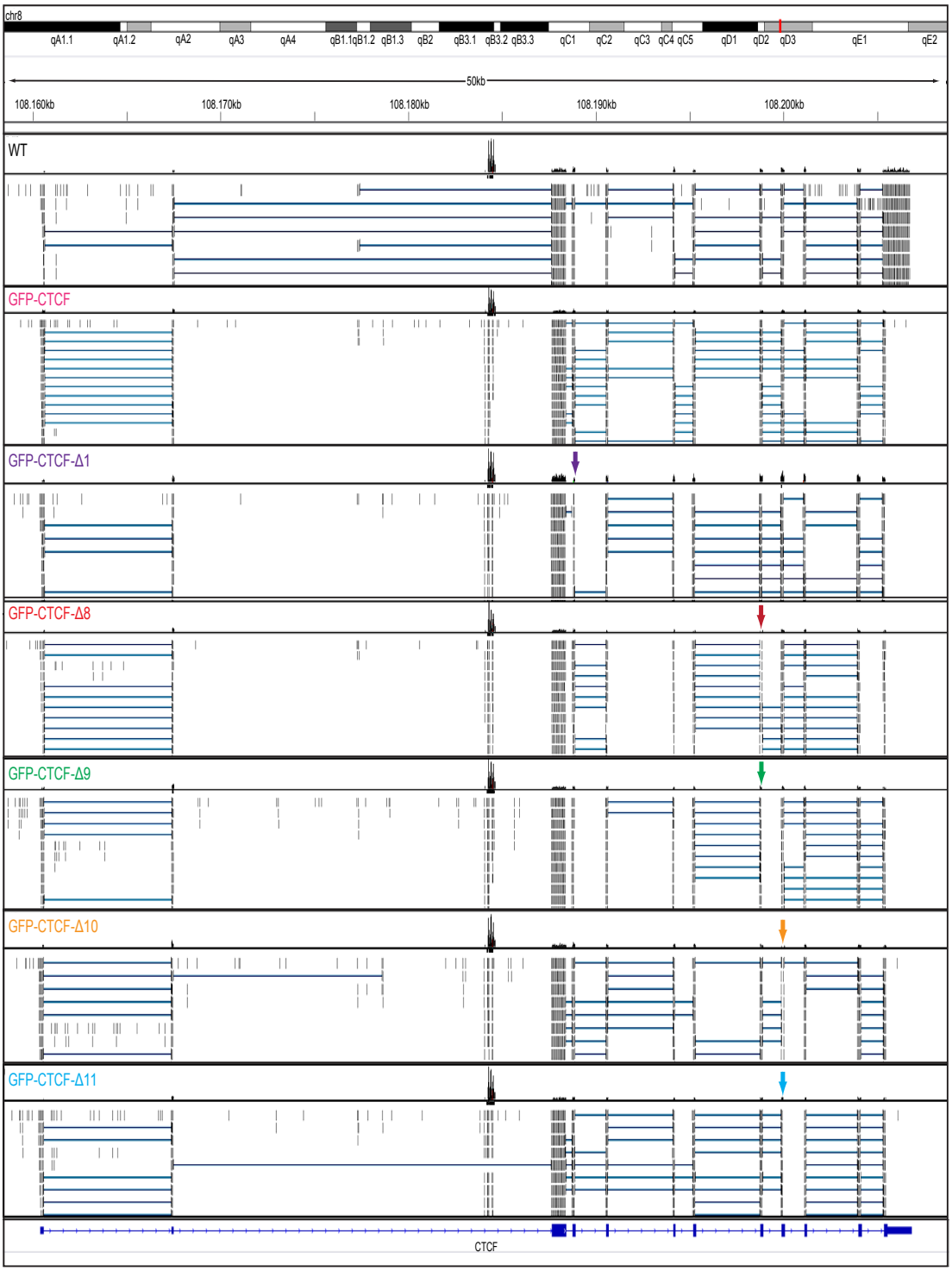
CTCF constructs	% Deleted (No. of colonies)	% Partially Deleted (No. of colonies)	% Not Deleted (No. of colonies)	Verified with sequencing	Functional CTCF substitution
GFP-CTCF	95% (40/42)	2% (1/42)	2% (1/42)	Yes	Yes
GFP-CTCF-Δ1	6.25% (1/16)	6.25% (1/16)	87.5% (14/16)	Yes	Yes
GFP-CTCF-Δ2	0% (0/4)	25% (1/4)	75% (3/4)	No	No
GFP-CTCF-Δ3	0% (0/9)	55% (5/9)	44%(4/9)	No	No
GFP-CTCF-Δ4	0% (0/6)	0% (0/6)	100% (6/6)	No	No
GFP-CTCF-Δ5	0% (0/7)	28% (2/7)	71% (5/7)	No	No
GFP-CTCF-Δ6	0% (0/5)	60% (3/5)	40% (2/5)	No	No
GFP-CTCF-Δ7	40% (2/5)	0% (0/5)	60% (3/5)	No	No
GFP-CTCF-Δ8	81% (9/11)	9% (1/11)	9% (1/11)	Yes	Yes
GFP-CTCF-Δ9	83% (5/6)	0% (0/6)	16% (1/6)	Yes	Yes
GFP-CTCF-Δ10	83% (10/12)	0% (0/12)	16% (2/12)	Yes	Yes
GFP-CTCF-Δ11	75% (6/8)	0% (0/8)	25% (2/8)	Yes	Yes

D

CTCF constructs	Expression (FPKM)
WT	54.66
GFP-CTCF	31.49
GFP-CTCF-Δ1	52.43
GFP-CTCF-Δ8	23.64
GFP-CTCF-Δ9	25.61
GFP-CTCF-Δ10	29.09
GFP-CTCF-Δ11	29.38

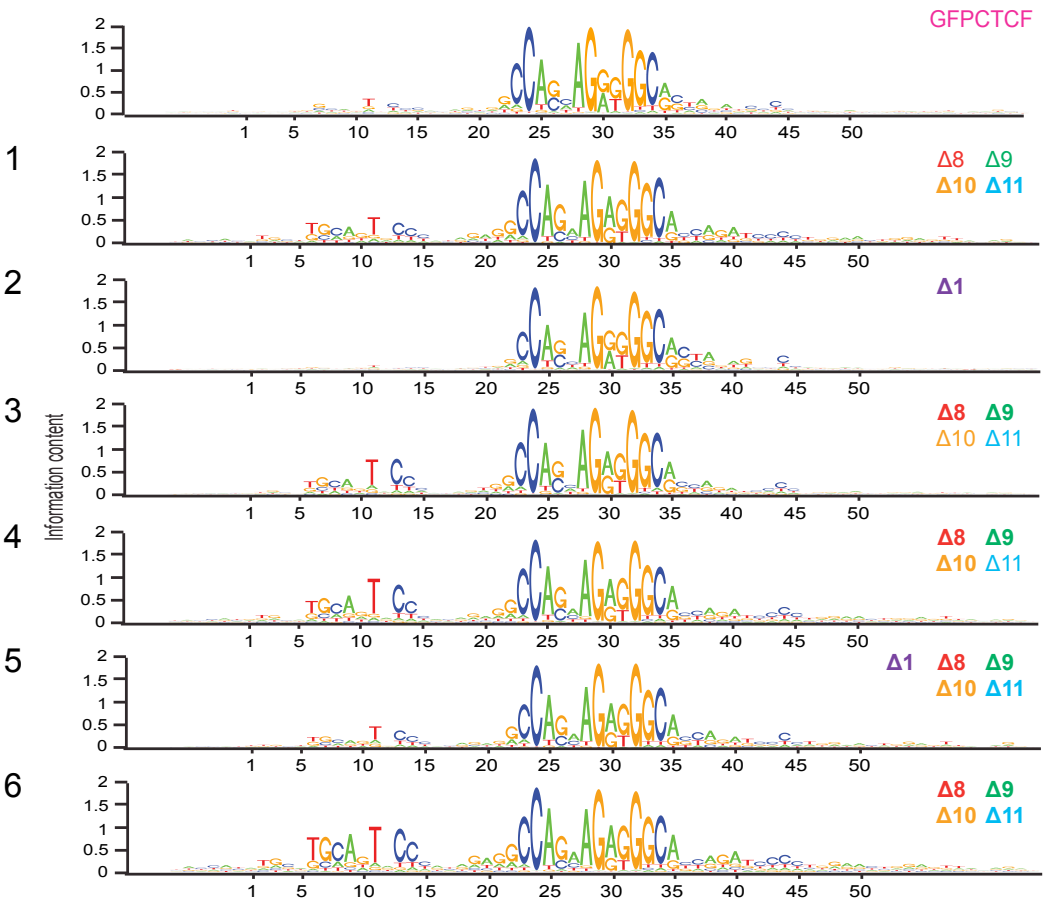
Supplemental Figure 1. Characterization of CTCF mutant ES cell lines
(A) Schematic representation of *Ctcf*^{wt} and *Ctcf*^{delneo} alleles. Exons are represented by rectangles and numbered. Exon 3 and 12 contain start and stop codon, respectively. The *Ctcf*^{delneo} allele contains a single loxP site at the 5'end (small triangle), and two more loxP sites (small triangles) flanking a PGK-puromycin cassette (puro) at the 3'end. A LacZ reporter cassette containing a splice acceptor (SA) site is also present at the 3'end of the gene.
(B) Southern blot analysis of *Ctcf*^{delneo} (WT) and *Ctcf*^{del} alleles in neomycin-resistant and GFP-positive clones. Rescued cell lines are marked in green.
(C) Overview of results of the ES cell rescue experiment. Percentage (and number) of colonies is shown in which both, one or none of the endogenous *Ctcf*^{delneo} alleles was deleted. Colonies were analyzed with DNA and RNA sequencing to verify mutation. The final conclusion regarding functional CTCF substitution is depicted in the right hand column.
(D) RNA levels of wild type and mutant CTCF encoding RNAs. RNA-sequencing was performed on RNA from ES cells expressing endogenous (WT) or the indicated GFP-CTCF proteins. Values are depicted in FPKM (Fragments Per Kilobase of transcripts per Million mapped reads).

Supplemental Figure S2



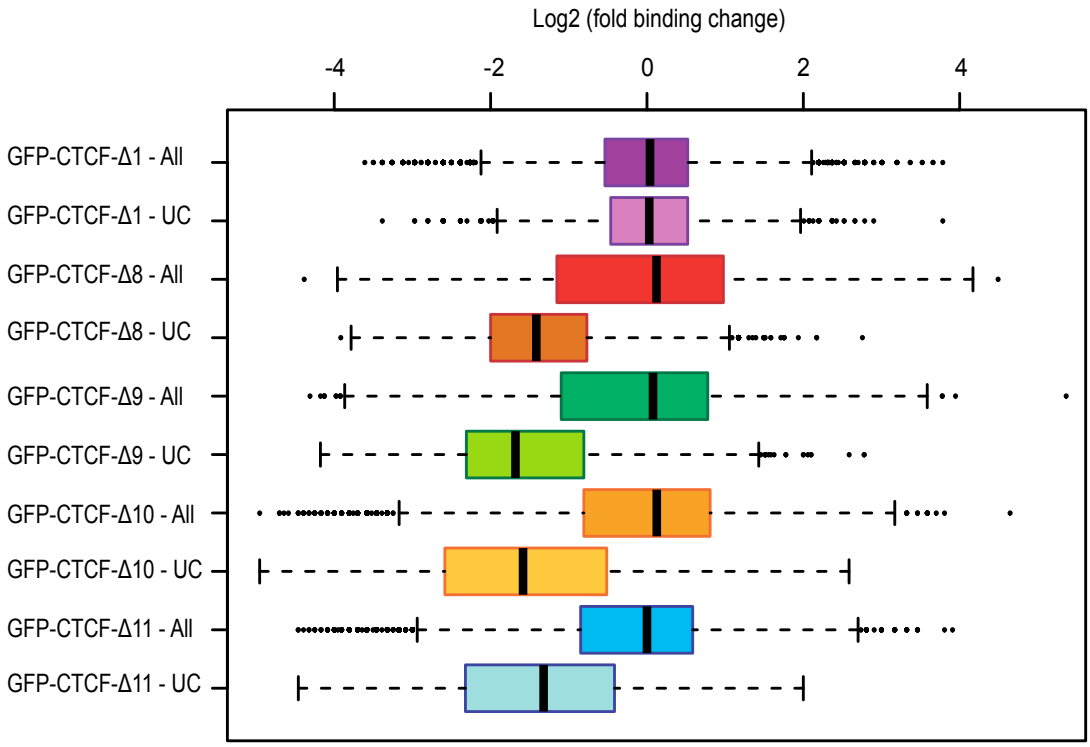
Supplemental Figure 2. RNA-sequencing profile shows zinc finger deletion
IGV profile of the *Ctcf* gene together with the mapped reads of each GFP-CTCF mutant. Arrows show the location of each ZF deletion confirming that each mutant cell line deleted endogenous *Ctcf* alleles and express only mutant CTCF.

Supplemental Figure S3



Supplemental Figure 3. Characterization of GFP-CTCF-ZF binding
Motif analysis of random GFP-CTCF binding sites and the 6 subgroups shown in **Figure 3**. This reveals a strong upstream motif in clusters 3, 4, and 6 (affected binding sites of GFP-CTCF-Δ8-11) and nucleotides downstream of the core motif in clusters 1, 2 (affected binding sites of GFP-CTCF-Δ1), and 6. Bold text indicates most affected mutant(s) in each cluster. Red, green, yellow and blue represent T, A, G, and C nucleotides, respectively.

Supplemental Figure S4



Supplemental Figure 4. Reduced binding of GFP-ZFΔ8-11 to (UC) motif-containing CTCF sites
Fold change in binding to all CTCF target sites (All) and upstream motif-containing CTCF sites (UC) is depicted for each mutant. Log2 fold binding changes for these two groups are plotted (Y-axis). ZF8-11 show strongly reduced binding to UC sites whereas ZF1 does not.

Supplemental Table S1

ZF deletion	Deleted amino acid sequence	AA position	No of AA
ZF1	TFQCELCSYTCPRRSNLDRHMKSHTD	265-290	26
ZF2	RPHKCHLCGRAFRTVTLLRNHLNTHTGT	292-319	28
ZF3	RPHKCPDCDMAFVTSGELVRHRRYKHTHE	320-348	29
ZF4	KPFKCSMCDYASVEVSKLKRHIRSHTGE	349-376	28
ZF5	RPFQCSLCSYASRDTYKLKRHMRTHSGE	377-404	28
ZF6	EKPYECYICHARFTQSGTMKMHILQKHTEN	404-433	30
ZF7	VAKFHCPHCDTVIARKSDLGVHLRKQHSYI	434-463	30
ZF8	QGKKCRYCDAVFHERYALIQHQKSHKNE	465-492	28
ZF9	RFKCDQCDYACRQERHMIMHKRTHTGEK	494-521	28
ZF10	KPYACSHCDKTFRQKQLLDMHFKRYHDPNFV	521-551	31
ZF11	PAAFVCSKCGKTFTRRNTMARHADNCAG	552-579	28

Supplemental Table 1. Zing finger amino acid sequence
Amino acid sequence and position number of each zinc finger. Number of amino acids deleted in each zinc finger are described in outer right column.

Supplemental Table S2

CTCF constructs	CTCF Expression (FPKM)	Pou5f1 Expression (FPKM)	Nanog Expression (FPKM)	Sox2 Expression (FPKM)	Alpl Expression (FPKM)
WT	54.66	1173.69	173.85	195.15	73.7
GFP-CTCF	31.49	1005.14	193.06	216.97	56.3
GFP-CTCF-Δ1	52.43	1343.45	222.64	186.37	67.24
GFP-CTCF-Δ8	23.64	1187.15	95.85	178.31	93.77
GFP-CTCF-Δ9	25.61	890.61	129.89	183.33	71.31
GFP-CTCF-Δ10	29.09	1149.04	119.72	185.74	74.53
GFP-CTCF-Δ11	29.38	897	86.7	120.88	40.95

Supplemental Table 2. Gene expression levels from RNA-Sequencing
Gene expression levels of *Ctcf*, *Pou5f1* (OCT4), *Nanog*, *Sox2* and *Alpl* by RNA-Sequencing. Values are depicted in FPKM (Fragments Per Kilobase of transcripts per Million mapped reads). Full excel sheet is available upon request.

Supplemental Table S3

Zinc finger primer name	Primer sequence
Zinc finger 1 F	TTAAAAAAAAAGGTGTAAAGAAAGAGAGACCACACAAATGCCAC
Zinc finger 1 R	GTGGCATTTGTGTGGTCTCTCTTTCTTTACACCTTTTTTTTTAA
Zinc finger 2 F	CATGAAAAGCCACACTGATGAGCGTCCTCACAAGTGCCCAGAC
Zinc finger 2 R	GTCTGGGCACTTGTGAGGACGCTCATCAGTGTGGCTTTTCATG
Zinc finger 3 F	ATCTGAACACACACACAGGTACTAAACCATTTAAGTGTTCCATG
Zinc finger 3 R	CATGGAACACTTAAATGGTTTAGTACCTGTGTGTGTGTTCAGAT
Zinc finger 4 F	CGTTATAAACACACTCATGAGCGCCCGTTCCAGTGCAGTTTG
Zinc finger 4 R	CAAACTGCACTGGAACGGGCGCTCATGAGTGTGTTTATAACG
Zinc finger 5 F	CATTCGCTCTCATACTGGAGAGTGTTATATTTGTCACGCTCGG
Zinc finger 5 R	CCGAGCGTGACAAATATAACACTCTCCAGTATGAGAGCGAATG
Zinc finger 6 F	CATTCAGGGGAAAAACCTTATGTGGCCAAATTTCATTGTCCC
Zinc finger 6 R	GGGACAATGAAATTTGGCCACATAAGGTTTTTCCCCTGAATG
Zinc finger 7 F	TTACAGAAGCACACAGAAAATGAACAGGGCAAAAAATGTCGC
Zinc finger 7 R	GCGACATTTTTTGCCCTGTTCATTTTCTGTGTGCTTCTGTAA
Zinc finger 8 F	AAGCAGCATTCCTATATTGAAAAGCGCTTCAAGTGTGACCAG
Zinc finger 8 R	CTGGTCACACTTGAAGCGCTTTTCAATATAGGAATGCTGCTT
Zinc finger 9 F	TCAAAAATCACACAAAAATGAGAAGCCTTATGCCTGCAGCCAC
Zinc finger 9 R	GTGGCTGCAGGCATAAGGCTTCTCATTTTTGTGTGATTTTTGA
Zinc finger 10 F	CAAGCGCACTCACACGGGGGAGCCTGCTGCCTTTGTCTGTTCC
Zinc finger 10 R	GGAACAGACAAAGGCAGCAGGCTCCCCCGTGTGAGTGCGCTTG
Zinc finger 11 F	CTATCATGATCCCAACTTTGTCCCAGATGGCGTAGAGGGGGAA
Zinc finger 11 R	TTCCCCCTCTACGCCATCTGGGACAAAGTTGGGATCATGATAG
Znmutforward	CACGCTCGAGCATGGAAGGTGAGGCGGTTGAAG
Znmutreverse	TCCGCTCGAGTCACCGGTCCATCATGCTGAGG

Supplemental Table 3. Fusion primers used to generate zinc finger deletions
Green: XhoI rescrtiction site, red: start codon

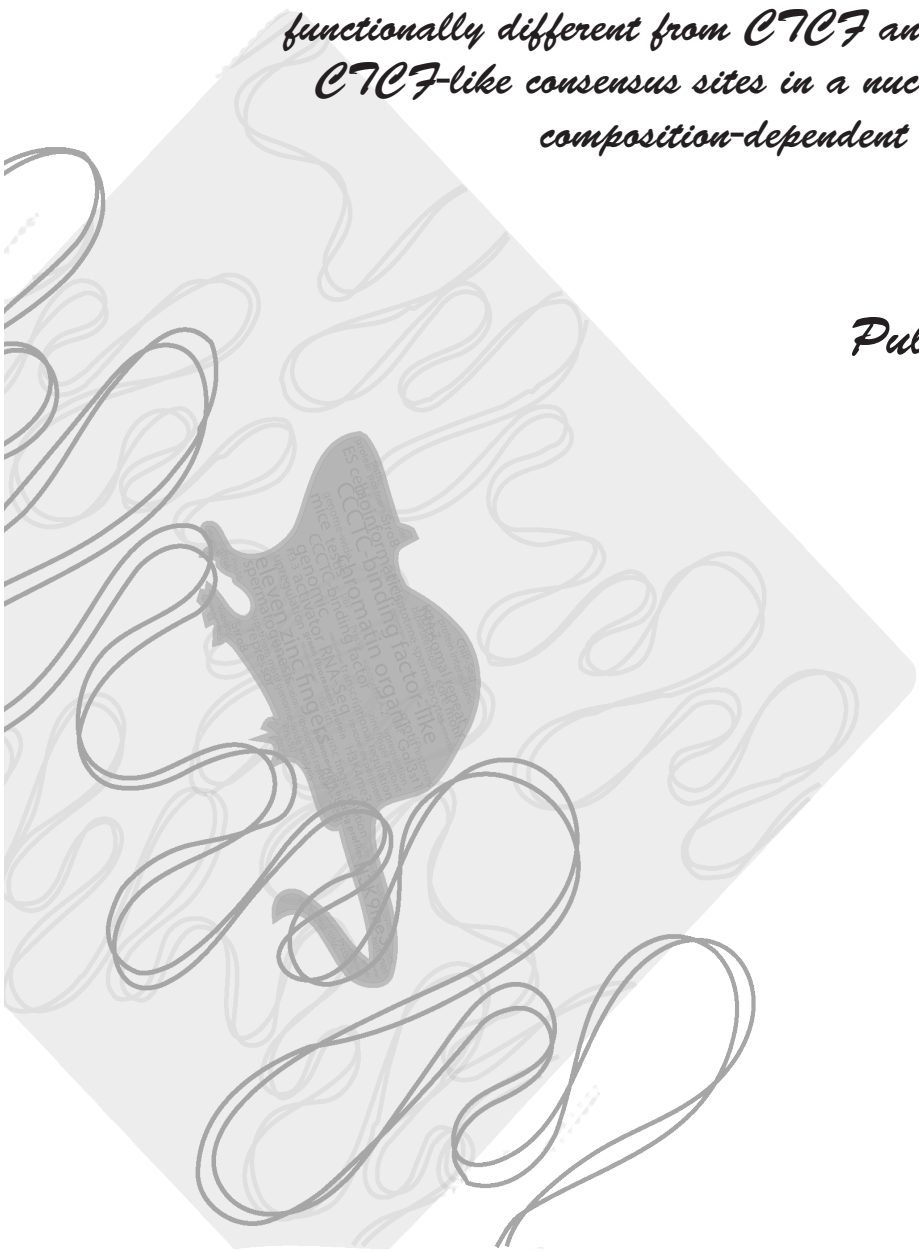
Supplemental Table S4

ChIP primer name	Primer sequence
Amylase_ChIP_F	AATTCTCCTTGTACGGGTTGGTG
Amylase_ChIP_R	TAGCAATGATGTGCACAGCTGAA
Pcdha6_ChIP_F	TAATGATAACCCGCCTGTGTTC
Pcdha6_ChIP_R	TCGCTGGGACTGAGTTTATAGG
Pcdhac1_ChIP_F	GTCCAGAGCCAGACAAAGGAC
Pcdhac1_ChIP_R	CATTGCAAGCTCCAAGTGTTG
Hs5-1_ChIP-F	TGGCCATGGAGATTTTCTTTAC
Hs5-1_ChIP_R	CTAGCCTTTCCTCATCTTCCAG
Pcdhb3_ChIP_F	GATCAAAACCAGTGACAAAACC
Pcdhb3_ChIP_R	ATCTACCCAGGTTCACACAAGC
Pcdhga3_ChIP_F	AACAAAGGGTAACTTGGGTTTG
Pcdhga3_ChIP_R	TAGCTTTACAGCAGGGAAATGG
Pcdhga12_ChIP_F	GGTTCACAGACCAAAAATCCTC
Pcdhga12_ChIP_R	ATCTGTCCGCATTGTGTTTTAG
Pcdhgc3_promoter _ChIP_F	CGTGTCCTGCTCAAAGAACC
Pcdhgc3_promoter _ChIP_R	TGTTCAGTCTCTGCCTCATCC
Pcdhgc3_downstream _ChIP_F	GCGAAGTTGTGATCCTGTGTTC
Pcdhgc3_downstream _ChIP_R	GACCCATACCTCTGTGAAGGAG

Supplemental Table 4. ChIP-pPCR primers

Chapter 4

The male germ cell gene regulator CTCFL is functionally different from CTCF and binds CTCF-like consensus sites in a nucleosome composition-dependent manner



Published

Sleutels et al. Epigenetics & Chromatin 2012, 5:8
http://www.epigeneticsandchromatin.com/content/5/1/8

Epigenetics & Chromatin

RESEARCH Open Access

The male germ cell gene regulator CTCFL is functionally different from CTCF and binds CTCF-like consensus sites in a nucleosome composition-dependent manner

Frank Sleutels[1*], Widia Soochit[1], Marek Bartkuhn[2], Helen Heath[1], Sven Dienstbach[2], Philipp Bergmaier[2], Vedran Franke[3], Manuel Rosa-Garrido[4,5], Suzanne van de Nobelen[1], Lisa Caesar[1], Michael van der Reijden[1], Jan Christian Bryne[3], Wilfred van IJcken[6], J Anton Grootegoed[7], M Dolores Delgado[4], Boris Lenhard[3], Rainer Renkawitz[2], Frank Grosveld[1,8,9] and Niels Galjart[1,8,9*]

Abstract

Background: CTCF is a highly conserved and essential zinc finger protein expressed in virtually all cell types. In conjunction with cohesin, it organizes chromatin into loops, thereby regulating gene expression and epigenetic events. The function of CTCFL or BORIS, the testis-specific paralog of CTCF, is less clear.

Results: Using immunohistochemistry on testis sections and fluorescence-based microscopy on intact live seminiferous tubules, we show that CTCFL is only transiently present during spermatogenesis, prior to the onset of meiosis, when the protein co-localizes in nuclei with ubiquitously expressed CTCF. CTCFL distribution overlaps completely with that of Stra8, a retinoic acid-inducible protein essential for the propagation of meiosis. We find that absence of CTCFL in mice causes sub-fertility because of a partially penetrant testicular atrophy. CTCFL deficiency affects the expression of a number of testis-specific genes, including Gal3st1 and Prss50. Combined, these data indicate that CTCFL has a unique role in spermatogenesis. Genome-wide RNA expression studies in ES cells expressing a V5- and GFP-tagged form of CTCFL show that genes that are downregulated in CTCFL-deficient testis are upregulated in ES cells. These data indicate that CTCFL is a male germ cell gene regulator. Furthermore, genome-wide DNA-binding analysis shows that CTCFL binds a consensus sequence that is very similar to that of CTCF. However, only ~3,700 out of the ~5,700 CTCFL- and ~31,000 CTCF-binding sites overlap. CTCFL binds promoters with loosely assembled nucleosomes, whereas CTCF favors consensus sites surrounded by phased nucleosomes. Finally, an ES cell-based rescue assay shows that CTCFL is functionally different from CTCF.

Conclusions: Our data suggest that nucleosome composition specifies the genome-wide binding of CTCFL and CTCF. We propose that the transient expression of CTCFL in spermatogonia and preleptotene spermatocytes serves to occupy a subset of promoters and maintain the expression of male germ cell genes.

Keywords: CTCF, CTCFL, Gametogenesis, Genome-wide binding, Nucleosome

* Correspondence: f.sleutels@erasmusmc.nl; n.galjart@erasmusmc.nl
[1]Department of Cell Biology Erasmus Medical Center, Rotterdam, The Netherlands
[8]Cancer Genomics Center, NGI, Rotterdam, The Netherlands
Full list of author information is available at the end of the article

Background

Three-dimensional folding of the eukaryotic genome occurs in a highly organized manner so as to compact chromatin while allowing temporal and spatial expression of genes. The genome contains regulatory elements, such as promoters, enhancers, locus control regions, insulators and enhancer blockers, that can orchestrate chromatin folding and gene activity over short and long distances, both in cis and in trans [1]. CTCF is a key coordinator of three-dimensional chromatin structure, allowing loop formation and specific chromatin compositions [2,3]. Gene activity is controlled in a positive or negative manner depending on the regulatory sequences present in the loops that are formed. The importance of CTCF in chromatin organization is emphasized by its evolutionary conservation, its ubiquitous expression, and its essential role in virtually all cells and tissues examined [4,5]. Hence, CTCF has been termed the "master weaver" of the genome [3].

The genome-wide binding by CTCF has been studied by different groups (see, for example, [6-10]). This has revealed ~35,000 CTCF-binding sites in the mammalian genome, of which more than 70% are shared between cell types. A relatively long consensus-binding motif for CTCF has been determined, which displays variability when compared to sites of transcription factors like KLF4, SOX2 and MYC [7]. The majority of CTCF binding-sites are found near genes, and ~8% is in the vicinity of transcription start sites (TSSs). Arrays of positioned (or 'phased') nucleosomes are found surrounding the nucleosome-free CTCF-binding sites [11-13], suggesting that CTCF binding promotes the ordered positioning of histones in its vicinity. CTCF has also been proposed to regulate the positioning of variant histones, such as H2A.Z [6,14]. Interestingly, the cohesin complex binds at the same position as CTCF in a CTCF-dependent manner. Together with CTCF, cohesin is essential for a proper three-dimensional chromatin structure and correct gene regulation [15-17].

CTCF-dependent loop formation is of crucial importance at imprinted loci. A well-studied example is the imprinted *Igf2-H19* locus, in which *Igf2* is expressed from the paternal and *H19* from the maternal allele [18]. The imprinting control region (ICR) located in between the *Igf2* and *H19* genes is methylated on the paternal allele, preventing CTCF binding. As a result the enhancer downstream of the *H19* gene can interact with the *Igf2* promoter and drive expression of this gene. On the non-methylated maternally derived ICR, CTCF does bind, thereby preventing enhancer-*Igf2* interaction, resulting in a chromatin loop that allows enhancer-*H19* association and *H19* expression. By binding the ICR, CTCF therefore acts as a regulator of imprinted sites.

The CTCF-like (CTCFL) protein, or Brother Of the Regulator of Imprinted Sites (BORIS) [19], has a central domain of 11 zinc fingers (ZFs) that is very similar to that of CTCF and that is essential for DNA binding. The N- and C-terminal domains of CTCF and CTCFL are not homologous. CTCFL is less conserved across species, and the protein arose later in evolution, as it is detected in amniotes only [20]. Furthermore, expression of CTCFL is restricted to testis, several types of cancers and a number of cell lines [21-23].

Studies of CTCF and CTCFL protein distribution in the testis have yielded contradictory results. Initially, a mutually exclusive expression pattern of CTCFL and CTCF was described [19], with CTCF being present in round spermatids (i.e. after meiosis) and CTCFL in primary spermatocytes (i.e. during meiotic prophase).

Surprisingly, CTCFL was reported to be more abundant in the spermatocyte cytoplasm than in the nucleus. This led to the hypothesis that during germ cell development, CTCFL substitutes for the absence of CTCF and might be involved in reprogramming of DNA methylation in the male germ line. CTCFL was later reported to be present in gonocytes during embryonic development and, after birth, in spermatogonia, whereas CTCF was reported to localize to the supporting Sertoli cells [24]. In the same study CTCFL, together with the protein methyltransferase PRMT7, was suggested to regulate DNA methylation of imprinted genes in the male germline. However, defects in imprinting often result in embryonic phenotypes [25], whereas Ctcfl knockout mice were shown to display a phenotype only in the testis [26]. Recently, enrichment of *Ctcfl* mRNA in round spermatids was reported, adding perplexity to the localization and expression of CTCFL [26,27].

While the whole genome DNA-binding profile for CTCF has been elucidated, this has not been done for CTCFL. It therefore remains unclear how CTCFL binding relates to that of CTCF. In addition, it is unknown how these proteins are related functionally and mechanistically. To address these issues, we examined CTCFL function and localization with respect to CTCF, and identified the genome-wide binding sites of CTCFL and CTCF. We show that CTCF and CTCFL are functionally different proteins that co-localize within the nuclei of pre-meiotic germ cells. CTCFL acts as a male germ cell gene regulator, preferably binding near promoters with active chromatin marks. Interestingly, CTCF and CTCFL bind a highly similar DNA motif; nevertheless, only two-third of the ~5,700 CTCFL-binding sites are bound by CTCF. Conversely, the vast majority of CTCF sites are not bound by CTCFL. We find that nucleosome composition specifies CTCF and CTCFL binding. In contrast to CTCF, CTCFL associates with relatively "open" chromatin, and we propose that CTCFL promotes the maintenance of the epigenetic state of a subset of gene promoters and hence gene expression during spermatogenesis.

Results

CTCFL and CTCF co-localize transiently in pre-meiotic male germ cells

To resolve the localization of CTCF and CTCFL in testis, and to address CTCFL function, we generated *Ctcfl* knockout and GFP-CTCF- and GFP-CTCFL-expressing knockin mice. To obtain information about the organization of the *Ctcfl* gene, we mapped its 5' end and examined *Ctcfl* expression (Figure 1A, B). We next generated three separate alleles using homologous recombination in ES cells: a *Ctcfl* knockout allele ($Ctcfl^{del}$), in which exons 1–8 of the Ctcfl gene are deleted, and Ctcfl and Ctcf knockin alleles ($Ctcfl^{gfp}$ and $Ctcf^{gfp}$, respectively), to express GFP-CTCFL and GFP-CTCF instead of CTCFL and CTCF, respectively (Figure 1C-I).

Mice were generated, and the distribution of CTCFL was investigated by immunocytochemistry in sections of seminiferous tubules from wild-type and Ctcfl knockout mice. CTCFL was present in wild-type testis in cells lining the basal lamina (Figure 2A, B). Not all cells lining the lamina were CTCFL-positive, and in some tubules no CTCFL-positive cells were detected. Importantly, no signal was detected on sections derived from CTCFL-deficient mice (Figure 2C, see also Figure 2F), showing that the

CTCFL staining in wild-type sections is specific. The localization of the CTCFL-positive cells in the basal compartment of the seminiferous tubules indicates that these cells are spermatogonia or preleptotene spermatocytes, as only upon progression in meiotic prophase do spermatocytes become disconnected from the basal lamina and move through the Sertoli cell barrier into the adluminal compartment of the seminiferous tubules.

The localization of CTCFL appeared reminiscent of STRA8 (STimulated by Retinoic Acid), which is expressed transiently from B spermatogonia to preleptotene spermatocytes and is essential for retinoic acid-induced commitment to meiosis [29-32]. An absolute overlap between STRA8 and CTCFL was confirmed using dual-color immunofluorescence (Figure 2D, E). Immunofluorescent staining experiments did not reveal an obvious change in the number of STRA8-positive tubules in CTCFL-deficient testis (Figure 2E, G, and data not shown). Thus, absence of the CTCFL signal in Ctcfl knockout sections is not due to the disappearance of a cell type.

To confirm CTCFL localization and compare its distribution to that of CTCF, we next analyzed expression and localization of the two proteins ex vivo. We isolated intact seminiferous tubules from the testes of $Ctcfl^{gfp}$ and $Ctcf^{gfp}$ male mice, which were injected with Hoechst via the rete testis to stain nuclei of cells at the adluminal compartment of the tubule. We then visualized GFP-CTCF(L) and Hoechst concomitantly using a multiphoton confocal laser scanning microscope setup [33]. Three-dimensional reconstruction of images taken longitudinally through the seminiferous tubules yielded an organizational view of the tubule, and the position and type of the GFP-positive cells (Figure 2H-P). GFP-CTCFL was detected in the nucleus of clustered cells representing a minor fraction of the total testis tubule (Figure 2H-m, and Additional file 1: Movie S1). These cells stained negative for Hoechst, and since the luminally injected Hoechst does not cross the Sertoli cell barrier, the GFP-CTCFL-positive cells must reside on the basal side of this barrier. Sertoli cells, which form the tight junctions of the Sertoli cell barrier, were Hoechst-positive (Figure 2I, J, L, M and Additional file 1: Movie S1). Primary spermatocytes pass this barrier in the preleptotene and leptotene stage [34]. Based on Hoechst staining, morphology, size and location, we conclude that the GFP-CTCFL positive cells represent spermatogonia and preleptotene spermatocytes. The ex vivo GFP/Hoechst results obtained in live tissue are consistent with our data obtained in fixed paraffin-embedded sections of the testis stained with the CTCFL antibodies (Figure 2A-C). Together with the STRA8 colocalization data, they strongly suggest that in the adult testis CTCFL is transiently expressed in late spermatogonia and preleptotene germ cells.

In contrast to GFP-CTCFL, GFP-CTCF was present in the nucleus of all cell types of the seminiferous tubule, including all germ cells prior to spermiogenesis (Figure 2N-P, and data not shown). GFP-CTCF was also expressed in round spermatids, albeit at lower levels. This is consistent with a primary role for CTCF in cells with histone-based chromatin. Thus, live imaging in seminiferous tubules shows that CTCF and CTCFL are co-expressed within late spermatogonia and preleptotene spermatocytes. Measurement of GFP fluorescence intensities indicate that in cells where both proteins are expressed, the level of CTCF is somewhat higher than that of CTCFL.

CTCFL is important for spermatogenesis

To study the role of CTCFL in the male germ line, we analyzed $Ctcfl^{del/+}$ and $Ctcfl^{del/del}$ mice. These mice demonstrated no gross phenotypic defects and appeared normal. Heterozygous and homozygous $Ctcfl^{del}$ females showed normal fertility and yielded offspring with expected ratios (data not shown), consistent with a role for CTCFL in spermatogenesis only. Heterozygous $Ctcfl^{del/+}$ males generated offspring, and demonstrated normal fertility (Table 1). However, homozygous $Ctcfl^{del}$ male littermates generated offspring in only half (14 out of 27) of the breedings (Table 1). Breeding with $Ctcfl^{del/del}$ males yielded significantly ($p = 0.01$; chi test) fewer litters than $Ctcfl^{del/+}$ males, but not a different litter size ($p = 0.11$; *t*-test). These data indicate that CTCFL is important for male fertility.

To further investigate the CTCFL deficiency, we weighed testes from 90-day-old $Ctcfl^{del/del}$ and $Ctcfl^{del/+}$ mice and plotted weight distributions. We found that, on average, the

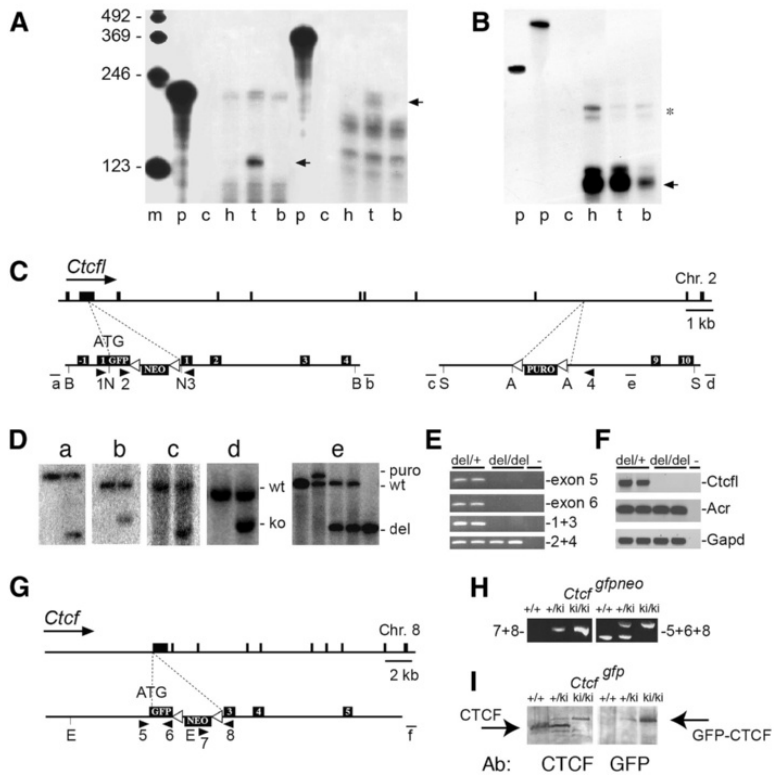


Figure 1 Ctcfl and Ctcf expression and targeting. A, **B** RNAse protection analysis of Ctcfl and Ctcf. For Ctcfl (**A**) RNase protection analysis (RPA) was performed on polyA purified mRNA with probes covering parts of Ctcfl exon 8 and 9 (*left*, small fragment) or a 5'end RACE product (*right*, large fragment). For Ctcf (**B**) the RPA was performed on total RNA with probes protecting Ctcf exon 2. The positions of the respective protected fragments are indicated with *arrows*. *Ctcfl* mRNA mRNA can only be detected in polyA purified mRNA from testis (*t*), whereas Ctcf is identified in total RNA from all three tissues tested. *M*, marker; *p*, input probe; *c*, tRNA control; *h*, heart; *t*, testis; *b*, brain. Aprt exon 3 is used as loading control and marked by an *asterisk* [28]. This analysis identifies the first exon containing the ATG translation initiation codon in Ctcfl and shows that Ctcfl is expressed in testis. **C** Schematic overview of the modified Ctcfl alleles and targeting constructs. The Ctcfl locus is shown on scale, with the constructs (not on scale) used for homologous recombination in ES cells underneath. Targeting at the 5'end of Ctcfl yielded the $Ctcfl^{gpf-\ neo}$ allele. Cre-mediated excision of the LoxP-embedded neomcyin resistance gene yielded the $Ctcfl^{gfp}$ allele (not shown). The 3'end targeting was performed on the $Ctcfl^{gpf-\ neo}$ allele, and yielded the $Ctcfl^{gfp\text{-neo-puro}}$ allele. Cre-mediated excision of the sequence in between the outermost LoxP sites yielded the $Ctcfl^{del}$ allele, in which exons 1–8 of the Ctcfl gene are deleted (not shown). A major difference between the $Ctcfl^{del}$ allele described here and the Ctcfl knockout published earlier [26] is that in the $Ctcfl^{del}$ allele the GFP coding sequence is fused in frame with the CTCFL coding sequence. *Black boxes* represent exons, GFP tag, neomycin and puromycin cassettes. Probes a, b, c, d and e are indicated by *lines*. Oligos 1, 2, 3 and 4 are represented by *arrowheads*. *White triangles* are LoxP sites. *B* = BglII; *N* = NcoI; *S* = SpeI; *A* = AvrII. **D** DNA blot showing Ctcfl targeting. Probes a and b were used on DNA blots from ES cells for identification of the $Ctcfl^{gfp\text{-neo}}$ allele and probes c and d for the $Ctcfl^{puro}$ allele. Probe e identifies the $Ctcfl^{del}$ allele from $Ctcfl^{gfp\text{-neo-puro}}$ mice that were crossed to a chicken Actin-Cre transgene. Probe a, HindIII digest (wt 8.9 kb, ko 5.7 kb); probe b, EcoRI digest (wt 14 kb; ko 11 kb); probe c, BamHI digest (wt 16.1 kb; ko 6.8 kb); probe d, BamHI digest (wt 16.1 kb; ko 11.1 kb). **E** Absence of Ctcfl DNA in the $Ctcfl^{del}$ allele. PCR on tail DNA indicates that $Ctcfl^{del/del}$ mice are deleted for exons 1–8 (*top three panels*) and are positive for GFP (oligos 2 and 4). **F** Absence of Ctcfl RNA in Ctcfl mutant mice. PCR on cDNA derived from testis mRNA shows that Ctcfl is absent from $Ctcfl^{del/del}$ mice. Acrosin and Gapd function as positive controls. **G** Schematic overview of the Ctcf allele and targeting strategy for the $Ctcf^{gfp\text{-neo}}$ allele. The Ctcf locus is shown on scale, with the construct (not on scale) used for homologous recombination in ES cells underneath. Cre-mediated excision of the LoxP-embedded neomcyin resistance gene yielded the $Ctcf^{gfp}$ (or $Ctcf^{ki}$) allele (not shown). *Black boxes* represent exons, GFP tag and neomycin cassette. Oligos 5, 6, 7 and 8 are represented by *arrowheads*. *White triangles* are LoxP sites. *E* = EcoRI. **H** PCR confirming $Ctcf^{gpf\text{-neo}}$ allele. Identification of the $CTCF^{gfp\text{-neo}}$ (or $Ctcf^{ki}$) allele by PCR with oligos 7 and 8 or oligos 5, 6 and 8 (see panel **G**). **I** Western blot confirming GFP-CTCF expression from the $Ctcf^{gfp}$ allele. We isolated MEFS from E13.5 day wild-type (+/+), heterozygous $Ctcf^{gfp/+}$ (or $Ctcf^{ki/+}$) or homozygous $Ctcf^{gfp/gfp}$ (or $Ctcf^{ki/ki}$) embryos, and identified the GFP-CTCF fusion protein by Western blot of MEF extracts using anti-CTCF or anti-GFP antibodies. Note the increased size of the GFP-CTCF protein compared to the CTCF protein due to the GFP tag.

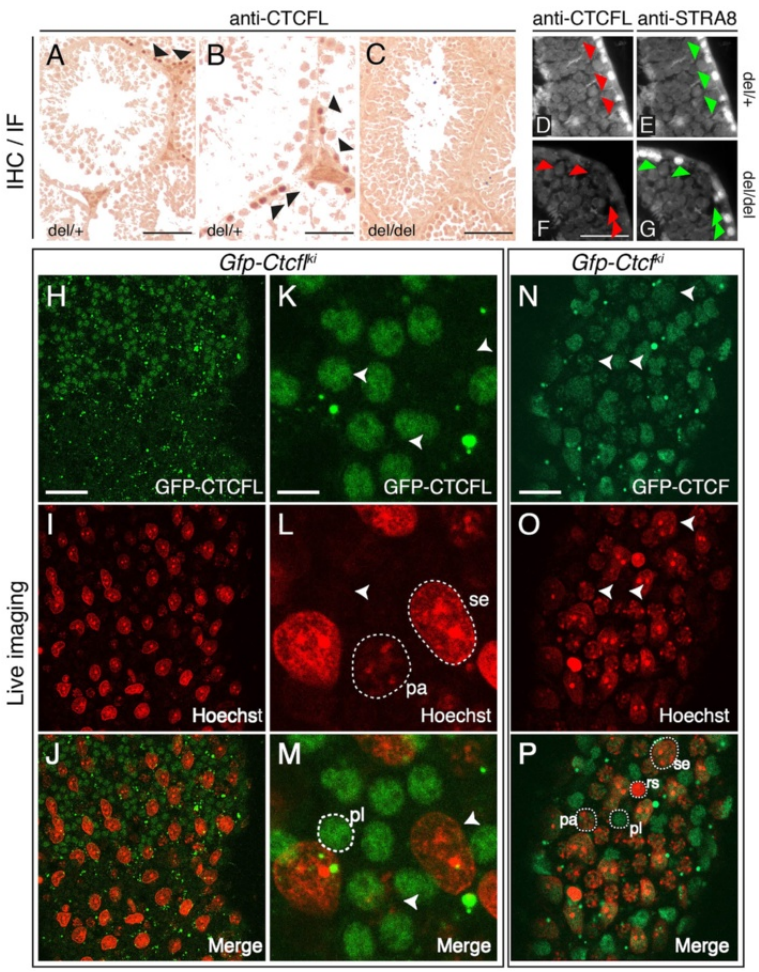


Figure 2 Expression of CTCFL and CTCF in the testis. A-**C** Immunohistochemical staining of testis sections. Paraffin-embedded sections from day 90 testes from heterozygous (del/+) and homozygous (del/del) Ctcfl mutant mice were stained with anti-CTCFL, followed by diaminobenzidine (*DAB*) coloring. Some of the CTCFL-positive cells are indicated with *black arrowheads*. Scale bars **A**, **C**: 100 μm, **B**: 50 μm. **D**-**G** Immunofluorescence staining of testis sections. Sections as described in **A**-**C** were stained with CTCFL (**D** and **F**) or STRA8 (**E** and **G**) antibodies. STRA8-positive cells in panels **E** and **G** are indicated with *green arrowheads*; the same cells are indicated with *red arrowheads* in the sections stained with anti-CTCFL antibodies (panels **D** and **F**). In Ctcfl mutant mice, STRA8 distribution is not changed. Scale bar is 50 μm. **H**-**P** Ex vivo confocal and multiphoton imaging of intact seminiferous tubules. Testis tubules were dissected from GFP-CTCFL- (**H**-**M**) or GFP-CTCF- (**N**-**P**) expressing mice, exposed to Hoechst at the adluminal side of the seminiferous tubule, and analyzed with a confocal/multiphoton microscope (GFP-CTCFL and GFP-CTCF, *green*; Hoechst, *red*). Panel **H**-**J** (see also Movie S1) shows a low magnification view of GFP-CTCFL distribution. Notice the presence of GFP-CTCFL-positive cells in the upper half of the tubule and their absence in the bottom half, indicating a transient population of cells. In (**K**-**M**) a high-magnification view of the same GFP-CTCFL-positive cells is shown. Notice the non-homogenous distribution of GFP-CTCFL in the nucleus. In (**N**-**P**) GFP-CTCF staining is shown. For clarity, some of the cell types are encircled, and their position is indicated in the other panels using *white arrowheads*. *Pl* = preleptone spermatocyte; *rs* = round spermatid; *pa* = pachytene spermatocyte; *se* = Sertoli cell. Bars, **H**-**J**: 70 μm, **K**-**M**: 10 μm, **N**-**P**: 25 μm.

CTCFL-deficient testes weighed significantly less compared to testes from heterozygous littermates (Figure 3A). In addition, we found that lower testes weights coincided with infertile males (Figure 3A). The weight distribution shows that there are also normal testes in the $Ctcfl^{del/del}$ population. Still, on average, the epididymides from homozygous $Ctcfl^{del}$ mice contained only 15% of sperm compared to heterozygous littermates (Figure 3B).

Table 1 Sub-fertility of CTCFL mutant mice

Genotype	*Ctcfl*$^{del/+}$	*Ctcfl*$^{del/del}$
Number (percentage) of breedings w/o offspring	3/60 (5%)	14/27 (51.9%)
Average number of offspring per litter (± SD)	7.7 ± 2.6	6.4 ± 2.8

From day 28 onwards Ctcfl mutant mice displayed loss of germ cells by apoptosis and an increasing level of atrophy that increased with age (Figure 3C, D). Mitotic spermatogonia, staining positive for BrdU incorporation, were still often observed (Figure 3E), whereas SCP3, a marker for spermatocytes, revealed severe tubule disorganization (Figure 3F). In fact, the level of atrophy and disorganization between individual mice and between individual

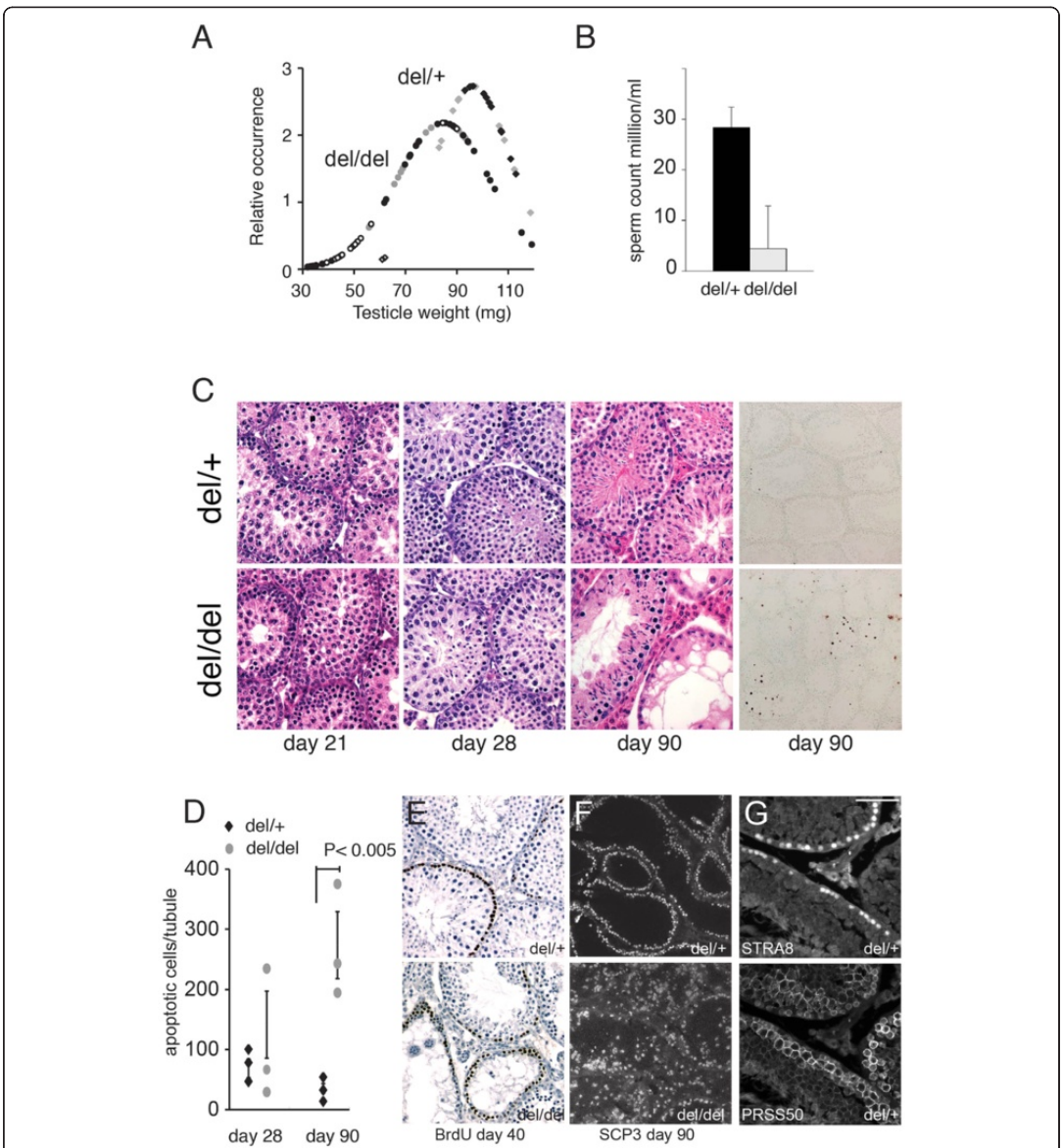


Figure 3 (See legend on next page.)

(See figure on previous page.)
Figure 3 CTCFL is important for spermatogenesis. A Testicular weight distribution. The testicular weight of Ctcfl heterozygous (*Ctcfl*$^{del/+}$; *diamonds*) and homozygous (*Ctcfl*$^{del/del}$; *circles*) mice was measured and plotted as a normalized probability distribution (i.e., the surface under the curve represents a total probability of 1). Testes of knockout mice are significantly smaller ($p < 0.0005$, *t*-test). *White symbols* represent infertile males, *black symbols* are fertile males, and *grey symbols* correspond to males not tested for fertility. **B** Ctcfl mutant mice display reduced fertility. Epididymal sperm count from Ctcfl heterozygous (*Ctcfl*$^{del/+}$; *black bar*) and homozygous (*Ctcfl*$^{del/del}$; *grey bar*) mice. Standard deviation is plotted ($p = 0.0002$, $n = 4$). **C** Testis histology. In the *left three panels* a timed series of HE-stained testicle sections is shown (postnatal day 21, 28, 90), while in the *right hand panel* an apoptosis assay (TUNEL staining) of testicle sections at day 90 is shown. Note that in CTCFL-deficient testes some seminiferous tubules appear normal, whereas others (which can be adjacent to the normal ones) have lost most of their meiotic and post-meiotic germ cells, leaving only mitotic spermatogonia (that stain positive for BrdU incorporation, see panel **E**) and Sertoli cells. Yet other tubules contain disorganized spermatocytes, and some of them even elongated sperm. Thus, there is no absolute block in differentiation or progression of germ cell development, but the incomplete penetrance of the infertility phenotype is however directly linked to the testicle weight (panel **A**) and to the degenerative level of the seminiferous tubules. **D** Apoptosis plot. Number of TUNEL-positive apoptotic cells per 100 seminiferous tubules counted at day 28 and day 90. Standard deviation of three animals per genotype and time point is indicated. **E** DNA synthesis marked by a 1-h pulse of BrdU in day 40 testicles reveals that mitotic spermatogonia are still present in degenerated tubules. Counterstaining with hematoxylin. **F** SCP3 staining in spermatocytes of day 90 testicles as a marker for tubule organization. **G** PRSS50 co-localizes only partially with STRA8. Immunofluorescence staining with a STRA8 antibody (*top panel*) or PRSS50 antibody (*bottom panel*) of adult testicle sections shows that PRSS50 and STRA8 expression overlaps only partially. Scale bars are 50 μm.

seminiferous tubules within mice was variable, and normal seminiferous tubules could even be adjacent to abnormal ones. Thus, the penetrance of the atrophic testes and sterility phenotype in CTCFL deficiency is incomplete and differs considerably per mouse.

Next we performed a microarray analysis on day 23 testis mRNA (a time point that precedes the start of apoptosis and degeneration in the testes of CTCFL-deficient mice) on *Ctcfl*$^{del/del}$ and *Ctcfl*$^{del/+}$ littermates. This revealed several affected genes in *Ctcfl*$^{del/del}$ testes (Figure 4A). The Prss50 and Gal3st1 genes were most downregulated (~1.5 fold), which matches results from another study [26]. Real-time RT-PCR verified results from the microarray (Figure 4B).

GAL3ST1 is crucial for spermatogenesis as mutant mice are infertile because of an arrest at the end of the meiotic prophase [35]. PRSS50 (Testis Specific Protease) has an exclusively testicular expression pattern, and is detected both in CTCFL-positive cells as well as in later stages of spermatogenesis [36], including STRA8/CTCFL-negative pachytene spermatocytes (Figure 3G). Since a CTCFL-deficiency affects differentiation of cells subsequent to the preleptotene stage, the reduction in Prss50 and Gal3st1 mRNA may be the result of a reduction in the number of cells going through meiosis or a reduction in Prss50 and Gal3st1 mRNA per cell. We therefore investigated PRSS50 expression in sections of wild-type and CTCFL-mutant mice and noted reduced protein levels per cell (Figure 4C).

CTCFL regulates testis-specific gene expression

The whole genome DNA-binding profile for CTCF has been elucidated in several cell systems (see, for example, [6-9]). We sought to compare the DNA-binding profiles for CTCF and CTCFL in the same cell type. Since CTCF is ubiquitously expressed in the testis, whereas the presence of CTCFL is highly restricted, genome-wide DNA-binding patterns derived from whole or partially purified testis preparations cannot be compared (see Discussion). We therefore generated ES cells, a cell type closely related to germ cells [37-40], in which expression of a V5- and GFP-tagged CTCFL protein could be induced (Figure 5A), thereby mimicking the CTCFL-positive germ cells that express both CTCFL and CTCF. Advantages of this system are furthermore the unlimited source of cells and the possibility to sort for GFP-positive cells that express the fusion protein to obtain a pure population of cells. In addition, the V5 tag permits stringent and exclusive immunoprecipitation of CTCFL. Thus, with this system genome-wide RNA (micro-arrays) and DNA-binding studies (ChIP-Seq) were carried out (Figures 5B).

Comparison of the expression of all genes on the microarray to expression of genes bound by CTCF or CTCFL revealed that the CTCFL-bound genes were, on average, more abundantly expressed (Figure 5C). The same held true when CTCFL-bound genes were compared to random gene sets (not shown). These data indicate that CTCFL associates with active genes. Several genes upregulated in CTCFL-GFP-V5-induced ES cells were also detected in the list of genes downregulated in CTCFL-deficient testes. Real-time RT-PCR confirmed that Gal3st1, Prss50 and even Stra8 expression were upregulated in CTCFL-induced ES cells (Figure 5D). Thus, CTCFL can act on male-specific germ cell genes in ES cells, and two of the most downregulated genes in CTCFL-deficient testis are upregulated in CTCFL-GFP-V5-expressing ES cells. These data underscore the notion that ES cells resemble germ cells and indicate that CTCFL acts as a male germ cell gene regulator.

Genome-wide binding of CTCFL and CTCF

To determine the genome-wide binding pattern of CTCFL, we used GFP-sorted CTCFLV5-GFP-induced ES cells, which express both CTCFL and CTCF. The V5

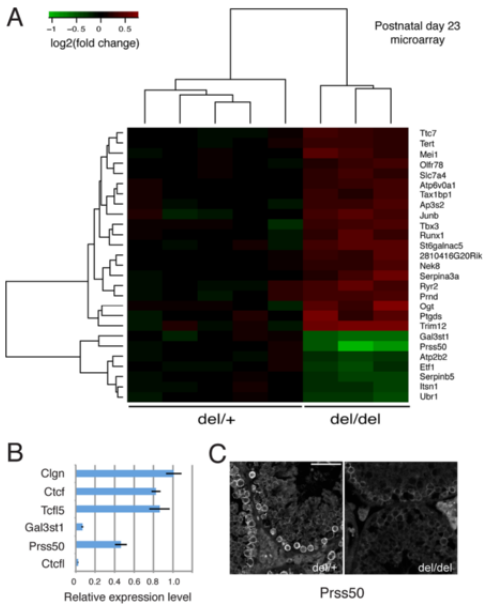


Figure 4 Regulation of testis-specific genes by CTCFL. A Heatmap representation of microarray data. We compared five samples from heterozygous and three samples from homozygous Ctcfl knockout mice. Depicted are the top 27 deregulated genes, where the log2-transformed fold change compared to the average expression in heterozygous testis is shown. **B** Expression analyses in Ctcfl mutant testes. Real-time RTPCR expression analyses on testis RNA from Ctcfl mutant mice relative to wild type using Ccna1 as reference. Genes were examined in 90-day-old testes. **C** Reduced PRSS50 expression in Ctcfl mutant testis. Immunofluorescence analysis of testis sections from heterozygous (del/+) and homozygous (del/del) Ctcfl knockout mice, using antibodies against PRSS50.

antibody was used for ChIP of CTCFL, with non-induced ES cells as control. Normal ES cells and a rabbit polyclonal antibody to CTCF [4] were used for ChIP-sequencing of CTCF. ChIP- sequencing revealed 5707 CTCFL and 37691 CTCF-binding sites (Figure 5E). To validate our data, we compared the number and position of CTCF sites determined by us with published data from the same cell type [7] and found a very high overlap (Figure 5F). Sorting the CTCFL-binding sites on the number of unique sequence reads yielded a list of genes that was headed by Stra8 and Prss50 (Table 2), two genes that are upregulated in CTCFL-inducible ES cells. Thus, the most prominent CTCFL sites locate at genes that are important for germ cells.

Interestingly, only 64% (3677) of CTCFL sites overlap with those of CTCF; conversely, only ~10% of CTCF sites are bound by CTCFL (Figure 5E). Despite their partial overlap, CTCFL and CTCF bind almost identical consensus sequences (Figure 5G). The most notable differences in the DNA-binding motif are the lower prevalence of a C at positions 1 and 2, the absence of A at position 3 and a lower prevalence for A at position 6, as well as a higher prevalence of G at positions 8 and 11, for the CTCFL motif relative to the one of CTCF. Whether subtle motif differences relate to differences in numbers of binding sites or to effects mediated through CTCFL and/or CTCF are questions currently under investigation. We also noted that, similar to CTCF [6,9,41], not all binding sites for CTCFL contain a consensus motif (Figure 5F).

Table 2 Top ten CTCFL-binding sites in induced ES cells

	Chr	Position	Gene	CTCFL bound*	CTCF bound
1	6	34,872,000	Stra8	134	Yes
2	9	110,760,000	Tsp50	113	Yes
3	14	103,450,500	Irg1	105	Yes
4	9	106,114,000	Twf2	101	Yes
5	9	50,260,500	-	100	Yes
6	5	125,061,000	Tctn2	97	Yes
7	8	107,058,500	Nae1	95	Yes
8	2	29,475,000	Rapgef1	94	Yes
9	9	108,838,500	Uqcrc1	93	Yes
10	12	112,970,500	Bag5	92	Yes

*Sites are ranked based on the number of ChIP-sequence reads, filtered for duplicates.

Nucleosome occupancy specifies binding of CTCFL versus CTCF

Further analysis of the genome-wide binding of CTCF(L) revealed that CTCFL binds almost exclusively to CTCF consensus sites near promoter areas, in contrast to CTCF (Figure 6A, B). We next split CTCF(L)-binding sites into three groups, i.e., CTCFL-only sites, CTCFL + CTCF sites and CTCF-only sites, and compared CTCF(L) binding to published data sets of transcription factors and other chromatin constituents. Binding sites are shown as heatmaps, which represent individual ChIP-Seq profiles from -2 kb to +2 kb relative to the center (peak maximum) of the analyzed peaks (Figure 6C) and as cumulative profiles (Figure 6D, E), which represent average ChIP-Seq profiles. Sites were sorted for binding strength within the three subsets. This comparative analysis revealed, for example, that CTCFL colocalizes with cohesin at CTCF consensus sites that are not occupied by CTCF (Figure 6C, D). In addition, CTCFL is enriched at transcriptionally active promoters, which are marked by H3K4me3 and PolII phosphorylation on serine 5 (Figure 6C, D). By contrast, CTCF-only sites are not associated with these marks. These data confirm the observation that CTCFL associates with transcriptionally active genes (see Figure 5C).

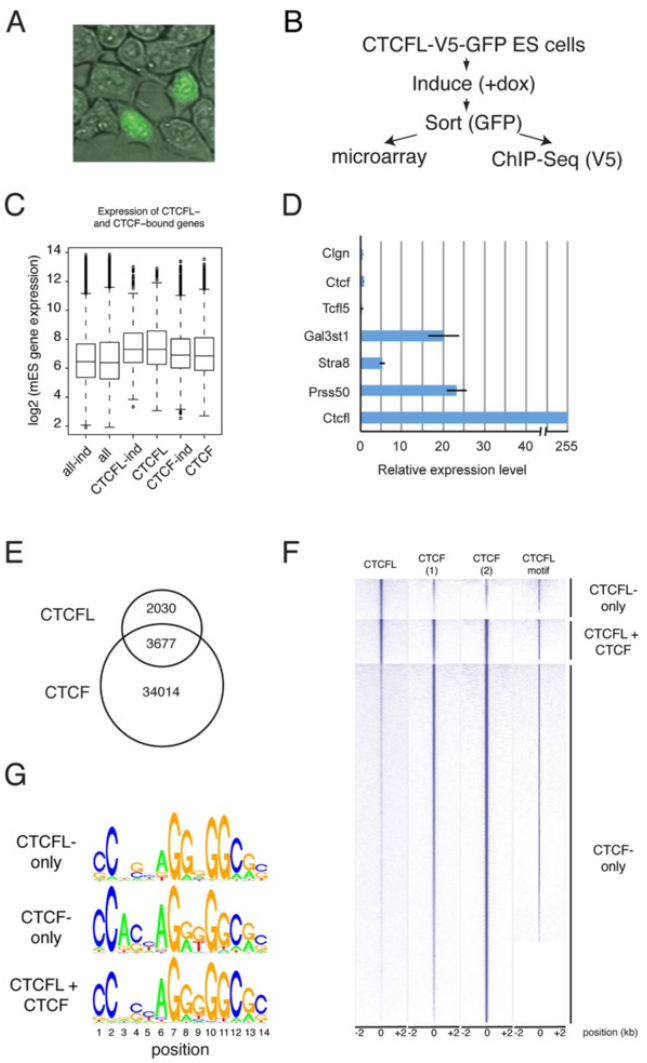


Figure 5 Genome-wide analysis of CTCFL expression in ES cells. A Inducible expression of CTCFL-V5-GFP in ES cells. Notice the nuclear localization of CTCFL-V5-GFP in cells expressing the protein. **B** Flow chart of experiments. ES cells with a Tet-on inducible expression of a CTCFLV5-GFP transgene were sorted for GFP and used for microarray and ChIP-Seq analyses. **C** CTCFL expression and DNA binding are associated with elevated gene expression levels. We plotted gene expression levels, as determined by microarray analysis of induced (*ind*) or non-induced ES cells, for all genes (*all*), or those bound by CTCF, or CTCFL, to the respective promoter region (–2 k to +1 kb around TSS). Differences are highly significant (*p*-value CTCF-ind versus CTCFL-ind: $5.1 \times e^{-14}$; *p*-value CTCF versus CTCFL: $5.9 \times e^{-13}$). **D** Transcript analyses in ES cells expressing CTCFL-V5-GFP. Real-time RT-PCR expression analyses of CTCFL-V5-GFP-induced and GFP-sorted ES cells, relative to non-induced ES cells, for the indicated genes, referenced to Cdk2 expression. **E** Venn diagram of DNA-binding sites for CTCFL and CTCF. **F** Clustered heatmap representation of three classes of CTCF/CTCFL-binding sites. Shown are the binding profiles of CTCFL and CTCF (*1*: our own data; *2*: [7]) across all CTCF/CTCFL-binding sites identified in mES cells. Sites are grouped into CTCFL-only, CTCF-only, and combined CTCFL and CTCF sites. Within the three classes, data sets were sorted decreasingly from *top* to *bottom* for average binding across the interval from 2 kb to +2 kb around the identified binding peak *center* positions. Additionally the occurrences of predicted CTCFL motifs within these intervals are plotted. **G** Motif comparison of CTCF and CTCFL. DNA-binding motif for CTCFL-only (*top panel*), CTCF + CTCFL (*middle panel*) and CTCF-only binding sites (*bottom panel*).

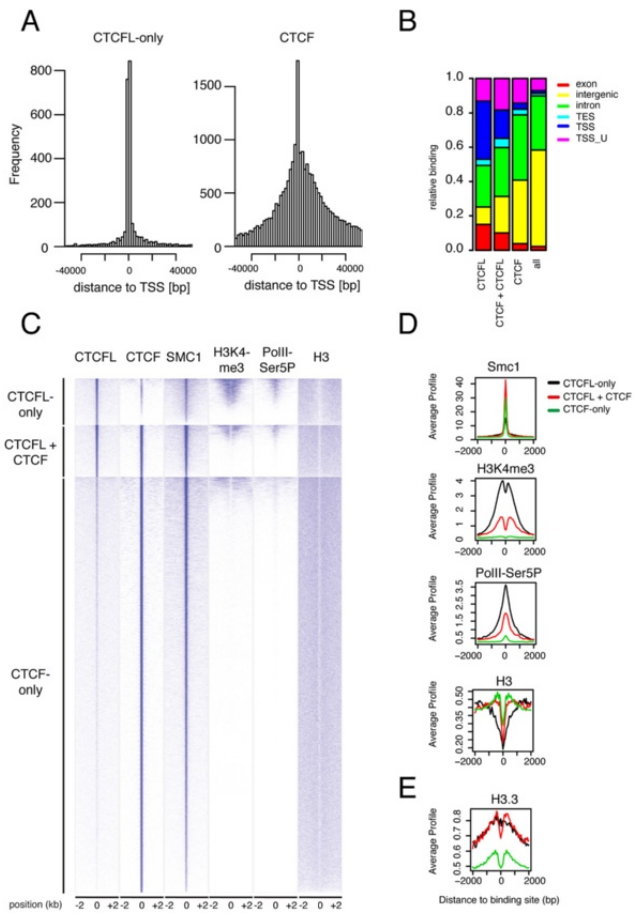


Figure 6 Characterization of CTCFL and CTCF binding. A A large fraction of CTCFL-binding sites is located close to promoters. We determined for each CTCFL-only-binding site the distance to the nearest transcriptional start site (TSS) and plotted the frequencies of binding sites in the depicted window from −40 kb to +40 kb around the center of CTCFL-binding sites. CTCF is plotted as comparison. **B** Comparison of the genomic distribution of CTCF- and CTCFL-binding sites. Sites are separated into CTCF-only, CTCFL-only and (CTCF + CTCF). The entire genome is also plotted (*all*). The binding location is separated into exon, intron, intergenic, transcription start site (*TSS*) and transcriptional end sites (*TES*), and plotted as frequencies of total (Y ax). **C** Clustered heatmap representation of the three different classes of CTCF/CTCFL-binding sites with respect to chromatin context. We compared CTCF and CTCFL binding to published ChIP-sequencing data sets for the cohesin complex subunit Smc1, H3K4me3, a phosphorylated form (serine 5) of RNA PolII, (PolIISer5P) and histone H3 [8,42,43]. **D** Cumulative profiles across the three different classes of CTCF/CTCFL-binding sites with respect to chromatin context. The average ChIP-sequencing profiles are shown for the same data sets as in (**C**). **E** Cumulative profiles across the three different classes of CTCF/CTCFL-binding sites with respect to H3.3.

When we compared binding of CTCF(L) to that of histone H3 we noted that CTCFL preferentially binds large H3-depleted areas (Figure 6C, D). By contrast, CTCF is enriched on sites that display H3 phasing around the CTCF-binding site (Figure 6D). These sites, in turn, do not attract CTCFL. Shared CTCFL/CTCF sites associate with "intermediate" H3-free regions (Figure 6C, D). As the H3-binding site analysis was performed in ES cells that do not express CTCFL, we conclude that the H3 depletion in these cells is not caused by CTCFL, but that H3-depleted regions appear to attract CTCFL.

It has been observed that many "H3-free" regions in the genome in actual fact do contain histones, but that these are loosely assembled and are lost upon DNA extraction with high salt [44]. The variant histone H3.3 has been shown to occupy these areas, often together with another variant histone, H2A.Z [45]. We therefore compared CTCFL binding to that of H3.3 (for which data

are available in mouse ES cells [46]) and found that these two proteins colocalize (Figure 6E), whereas CTCF does not associate with H3.3-enriched regions. We conclude that in addition to nucleotide sequence, nucleosome occupancy and composition specify the genome-wide binding of CTCFL and, surprisingly, of CTCF.

Competition between CTCFL and CTCF on distinct sites

ChIP-sequencing and direct ChIP experiments showed that CTCF and CTCFL bind the same site within the Stra8 and Prss50 promoters, but not in the Gal3st1 promoter (Figure 7A). We therefore tested the idea that these two proteins compete for binding on selected sites. Using in vitro band shift assays, we confirmed that CTCFL and CTCF bind the Stra8 and Prss50 promoters (Figure 7B, C). When proteins were added together we did not observe a higher band, indicating that CTCF and CTCFL do not interact to bind a probe simultaneously (Figure 7C). Instead, with increasing amounts of CTCFL, the amount of bound CTCF diminished (Figure 7C), suggesting that CTCFL and CTCF compete for binding sites in vitro.

To examine whether competition occurs in vivo, we used ChIP analysis on CTCFL-induced and -non-induced ES cells using a selected number of sites. In the presence of CTCFL, the amount of bound CTCF was reduced for both Stra8 and Prss50 (Figure 7D). CTCFL induction had no effect on CTCF in the CTCF-only-binding site within the Chr10 locus (Figure 7D). However, for the shared CTCFL/CTCF-binding site at Vps18 we also saw no effect on CTCF binding (Figure 7D). These ChIP results indicate that CTCFL can compete with CTCF, but only at specific sites. To test this hypothesis on a genome-wide level, we transiently transfected GFP-CTCFL into ES cells and examined differences in CTCF binding in ES cells expressing GFP-CTCFL compared to cells not expressing this protein. As shown in Figure 7E (*left panel*), in the presence of CTCFL, binding of CTCF was reduced on ~1,100 sites, whereas binding on ~100 sites was increased. Binding of CTCF to the affected sites was significantly reduced as compared to all CTCF-binding sites (Figure 7E, *right panel*). Among sites displaying reduced CTCF binding were the Prss50 and Stra8 promoters, but not Vps18 and chromosome 10 binding sites (not shown). These results are consistent with the ChIP data (Figure 7D). We conclude that in ES cells, CTCFL and CTCF compete on distinct sites.

To estimate the physiological relevance of our genome-wide ES cell data we performed ChIP using antibodies against CTCFL and CTCF on selected sites in cells isolated from wild-type testis by elutriation. Results show preferential binding of CTCFL to the Stra8 and Prss50 promoters, and preferential binding of CTCF to Vps18 (Figure 7F), indicating that the differential binding pattern of CTCF and CTCFL observed in ES cells is present in testis as well.

CTCFL is functionally different from CTCF

The co-expression of CTCFL and CTCF in late spermatogonia and preleptotene spermatocytes combined with their differential genome-wide binding patterns raises the question whether CTCFL and CTCF are functionally redundant, complementary or antagonistic. To test whether CTCFL can functionally substitute for CTCF, we designed an ES cell rescue assay (Figure 8A). ES cell lines were derived from mice carrying the conditional Ctcf knockout allele ($Ctcf^{lox/lox}$ [4]). Ctcf is deleted upon lentivirus-mediated Cre recombination, and these cells fail to form colonies because CTCF-deficient cells ($Ctcf^{del/del}$) do not survive. A rescue of cell death by concurrent introduction of CTCFL would show that the two proteins compensate for each other.

Using this strategy we co-expressed Cre with GFP-tagged mouse CTCF, YFP-tagged chicken CTCF or GFP-tagged mouse CTCFL in $Ctcf^{lox/lox}$ ES cells (Figure 8A). Resulting colonies were analyzed on the DNA level for Cre-mediated CTCF deletion of $Ctcf^{lox/lox}$ into $Ctcf^{del/del}$ (Figure 8B) and on the protein level for expression of endogenous or exogenous protein (Figure 8C). A few surviving colonies transduced with Cre-only were observed (Table 3), but these had not performed the Cre-mediated CTCF deletion completely and thus still expressed endogenous CTCF (Figure 8B, C). However, nearly all colonies transduced with N- or C-terminally tagged mouse CTCF, or with C-terminally tagged chicken CTCF, had deleted endogenous $Ctcf^{lox/lox}$, and expressed fluorescently tagged exogenously introduced protein (Figure 8B, C, Table 3). Thus GFP-tagged mouse CTCF and even chicken CTCF, which is 96% identical at the amino acid level to mouse, can functionally substitute for endogenous CTCF.

Strikingly, rescue experiments with GFP-tagged mouse CTCFL yielded no ES cells in which both endogenous $Ctcf^{lox/lox}$ alleles were deleted and wild-type protein was replaced (Figure 8B, C, Table 3). These data indicate that either CTCFL and CTCF are not interchangeable or that GFP-CTCFL is not a functional protein. To demonstrate that GFP-CTCFL is functional, we transiently transfected the protein into ES cells and examined DNA binding of GFP-CTCFL on selected sites and the induction of expression of testis-specific genes. ChIP experiments showed that GFP-CTCFL binds the three selected sites (Figure 8D) and that Gal3st1, Stra8 and Prss50 expression is induced inES cells expressing this fusion protein (Figure 8E). These data demonstrate that GFP-CTCFL is functional.

Discussion

We have used a combination of approaches and technologies to unravel the physiological function of the testis-specific paralog of CTCF, called CTCFL or BORIS. We find that CTCFL is only expressed in late spermatogonia and preleptotene spermatocytes, and that CTCFL-deficient mice

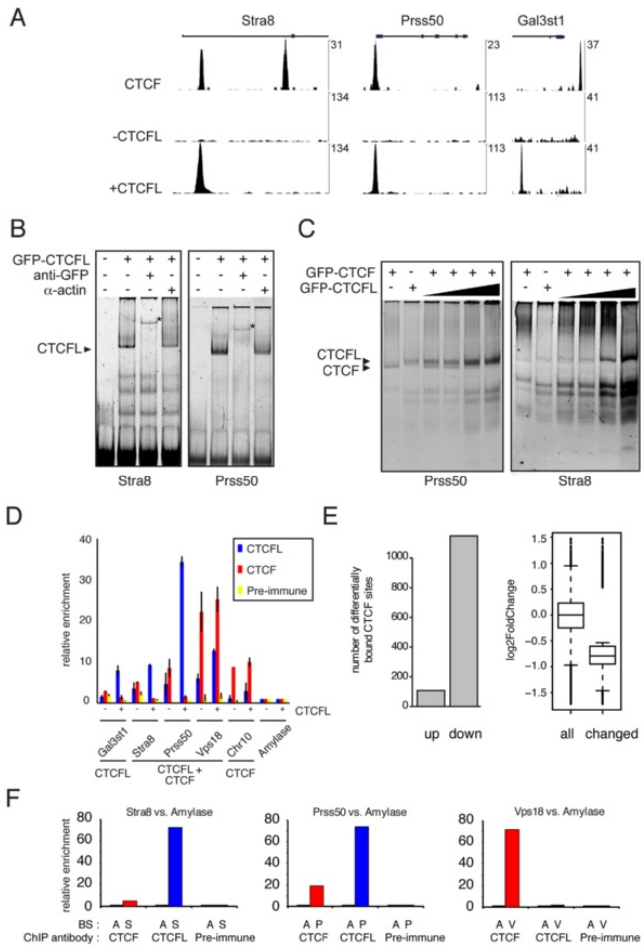


Figure 7 Characterization of CTCF and CTCFL binding. A Examples of CTCF- and CTCFL-binding site location. The genomic location of CTCF (*upper part*) and CTCFL (*middle* and *bottom* parts) binding sites in the absence (–CTCFL, *middle*) or presence (+CTCFL, *bottom*) of CTCFL, within the Stra8, Prss50 and Gal3st1 genes. The vertical axes show the number of unique sequence reads. **B** CTCFL binds to Stra8 and Prss50. Band shift analyses of GFP-CTCFL on Stra8 and Prss50 fragments. GFP-CTCFL binding can be super shifted (marked with *asterisks*) with anti-GFP, but not with an Actin antibody. Band shifts were performed under excess probe conditions. **C** In vitro effect of CTCFL on CTCF binding. Band shift analyses with GFP-CTCF and/or GFP-CTCFL on Prss50- and Stra8-bindings sites. GFP-CTCFL is added in increasing amounts (1-, 2-, 5- and 10-fold compared to GFP-CTCF). To allow competition, the band shift was performed under probe-limiting concentrations. **D** Cellular effect of CTCFL on CTCF binding. ChIP analyses with CTCFL (*blue*), CTCF (*red*) and pre-immune (*yellow*) antisera in ES cells that were either non-transfected (–) or transiently transfected CTCFL-V5-GFP (+). According to ChIP-sequencing data, Prss50, Stra8 and Vps18 bind both CTCF and CTCFL, whereas Gal3st1 only binds CTCFL, and Chr10 only binds CTCF. A CTCF- and CTCFL-negative site within the Amylase gene is used as reference and set to 1. *Error bars* represent standard deviations of biological replicates. **E** Competition between CTCF and CTCFL in ES cells. Genome-wide binding of CTCF was compared to that of CTCFL by ChIP-Seq using non-transfected ES cells and ES cells transiently transfected with GFP-CTCFL. The *left hand panel* shows the effect of CTCFL binding on shared CTCFL/CTCF sites that showed >1.5 fold difference in CTCF binding. The effect is categorized into sites with increased (*up*) or decreased (*down*) CTCF binding. The *right hand panel* shows a more general effect of CTCFL binding on CTCF binding. Here, we examined the change in CTCF binding in all shared CTCF(L)-binding sites (*all*) compared to those shared sites that were significantly changed in CTCF binding (*changed*). The effect on CTCF binding is plotted as log2-fold difference. **F** In vivo CTCF(L) binding. ChIP was performed using anti-CTCF (*red*) or anti-CTCFL (*blue*) antibodies, or pre-immune serum, on the indicated sites (*A*: Amylase, *S*: Stra8, *P*: Prss50, *V*: Vps18) in nuclei from dissociated seminiferous tubules, partly purified by elutriation. Relative enrichment is shown compared to Amylase.

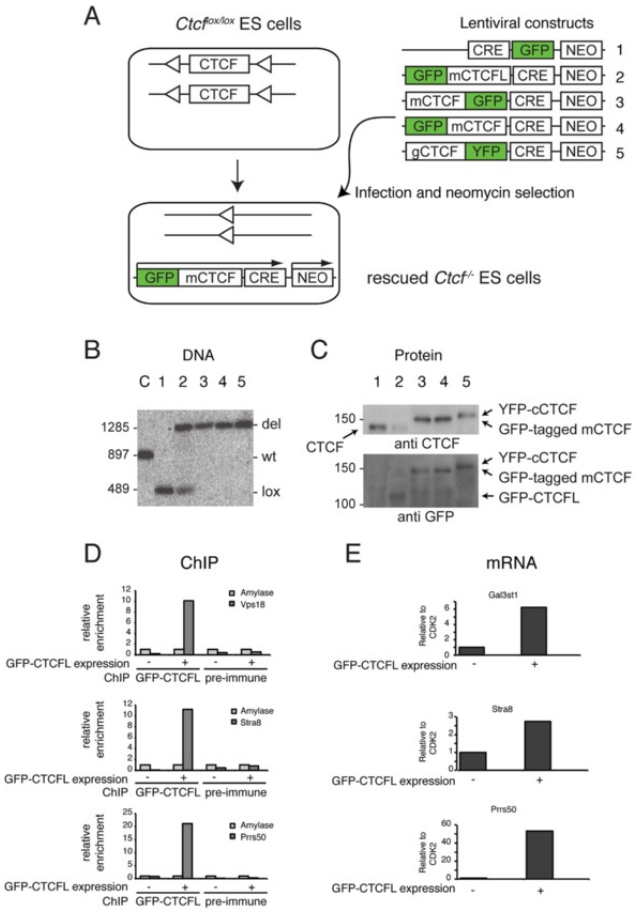


Figure 8 CTCFL is functionally different from CTCF A Strategy for the rescue of CTCF-depleted ES cells. Ctcf$^{lox/lox}$ ES cells were infected with lentivirus containing the Cre recombinase and/or fluorescently tagged CTCF(L) proteins. After infection neomycin-resistant colonies were picked and analyzed. *m* = mouse, *g* = chicken. **B** Analysis of Ctcf$^{lox/lox}$ deletion. After infection with CRE-containing constructs, Ctcf$^{lox/lox}$ ES cells were scored for the status of the conditional Ctcf alleles by DNA blot. The position of wild-type (wt), deleted (Ctcf^{del}, or del) and conditional (Ctcf^{lox}, or lox) loci in control ES cells (**C**), non-treated Ctcf$^{lox/lox}$ ES cells (*1*) and lentivirally transduced Ctcf$^{lox/lox}$ ES cells (*2–5*, see panel **A** for numbering of constructs) is indicated. Cells are considered rescued when both conditional CTCF alleles have been deleted. **C** Analysis of CTCF protein expression. Neomycin-resistant colonies were grown and analyzed by Western blot for CTCF (*upper panel*) or GFP (*lower panel*) expression. Note that rescued cells are negative for endogenous CTCF. **D**, **E** GFP-CTCFL is a functional protein. ES cells were transiently transfected and harvested after 1 day. ChIP (DNA, **D**) and RT-PCR (mRNA, **E**) analyses revealed that GFP-CTCFL binds Vps18, Stra8 and Prss50 promoters (**D**) and is able to induce expression of Gal3st1, Stra8 and Prss50 (**E**).

have defects in spermatogenesis. CTCFL and CTCF are functionally different proteins. CTCFL therefore has a unique role in the adult testis. It has been proposed that CTCFL is involved in genomic imprinting of the *Igf2-H19* locus and other sites [19,24]. However, imprint-related mutations often have embryonic phenotypes [25]. We did not observe this in *Ctcfl*$^{del/del}$ mice, and despite their reduced fertility *Ctcfl*$^{del/del}$ mice could be bred through multiple generations. Furthermore, we have not been able to detect DNA methylation aberrations in specific loci in *Ctcfl*$^{del/del}$ mutant mice and in CTCFL-over-expressing cells (data not shown). This makes a role for CTCFL in DNA methylation-dependent genomic imprinting unlikely. The combined microarray data from CTCFL-deficient testis and CTCFL-expressing ES cells, and the preference of CTCFL for promoters instead suggest a function as a transcriptional regulator, required for the proper expression of a subset of male germ cell genes.

Table 3 Rescue of CTCF-deficient ES cells by exogenously introduced GFP-CTCF(L)

Species	Construct	Deletion in ES cells**	Partial deletion in ES cells**	No deletion in ES cells**	Functional CTCF substitution
N.a.*	GFP-CRE	0% (0/65)	28% (18/65)	72% (47/65)	No
Mouse	GFPCTCFL-	0% (0/18)	50% (9/18)	50% (9/18)	No
Mouse	CTCFGFP-CRE	83% (19/23)	0% (0/23)	17% (4/23)	Yes
Mouse	GFPCTCF-CRE	95% (40/42)	2% (1/42)	2% (1/42)	Yes
Chicken	CTCFYFP-CRE	90% (37/41)	2% (1/41)	7% (3/41)	Yes

*: Not applicable.
**: Percentage and number of clones (between brackets) in which the conditional CTCF alleles were deleted, partially deleted or not deleted are shown.

The most prominent CTCFL-binding sites in ES cells are on the promoters of the testis-specific Stra8 and Prss50 genes. The expression of these genes, and of Gal3st1, is upregulated in ES cells expressing CTCFL. Conversely, expression of Prss50 and Gal3st1 is downregulated in germ cells lacking CTCFL, at all ages examined, whereas Stra8 expression is affected at some but not all ages (data not shown). We speculate that the combined transcriptional deregulation of genes causes the testicular degeneration and reduction in fertility in Ctcfl knockout mice. Note that the expression of these genes is not completely hampered, which explains why the testicular phenotype in the knockouts is milder than the fully sterile phenotype described, for example, for STRA8- and GAL3ST1-deficient mice [29,30,35].

The phenotype of the *Ctcfl*$^{del/del}$ mice reported here only partly matches a recent report on another strain of CTCFL-deficient mice, in which exons 1 to 8 of Ctcfl were also deleted [26]. For example, the effect of a Ctcfl deletion on the average testicular size and on Gal3st1 and Prss50 expression is similar. However, our analysis also reveals a reduction in fertility in the *Ctcfl*$^{del/del}$ mice not noted previously [26]. In addition, the fact that some *Ctcfl*$^{del/del}$ mice have normal testis size and others have a combination of normal and abnormal seminiferous tubules was also not described. This is relevant, as this incomplete penetrance of the Ctcfl phenotype, even within a single testis, suggests that a stochastic mechanism determines whether CTCFL-deficient tubules degenerate or not. Finally, CTCFL was proposed to be present in round spermatids and to function during meiosis based on mRNA expression data [26]. By contrast, our data show that CTCFL is expressed earlier, just prior to the onset of meiosis, and we conclude that CTCFL protein expression precedes the developmental germ cell stages that show the major phenotypes in Ctcfl knockout mice. We propose that in the absence of CTCFL, epigenetic marks controlled by this protein gradually break down in a stochastic manner. Spermatogonia and primary spermatocytes exist in syncitia, in which each cell is connected with the other cells at the same step of development via intercellular bridges. Only in syncytia where the expression of CTCFL-controlled genes has been affected beyond a specific threshold will degeneration become apparent.

Neither CTCFL nor CTCF is saturating all consensus-binding sites present in the genome, and thus the DNA sequence is not the sole determinant of CTCF(L) binding. DNA methylation and hydroxymethylation are not a decisive aspect, as comparisons of DNA (hydroxy)methylation data sets to our CTCF(L)-binding sites does not provide an explanation for why CTCFL and CTCF occupy different binding sites (data not shown) [47]. Instead, the data suggest that binding of CTCFL and of the "master weaver" CTCF is specified by nucleosome occupancy and composition. We find that CTCFL prefers CTCF consensus sites in promoters that are embedded in regions that appear to be nucleosome-free. By contrast, CTCF is enriched on distinct sites, which are devoid of histone H3 on the binding site itself, but which are surrounded by ordered, or "phased," nucleosomes. This preference of CTCF has already been described [11-13].

It has recently been shown that unstable nucleosomes are lost when histones are prepared with conventional conditions; thus, regions containing these histones appear as nucleosome-free in the analysis, but are in reality not free [45]. Nucleosomes containing the variant histone H3.3 are quite unstable, and those containing both H3.3 and H2A.Z even less [44]. Since we find a correlation between CTCFL binding and H3.3 occupancy in ES cells, H3.3 and H3.3/H2A.Z might be determinant factors able to attract CTCFL and evict CTCF. It is important to realize that in ES cells H3.3-enriched genomic regions do not require CTCFL to be set up, yet the protein prefers such areas after its induction. A similar situation may exist in testis, i.e., specific H3.3/H2A.Z-containing regions might be set up during early phases of spermatogenesis; upon its expression, CTCFL "lands" on these regions, possibly evicting CTCF from some promoters. Notably, during male meiosis, and thus subsequent to CTCFL expression, H3.3 is incorporated into unsynapsed chromatin, which is transcriptionally inactive [48]. The function of CTCFL might be to ensure the expression status of genes by distinguishing specific promoter-associated H3.3 domains from whole chromosome domains that also contain H3.3. Through its interaction with SET1A [49], CTCFL might enhance H3K4 trimethylation at a subset of its binding sites.

The cohesin complex has a role in chromosome segregation, DNA-damage repair and gene regulation [50]. Although cohesin does not have a typical DNA-binding motif, it was shown to bind primarily to CTCF consensus sites [16,17,51]. Moreover, the SA2 subunit of cohesin directly interacts with the C-terminus of CTCF [52].

Cohesin's role in gene regulation therefore seems tied to that of CTCF. Recent studies revealed that also in ES cells cohesin binding largely overlaps with that of CTCF; however, there are ~2,000 cohesin sites with a CTCF motif that do not bind CTCF, while ~270 other cohesin sites do not have a CTCF consensus site [10]. Our data suggest that CTCFL binds these ~2,000 cohesin sites in CTCFL-GFP-V5-expressing ES cells.

However, in normal ES cells CTCFL is not expressed, raising the questions how a specific nucleosome composition and occupancy can be built around CTCF consensus sites that appear not to be occupied by CTCF, and how cohesin can stably bind these very same sites. We hypothesize that these sites might be bound by a modified form of CTCF, such as poly(ADP-ribosyl)ated CTCF [53]. This protein would not be able to bind DNA tightly and could be replaced very efficiently by CTCFL. Perhaps another molecular function of CTCFL in the testes is to interfere with and/or change the dynamics of CTCF and cohesin-mediated chromatin looping.

We observed competition between CTCF and CTCFL in ES cells, but only on a small subset of all CTCF-binding sites. Nucleosome occupancy and composition, CTCF(L) expression levels and posttranslational modifications on CTCF(L) could all determine whether competition between the proteins occurs on a given site. Our data reveal that CTCF and CTCFL co-localize within the nuclei of late spermatogonia and preleptotene spermatocytes, and the proteins might therefore also compete in vivo. ChIP experiments in testis extracts indeed reveal preferential binding of CTCFL at the Stra8 and Prss50 promoters and exclusive binding of CTCF to the Vps18 site. These data are consistent with binding profiles in ES cells. If competition on the Stra8 and Prss50 genes does occur in vivo, then CTCFL could be a gene activator by preventing the binding of CTCF. In Ctcfl knockout mice binding of CTCF to these genes might actually diminish their expression. However, CTCF is ubiquitously expressed in the testis, whereas CTCFL is only transiently present in spermatogonia and preleptotene germ cells. One would expect to see significant binding of CTCF to the Stra8 and Prss50 sites in the testicular extracts that we used, since most cells in these extracts contain CTCF and not CTCFL. The questions why CTCF is not highly enriched on the Stra8 and Prss50 promoters in testis, and whether these proteins compete in vivo can only be answered once there are tools available to isolate CTCFL-positive and -negative cell populations from testis so that genome-wide analyses can be performed on purified testicular fractions.

In human germ cell tumors, CTCFL is specifically upregulated in spermatocytic seminomas, which are benign testicular tumors originating from a spermatogonium or primary spermatocyte [54]. This fits with our observed cellular localization of CTCFL and could potentially point to an oncogenic role for CTCFL in these tumors. In fact, CTCFL belongs to the group of cancer testis antigens (CTAs), genes that are normally expressed in testis yet aberrantly expressed in a variety of cancers. One model holds that competition between CTCF and CTCFL plays a role in tumorigenesis, i.e., aberrant CTCFL expression would displace CTCF, and affect DNA methylation and the expression of other CTAs, including the NY-ESO-1 and MAGE-A1 genes [22,23], and even other important genes, such as the TERT gene, which encodes telomerase [55]. However, while there might be a relationship between DNA demethylation and the expression of CTAs [56], recent reports have shown that expression of CTCFL alone is not sufficient to induce expression of CTAs [27,57]. Furthermore, our data in CTCFL- deficient testis indicate that, if anything, CTCFL represses the Tert gene instead of activating it. To address a potential role of CTCFL in cancer, a correlation analysis of CTCFL binding, nucleosome occupancy and composition, and CTA expression in different types of cancers might be more revealing.

Conclusions

The three-dimensional folding of the eukaryotic genome serves to compact DNA while allowing gene expression. CTCF has been termed the "master weaver" of the genome, since this protein is a key coordinator of chromatin loop formation. In this study we have analyzed the physiological function and DNA-binding profile of CTCFL, a protein that is highly similar to CTCF but that is only expressed in the male germ line. Using a combination of cell biological, biochemical and bioinformatics approaches, we show that CTCF and CTCFL are functionally different proteins that bind to similar sites in the genome, but whose binding does not overlap completely. Our data suggest that nucleosome composition specifies the genome-wide binding of both CTCFL and CTCF. We show that CTCFL is only transiently expressed, in spermatogonia and preleptotene spermatocytes, prior to male meiosis. We propose that during its expression CTCFL occupies a subset of promoters and thereby maintains the expression of selected male germ cell genes.

Methods

RACE PCR and RNase protection assay

Human CTCFL was shown to consist of 23 isoforms with variations in N- and C-termini and zinc finger modules with different DNA-binding and transcriptional characteristics [58]. To analyze the genomic organization of the murine Ctcfl gene, we cloned the 5' end of the Ctcfl cDNA by a rapid amplification of cDNA ends-polymerase chain reaction (RACE-PCR) procedure, using first choice RACE ready testicular cDNA from Balb/c mice (Ambion) and nested oligos (see Table 4 for sequence). Compared to the

published murine Ctcfl sequence [19], the RACE-PCR-derived first Ctcfl exon is smaller and lacks an upstream ATG, and it is preceded by an intron of 489 bp and an additional exon of 130 bp (see also panel 1C). The sequence of the Ctcfl 5'end product has been submitted to Genbank (accession no.: EU154995). Our cDNA structure matches the HAVANA/VEGA curated sequences in Ensembl, Build 36 [59].

RNase protection assay (RPA) was performed according to the manufacturer's instructions (RPAII, Ambion). For Ctcfl the RPA was performed on poly A + purified RNA with probes spanning Ctcfl exons 8 and 9 (protecting 146 bp), spanning bp 1–220 of the Ctcfl race PCR product (up to oligo 3, see figure S1C) or, alternatively, spanning bp131-220 (protecting 89 bp), to detect the existence of another start site [19]. For Ctcf the RPA was performed on total RNA with a probe protecting 99 bp of Ctcf exon 2. The RPA with the 5'end RACE confirms that *Ctcfl* mRNA mRNA contains the additional upstream exon as identified by the RACE PCR. We found no evidence for alternative splicing in murine Ctcfl.

4

Mouse models

To generate the Ctcfl and Ctcf knockin alleles we inserted a Gfp-encoding cDNA, followed by a Loxp-flanked neomycin selection cassette, in the Ctcfl and Ctcf exons, respectively, that contain the ATG start codons. Insertion of GFP immediately downstream of the translational start sites yielded $Ctcfl^{gfp\text{-}neo}$ and $Ctcf^{gfp\text{-}neo}$ ES cells. Homology arms were generated by cloning from the RPCI21 129 PAC library (Geneservice). Constructs were sequenced, electroporated into isogenic ES cells (129/IB10) and neomycin- (or, later on, puromycin-) selected, analyzed by Southern blot and PCR, and injected into C57/Bl6 blastocysts.

Mice generated from $Ctcfl^{gfp\text{-}neo}$ and $Ctcf^{gfp\text{-}neo}$ ES cells were crossed to transgenic mice expressing Cre to delete the LoxP-flanked neomycin cassette. This yielded $Ctcfl^{gfp}$ mice in which the GFP is fused in frame to CTCFL and $Ctcf^{gfp}$ mice where GFP is fused to CTCF. These mice are phenotypically normal and fertile (data not shown).

The $Ctcfl^{gfp\text{-}neo}$ ES cells were retargeted with a LoxP-flanked puromycin cassette downstream of exon 8. Mice were generated using the $Ctcfl^{gfp\text{-}neo\text{-}puro}$ ES cells. The $Ctcfl^{del}$ mice were subsequently generated by crossing $Ctcfl^{gfp\text{-}neo\text{-}puro}$ mice to Cre-expressing mice. This resulted in the in vivo deletion of Ctcfl exons 1–8 and both selection cassettes as these were in between the LoxP sites.

Mice were maintained on a C57/Bl6 background at the Erasmus MC animal care facility under specific pathogen-free conditions. Animal experiments were reviewed and approved by the Erasmus University committee of animal experiments.

Cell culture, transfection and infection

The $Ctcf^{lox/lox}$ ES cells were isolated de novo from CTCF conditional mice [4] and grown on plastic in the presence of LIF. Lentiviral constructs were generated with Ctcf and Ctcfl cDNAs driven from a CAG promoter (CMV early enhancer/chicken β actin), followed by an IRES sequence and the Cre recombinase. Expression of a neomycin selection cassette was driven by a PGK promoter. Lentivirus particles were produced as described (Addgene). ES cells were infected in suspension for 4 h, plated and selected with G418 the next day. Clones were analyzed by Southern blot for the status of the $Ctcf^{lox/lox}$ conditional allele [4] and by Western blot using GFP (Abcam 32146) or CTCF antibody (BD Bioscience).

For the inducible CTCFL expressing ES cells, ROSA26-rtTA ES cells were Lipofectamine transfected (Invitrogen) with a TRE-mCTCFL-V5-GFP-neomycin construct and selected with G418. Clones were analyzed for the induction and expression of CTCFL-V5-GFP by flow cytometry for GFP (FACSAria, BD Biosciences), and by Western blot and immunofluorescence using rat monoclonal anti-CTCFL antibodies raised against mouse CTCFL (AA 1–113 and AA 569–635) and V5 antibody (Sigma, V8012). Transient transfections of mCTCFL-V5-GFP and of GFP-mCTCFL in ES cells were done with Lipofectamine 2000 (Invitrogen).

Chromatin immunoprecipitation (ChIP)

\ChIP was performed as described [14] or according to the Magnify system procedure (Invitrogen). Briefly, preparation of cross-linked chromatin (2×10^7 cells treated with 1% formaldehyde for 10 min at room temperature), sonication of chromatin to yield fragments up to 800 bp, and immunoprecipitation with V5-Agarose beads (Sigma, A7345) or with polyclonal CTCF(L) antibodies [4] were performed as described in the Upstate protocol (http://www.upstate.com). Ct values from real-time PCR were normalized to input measurements, and enrichment was calculated relative to the Amylase gene. For oligos used, see Table 4. ChIP was performed on nuclei derived from induced or transiently transfected ES cells (see above) or from seminiferous tubules in which multiple testicular cell populations were first dissociated by enzymatic digestion of seminiferous tubules and subsequently isolated by elutriation.

ChIP-sequencing and analysis

For ChIP-sequencing a DNA library was prepared from the ChIPped DNA according to the Illumina protocol (www.illumina.com). Briefly, 10 ng of end-repaired DNA was ligated to adapters, size selected on gel (200 ± 25-bp range) and PCR amplified using Phusion polymerase as follows: 30 s at 98 °C, 18 cycles of (10 s at 98 °C, 30 s at 65 °C, 30 s at 72 °C) and 5 min at 72 °C final extension.

Table 4 Oligos

Oligos used for RACE PCR

Name	Forward	Backward
RACE	GGACACTCGTATTTGGGCACATTC	CACAGGGAGCACTTGAAGGGCTTC

Oligos used in real-time PCR for ChIP

Name	Forward	Backward
Gal3st1	TCCTGGGTGAGGTCAGGAAG	GGAACTCCGAGTAGCTTCAATG
Stra8	TCCTAGAGAAGGGGGTGTTACC	AGCTGACCACCACACGTTTTC
Prss50	AGAGGAGGGTAGGGGTATCGAC	TCGCCTCAGCTAATTTCTAAGC
Vps18	CTGCTGCAGTTCCTCATGTTG	GTGTGACAGATGGAGGAGCAC
Chr10	AAGGTTGGTAGCTCTGCTTGGACTGCTCG	AATGTCACAAGCAAAGAAAAGCACGCAAAT
Amylase	AATTCTCCTTGTACGGCTTCGTG	TAGCAATGATGTGCACAGCTGAA

Oligos used in real-time PCR for RNA expression

Name	Forward	Backward
Clgn	TGTCTTCCTTACTCTTCTCTTCCG	GAAGCCAGGTGAAGCTGAGGTC
Ctcf	CCACCTGCCAAGAAGAGAAGG	GCACCTGTATTCTGATCTTCGAC
Tcfl5	ACGAGATAGGAGGCGCAGAATC	GTTGTTGCTTTATCTGTCTCCG
Gal3st1testis-specific form	GCTACTCGGAGTTCCGGAAA	GACTTGCAGGGCTTCTTTGG
Gal3st1	ACTGTATCCCAACATGGCCTTC	ATATCTCGCCGAGGTTGACAC
Stra8	GGCAGTTTACTCCCAGTCTGATA	CAACTTATCCAGGCTTTCTTCCT
Prss50	GACAGTTCTCTCTGCACTGTGAC	CACATTTCTTGCTGTTCAGGATA
Ctcfl	GCTCTGGCTGTGCACCTTACG	CCCACTGTGCCACCATCATC
Ccna1	GAGTATGCAGAGGAGATTCATCG	TCATGTAGTGAGCCTTGGGTCTG
lpcat2	AGCACCCAGTGAGGAAGAGA	TTCGTAGGTGTGATCCGTCA
itfg3	ACGAGGTGTCTTCTGCCTGT	GTTCCCACTAAAGCTGCTGG
dio2	TGCAGATCCTGCCAGTCTTT	CACACTGGAATTGGGAGCAT
hgf	GATGAGTGTGCCAACAGGTG	GGTCAAATTCATGGCCAAAC
akr1c18	CCAGGCCATTCTAAGCAAGA	TCAGGGAATTTTCCAAGCTG

Oligos used in EMSA

Name	Forward	Backward
Stra8	GGATCTGTGCTGTGTGTCCTCCTCGACTCCT	CCTCTAGGAGTCGCAAGTGACCCACACATGCATGC
	AGAGCATGCATGTGTGGGTCACTTGCGACTC	
	CTAGAGGA	TCTAGGAGTCGAGGAGGACACACAGCACAGATCCT
Prss50	ATCTAGGGGGCGCCACGCAGGCTGGGCACC	CCACAATGGCGCCCTCCATCGGGCGCCTCATGGT
	AGCGCACCATGAGGCGCCCGATGGAGGGCG	GCGCTGGTGCCCAGCCTGCGTGGCGCCCCCTAGATG
	CCATTGTGGA	

Oligos used for generating transgenic mice and cells

Name	
Oligo 1	5-TCTTTTTCCATCAGGGGTCGTCAC-3
Oligo 2	5-GAGAAGCGCGATCACATGGTCCTG-3
Oligo 3	5-GCACCGTTTGCAGGGTCAGGATC-3
Oligo 4	5-TCCAAATCACAGCGCCACCTACAG-3
Oligo 5	5-GGTTCTTAGAGATAGGGTTTCTCTG-3
Oligo 6	5-GGTGTTCTGCTGGTAGTGGTC-3
Oligo 7	5-CGGCATCAGAGCAGCCGATTG-3
Oligo 8	5-GTTATGATCTGGGTATCGTCCACTG-3

Cluster generation was performed using the Illumina Cluster Reagents preparation, and the library was sequenced on the Illumina Genome Analyzer IIx platform to generate 36-bp reads. Images were recorded and analyzed by the Illumina Genome Analyzer Pipeline (GAP) and processed using the IPAR (Integrated Primary Analysis Reporting Software) and the GAP. The resultant sequences were mapped against NCBI Build 37.1 of the mouse genome using the ELAND alignment software (Illumina).

Published data sets generated for mouse ES cells were downloaded from NCBI's gene expression omnibus (GEO). We used the following data sets: H3: GSM587479, CTCF: GSM288351 [7], Smc1: GSM560341 and GSM560342 [8], H3K4me3: GSM594581 [42], PolIIser5p: GSM515662 [43] and H3.3-HA: GSM423355 [46]. Reads were converted to the fastq format and aligned to a precompiled mm9 reference index with BOWTIE [60]. In case multiple sequencing lanes were available, fastq files were merged before alignment. Unambiguously mapped and unique reads were kept for subsequent generation of binding profiles and calling of peaks using MACS with an fdr < 0.05 [61]. Downstream analysis was done in R/BioConductor (http://www.bioconductor.org), partly according to published strategies [62].

For comparative ChIP-Seq analysis mapped reads were transformed to continuous binding profiles. Those were used to collect data in 4-kb windows spanning CTCF and CTCFL binding sites. The binding sites were grouped into three classes based on intersection analysis: sites bound by CTCF only, CTCFL only, or both CTCF and CTCFL. The binding data were binned across binding sites in 50-bp windows, and the mean was calculated at each position in order to generate cumulative binding profiles. Alternatively the complete data were plotted in heatmaps. The identified CTCFL-binding motif was used to scan the mm9 genome using the Patser tool [63] and plotted as a heatmap after the motif data had been binned, as explained for the binding profiles.

RNA analyses

Expression analyses by real-time PCR were performed as follows: total RNA was isolated with RNA-Bee (Tel-Test Inc.). RNA was reverse transcribed (RT) with a combination of random and oligo-dT primers by Superscript reverse transcriptase (Invitrogen), and real-time RT-PCR was performed with a Sybrgreen platform on a Bio-Rad CFX Cycler. For oligos used, see Supplemental Information.

For testis and ES cell microarray analysis, the purity and quality of the isolated RNA were assessed by RNA 6000 Nano assay on a 2100 Bioanalyzer (Agilent Technologies). Then 5 ug testes RNA was used for the production of cRNA. Labeled cRNA was hybridized to the GeneChip Mouse Genome 430 2.0 array oligonucleotide microarray (Affymetrix) according to manufacturer's recommendations; 300 ng ES cell RNA was used for production of end-labeled biotinylated ssDNA. Labeled ssDNA was hybridized to the Mouse Gene 1.0ST array (Affymetrix) according to manufacturer's recommendations. Measured intensity values were analyzed using the Gene Expression Console (Affymetrix) and normalized by quantile normalization.

Scanned microarray data were processed using R/Bioconductor using standard procedures. Normalization and background correction were done by RMA. Differentially expressed genes were determined using the limma package within R [64]. For visualization the mean expression was determined across the heterozygous samples, which was then subtracted from the expression levels for the individual samples. For the analysis of the association between gene expression and CTCFL/CTCF binding, RefSeq genes were downloaded from the UCSC genome browser homepage. For each gene represented on the MoGene 1.0 ST array, the nearest CTCF or CTCFL site was calculated. Genes with a binding site within an interval from −2 kb to +1 kb around the transcriptional start sites were determined as bound. Log2-transformed expression values derived from Affymetrix analysis of mES cells was then plotted for the identified genes.

All Chip-seq and Microarray datasets are available at NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) under accessions: GSE34091, GSE34092, GSE34093 and GSE34094.

Electrophoretic mobility shift analysis (EMSA) or band shift analysis

Nuclear extracts were obtained from mock-transfected HEK 293 T cells and HEK 293 T transfected with pEGFP, pEGFP-mCTCF or pEGFP-mCTCFL. After 24 h, cells were harvested, washed with cold PBS, resuspended in buffer 1 [10 mM HEPES; 10 mM KCl; 0.25 mM EDTA pH 8, 0.125 M EGTA-K pH 8, 0.5 mM Spermidin, 0.1%, NP40, 1 mM DTT, protease inhibitor cocktail set I (Calbiochem)] and incubated for 10 min on ice. Cells were then centrifuged 5 min at 1,500 rpm. The supernatant was removed, and the pellet was resuspended in 50 μl of buffer 2 [20 mM HEPES; 0.4 M NaCl; 0.25 mM EDTA, 1.5 mM MgCl2, 0.5 mM DTT, protease inhibitor cocktail set I (Calbiochem)] and incubated 1 h at 4 °C. Samples were centrifuged for 30 min at 15,000 rpm, and the supernatant (nuclear extract) was frozen at −70 °C until used.

Radiolabeled probes were generated by PCR of genomic DNA (for oligos used, see Table 4). In all cases the PCR was performed in a final volume of 50 μl containing 3 μl of [α-32P]dCTP (Hartman Analytic), 20 ng of genomic DNA (from K562 cells), 0.2 mM each dNTP, 0.5 μM of each primer and 1U of DFS-Taq DNA polymerase (BIORON). The PCR fragments were purified using the Wizard SV Gel and PCR CleanUp System (Promega).

The EMSA reaction was performed by mixing 10 μg of the nuclear extract with 6 μl of EMSA buffer (1.5 μg of poli-dIdC, 20 mM HEPES pH 7.5, 50 mM KCl, 5% glycerol, 0.175 mM EDTA) in a volume of 19 μl. Mixtures were pre-incubated at 25 °C for 10 min. Then 162 μl of the radiolabeled probe was added to each condition, and the resulting mixture was incubated for 30 additional min at 25 °C. For competition, 10 μg of CTCF nuclear extract, followed by increasing amounts of competing extracts, was added to the binding reaction. Then, the mixtures were pre-incubated as previously described. For supershift experiments, 1 μl of anti-CTCF mouse monoclonal antibody (BD Biosciences) or 1 μl of anti-actin (Santa Cruz, sc-1616), used as a non-specific antibody, was added to the binding reaction prior to the radiolabeled probe. Complexes were analyzed by electrophoresis on a 4% polyacrylamide gel at 160 V for 2 h with 0.5×Tris-borate-EDTA buffer. Gels were fixed using 10% acetic acid for 10 min and then dried for 30 min using a Gel Dryer (Bio-Rad). Radioactive complexes were revealed using a Molecular Imager Fx (Bio-Rad).

Pathological analysis of ctcfl knockout mice

Testis weight was determined immediately after dissection. Weights were measured within the tunica albuginea, excluding the cauda epididymis. Sperm analysis and counts were performed as described [33]. The epididymis was dissected and transferred into a small conical glass grinder, and homogenized by hand. The total number of sperm present in the epididymis was counted using a Neubauer hemocytometer and a phase contrast microscope (magnification 400×). At least 200 sperm in two different samples were counted. Fertility of mice was determined by breeding the mice to multiple mates and scoring the number of offspring.

Immunofluorescence and immunohistochemistry techniques

For BrdU incorporation, mice were injected intraperitoneally with 1.2 μg BrdU. One hour after injection, testes were dissected, fixed in 4% paraformaldehyde or Bouin, paraffin embedded and sectioned. For H/E staining, Bouin or PFA fixed testes were fixed overnight at 4 °C, washed and dehydrated using ethanol and xylene, and embedded in paraffin. Sections of 10 μm were cut, mounted on SuperFrost Plus slides (Menzel-gläser), rehydrated and stained with H/E.

For immunofluorescence analyses, Bouin or PFA fixed and paraffin embedded testes were sectioned, treated with the microwave (three times for 5 min, 750 W) in 10 mM NaCitrate buffer (pH 6.0) to expose antigens and stained using standard procedures. Antibodies used: Rat monoclonal anti-CTCFL (see abobe), PRSS50 (Abcam 49405) and STRA8 (Abcam 49405).

Live imaging in seminiferous tubules

Imaging of testis tubules was performed as described [33]. Briefly, testis were injected through the rete testis with Hoechst 33342 and Trypan blue (Sigma) in 3–5 μl of PBS, 1 h prior to testis dissection, to allow spreading of the vital DNA stain throughout the adluminal compartment of the testis tubules and uptake by nuclei on the adluminal side of the Sertoli cell barrier. Trypan blue served as a marker for injected tubules. Individual seminiferous tubules were isolated from testes using a collagenase and hyaluronidase method, and Trypan blue positive tubules were transferred into a drop of PBS + with 0.2% BSA in a live-cell chamber overlaid with PBS-saturated mineral oil. The testis tubules were examined at 33 °C, using a Zeiss LSM510NLO confocal/multiphoton setup, to allow simultaneous acquisition of phase-contrast, GFP and Hoechst images.

Additional file

Additional file 1: Movie S1. Live imaging of GFP-CTCFL. Ex vivo confocal imaging of a live seminiferous tubule derived from a *Ctcfl*^{gfp} knockin mouse. Images were acquired throughout tubules using a combined multiphoton (Hoechst) and confocal laser (GFP) scanning microscope setup. Images were assembled for 3D reconstruction afterwards. The GFP-CTCFL fusion protein is shown in green. The DNA stain Hoecht, which was injected at the adluminal site of the testis tubule, is shown in red. Hoechst-positive cells represent Sertoli cells, leptotene stage IX and later stage spermatocytes, and spermatids, all of which are negative for GFP-CTCFL. Notice the presence of the GFP-CTCFL-positive cells on one side (basal lamina) of the tubule.

Abbreviations

BrdU: bromodeoxyuridine; ChIP: chromatin immunoprecipitation; DAB: diaminobenzidine; dCTP: deoxycytidine triphosphate; dNTP: deoxynucleoside triphosphates; DTT: dithiothreitol; EDTA: ethylenediaminetetraacetic acid; EMSA: electrophoretic mobility shift analysis; H/E: hematoxylin/eosin; HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; PCR: polymerase chain reaction; PBS: phosphate-buffered saline; PFA: paraformaldehyde; RMA: robust multi-array average; RPA: RNase protection analysis; TSS: transcription start site.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work was supported by the Earth and Life Sciences (ALW) and Medical Sciences (ZonMw) divisions of the Netherlands Organization for Scientific Research (NWO), the Dutch Cancer Society (KWF), the Dutch Cancer Genomics Centre (CGC) and Centre for Biomedical Genetics (CBG), an EC Integrated Project (EuTRACC), and the Spanish Fondo Investigaciones Sanitarias.

Author details

[1]Department of Cell Biology Erasmus Medical Center, Rotterdam, The Netherlands. [2]Institut für Genetik, Justus-Liebig-Universität, Giessen, Heinrich-Buff-Ring 58-62, 35392, Giessen, Germany. [3]Computational Biology Unit, Bergen Center for Computational Science, University of Bergen, Thormøhlensgate 55, N-5008, Bergen, Norway. [4]Department of Molecular Biology, Instituto de Biomedicina y Biotecnología de Cantabria, IBBTEC Universidad de Cantabria-CSIC-IDICAN, Santander, Spain. [5]Cell Cycle, Stem Cell Fate and Cancer Laboratory, Fundación Marqués de Valdecilla (IFIMAV), Santander, Spain. [6]Center for Biomics Erasmus Medical Center, Rotterdam, The Netherlands. [7]Department of Reproduction and Development Erasmus

Medical Center, Rotterdam, The Netherlands. [8]Cancer Genomics Center, NGI, Rotterdam, The Netherlands. [9]Center for Biomedical Genetics, Rotterdam, The Netherlands.

Authors' contributions

FS, WS, MB, HH, SD, PB, VF, MR-G, SvdN, LC, MvdR, JCB, WvIJ, BL, MDD, RR, and NG contributed to acquisition of data. FS, WS, MB, SD, PB, MR-G, WvIJ, AG, MDD, RR, FG, and NG designed the experiments, and analyzed and interpreted the data. FS, WS, MB, RR, and NG drafted the manuscript which wasapproved by all authors.

Received: 23 April 2012 Accepted: 18 June 2012
Published: 18 June 2012

References

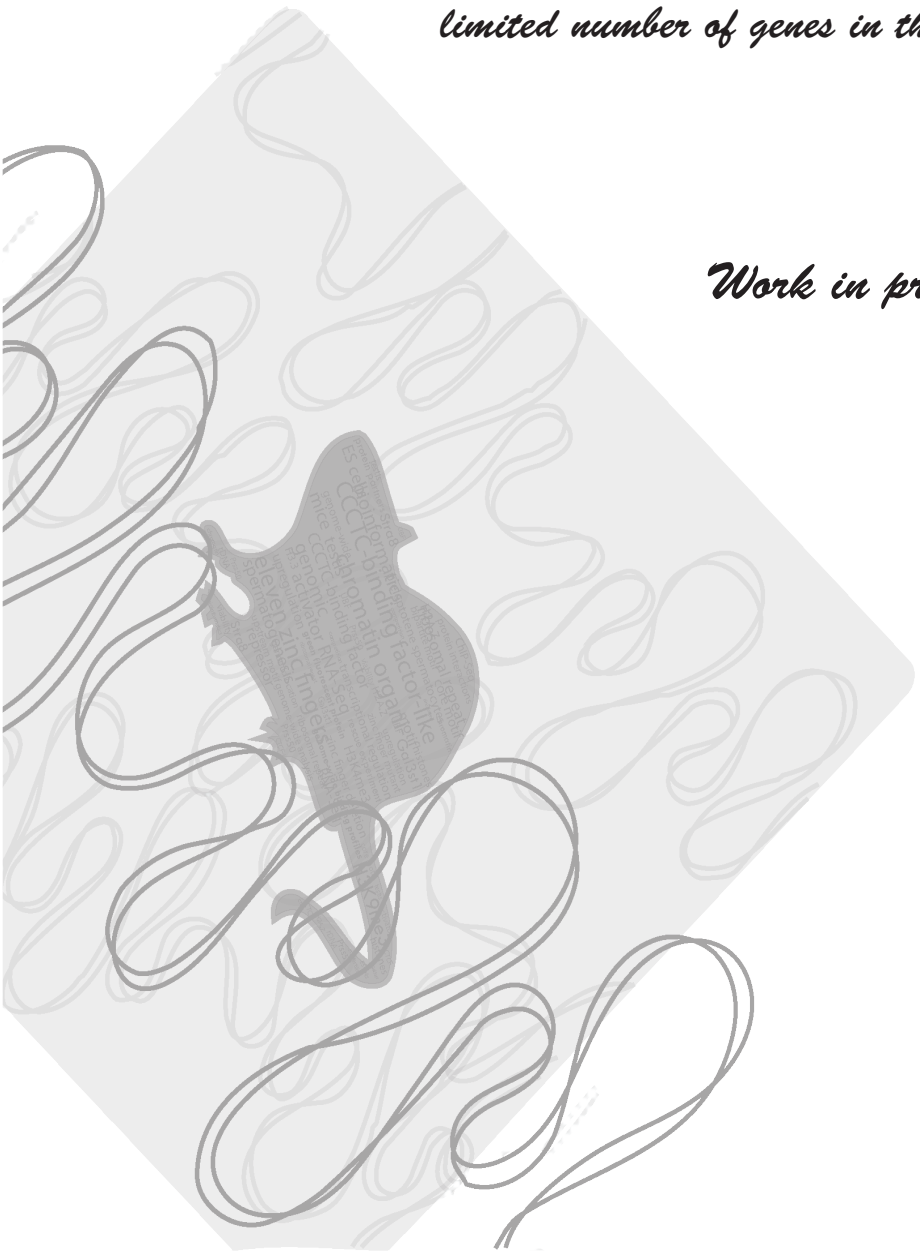
1. Bartkuhn M, Renkawitz R: **Long range chromatin interactions involved in gene regulation.** *Biochimica et biophysica acta* 2008, **1783:**2161–2166.
2. Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, Lee CW, Ye C, Ping JL, Mulawadi F, *et al*: **CTCF-mediated functional chromatin interactome in pluripotent cells.** *Nature Genetics* 2011, **43:**630–638.
3. Phillips JE, Corces VG: **CTCF: master weaver of the genome.** *Cell* 2009, **137:**1194–1211.
4. Heath H, Ribeiro De Almeida C, Sleutels F, Dingjan G, Van De Nobelen S, Jonkers I, Ling KW, Gribnau J, Renkawitz R, Grosveld F, *et al*: **CTCF regulates cell cycle progression of alphabeta T cells in the thymus.** *Embo J* 2008, **27:**2839–2850.
5. 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: **Critical role for the transcription regulator CCCTC-binding factor in the control of Th2 cytokine expression.** *J Immunol* 2009, **182:**999–1010.
6. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K: **High-resolution profiling of histone methylations in the human genome.** *Cell* 2007, **129:**823–837.
7. Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, *et al*: **Integration of external signaling pathways with the core transcriptional network in embryonic stem cells.** *Cell* 2008, **133:**1106–1117.
8. Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, *et al*: **Mediator and cohesin connect gene expression and chromatin architecture.** *Nature* 2010, **467:**430–435.
9. Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenkov VV, Ren B: **Analysis of the vertebrate insulator protein CTCF- binding sites in the human genome.** *Cell* 2007, **128:**1231–1245.
10. Nitzsche A, Paszkowski-Rogacz M, Matarese F, Janssen-Megens EM, Hubner NC, Schulz H, de Vries I, Ding L, Huebner N, Mann M, *et al*: **RAD21 cooperates with pluripotency transcription factors in the maintenance of embryonic stem cell identity.** *PloS one* 2011, **6:**e19470.
11. Cuddapah S, Jothi R, Schones DE, Roh TY, Cui K, Zhao K: **Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains.** *Genome research* 2009, **19:**24–32.
12. Fu Y, Sinha M, Peterson CL, Weng Z: **The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome.** *PLoS genetics* 2008, **4:**e1000138.
13. Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A: **Determinants of nucleosome organization in primary human cells.** *Nature* 2011, **474:**516–520.
14. van de Nobelen S, Rosa-Garrido M, Leers J, Heath H, Soochit W, Joosen L, Jonkers I, Demmers J, van der Reijden M, Torrano V, *et al*: **CTCF regulates the local epigenetic state of ribosomal DNA repeats.** *Epigenetics & chromatin* 2010, **3:**19.
15. Nativio R, Wendt KS, Ito Y, Huddleston JE, Uribe-Lewis S, Woodfine K, Krueger C, Reik W, Peters JM, Murrell A: **Cohesin is required for higher-order chromatin conformation at the imprinted *Igf2-H19* locus.** *PLoS Genet* 2009, **5:**e1000739.
16. Parelho V, Hadjur S, Spivakov M, Leleu M, Sauer S, Gregson HC, Jarmuz A, Canzonetta C, Webster Z, Nesterova T, *et al*: **Cohesins functionally associate with CTCF on mammalian chromosome arms.** *Cell* 2008, **132:**422–433.
17. Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T, *et al*: **Cohesin mediates transcriptional insulation by CCCTC-binding factor.** *Nature* 2008, **451:**796–801.
18. Kurukuti S, Tiwari VK, Tavoosidana G, Pugacheva E, Murrell A, Zhao Z, Lobanenkov V, Reik W, Ohlsson R: **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* 2006, **103:**10684–10689.
19. Loukinov DI, Pugacheva E, Vatolin S, Pack SD, Moon H, Chernukhin I, Mannan P, Larsson E, Kanduri C, Vostrov AA, *et al*: **BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma.** *Proc Natl Acad Sci U S A* 2002, **99:**6806–6811.
20. Hore TA, Deakin JE, Marshall Graves JA: **The evolution of epigenetic regulators CTCF and BORIS/CTCFL in amniotes.** *PLoS Genet* 2008, **4:**e1000169.
21. D'Arcy V, Abdullaev ZK, Pore N, Docquier F, Torrano V, Chernukhin I, Smart M, Farrar D, Metodiev M, Fernandez N, *et al*: **The potential of BORIS detected in the leukocytes of breast cancer patients as an early marker of tumorigenesis.** *Clin Cancer Res* 2006, **12:**5978–5986.
22. Vatolin S, Abdullaev Z, Pack SD, Flanagan PT, Custer M, Loukinov DI, Pugacheva E, Hong JA, Morse H 3rd, Schrump DS, *et al*: **Conditional expression of the CTCF-paralogous transcriptional factor BORIS in normal cells results in demethylation and derepression of MAGE-A1 and reactivation of other cancer-testis genes.** *Cancer Res* 2005, **65:**7751–7762.
23. Hong JA, Kang Y, Abdullaev Z, Flanagan PT, Pack SD, Fischette MR, Adnani MT, Loukinov DI, Vatolin S, Risinger JI, *et al*: **Reciprocal binding of CTCF and BORIS to the NY-ESO-1 promoter coincides with derepression of this cancer-testis gene in lung cancer cells.** *Cancer Res* 2005, **65:**7763–7774.
24. Jelinic P, Stehle JC, Shaw P: **The testis-specific factor CTCFL cooperates with the protein methyltransferase PRMT7 in *H19* imprinting control region methylation.** *PLoS Biol* 2006, **4:**e355.
25. Tomizawa S, Sasaki H: **Genomic imprinting and its relevance to congenital disease, infertility, molar pregnancy and induced pluripotent stem cell.** *J Hum Genet* 2012, **57:**84–91.
26. Suzuki T, Kosaka-Suzuki N, Pack S, Shin DM, Yoon J, Abdullaev Z, Pugacheva E, Morse HC 3rd, Loukinov D, Lobanenkov V: **Expression of a testis-specific form of Gal3st1 (CST), a gene essential for spermatogenesis, is regulated by the CTCF paralogous gene BORIS.** *Mol Cell Biol* 2010, **30:**2473–2484.
27. Kosaka-Suzuki N, Suzuki T, Pugacheva EM, Vostrov AA, Morse HC 3rd, Loukinov D, Lobanenkov V: **Transcription factor BORIS (Brother of the Regulator of Imprinted Sites) directly induces expression of a cancer-testis antigen, TSP50, through regulated binding of BORIS to the promoter.** *J Biol Chem* 2011, **286:**27378–27388.
28. Sleutels F, Zwart R, Barlow DP: **The non-coding Air RNA is required for silencing autosomal imprinted genes.** *Nature* 2002, **415:**810–813.
29. Anderson EL, Baltus AE, Roepers-Gajadien HL, Hassold TJ, de Rooij DG, van Pelt AM, Page DC: **Stra8 and its inducer, retinoic acid, regulate meiotic initiation in both spermatogenesis and oogenesis in mice.** *Proc Natl Acad Sci U S A* 2008, **105:**14976–14980.
30. Mark M, Jacobs H, Oulad-Abdelghani M, Dennefeld C, Feret B, Vernet N, Codreanu CA, Chambon P, Ghyselinck NB: **STRA8-deficient spermatocytes initiate, but fail to complete, meiosis and undergo premature chromosome condensation.** *J Cell Sci* 2008, **121:**3233–3242.
31. Oulad-Abdelghani M, Bouillet P, Decimo D, Gansmuller A, Heyberger S, Dolle P, Bronner S, Lutz Y, Chambon P: **Characterization of a premeiotic germ cellspecific cytoplasmic protein encoded by Stra8, a novel retinoic acid-responsive gene.** *The Journal of cell biology* 1996, **135:**469–477.
32. Zhou Q, Nie R, Li Y, Friel P, Mitchell D, Hess RA, Small C, Griswold MD: **Expression of stimulated by retinoic acid gene 8 (Stra8) in spermatogenic cells induced by retinoic acid: an in vivo study in vitamin A-sufficient postnatal murine testes.** *Biol Reprod* 2008, **79:**35–42.
33. Akhmanova A, Mausset-Bonnefont AL, van Cappellen W, Keijzer N, Hoogenraad CC, Stepanova T, Drabek K, van der Wees J, Mommaas M, Onderwater J, *et al*: **The microtubule plus-end-tracking protein CLIP-170 associates with the spermatid manchette and is essential for spermatogenesis.** *Genes Dev* 2005, **19:**2501–2515.
34. Cheng CY, Mruk DD: **Cell junction dynamics in the testis: Sertoli-germ cell interactions and male contraceptive development.** *Physiol Rev* 2002, **82:**825–874.

35. Honke K, Hirahara Y, Dupree J, Suzuki K, Popko B, Fukushima K, Fukushima J, Nagasawa T, Yoshida N, Wada Y, Taniguchi N: **Paranodal junction formation and spermatogenesis require sulfoglycolipids.** *Proc Natl Acad Sci U S A* 2002, **99:**4227–4232.
36. Xu HP, Yuan L, Shan J, Feng H: **Localization and expression of TSP50 protein in human and rodent testes.** *Urology* 2004, **64:**826–832.
37. Geijsen N, Horoschak M, Kim K, Gribnau J, Eggan K, Daley GQ: **Derivation of embryonic germ cells and male gametes from embryonic stem cells.** *Nature* 2004, **427:**148–154.
38. Kanatsu-Shinohara M, Shinohara T: **Culture and genetic modification of mouse germline stem cells.** *Ann N Y Acad Sci* 2007, **1120:**59–71.
39. Sharova LV, Sharov AA, Piao Y, Shaik N, Sullivan T, Stewart CL, Hogan BL, Ko MS: **Global gene expression profiling reveals similarities and differences among mouse pluripotent stem cells of different origins and strains.** *Dev Biol* 2007, **307:**446–459.
40. Zwaka TP, Thomson JA: **A germ cell origin of embryonic stem cells?** *Development* 2005, **132:**227–233.
41. Chen Q, Lin L, Smith S, Huang J, Berger SL, Zhou J: **CTCF-dependent chromatin boundary element between the latency-associated transcript and ICP0 promoters in the herpes simplex virus type 1 genome.** *Journal of virology* 2007, **81:**5192–5201.
42. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, *et al*: **Histone H3K27ac separates active from poised enhancers and predicts developmental state.** *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107:**21931–21936.
43. Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, Sharp PA, Young RA: **c-Myc regulates transcriptional pause release.** *Cell* 2010, **141:**432–445.
44. Jin C, Felsenfeld G: **Nucleosome stability mediated by histone variants H3.3 and H2A.Z.** *Genes & development* 2007, **21:**1519–1529.
45. Jin C, Zang C, Wei G, Cui K, Peng W, Zhao K, Felsenfeld G: **H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions.** *Nature Genetics* 2009, **41:**941–945.
46. Goldberg AD, Banaszynski LA, Noh KM, Lewis PW, Elsaesser SJ, Stadler S, Dewell S, Law M, Guo X, Li X, *et al*: **Distinct factors control histone variant H3.3 localization at specific genomic regions.** *Cell* 2010, **140:**678–691.
47. Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PA, Rappsilber J, Helin K: **TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity.** *Nature* 2011, **473:**343–348.
48. Van der Heijden GW, Derijck AA, Posfai E, Giele M, Pelczar P, Ramos L, Wansink DG, Van der Vlag J, Peters AH, De Boer P: **Chromosome-wide nucleosome replacement and H3.3 incorporation during mammalian meiotic sex chromosome inactivation.** *Nature Genetics* 2007, **39:**251–258.
49. Nguyen P, Bar-Sela G, Sun L, Bisht KS, Cui H, Kohn E, Feinberg AP, Gius D: **BAT3 and SET1A form a complex with CTCFL/BORIS to modulate H3K4 histone dimethylation and gene expression.** *Mol Cell Biol* 2008, **28:**6720–6729.
50. Peters JM, Tedeschi A, Schmitz J: **The cohesin complex and its roles in chromosome biology.** *Genes & development* 2008, **22:**3089–3114.
51. Rubio ED, Reiss DJ, Welcsh PL, Disteche CM, Filippova GN, Baliga NS, Aebersold R, Ranish JA, Krumm A: **CTCF physically links cohesin to chromatin.** *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105:**8309–8314.
52. Xiao T, Wallace J, Felsenfeld G: **Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity.** *Molecular and cellular biology* 2011, **31:**2174–2183.
53. Yu W, Ginjala V, Pant V, Chernukhin I, Whitehead J, Docquier F, Farrar D, Tavoosidana G, Mukhopadhyay R, Kanduri C, *et al*: **Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation.** *Nat Genet* 2004, **36:**1105–1110.
54. Looijenga LH, Stoop H, Hersmus R, Gillis AJ, Wolter Oosterhuis J: **Genomic and expression profiling of human spermatocytic seminomas: pathogenetic implications.** *Int J Androl* 2007, **30:**328–335. discussion 335–326.
55. Renaud S, Loukinov D, Alberti L, Vostrov A, Kwon YW, Bosman FT, Lobanenkov V, Benhattar J: **BORIS/CTCFL-mediated transcriptional regulation of the hTERT telomerase gene in testicular and ovarian tumor cells.** *Nucleic Acids Res* 2011, **39:**862–873.
56. Woloszynska-Read A, Zhang W, Yu J, Link PA, Mhawech-Fauceglia P, Collamat G, Akers SN, Ostler KR, Godley LA, Odunsi K, Karpf AR: **Coordinated cancer germline antigen promoter and global DNA hypomethylation in ovarian cancer: association with the BORIS/CTCF expression ratio and advanced stage.** *Clinical cancer research: an official journal of the American Association for Cancer Research* 2011, **17:**2170–2180.
57. Woloszynska-Read A, James SR, Song C, Jin B, Odunsi K, Karpf AR: **BORIS/CTCFL expression is insufficient for cancer-germline antigen gene expression and DNA hypomethylation in ovarian cell lines.** *Cancer Immun* 2010, **10:**6.
58. Pugacheva EM, Suzuki T, Pack SD, Kosaka-Suzuki N, Yoon J, Vostrov AA, Barsov E, Strunnikov AV, Morse HC 3rd, Loukinov D, Lobanenkov V: **The structural complexity of the human BORIS gene in gametogenesis and cancer.** *PloS one* 2010, **5:**e13872.
59. Wilming LG, Gilbert JG, Howe K, Trevanion S, Hubbard T, Harrow JL: **The vertebrate genome annotation (Vega) database.** *Nucleic Acids Res* 2008, **36:**D753–760.
60. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol* 2009, **10:**R25.
61. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS: **Model-based analysis of ChIP-Seq (MACS).** *Genome Biol* 2008, **9:**R137.
62. Soler E, Andrieu-Soler C, Boer E, Bryne JC, Thongjuea S, Rijkers E, Demmers J, Ijcken W, Grosveld F: **A systems approach to analyze transcription factors in mammalian cells.** *Methods* 2011, **53:**151–162.
63. Hertz GZ, Stormo GD: **Identifying DNA and protein patterns with statistically significant alignments of multiple sequences.** *Bioinformatics* 1999, **15:**563–577.
64. Smyth GK: **Linear models and empirical Bayes methods for assessing differential expression in microarray experiments.** *Stat Appl Genet Mol Biol* 2004, **3**. Article3 http://www.degruyter.com/view/j/sagmb.2004.3.1/sagmb.2004.3.1.1027/sagmb.2004.3.1.1027.xml.

doi:10.1186/1756-8935-5-8
Cite this article as: Sleutels *et al.*: The male germ cell gene regulator CTCFL is functionally different from CTCF and binds CTCF-like consensus sites in a nucleosome composition-dependent manner. *Epigenetics & Chromatin* 2012 **5**:8.

Chapter 5

CTCFL activates transcription of a limited number of genes in the testis



Work in progress

CTCFL activates transcription of a limited number of genes in the testis

Widia Soochit[1], Marek Barthkuhn[2], Zhe Liu[1], Jos Hoogerbrugge[3], Reinier van der Linden[4], Frank Grosveld[1,5], Rainer Renkawitz[2], Frank Sleutels[1] and Niels Galjart[1,5,6]

1: Department of Cell Biology Erasmus Medical Center, Rotterdam, The Netherlands
2: Institut für Genetik, Justus-Liebig-Universität, Heinrich-Buff-Ring 58-62, 35392, Giessen, Germany
3: Department of Reproduction and Development Erasmus Medical Center, Rotterdam, The Netherlands
4: Erasmus Medical Center Stem Cell Institute, Rotterdam, The Netherlands
5: Epigenetics Consortium, NGI, Erasmus Medical Center, Rotterdam, The Netherlands
6: Corresponding author: n.galjart@erasmusmc.nl

Abstract

CTCFL, the testis-specific paralogue of the highly conserved chromatin organizer CTCF, is expressed in type B spermatogonia and pre-leptotene spermatocytes of adult male mice. CTCFL binds to a 20 bp consensus sequence similar to that for CTCF, but is mainly located on promoters, and was proposed to regulate expression of the germ cell-specific factors *Prss50*, *Stra8* and *Gal3st1*. To examine the transcriptional function of CTCFL genome-wide and in more detail, we used a novel FACS-based approach to separate CTCFL-expressing cells from non-expressing cells. We have shown previously that CTCFL and STRA8 are present in the same cells in the adult testis. We therefore sorted GFP$^+$ and GFP$^-$ cell populations from a *Stra8-Gfp* transgenic mouse maintained in either a wild type or *Ctcfl* knock out background. RNA-Sequencing revealed specific enrichment of *Stra8*, *Prss50* mRNAs in the GFP$^+$ fractions, validating our purification method. Many more genes were down-regulated than up-regulated genes in the GFP$^+$ *Ctcfl* knock out fraction, confirming the hypothesis that CTCFL is a transcriptional activator. Aberrant expression of *Stra8* was observed in ES cells expressing a mutated form of CTCF and a higher level of *Stra8* expression could be obtained upon exogenous expression of CTCFL in these ES cells compared to CTCFL induction in wild type ES cells. Combined our data suggest that CTCFL activates transcription of a limited number of genes in the testis and that CTCF acts as repressor of these genes. This suggests that CTCFL competes with CTCF to maintain proper gene expression in the testis.

Introduction

CTCF (CCCTC-binding factor) is a highly conserved and ubiquitously expressed nuclear protein that coordinates higher-order chromatin structures together with cohesin and other factors (for review see *(Phillips and Corces, 2009)*). CCCTC-binding factor-Like (CTCFL), also known as BORIS (Brother Of the Regulator of Imprinted Sites), is less conserved and is only present in certain vertebrates, including amphibians, reptiles, monotremes and mammals *(Hore et al., 2008; Loukinov et al., 2002)*. Both proteins have N- and C-terminal regions surrounding an eleven zinc finger (ZF) domain. Whereas the N- and C-termini of CTCFL and CTCF are not homologous, the ZF domains of these proteins are 71% identical *(Loukinov et al., 2002)*.

In mice and humans CTCFL is virtually only expressed in the testis, suggesting a role for CTCFL in gametogenesis *(Loukinov et al., 2002)*. In adult mice CTCFL is detected during early spermatogenesis, where it localizes in the nucleus of late spermatogonia and pre-leptotene spermatocytes *(Sleutels et al., 2012)*. *Ctcfl* knock out mice show defects in spermatogenesis, which results in small testicles and reduced fertility *(Sleutels et al., 2012; Suzuki et al., 2010)*. CTCFL has been proposed to activate the *Prss50* (*Protease serine 50*), *Gal3st1* (*Galactose-3-O-Sulfotransferase 1*), and *Stra8* (*Stimulated by retinoic acid*) genes *(Kosaka-Suzuki et al., 2011; Sleutels et al., 2012; Suzuki et al., 2010)*. *Prss50* is expressed during early spermatogenesis from spermatogonia to pachytene spermatocytes *(Sleutels et al., 2012)*, but not much is known about the function of this protein. By contrast, *Gal3st1* is crucial for spermatogenesis, since *Gal3st1* knock out mice are infertile due to an arrest at the end of the meiotic prophase *(Honke et al., 2002)*. *Stra8* is essential for initiation of meiosis in both males and females *(Anderson et al., 2008; Baltus et al., 2006; Mark et al., 2008)*.

Genome-wide binding studies in embryonic stem cells (ES) cells overexpressing CTCFL revealed ~5700 sites *(Sleutels et al., 2012)*. Of these sites ~3700 could also be bound by CTCF, while ~2000 sites were only bound by CTCFL. CTCFL recognizes a similar consensus sequence as CTCF and is able to compete with CTCF on specific genes to regulate their expression. CTCFL binding sites are enriched on promoters and are marked by the active chromatin marks H3.3 and H3K4me3 *(Sleutels et al., 2012)*. Altogether these data suggest that the main function of CTCFL is to regulate transcription of genes in a germ cell-specific manner by binding to their promoters.

STRA8 and CTCFL showed a completely overlapping protein expression pattern in the testis of adult mice *(Sleutels et al., 2012)*. Recently, a transgenic mouse strain was generated in which GFP expression is regulated by a 1.4 kb region encompassing the *Stra8* promoter *(Nayernia et al., 2004)*. Here, we explored the feasibility of using this mouse line, on wild type or *Ctcfl*-negative background, to specifically isolate the CTCFL-expressing cell population from seminiferous tubules of the testis using FACS. We show that this method enables the fractionation of *Stra8* and *Prss50* RNAs into a GFP^+ cell fraction. RNA-Sequencing (RNA-Seq) reveals that ~600 genes are down-regulated in the *Ctcfl* knock out GFP^+ cell fraction, whereas ~100 genes are upregulated. These data suggest that CTCFL is a transcriptional activator in testis that competes with CTCF as shown by functional experiments using ES cells.

Results

Genome-wide transcription profiles of STRA8-GFP$^+$ and STRA8-GFP$^-$ germ cell fractions

STRA8 and CTCFL protein distribution completely overlaps in the testis *(Sleutels et al., 2012)*, which makes STRA8 a suitable marker to discriminate between CTCFL-expressing and non-expressing cells. Therefore, *Ctcfl*$^{wt/del}$ mice were crossed with *Stra8-Gfp* transgenic mice to mark the CTCFL-expressing cells with GFP *(Nayernia et al., 2004)*. The *Stra8-Gfp* fusion construct was generated harboring the 1.4 kb promoter region of *Stra8* followed by a *Gfp* cassette. This 1.4 kb region is sufficient to direct gene expression to the germinal stem cells in testis of transgenic mice *(Giuili et al., 2002; Nayernia et al., 2004)*. Furthermore, this region does not contain the shared CTCF and CTCFL site located in the first intron of *Stra8* as described previously *(Sleutels et al., 2012)*.

By crossing the *Ctcfl*$^{wt/del}$ mice with the *Stra8-Gfp* transgenic mice we generated *Ctcfl*$^{wt/wt}$*-Stra8-Gfp*, *Ctcfl*$^{wt/del}$*-Stra8-Gfp* and *Ctcfl*$^{del/del}$*-Stra8-Gfp* double mutant mice. Testes from double mutant mice were isolated and germ cell preparation was performed in order to enrich for spermatogonia. After germ cell preparation testis extracts were either directly used in following assays or FACS-sorted using the GFP signal. This experimental set-up produced 9 distinct cell populations; spermatogonia enriched samples from *Ctcfl*$^{wt/wt}$, *Ctcfl*$^{wt/del}$, *Ctcfl*$^{del/del}$ mice, and spermatogonia enriched samples followed by FACS-sorting of *Ctcfl*$^{wt/wt}$, *Ctcfl*$^{wt/del}$, *Ctcfl*$^{del/del}$ mice, divided into GFP$^+$ and GFP$^-$ (i.e. STRA8/CTCFL$^+$ and STRA8/CTCFL$^-$) fractions.

Figure 1

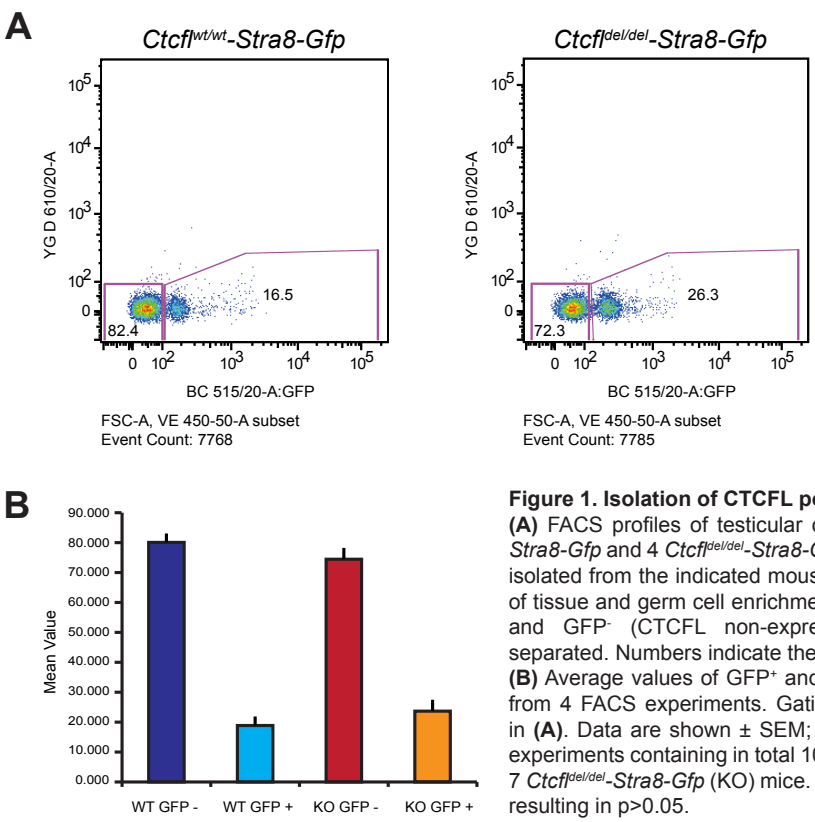


Figure 1. Isolation of CTCFL positive cells from testis
(A) FACS profiles of testicular cells derived from 2 *Ctcfl*$^{wt/wt}$*-Stra8-Gfp* and 4 *Ctcfl*$^{del/del}$*-Stra8-Gfp* mice. Testicular cells were isolated from the indicated mouse lines and after dissociation of tissue and germ cell enrichment GFP$^+$ (CTCFL expressing) and GFP$^-$ (CTCFL non-expressing) cell fractions were separated. Numbers indicate the percentages of gated cells.
(B) Average values of GFP$^+$ and GFP$^-$ cell fractions obtained from 4 FACS experiments. Gating was performed as shown in **(A)**. Data are shown ± SEM; 4 independent FACS sorting experiments containing in total 10 *Ctcfl*$^{wt/wt}$*-Stra8-Gfp* (WT) and 7 *Ctcfl*$^{del/del}$*-Stra8-Gfp* (KO) mice. Student T-test was performed resulting in p>0.05.

Interestingly, our first FACS sorting experiment using 4 *Ctcfl*$^{del/del}$ mice (8 testes) and 2 *Ctcfl*$^{wt/wt}$ mice (4 testes) revealed a general increase in the percentage of GFP^{+} cells in the *Ctcfl*$^{del/del}$ mice **(Figure 1A)**. We subsequently analyzed more mice, total 10 *Ctcfl*$^{del/del}$*-Stra8-Gfp* and 7 *Ctcfl*$^{wt/wt}$*-Stra8-Gfp* (14 testes), in 4 independent FACS sorting experiments. Although not statistically significant, *Ctcfl*$^{del/del}$ mice showed a ~1.25 times increase of the percentage of GFP^{+} cells **(Figure 1B)**. These results suggest that deletion of *Ctcfl* causes a developmental delay in the differentiation of late spermatogonia/pre-leptotene spermatocytes, leading to an increase in the fraction of GFP^{+} CTCFL^{-} cells in the knock out.

Figure 2

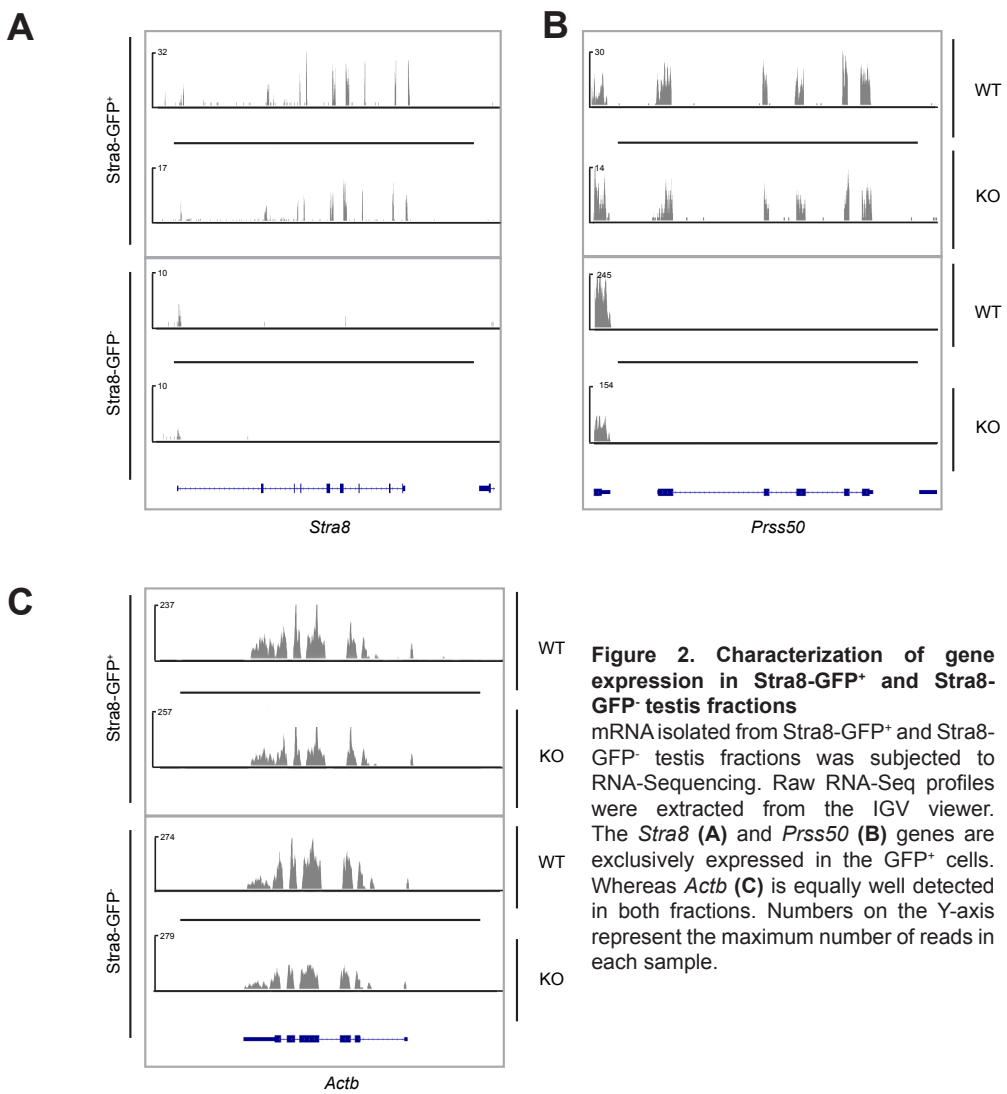


Figure 2. Characterization of gene expression in Stra8-GFP^{+} and Stra8-GFP^{-} testis fractions
mRNA isolated from Stra8-GFP^{+} and Stra8-GFP^{-} testis fractions was subjected to RNA-Sequencing. Raw RNA-Seq profiles were extracted from the IGV viewer. The *Stra8* **(A)** and *Prss50* **(B)** genes are exclusively expressed in the GFP^{+} cells. Whereas *Actb* **(C)** is equally well detected in both fractions. Numbers on the Y-axis represent the maximum number of reads in each sample.

CTCFL effects on transcription were examined by RNA-Sequencing of germ cell enriched and FACS-sorted GFP^{+} and GFP^{-} cell populations derived from *Ctcfl*$^{wt/wt}$*-Stra8-Gfp* and *Ctcfl*$^{del/del}$*-Stra8-Gfp* mice. We performed two independent experiments using in each experiment 1 *Ctcfl* wild type (2 testes) and 1 *Ctcfl* knock out (2 testes) mice. We analyzed

changes in gene expression by comparing the *Ctcfl*$^{wt/wt}$ GFP^{+} and the *Ctcfl*$^{del/del}$ GFP^{+} cells within each experiment with each other (further referred to as dataset 1 and dataset 2). To examine whether the FACS sorting worked properly, we checked genes that were known to be expressed in either GFP^{+} or GFP^{-} cell population. In both experiments *Stra8* and *Prss50* RNAs were exclusively found in the GFP^{+} cell fractions, both in wild type as well as in *Ctcfl* knockout mice **(Figure 2A, B,** and data not shown**)**. By contrast, *Actb*, encoding ACTIN, was found in both the GFP^{+} and GFP^{-} fractions **(Figure 2C)**.

Figure 3

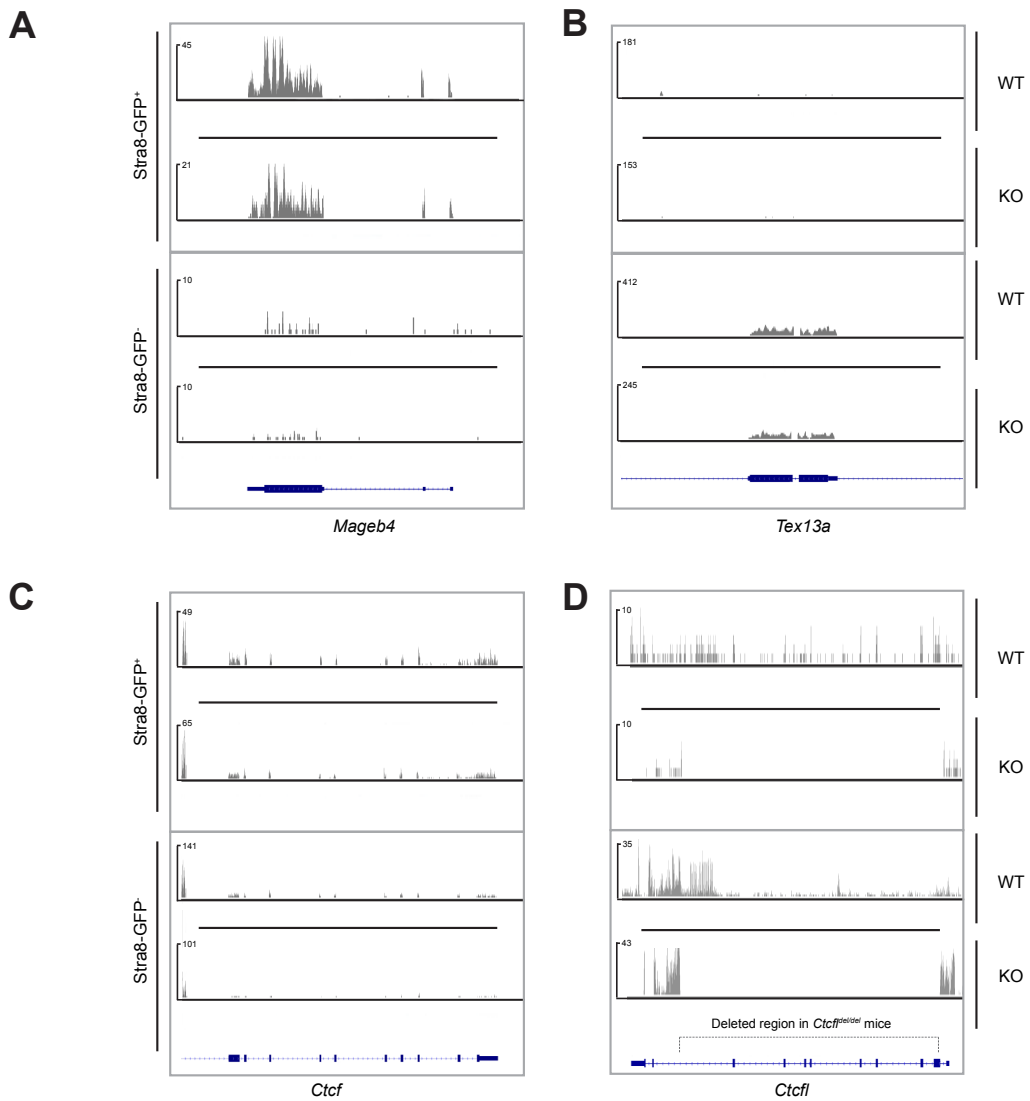


Figure 3. Characterization of gene expression in Stra8-GFP^{+} and Stra8-GFP^{-} testis fractions
mRNA isolated from Stra8-GFP^{+} and Stra8-GFP^{-} testis fractions was subjected to RNA-Sequencing. Raw RNA-Seq profiles were extracted from the IGV viewer. The *Mageb4* **(A)** and *Tex13a* **(B)** genes are mutually exclusive expressed in the GFP^{+} and GFP^{-} cells, respectively. In contrast, *Ctcf* mRNA **(C)** is not enriched in one of the GFP fractions. *Ctcfl* **(D)** is not abundantly expressed in WT, making its detection with RNA-Seq, and its localization to a certain fraction, less accurate. *Ctcfl* transcripts are not detected in KO. Grey dashed lines mark the deleted area in the *Ctcfl*$^{del/del}$ mice. Numbers on the Y-axis represent the maximum number of reads in each sample.

Further visual inspection of the GFP$^+$ and GFP$^-$ fractions from wild type and *Ctcfl* knock out mice revealed that the *Mageb4* **(Figure 3A)** and *Tex13a* **(Figure 3B)** mRNAs, which encode a member of the cancer testis antigen (CTA) family and a testis-expressed sequence respectively, are enriched in either the GFP$^+$ or the GFP$^-$ fraction. *Ctcf* mRNA is found in both fractions **(Figure 3C)**, like *Actb*. *Ctcfl* is expressed at lower levels compared to *Ctcf*, nevertheless its segregation into the GFP$^+$ fraction of wild type mice is obvious **(Figure 3D)**. These data show that our FACS-based RNA-Seq approach is a valid method to separate *Ctcfl*$^+$ and *Ctcfl*$^-$ cells from each other.

Genome-wide analysis of the RNA-Seq data was performed to determine the degree of similarity of the two independent experiments. This revealed 691 and 787 deregulated genes between wild type and *Ctcfl* knock out in the GFP$^+$ fractions of dataset 1 and dataset 2, respectively. In dataset 1 574 and 117 genes were down- and up-regulated, respectively, in *Ctcfl* knock out GFP$^+$ cells, whereas in dataset 2 613 and 174 genes were down- and up-regulated, respectively. The two datasets did not share any of the up-regulated genes. However, 88 genes were down-regulated in both datasets **(Table S1)**. This analysis revealed that only a subset (~11% of all genes) of deregulated genes was similar between the two datasets.

Next, we examined *Ctcfl*, along with the *Ctcfl*-responsive genes *Gal3st1*, *Prss50* and *Stra8* expression. Expression of these genes were reduced in *Ctcfl*$^{del/del}$ GFP$^+$ cells. Of all down-regulated genes, only *Gal3st1* and the pseudogene *Gm6525* showed a >10-fold fold change in expression, indicating that these genes depend on CTCFL confirming previous data. The other genes, including *Prss50* and *Stra8*, showed less reduction in mRNA levels in the GFP$^+$ *Ctcfl* knock out fractions compared to *Gal3st1* and *Gm6525* **(Table S1)**. This suggests that CTCFL regulates these genes but is not the only (direct) regulator.

In addition to the known *Ctcfl*-responsive genes we detected other deregulated genes involved in gametogenesis. In *Ctcfl*$^{del/del}$ Stra8$^+$ cells *Dmrt1* **(Table S1)** and *Kit* (only in dataset 1, data not shown) were down-regulated. Both genes are essential for spermatogonial differentiation *(Brannan et al., 1992; Schrans-Stassen et al., 1999)*. Furthermore, DMRT1 represses *Stra8* during spermatogenesis but stimulates *Stra8* during oogenesis *(Krentz et al., 2011; Matson et al., 2010)*.

CTCFL is aberrantly expressed in multiple types of cancer and regulates transcription of several genes including the multi-gene family Mage. Interestingly, a subset of genes from the Mage-family was down-regulated in *Ctcfl*$^{del/del}$ STRA8$^+$ cells **(Table 1).** Additionally, we noticed multiple other genes involved in carcinogenesis e.g. *Brca1*, *Atm* and *Igf1r* **(Table S1)**. Overall these data show that CTCFL is a transcriptional activator. They also reveal a possible connection between genes involved in tumorigenesis and their regulation by CTCFL.

Table 1

	Dataset 1					Dataset 2			
gene	wt/wt Expression (FPKM)	del/del Expression (FPKM)	fold change (log2)	p-value	gene	wt/wt Expression (FPKM)	del/del Expression (FPKM)	fold change (log2)	p-value
Magea10	1.53978	0.52578	-1.5502	0.0477811	Magea10	2.56129	0.788557	-1.69958	0.0178293
Mageb4	13.7945	8.08269	-0.771181	0.0306431	Mageb4	24.3835	13.3884	-0.864923	0.0221595
Maged2	4.428	1.76505	-1.32694	0.021447	Magea5	3.50843	1.07114	-1.71168	0.0141532
Maged1	30.0477	20.1085	-0.57945	0.0426714	Magea8	3.65869	1.40882	-1.37683	0.0365106
Magea3	1.54297	0.343443	-2.16757	0.0433912					
Magea6	1.5538	0.387263	-2.00442	0.0446153					

Table 1. RNA-Seq expression profiles of *Ctcfl*$^{wt/wt}$*-Stra8-Gfp* and *Ctcfl*$^{del/del}$*-Stra8-Gfp mice
Gene expression levels of deregulated Mage genes by RNA-sequencing in *Ctcfl*$^{wt/wt}$*-Stra8-Gfp* and *Ctcfl*$^{del/del}$*-Stra8-Gfp* mice from two independent experiments. Values are depicted in FPKM (Fragments per Kilobase of transcripts per Million mapped reads). Fold change (log2) between wild type and knock out mice are depicted together with the p-value.

Figure 4

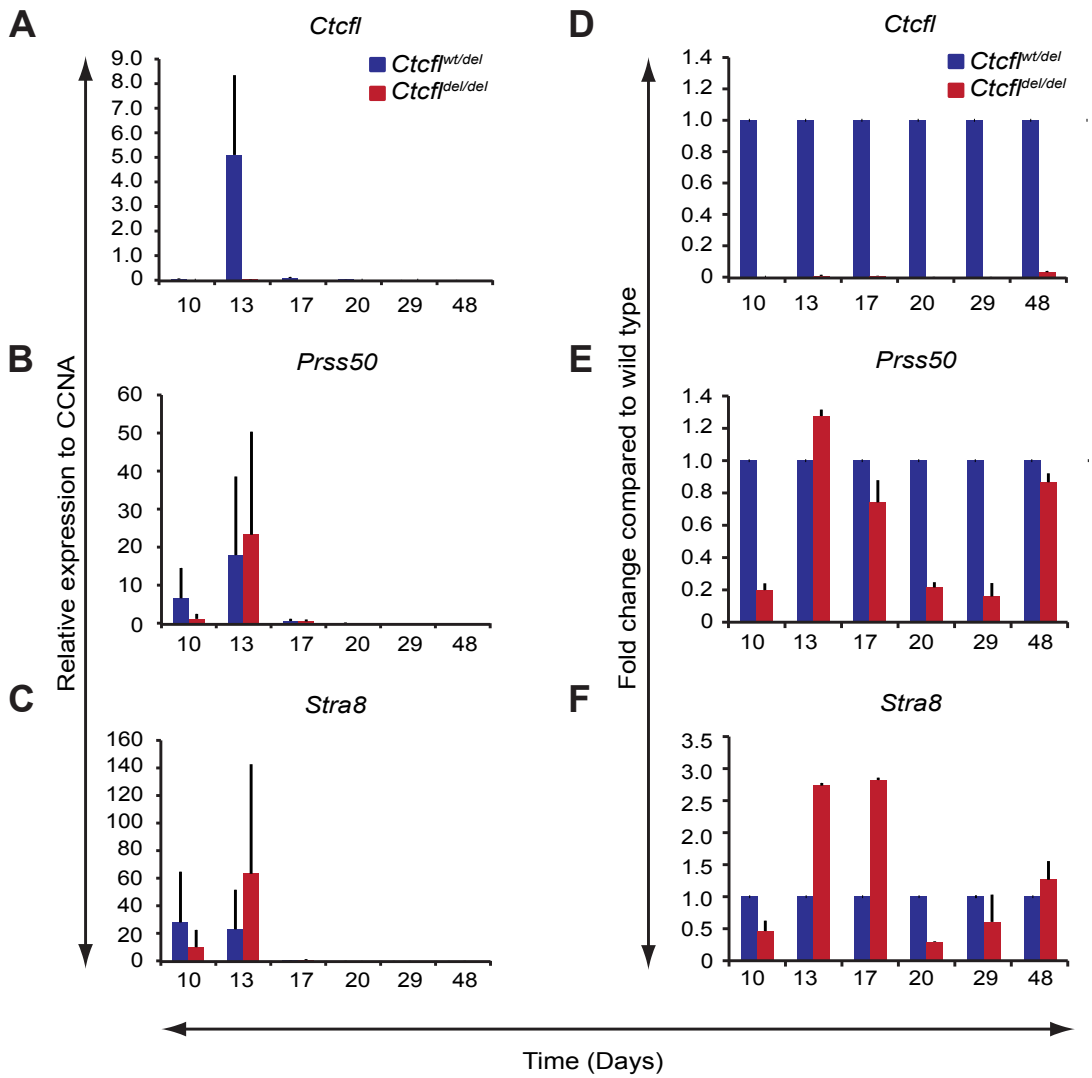


Figure 4. Gene expression analysis in *Ctcfl*$^{wt/del}$ and *Ctcfl*$^{del/del}$ testis
RNA was isolated from testis of *Ctcfl*$^{wt/del}$ and *Ctcfl*$^{del/del}$ mice of the indicated postnatal ages. Expression levels of *Ctcfl* **(A, D)**, *Prss50* **(B, E)** and *Stra8* **(C, F)** were measured by qPCR. Expression levels were normalized to the housekeeping gene *Ccna*. In **(D-F)** expression is shown relative to wild type level. Data are shown ± SEM (N=2).

To gain more insight in the developmental expression pattern of *Stra8* and *Prss50* genes and their regulation by CTCFL, we examined *Ctcfl*, *Stra8* and *Prss50* expression in whole testis of *Ctcfl*$^{wt/del}$ and *Ctcfl*$^{del/del}$ mice at various time points after birth. As described before *(Suzuki et al., 2010)*, there is a massive *Ctcfl* expression at postnatal day 13; levels drop down before and after this point, and *Ctcfl* was undetectable in *Ctcfl*$^{del/del}$ mice **(Figure 4A, D)**. *Prss50* and *Stra8* levels peaked with *Ctcfl* **(Figure 4B, C)**. *Prss50* expression was generally reduced in *Ctcfl*$^{del/del}$ testis with the exception of day 13 **(Figure 4B, E)**. *Stra8* expression was also reduced in *Ctcfl*$^{del/del}$ mice, except for day 13, 17 and 48 **(Figure 4C, F)**. Developmental defects in *Ctcfl* knock out mice may lead to a delay in differentiation and could give rise to increased populations of *Stra8*-positive cells. This would explain the higher levels of *Stra8* and *Prss50*

mRNAs (and of GFP+ cells in FACS experiments) in *Ctcfl* knockout testes at certain periods of development. However, CTCFL is clearly not the only factor regulating expression of *Stra8* and *Prss50*. Our results support previous findings that revealed early CTCFL expression during spermatogenesis and down-regulation of *Prss50* and *Stra8* at day 23 in *Ctcfl*$^{del/del}$ mice *(Sleutels et al., 2012)*. Overall, these results indicate that the regulation of *Stra8* and *Prss50* by CTCFL during development is complex.

Figure 5

A

CTCF constructs	Ctcfl Expression (FPKM)	Stra8 Expression (FPKM)	Prss50 Expression (FPKM)	Gal3st1 Expression (FPKM)
WT	0.07	0.85	2.04	1.36
GFPCTCF	0.00	0.74	1.26	0.37
GFPCTCFΔ1	0.21	6.11	2.35	0.76
GFPCTCFΔ8	5.59	38.05	21.83	2.40
GFPCTCFΔ9	0.09	7.28	2.68	0.88
GFPCTCFΔ10	0.04	5.44	3.35	1.18
GFPCTCFΔ11	0.07	1.44	2.86	1.52

B

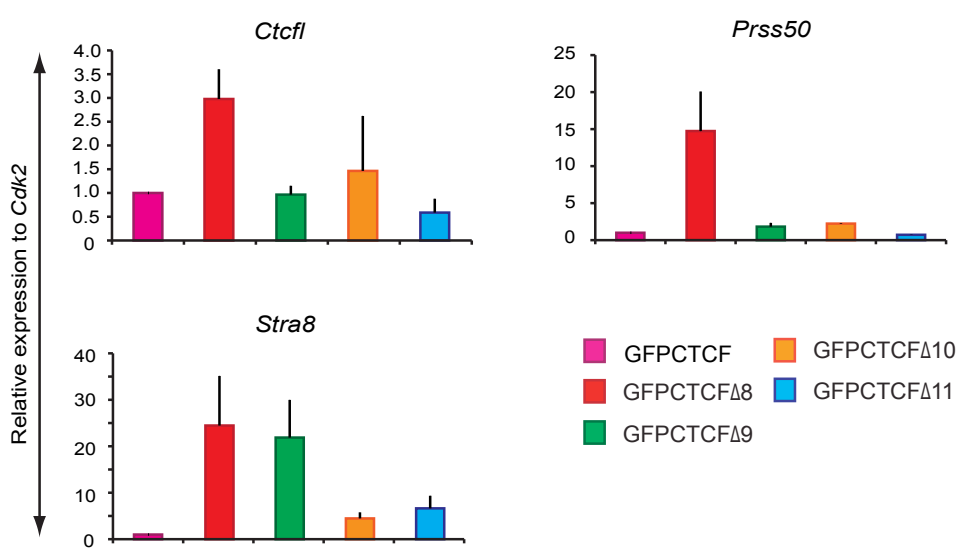


Figure 5. Gene expression analysis of testis specific factors in ES cells
(A) mRNA levels of testis specific factors *Ctcfl*, *Stra8*, *Prss50* and *Gal3st1* in ES cells expressing wild type CTCF (WT), GFP-CTCF, or the indicated GFP-tagged CTCF zinc finger mutants. Values are depicted in FPKM (Fragments Per Kilobase of transcripts per Million mapped reads).
(B) *Ctcfl*, *Prss50*, *Stra8* mRNA levels in the indicated ES cell lines as determined by qPCR. Expression levels are relative to the housekeeping gene *Cdk2* and were normalized to GFP-CTCF. All data are shown ± SEM (N=3).

Figure 6

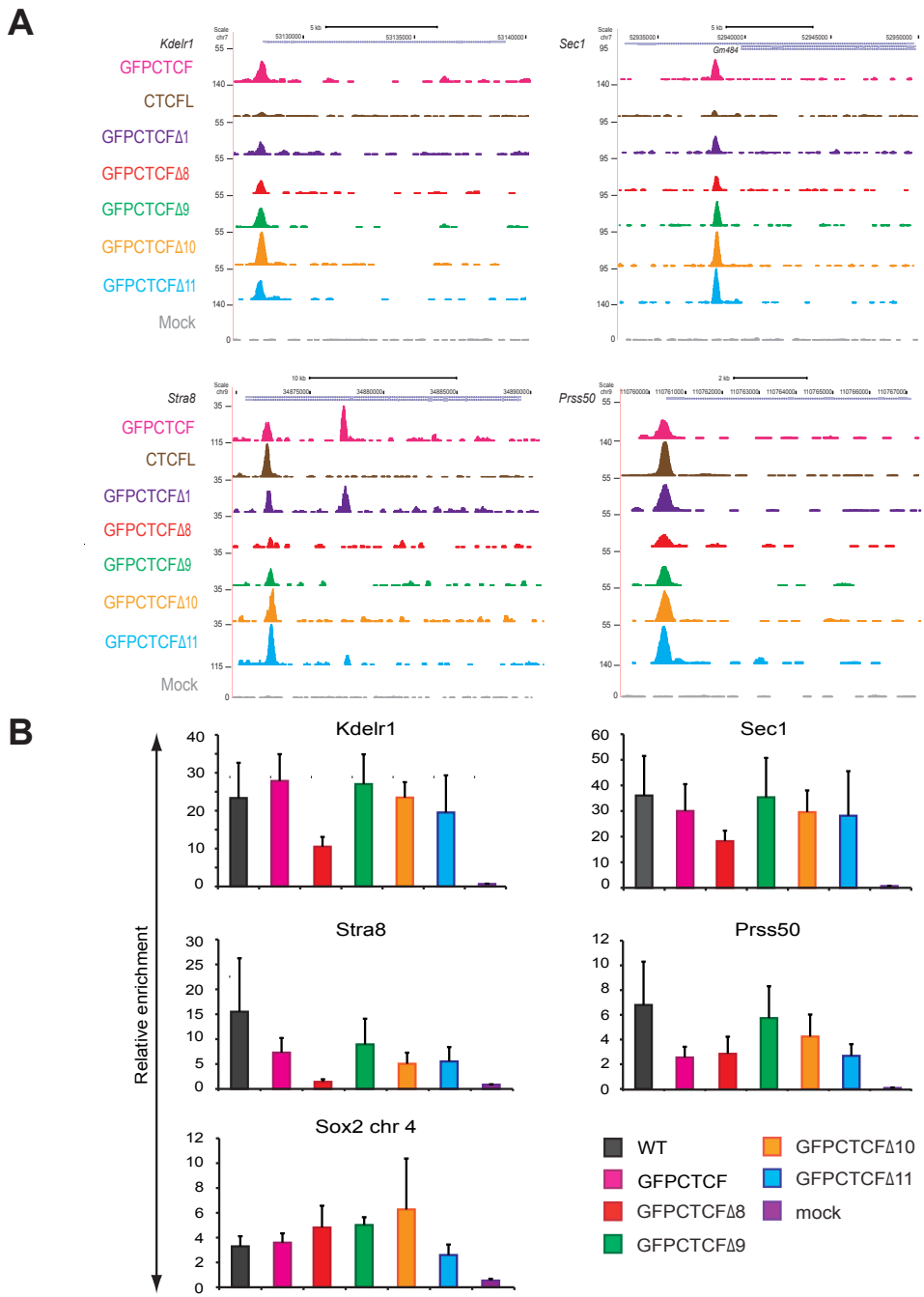


Figure 6. ChIP-Seq and ChIP-qPCR analysis in ES cells
(A) CTCF, CTCF zinc finger mutants and CTCFL ChIP-Seq profiles extracted from the UCSC genome browser for *Kdelr1*, *Sec1*, *Stra8* and *Prss50* genes.
(B) ChIP-qPCR validation experiment was performed with CTCF and Sox2 antibodies. Pre-immune serum was used as mock and amylase was used as non-CTCF binding site. CTCF enrichment was normalized to amylase. Data are shown ± SEM (N=3)

Expression of Ctcfl and Ctcfl responsive genes

In order to study how CTCF utilizes its zinc fingers to recognize its DNA binding site we used CTCF zinc finger mutants. We have generated ES cells that, instead of endogenous CTCF, express GFP-tagged wild type CTCF, or mutant CTCF proteins with deletions of ZFs 1, 8, 9, 10, or 11 (see chapter 3). RNA-Seq performed on these ES cells showed that the GFP-CTCF-Δ8 mutant cells expressed *Ctcfl*, *Prss50* and *Stra8* **(Figure 5A)**. Analysis of GFP-CTCF-Δ9 cells showed increased *Stra8* expression. These unexpected observations were confirmed by qPCR **(Figure 5B)**.

We hypothesized that the increased expression of these genes could be caused by reduction of CTCF binding due to mutations in the ZF domain, and in the case of the ES cells expressing GFP-CTCF-Δ8, to activation by CTCFL. We therefore examined CTCF and CTCFL genome-wide ChIP-Seq profiles in ES cells. Interestingly, focusing on the *Prss50*, and *Stra8* genes shows that GFP-CTCF-Δ8 binding was reduced at the *Stra8* promoter, as compared to wild type and GFP-CTCF binding **(Figure 6A)**. Binding to the *Prss50* promoter was less affected. Interestingly, CTCFL was also bound to the *Stra8* and *Prss50* promoters. As controls we examined binding at the *Kdelr1* and *Sec1* promoters, which were bound by CTCF and all the ZF mutants but not by CTCFL. These observations were confirmed by ChIP-qPCR, indicating that GFP-CTCF-Δ8 is less stably bound than GFP-CTCF **(Figure 6B)**. In addition to CTCF, endogenous SOX2 was also ChIPped, which is a CTCF-independent factor that should not change in binding profile upon expression of a ZF mutant. Indeed, SOX2 binding to a cognate site on chromosome 4 was unchanged irrespective of the ES cell line tested **(Figure 6B)**. Taken together these data suggest that CTCF acts as repressor of *Stra8* and CTCFL as an activator. Relative loss of binding of the GFP-CTCF-Δ8 and GFP-CTCF-Δ9 mutants to the Stra8 promoter, in combination with a gain of binding of CTCFL in ES cells expressing GFP-CTCF-Δ8, explains the selective up-regulation of *Stra8* in ES cells expressing the two mutant ZF proteins. Similarly, the induction of *Prss50* in GFP-CTCF-Δ8-expressing ES cells could be due to the expression of CTCFL in this cell line and the binding of the protein to the *Prss50* promoter.

Figure 7

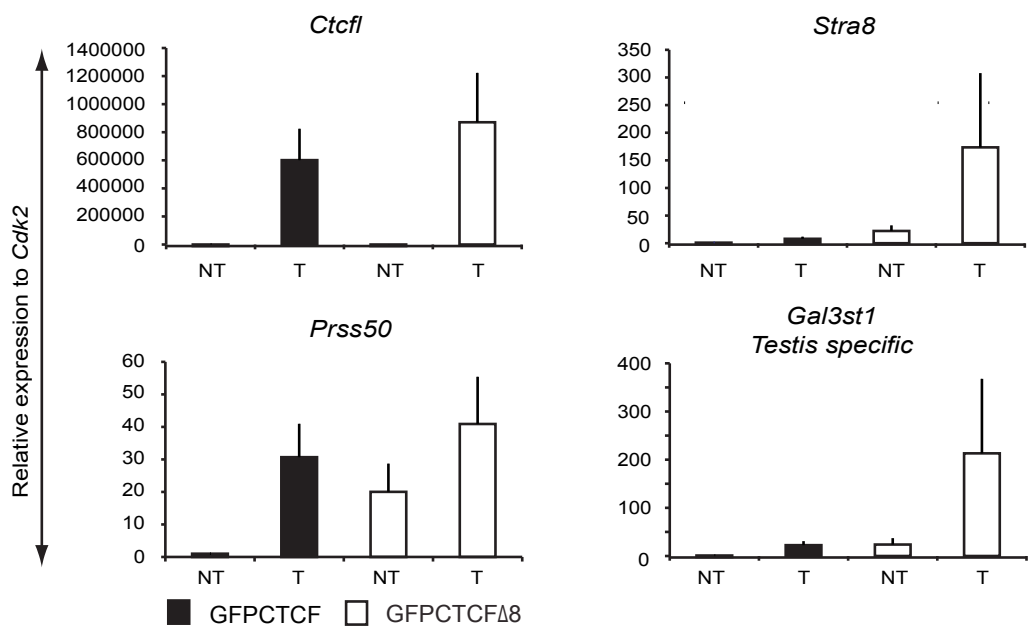


Figure 7. Gene expression analysis in ES cells overexpressing CTCFL-V5-GFP
Ctcfl, *Stra8*, *Prss50* and *Gal3st1* expression levels were analyzed in ES cells expressing GFP-CTCF or GFP-CTCF-Δ8. Cells were either not transfected (NT) or transfected (T) with CTCFL-RFP. Expression levels are relative to the housekeeping gene *Cdk2* and are normalized to not-transfected GFP-CTCF. Data are shown ± SEM (N=3)

If CTCFL competes with CTCF on shared binding sites CTCFL should be able to compete with a CTCF mutant more efficiently than with wild type CTCF. To test this CTCFL-RFP was transfected in GFP-CTCF- and GFP-CTCF-Δ8-expressing ES cells and sorted for GFP and RFP double positive cells. We tested the expression level of *Stra8*, *Prss50* and *Gal3st1*, using qPCR **(Figure 7)**. As control we tested *Ctcfl* expression **(Figure 7A)**. We found that *Stra8* expression was higher in GFP-CTCF-Δ8-expressing transfected ES cells than in GFP-CTCF-expressing transfected cells **(Figure 7B)**. Similar results were obtained with *Gal3st1* and *Prss50* **(Figure 7C, D)**, although for *Prss50* only a ~1.5 fold increase was observed in GFP-CTCF-Δ8-expressing transfected ES cells. However, GFP-CTCF-Δ8-expressing non-transfected ES cells already showed high levels of *Prss50* **(Figure 7C)** and perhaps this gene cannot be further induced. Our data confirm that CTCFL competes more efficiently with less stably bound CTCF.

Discussion

Spermatogenesis is a complex differentiation process that starts in the basal compartment of the seminiferous epithelium. In newborn mice the seminiferous epithelium contains gonocytes and Sertoli cells (for review see *(Zhou and Griswold, 2008)*). At day 6 of postnatal development germ cells are attached to the basement membrane and have differentiated to primitive type A spermatogonia, which form 16% of the total cell population in the seminiferous tubules *(Bellve et al., 1977; Nebel et al., 1961).* Type A and B spermatogonia are present at day 8 after birth and primary spermatocytes at the preleptotene and leptotene stages of meiotic prophase are present at day 10 of development *(Bellve et al., 1977; Nebel et al., 1961)*. *Ctcfl* and *Stra8* are expressed in type B spermatogonia and preleptotene spermatocytes and *Prss50* is both expressed in CTCFL-positive cells and CTCFL-negative pachytene spermatocytes *(Anderson et al., 2008; Baltus et al., 2006; Mark et al., 2008; Sleutels et al., 2012)*.

Here we demonstrate that *Ctcfl*, *Prss50* and *Stra8* expression peak at 13 days of postnatal development. Initiation of *Stra8* and *Prss50* expression coincides with the presence of type B spermatogonia and the onset of preleptotene spermatocytes. *Stra8* expression dropped at day 17 coinciding with the onset of pachytene spermatocytes and cells approaching the first meiotic metaphase. Unexpectedly, *Prss50* expression showed similar dynamics and declined at day 17 with the onset of pachytene spermatocytes. This suggests that the PRSS50 protein is present for a longer time period after transcription initiation during spermatogenesis. *Ctcfl* expression only started at day 13, which suggests that CTCFL is not involved in regulating the early expression of *Stra8* and *Prss50* but starts to regulate expression of these genes in a later phase. CTCFL clearly is not the only protein that regulates *Stra8*, since the *Stra8-GFP* transgene is expressed in a *Ctcfl*-negative background. Other factors, besides CTCFL, that could regulate *Stra8* and *Prss50* are, for example, DMRT1, WIN 18,466, P53 and SP1 *(Krentz et al., 2011; Matson et al., 2010; Wang et al., 2010; Xu et al., 2007)*. The perfect co-localization of CTCFL and STRA8 in the adult testis that we observed in our earlier work *(Sleutels et al., 2012)*, indicates that the timing of *Stra8* and *Ctcfl* mRNA expression is different from that of protein expression.

Ctcfl$^{del/del}$ mice contained more *Stra8* and *Prss50* mRNA at day 13 and 17 and more GFP^{+} (i.e. STRA8^{+}) cells compared to wild type. This could be due to a developmental delay in the *Ctcfl*$^{del/del}$ mice caused by an accumulation of early stage spermatogenic cell types. Both, an arrest of CTCFL-negative cells during early stages of spermatogenesis or enhanced apoptosis

at later stages would result in an overrepresentation of GFP$^+$ cells in the testis of *Ctcfl* knock out mice. To examine this further we have to discriminate the cell populations from each other during development by using cell surface markers at various time points of postnatal development.

Our strategy to cross *Ctcfl*$^{wt/del}$ mice with *Stra8-Gfp* transgenic mice and utilize FACS-sorting on GFP in offspring from these mice was very efficient to obtain pure STRA8-positive cell populations with or without CTCFL. The RNA-Seq-based databases that we generated may serve as a valuable repository for other researchers in the field of spermatogenesis. RNA sequencing revealed that virtually all mis-regulated genes in the STRA8-positive fraction of *Ctcfl* knock out testis were down-regulated. This implies that CTCFL acts as transcriptional activator, which supports our previous findings *(Sleutels et al., 2012)*. RNA-Seq showed reduced expression levels of many genes, including *Stra8*, *Prss50* and *Gal3st1,* in *Ctcfl*$^{del/del}$ mice. Of these, *Gal3st1* and *Gm6525* were reduced more than 10-fold, suggesting that these two genes are completely dependent on CTCFL. The fact that *Gal3st1* knock out mice show a much more severe phenotype in the testis than the *Ctcfl* knock outs *(Honke et al., 2002; Sleutels et al., 2012; Suzuki et al., 2010)* suggests that *Gal3st1* is more widely expressed in the testis than *Ctcfl*, and is regulated by other factors in testicular cell types that do not express CTCFL.

Remarkably, a subset of deregulated genes was related to carcinogenesis and was down-regulated in the GFP$^+$ testis fraction of *Ctcfl*$^{del/del}$ mice. CTCFL belongs to the group of cancer germline (CG) or cancer testis antigen (CTA) genes. CG genes are germ cell specific, but are reactivated and are aberrantly expressed in cancer *(Cheng et al., 2011)*. Aberrant *Ctcfl* expression has indeed been observed in various cancer types *(Hong et al., 2005; Looijenga et al., 2006; Renaud et al., 2005; Vatolin et al., 2005)*. CTCFL was proposed to induce expression of multiple genes within the MAGE-A family *(Bhan et al., 2011; Smith et al., 2009)*. We indeed found a number of *Mage-a* genes down-regulated in *Ctcfl*$^{del/del}$ mice. In addition, multiple cancer-related genes, e.g. *Brca1* and *Atm*, were also down-regulated in the *Ctcfl*$^{del/del}$ mice. It has been proposed that aberrant *Ctcfl* and *Mage-a* expression could promote cell growth *(Smith et al., 2009)*. Altogether these data suggest that CTCFL might promote carcinogenesis, since it positively regulates various cancer-related genes. So far only one study showed that patients with CTCFL expression in esophageal squamous cell cancer have a poor 5-years survival rate *(Okabayashi et al., 2012)*. Additional research is required to understand the putative role of CTCFL in cancer.

The analysis of the RNA-Seq data presented here revealed interesting candidate genes, which we can examine to understand how CTCFL regulates their transcription. However, the RNA-Seq data first requires further analysis before we are able to make solid conclusions. We need to pool the two datasets to examine significantly deregulated genes. In this way we are further able to perform additional downstream analysis, such as GO-analysis to reveal affected pathways in the *Ctcfl* knock out mice. Next, to CTCFL dependent transcriptional regulation it is interesting to examine were CTCFL and CTCF binds in the testis. ChIP-Seq on both proteins and histone modifications, e.g. H3K4me3, is suitable to address this question.

An unexpected observation was the expression of *Ctcfl* and the CTCFL-responsive genes *Stra8* and *Prss50* in GFP-CTCF-Δ8-expressing ES cells. *Stra8* was also expressed in GFP-CTCF-Δ9-expressing ES cells. In both cell lines we observed a relative loss of binding of mutant proteins to the *Stra8* promoter. This possible loss of function of mutant CTCF suggests that CTCF normally acts as repressor of this gene. Expression levels of CTCFL-responsive genes could be better stimulated by CTCFL when mutant CTCF was present instead of wild type protein. This suggests that on certain promoters there is a balance between the binding of the transcriptional repressor CTCF and the activator CTCFL.

Materials and Methods

Mouse models

Ctcfl$^{wt/del}$ *(Sleutels et al., 2012)* and S*tra8-Gfp* transgenic *(Nayernia et al., 2004)* mice were crossed to generate *Ctcfl*$^{wt/wt}$*-Stra8-Gfp*, *Ctcfl*$^{wt/del}$*-Stra8-Gfp* and *Ctcfl*$^{del/del}$*-Stra8-Gfp* mice. In total 7 *Ctcfl*$^{wt/wt}$*-Stra8-Gfp* (2 mice: 33 days, 3 mice: 121 days and 2 mice: 133 days), and 10 *Ctcfl*$^{del/del}$*-Stra8-Gfp* mice (2 mice: 33 days, 4 mice 77 days and 4 mice: 121 days) were used in 4 independent FACS sorting experiments. Two independent RNA-seq experiments were performed using in each experiment 1 *Ctcfl*$^{wt/wt}$*-Stra8-Gfp* and 1 *Ctcfl*$^{del/del}$*-Stra8-Gfp* mice. All mice used for RNA-seq were 33 days old. Whole testis extracts of *Ctcfl*$^{wt/del}$ and *Ctcfl*$^{del/del}$ mice at various ages (day 10, 13, 17, 20, 29 and 48) were used to examine gene expression profiles by qPCR.

Mice were maintained on a C57/Bl6 background at the Erasmus MC animal facility under specific pathogen–free conditions. Animal experiments were reviewed and approved by the Erasmus University committee of animal experiments.

Cell Culture and transfection

ES cells expressing CTCF zinc finger mutant (see chapter 3) were grown on plastic dishes coated with 0.2% gelatin (Merck) in the presence of ES cell medium containing: Dulbecco's Modified Eagle Medium (DMEM) (Lonza), 15% heat-inactivated fetal calf serum (FCS), Non Essential Amino Acids (Lonza), 100 U ml^{-1} penicillin and 100 mg ml^{-1} streptamycin, 0.1 mM β-mercaptoethanol (Sigma) and 1000 U/ml leukaemia inhibitory factor (LIF). ES cells were cultured at 37°C with 5% CO_2. Cells were passaged by trypsinization in 1xTE (trypsin/EDTA) for 5 minutes at 37°C and subsequently diluted and resuspended in ES medium. Transient transfection of CTCFL-RFP in ES cells was performed with Lipofectamine 2000 (invitrogen). Cells were harvested after 24 hours of transfection.

RNA-isolation whole testis and ES cells for qPCR analysis

Total RNA of whole testis of *Ctcfl*$^{wt/wt}$ and *Ctcfl*$^{del/del}$ mice (age: 10, 13, 17, 20, 29 and 48 days), CTCF zinc finger mutant ES cells and CTCF zinc finger mutant ES cells expressing CTCFL-RFP was isolated using Trizol-BCP (1-bromo-3-chloro propane). RNA in aqueous phase was precipitated with isopropanol and washed with 70% ethanol. RNA was resuspended in water and reverse transcribed (RT) with a combination of random and oligo-dT primers by Superscript reverse transcriptase (Invitrogen). qPCR was performed with a Sybrgreen platform on a Bio-Rad CFX Cycler. The relative quantification method was used to calculate ΔCt and ΔΔCt values *(Livak and Schmittgen, 2001)*. Housekeeping genes *Ccna* and *Cdk2* were used to normalize the qPCR data generated from the testes and ES cells, respectively. For primer sequences see **supplemental table 2**.

Generation of germ cell enriched populations

4 testes were incubated in 20 ml DPBS (Lonza) (volume was adjusted according to the amount of testes, 2 testes per 5 ml) with 1.1 mM Ca^{2+} and 0.52 mM Mg^{2+}. 20 mg / 200 μl Collagenase (Boehringer), 10 mg / 200 μl Hyaluronidase and 20 mg / 200 μl Trypsin were added, followed by 20 minutes incubation at 32-34°C at 90 rpm. Samples were centrifuged for 3 minutes at 1800 rpm at room temperature. Extracts were resuspended in 20 ml DPBS without Ca^{2+} and Mg^{2+} and incubated for 10 minutes at 32-34°C at 90 rpm. Clumps were removed and samples went through a 0.45 μm cell strainer.

Flow cytometric analysis

Testis extracts were sorted on GFP expression. In total 7 *Ctcfl*$^{wt/wt}$*-Stra8-Gfp* (2 mice: 33 days, 3 mice: 121 days and 2 mice: 133 days), and 10 *Ctcfl*$^{del/del}$*-Stra8-Gfp* mice (2 mice: 33 days, 4 mice 77 days and 4 mice: 121 days) were used in 4 independent FACS sorting experiments. GFP cell lines with CTCFL-RFP expression were sorted on GFP and RFP expression by BD FACS Aria Cell sorter. Data was analyzed with FlowJo software.

RNA-sequencing of testis and ES cells

Two independent RNA-seq experiments were performed using in each experiment 1 *Ctcfl*$^{wt/wt}$*-Stra8-Gfp* and 1 *Ctcfl*$^{del/del}$*-Stra8-Gfp* mice. Testes were enriched for germ cells and subsequently FACS sorted for GFP. Two independent RNA-Seq experiments were performed on ES cells expressing a CTCF mutant. Total RNA isolation was performed with Trizol-chloroform extraction. After Trizol addition samples were incubated for 5 minutes at 30°C. Chloroform was added and aquaeous phase was transferred after centrifugation. 100% ethanol was added and an RNeasy Mini Kit was subsequently used (Cat. no. 74104, Qiagen).

Purity and quality of the isolated RNA was assessed by RNA 6000 Nano assay on a 2100 Bioanalyzer (Agilent Technologies). 1 μg total RNA of each sample was used as starting material for Illumina Truseq sequencing. Samples were sequenced on a HiSeq 2000 generating 36 bp single reads and a 7 bp index read. Samples were de-multiplexed and aligned to the mouse reference genome (build mm9) using Tophat alignment software.

Aligned reads of testis RNA-Seq were visualized using the IGV viewer. Aligned reads of testis and CTCF zinc finger mutant ES cells were used to identify differentially expressed genes and to calculate FPKM values by Cufflinks *(Trapnell et al., 2013)*. Expression levels were normalized according to gene length and sequencing dept. *Ctcfl*$^{wt/wt}$ GFP^{+} samples were compared to the corresponding *Ctcf*$^{del/del}$ GFP^{+} samples. The false discovery rate was set to 0.1 and an internal t-test (p-value <0.05) was used. Comparisons with p>0.05 were excluded from further analysis. Manual analysis was performed to identify overlapping deregulated genes.

Chromatin immunoprecipitation ES cells

ChIP was performed as described *(van de Nobelen et al., 2010)*. Briefly, 40-80*10^{6} cells were harvested and cross-linked with 1% formaldehyde (Sigma) for 10 minutes at room temperature and quenched with Glycine (Sigma). Cell lysates were prepared with cell lysis buffer (10 mM Tris pH 8.0, 10 mM NaCl, 0.2% NP-40 (Sigma), Protease Inhibitor) followed by nuclei lysis buffer (50 mM Tris pH 8.0, 10 mM EDTA, 1% SDS, Protease Inhibitor). Sonication was performed with the bioruptor (Diagnode) to yield fragments up to 800 bp. Immunoprecipitation with CTCF antibody (N2.2, home made), SOX2 (goat, Santa cruz) or pre-immune serum rabbit (home made) was performed. ChIP-Seq was validated by three independent ChIP-qPCR experiments. Ct values from qPCR were normalized to input measurements, and enrichment was calculated relative to CTCF negative binding site amylase. For primer sequences see **supplemental table 3**.

ChIP-Sequencing

A ChIP DNA library was prepared according to the Illumina protocol (www.illumina.com). Briefly, 10 ng of end-repaired ChIPped DNA was ligated to adapters, size selected on gel (200±25 bp range), and PCR amplified using Phusion polymerase as follow: 30 sec at 98ºC, 18 cycles of (10 sec at 98ºC, 30 sec at 65ºC, 30 sec at 72ºC), 5 min at 72ºC final extension. Cluster generation was performed using the Illumina Cluster Reagents preparation. The library was sequenced on the Illumina HiSeq2000 systems to generate 36 bp reads and a 7 bp index read. Images were recorded and analyzed by the Illumina Genome Analyzer Pipeline (GAP) and processed using the IPAR (Integrated Primary Analysis Reporting Software). Samples were de-multiplexed and mapped against mouse build mm9 reference genome. ChIP-Seq reads

were aligned to mm9 reference genome using Bowtie2 *(Langmead and Salzberg, 2012)*. Datasets were uploaded into the UCSC genome browser.

References

Anderson, E.L., Baltus, A.E., Roepers-Gajadien, H.L., Hassold, T.J., de Rooij, D.G., van Pelt, A.M., and Page, D.C. (2008). Stra8 and its inducer, retinoic acid, regulate meiotic initiation in both spermatogenesis and oogenesis in mice. Proc Natl Acad Sci U S A *105*, 14976-14980.

Baltus, A.E., Menke, D.B., Hu, Y.C., Goodheart, M.L., Carpenter, A.E., de Rooij, D.G., and Page, D.C. (2006). In germ cells of mouse embryonic ovaries, the decision to enter meiosis precedes premeiotic DNA replication. Nat Genet *38*, 1430-1434.

Bellve, A.R., Cavicchia, J.C., Millette, C.F., O'Brien, D.A., Bhatnagar, Y.M., and Dym, M. (1977). Spermatogenic cells of the prepuberal mouse. Isolation and morphological characterization. J Cell Biol *74*, 68-85.

Bhan, S., Negi, S.S., Shao, C., Glazer, C.A., Chuang, A., Gaykalova, D.A., Sun, W., Sidransky, D., Ha, P.K., and Califano, J.A. (2011). BORIS binding to the promoters of cancer testis antigens, MAGEA2, MAGEA3, and MAGEA4, is associated with their transcriptional activation in lung cancer. Clin Cancer Res *17*, 4267-4276.

Brannan, C.I., Bedell, M.A., Resnick, J.L., Eppig, J.J., Handel, M.A., Williams, D.E., Lyman, S.D., Donovan, P.J., Jenkins, N.A., and Copeland, N.G. (1992). Developmental abnormalities in Steel17H mice result from a splicing defect in the steel factor cytoplasmic tail. Genes Dev *6*, 1832-1842.

Cheng, Y.H., Wong, E.W., and Cheng, C.Y. (2011). Cancer/testis (CT) antigens, carcinogenesis and spermatogenesis. Spermatogenesis *1*, 209-220.

Giuili, G., Tomljenovic, A., Labrecque, N., Oulad-Abdelghani, M., Rassoulzadegan, M., and Cuzin, F. (2002). Murine spermatogonial stem cells: targeted transgene expression and purification in an active state. EMBO Rep *3*, 753-759.

Hong, J.A., Kang, Y., Abdullaev, Z., Flanagan, P.T., Pack, S.D., Fischette, M.R., Adnani, M.T., Loukinov, D.I., Vatolin, S., Risinger, J.I.*, et al.* (2005). Reciprocal binding of CTCF and BORIS to the NY-ESO-1 promoter coincides with derepression of this cancer-testis gene in lung cancer cells. Cancer Res *65*, 7763-7774.

Honke, K., Hirahara, Y., Dupree, J., Suzuki, K., Popko, B., Fukushima, K., Fukushima, J., Nagasawa, T., Yoshida, N., Wada, Y.*, et al.* (2002). Paranodal junction formation and spermatogenesis require sulfoglycolipids. Proc Natl Acad Sci U S A *99*, 4227-4232.

Hore, T.A., Deakin, J.E., and Marshall Graves, J.A. (2008). The evolution of epigenetic regulators CTCF and BORIS/CTCFL in amniotes. PLoS Genet *4*, e1000169.

Kosaka-Suzuki, N., Suzuki, T., Pugacheva, E.M., Vostrov, A.A., Morse, H.C., 3rd, Loukinov, D., and Lobanenkov, V. (2011). Transcription factor BORIS (Brother of the Regulator of Imprinted Sites) directly induces expression of a cancer-testis antigen, TSP50, through regulated binding of BORIS to the promoter. J Biol Chem *286*, 27378-27388.

Krentz, A.D., Murphy, M.W., Sarver, A.L., Griswold, M.D., Bardwell, V.J., and Zarkower, D. (2011). DMRT1 promotes oogenesis by transcriptional activation of Stra8 in the mammalian fetal ovary. Dev Biol *356*, 63-70.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402-408.

Looijenga, L.H., Hersmus, R., Gillis, A.J., Pfundt, R., Stoop, H.J., van Gurp, R.J., Veltman, J., Beverloo, H.B., van Drunen, E., van Kessel, A.G.*, et al.* (2006). Genomic and expression profiling of human spermatocytic seminomas: primary spermatocyte as tumorigenic precursor and DMRT1 as candidate chromosome 9 gene. Cancer Res *66*, 290-302.

Loukinov, D.I., Pugacheva, E., Vatolin, S., Pack, S.D., Moon, H., Chernukhin, I., Mannan, P., Larsson, E., Kanduri, C., Vostrov, A.A.*, et al.* (2002). BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. Proc Natl Acad Sci U S A *99*, 6806-6811.

Mark, M., Jacobs, H., Oulad-Abdelghani, M., Dennefeld, C., Feret, B., Vernet, N., Codreanu, C.A., Chambon, P., and Ghyselinck, N.B. (2008). STRA8-deficient spermatocytes initiate, but fail to complete, meiosis and undergo premature chromosome condensation. J Cell Sci *121*, 3233-3242.

Matson, C.K., Murphy, M.W., Griswold, M.D., Yoshida, S., Bardwell, V.J., and Zarkower, D. (2010). The mammalian doublesex homolog DMRT1 is a transcriptional gatekeeper that controls the mitosis versus meiosis decision in male germ cells. Dev Cell *19*, 612-624.

Nayernia, K., Li, M., Jaroszynski, L., Khusainov, R., Wulf, G., Schwandt, I., Korabiowska, M., Michelmann, H.W., Meinhardt, A., and Engel, W. (2004). Stem cell based therapeutical approach of male infertility by teratocarcinoma derived germ cells. Hum Mol Genet *13*, 1451-1460.

Nebel, B.R., Amarose, A.P., and Hacket, E.M. (1961). Calendar of gametogenic development in the prepuberal male mouse. Science *134*, 832-833.

Okabayashi, K., Fujita, T., Miyazaki, J., Okada, T., Iwata, T., Hirao, N., Noji, S., Tsukamoto, N., Goshima, N., Hasegawa, H.*, et al.* (2012). Cancer-testis antigen BORIS is a novel prognostic marker for patients with esophageal cancer. Cancer Sci *103*, 1617-1624.

Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. Cell *137*, 1194-1211.

Renaud, S., Loukinov, D., Bosman, F.T., Lobanenkov, V., and Benhattar, J. (2005). CTCF binds the proximal exonic region of hTERT and inhibits its transcription. Nucleic Acids Res *33*, 6850-6860.

Schrans-Stassen, B.H., van de Kant, H.J., de Rooij, D.G., and van Pelt, A.M. (1999). Differential expression of c-kit in mouse undifferentiated and differentiating type A spermatogonia. Endocrinology *140*, 5894-5900.

Sleutels, F., Soochit, W., Bartkuhn, M., Heath, H., Dienstbach, S., Bergmaier, P., Franke, V., Rosa-Garrido, M., van de Nobelen, S., Caesar, L.*, et al.* (2012). The male germ cell gene regulator CTCFL is functionally different from CTCF and binds CTCF-like consensus sites in a nucleosome composition-dependent manner. Epigenetics Chromatin *5*, 8.

Smith, I.M., Glazer, C.A., Mithani, S.K., Ochs, M.F., Sun, W., Bhan, S., Vostrov, A., Abdullaev, Z., Lobanenkov, V., Gray, A.*, et al.* (2009). Coordinated activation of candidate proto-oncogenes and cancer testes antigens via promoter demethylation in head and neck cancer and lung cancer. PLoS One *4*, e4961.

Suzuki, T., Kosaka-Suzuki, N., Pack, S., Shin, D.M., Yoon, J., Abdullaev, Z., Pugacheva, E., Morse, H.C., 3rd, Loukinov, D., and Lobanenkov, V. (2010). Expression of a testis-specific form of Gal3st1 (CST), a gene essential for spermatogenesis, is regulated by the CTCF paralogous gene BORIS. Mol Cell Biol *30*, 2473-2484.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol *31*, 46-53.

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

Vatolin, S., Abdullaev, Z., Pack, S.D., Flanagan, P.T., Custer, M., Loukinov, D.I., Pugacheva, E., Hong, J.A., Morse, H., 3rd, Schrump, D.S.*, et al.* (2005). Conditional expression of the CTCF-paralogous transcriptional factor BORIS in normal cells results in demethylation and derepression of MAGE-A1 and reactivation of other cancer-testis genes. Cancer Res *65*, 7751-7762.

Wang, M., Bao, Y.L., Wu, Y., Yu, C.L., Meng, X.Y., Huang, Y.X., Sun, Y., Zheng, L.H., and Li, Y.X. (2010). Basic FGF downregulates TSP50 expression via the ERK/Sp1 pathway. J Cell Biochem *111*, 75-81.

Xu, H., Shan, J., Jurukovski, V., Yuan, L., Li, J., and Tian, K. (2007). TSP50 encodes a testis-specific protease and is negatively regulated by p53. Cancer Res *67*, 1239-1245.

Zhou, Q., and Griswold, M.D. (2008). Regulation of spermatogonia.

Supplemental Table S1

Supplemental Table S1. RNA-Seq expression profiles of *Ctcfl*$^{wt/wt}$*-Stra8-Gfp* and *Ctcfl*$^{del/del}$*-Stra8-Gfp* mice
Gene expression levels of common down-regulated genes by RNA-sequencing in *Ctcfl*$^{wt/wt}$*-Stra8-Gfp* and *Ctcfl*$^{del/del}$*-Stra8-Gfp* mice from two independent experiments. Values are depicted in FPKM (Fragments per Kilobase of transcripts per Million mapped reads). Fold change (log2) between wild type and knock out mice are depicted together with the p-value.

	Dataset 1				Dataset 2			
gene	wt/wt Expression (FPKM)	del/del Expression (FPKM)	fold change (log2)	p-value	wt/wt Expression (FPKM)	del/del Expression (FPKM)	fold change (log2)	p-value
Ctcfl	2.51	0.19	-3.70	2.33E-06	3.09	0.03	-6.52	1.57E-05
Gal3st1	4.73	0.26	-4.18	5.84E-05	4.05	0.38	-3.41	2.40E-04
Prss50	47.73	12.89	-1.89	6.24E-08	47.17	15.57	-1.60	7.03E-05
Stra8	33.19	12.60	-1.40	1.00E-04	42.10	23.00	-0.87	2.32E-02
Gm6525	7.88	0.53	-3.88	1.15E-02	15.46	0.93	-4.05	3.05E-03
Mid1	3.03	0.29	-3.40	5.44E-05	3.12	1.37	-1.19	3.43E-02
Tubb6	7.28	0.87	-3.06	1.24E-06	7.43	2.88	-1.37	1.51E-02
Tcf7l1	1.94	0.65	-1.59	2.06E-02	2.43	0.97	-1.32	3.89E-02
Magea10	1.54	0.53	-1.55	4.78E-02	2.56	0.79	-1.70	1.78E-02
Greb1l	1.31	0.45	-1.53	3.30E-03	1.24	0.37	-1.74	2.92E-03
Amot	0.85	0.31	-1.44	2.14E-02	1.22	0.48	-1.34	2.61E-02
1600021P15Rik	1.09	0.40	-1.43	4.48E-02	2.24	0.96	-1.23	4.71E-02
Casp7	2.68	1.00	-1.43	2.02E-02	5.04	2.27	-1.15	3.88E-02
Mycn	3.12	1.18	-1.40	1.43E-02	4.17	1.85	-1.17	3.83E-02
Cyp17a1	7.48	2.89	-1.37	6.55E-03	5.26	2.27	-1.21	3.92E-02
Art1	3.73	1.45	-1.36	3.85E-02	6.39	1.72	-1.89	3.08E-03
Snhg11	1.73	0.68	-1.35	1.03E-02	2.00	0.88	-1.19	3.05E-02
Glt25d1	11.71	4.61	-1.35	2.44E-04	15.37	8.51	-0.85	2.84E-02
Vgll3	1.32	0.52	-1.34	4.18E-02	1.20	0.32	-1.88	1.14E-02
Serpine2	3.50	1.43	-1.30	2.84E-02	8.60	3.91	-1.14	2.69E-02
Zmynd8	1.86	0.76	-1.29	1.47E-02	3.09	1.34	-1.21	1.86E-02
4933427D06Rik	6.96	2.87	-1.28	4.68E-03	10.45	5.49	-0.93	3.95E-02
Ctnnd2	1.47	0.61	-1.26	1.96E-02	1.91	0.90	-1.09	4.67E-02
Xlr4b	5.49	2.32	-1.24	4.19E-02	8.17	3.21	-1.35	2.39E-02
Pxdn	3.29	1.41	-1.22	3.20E-03	4.23	2.04	-1.05	1.78E-02
Fam199x	1.17	0.53	-1.14	2.72E-02	2.45	1.10	-1.16	1.62E-02
Col5a1	2.60	1.25	-1.06	8.72E-03	3.96	2.09	-0.92	2.82E-02
Dck	5.94	2.86	-1.06	1.56E-02	9.52	5.04	-0.92	3.75E-02
Sema6d	1.42	0.69	-1.04	4.59E-02	2.56	1.11	-1.21	3.21E-02
Atm	8.05	3.97	-1.02	3.39E-04	9.61	5.26	-0.87	8.20E-03

Supplemental Table S1-Part 2

Supplemental Table S1. RNA-Seq expression profiles of *Ctcfl*$^{wt/wt}$*-Stra8-Gfp* and *Ctcfl*$^{del/del}$*-Stra8-Gfp* mice
Gene expression levels of common down-regulated genes by RNA-sequencing in *Ctcfl*$^{wt/wt}$*-Stra8-Gfp* and *Ctcfl*$^{del/del}$*-Stra8-Gfp* mice from two independent experiments. Values are depicted in FPKM (Fragments per Kilobase of transcripts per Million mapped reads). Fold change (log2) between wild type and knock out mice are depicted together with the p-value.

	Dataset 1				Dataset 2			
gene	wt/wt Expression (FPKM)	del/del Expression (FPKM)	fold change (log2)	p-value	wt/wt Expression (FPKM)	del/del Expression (FPKM)	fold change (log2)	p-value
Mll1	5.19	2.56	-1.02	4.25E-04	5.84	3.10	-0.91	6.89E-03
Ppp1ca	47.61	23.59	-1.01	3.79E-03	75.24	43.41	-0.79	3.08E-02
Trim26	6.95	3.45	-1.01	1.30E-02	9.89	5.32	-0.89	3.81E-02
Usp22	8.93	4.43	-1.01	3.17E-03	13.15	7.94	-0.73	4.82E-02
Gm98	3.96	2.00	-0.98	1.40E-02	5.07	2.16	-1.23	6.39E-03
Nkain1	6.00	3.06	-0.97	3.43E-02	10.96	5.18	-1.08	1.76E-02
Tex15	27.22	13.94	-0.97	5.47E-04	28.95	16.31	-0.83	7.37E-03
Wdhd1	16.75	8.76	-0.94	1.79E-03	26.24	15.12	-0.80	1.65E-02
Mex3a	4.24	2.24	-0.92	1.74E-02	6.25	3.46	-0.85	3.84E-02
Kdm1b	3.18	1.69	-0.91	4.85E-02	5.83	2.80	-1.06	2.17E-02
Tex13	12.80	6.88	-0.89	2.05E-02	19.90	10.40	-0.94	2.30E-02
Tet3	3.54	1.91	-0.89	7.86E-03	5.06	2.93	-0.79	3.37E-02
Igf1r	3.09	1.66	-0.89	8.68E-03	4.24	2.32	-0.87	2.19E-02
Prep	11.20	6.11	-0.87	1.76E-02	18.66	10.29	-0.86	2.68E-02
Apaf1	2.41	1.32	-0.87	4.76E-02	3.43	1.71	-1.00	3.37E-02
Prr12	4.39	2.40	-0.87	1.59E-02	6.85	3.82	-0.84	2.94E-02
Zdhhc18	6.13	3.36	-0.87	4.54E-02	9.06	4.22	-1.10	1.75E-02
4930529F22Rik	6.65	3.65	-0.86	4.79E-02	7.10	2.71	-1.39	7.12E-03
Klf11	4.70	2.59	-0.86	3.93E-02	7.85	4.26	-0.88	4.00E-02
Dhcr24	8.79	4.87	-0.85	1.41E-02	13.43	7.65	-0.81	3.14E-02
Spnb3	4.52	2.51	-0.85	1.23E-02	6.31	3.48	-0.86	2.36E-02
Dmrt1	30.90	17.21	-0.84	5.16E-03	55.34	30.11	-0.88	8.01E-03
Gsr	15.21	8.47	-0.84	1.28E-02	26.89	15.70	-0.78	3.05E-02
Ptma	245.09	137.82	-0.83	2.83E-03	386.98	222.22	-0.80	9.43E-03
Abca2	10.59	5.98	-0.82	3.91E-03	13.86	6.87	-1.01	2.43E-03
Ubtf	10.96	6.20	-0.82	3.19E-02	16.22	9.10	-0.83	3.15E-02
Cdca7	10.47	5.94	-0.82	3.61E-02	19.20	9.71	-0.98	1.51E-02
Ctps	17.88	10.14	-0.82	1.14E-02	27.99	16.92	-0.73	4.01E-02
Taf1	7.92	4.50	-0.81	6.49E-03	11.39	6.35	-0.84	1.30E-02
Nmt2	45.01	25.62	-0.81	5.49E-03	67.62	41.34	-0.71	3.07E-02

Supplemental Table S1-Part 3

Supplemental Table S1. RNA-Seq expression profiles of *Ctcfl*$^{wt/wt}$*-Stra8-Gfp* and *Ctcfl*$^{del/del}$*-Stra8-Gfp* mice
Gene expression levels of common down-regulated genes by RNA-sequencing in *Ctcfl*$^{wt/wt}$*-Stra8-Gfp* and *Ctcfl*$^{del/del}$*-Stra8-Gfp* mice from two independent experiments. Values are depicted in FPKM (Fragments per Kilobase of transcripts per Million mapped reads). Fold change (log2) between wild type and knock out mice are depicted together with the p-value.

	Dataset 1				Dataset 2			
gene	wt/wt Expression (FPKM)	del/del Expression (FPKM)	fold change (log2)	p-value	wt/wt Expression (FPKM)	del/del Expression (FPKM)	fold change (log2)	p-value
Rlim	4.48	2.56	-0.81	2.01E-02	5.15	2.92	-0.82	4.28E-02
Msh2	9.17	5.27	-0.80	3.14E-02	16.68	9.68	-0.78	4.10E-02
Lrba	4.79	2.76	-0.80	1.48E-02	6.99	4.14	-0.76	3.63E-02
Marcksl1	53.06	30.71	-0.79	7.30E-03	83.30	49.13	-0.76	2.06E-02
Sgpl1	11.24	6.50	-0.79	1.44E-02	16.45	9.05	-0.86	1.72E-02
Esx1	28.15	16.35	-0.78	2.00E-02	37.55	20.03	-0.91	1.84E-02
Srp54a	11.39	6.63	-0.78	1.42E-02	11.23	6.28	-0.84	3.05E-02
Itpr3	3.47	2.02	-0.78	2.68E-02	6.36	3.10	-1.04	5.96E-03
Tubb5	14.68	8.56	-0.78	2.24E-02	26.03	15.20	-0.78	3.20E-02
Mageb4	13.79	8.08	-0.77	3.06E-02	24.38	13.39	-0.86	2.22E-02
Zfp36l2	7.59	4.52	-0.75	4.41E-02	10.92	5.77	-0.92	2.57E-02
Plod1	10.21	6.13	-0.74	3.61E-02	14.04	8.10	-0.79	4.32E-02
Hdac6	19.19	11.92	-0.69	3.77E-02	33.21	18.87	-0.82	3.32E-02
Huwe1	24.78	15.40	-0.69	2.15E-02	32.14	18.82	-0.77	1.35E-02
Mcm4	9.83	6.11	-0.69	4.50E-02	16.10	9.23	-0.80	3.11E-02
Hcfc1	10.37	6.47	-0.68	1.63E-02	14.46	8.46	-0.77	1.77E-02
Mcm2	16.78	10.46	-0.68	2.63E-02	30.05	16.36	-0.88	9.59E-03
Ccnd1	8.89	5.55	-0.68	4.93E-02	16.76	10.16	-0.72	4.64E-02
Fat1	5.73	3.58	-0.68	1.65E-02	7.86	4.36	-0.85	9.42E-03
Ncl	27.64	17.59	-0.65	2.03E-02	43.82	28.59	-0.62	4.63E-02
Bahcc1	4.32	2.75	-0.65	3.82E-02	8.63	4.38	-0.98	4.22E-03
Nxf2	20.22	12.90	-0.65	4.28E-02	29.62	15.97	-0.89	1.47E-02
Spnb2	8.59	5.48	-0.65	2.29E-02	11.45	6.05	-0.92	7.15E-03
Smc3	21.05	13.58	-0.63	2.30E-02	25.35	16.20	-0.65	4.56E-02
Brca1	6.87	4.44	-0.63	4.59E-02	10.52	6.14	-0.78	2.89E-02
Hspg2	4.92	3.22	-0.61	3.45E-02	6.81	3.60	-0.92	6.54E-03
Akap9	6.54	4.44	-0.56	4.87E-02	8.37	5.19	-0.69	3.69E-02
Lphn1	13.12	9.01	-0.54	4.87E-02	15.13	9.69	-0.64	4.54E-02

Supplemental Table S2

q-PCR primer name	Primer sequence
Ctcfl_F	GCTCTGGCTGTGCACCTTACG
Ctcfl_R	CCCACTGTGCCACCATCATC
Stra8_F	GGCAGTTTACTCCCAGTCTGATA
Stra8_R	CAACTTATCCAGGCTTTCTTCCT
Prss50_F	GACAGTTCTCTCTGCACTGTGAC
Prss50_R	CACATTTCTTGCTGTTCAGGATA
Gal3st1_F	ACTGTATCCCAACATGGCCTTC
Gal3st1_R	ATATCTCGCCGAGGTTGACAC
Ccna1_F	GAGTATGCAGAGGAGATTCATCG
Ccna1_R	TCATGTAGTGAGCCTTGGGTCTG
Cdk2_F	ACAGCCGTGGATATCTGGAG
Cdk2_R	TTAGCATGGTGCTGGGTACA
Gal3st1 testis specific form_F	GCTACTCGGAGTTCCGGAAA
Gal3st1 testis specific form_R	GACTTGCAGGGCTTCTTTGG

Supplemental Table S2. Primers used for gene expression analysis by qPCR

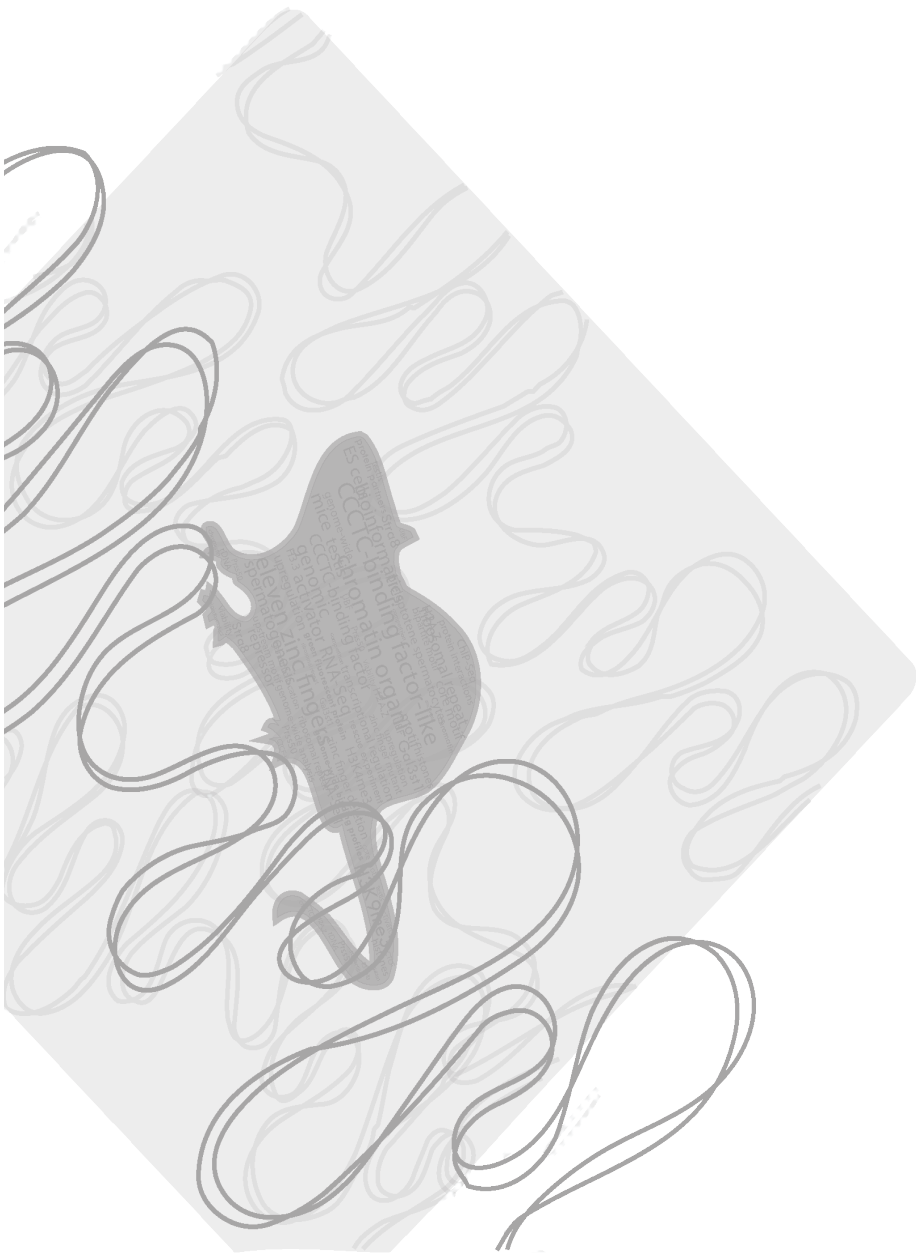
Supplemental Table S3

ChIP primer name	Primer sequence
Amylase_ChIP_F	AATTCTCCTTGTACGGCTTCGTG
Amylase_ChIP_R	TAGCAATGATGTGCACAGCTGAA
Stra8_ChIP_F	TCCTAGAGAAGGGGGTGTTACC
Stra8_ChIP_R	AGCTGACCACCACACGTTTTC
Prss50_ChIP_F	AGAGGAGGGTAGGGGTATCGAC
Prss50_ChIP_R	TCGCCTCAGCTAATTTCTAAGC
Kdelr1_ChIP_F	CCCTAAAGACCCTTCCATCAG
Kdelr1_ChIP_R	TAAAGCTGGCTGGGAGAAAAG
Sec1_ChIP_F	ACCTGGGCAGAGAAGAGAAAG
Sec_ChIP_R	AGCCTCTCCACTGAGTCTTCC
Sox2_Ch4_ ChIP_F	TCGGGATAGAAAGAGGGTAAGG
Sox2_Ch4_ChIP_R	AGCATCTGGGGAGTAACTGTTG

Supplemental Table S3. Primers used ChIP-qPCR

Chapter 6

General Discussion



General Discussion

Chromatin is hierarchically organized in order to compact DNA on the one hand and on the other hand to allow regulation of various cellular processes *(Misteli, 2007)*. CTCF is an important factor that regulates chromatin and spatial organization *(Phillips and Corces, 2009)*. It is important to understand how CTCF binds to DNA and interacts with other proteins in order to gain insight into the functionality of CTCF. It has only one homologous protein called CTCFL, which normally is only expressed in the testis but is also found in various cancer types. In order to understand the role of CTCFL in cancer pathology, it is important to gain an understanding about its basic function during spermatogenesis. In this thesis we investigated the role of CTCF and CTCFL in mice, by identifying interacting proteins, examining CTCF- and CTCFL-dependent gene regulation, and by analyzing the genome-wide distribution and intracellular localization of these proteins.

We used a mass-spectrometry based proteomics analysis of biotinylated CTCF to identify novel interacting partners. We showed that CTCF interacts with the rDNA factor UBF and that it regulates the local chromatin state at the rDNA spacer promoter **(Chapter 2)**.

CTCF mutants with individual ZF deletions were used to elucidate which specific nucleotides within the CTCF binding motif are recognized by a given ZF. We found that ZFs 2-7 are essential for the function of CTCF. By contrast, deletion of ZFs 1 and 8-11 did not affect ES cell viability and these ZFs are therefore less essential. Furthermore, we identified the full length CTCF binding motif and showed that ZFs 8-11 are important to bind a CTCF bipartite motif. This particular motif is excluded from transcription start sites and exons and is associated with the repressive chromatin mark H3K9me3 **(Chapter 3)**.

Finally, we used genome-wide approaches to identify CTCFL function. We identified a CTCFL consensus sequence, which is highly similar to the 20 bp consensus sequence of CTCF. In addition we found that CTCFL binds to sites that are also occupied by CTCF as well as to CTCFL only-sites. CTCFL competes with CTCF on specific sites to maintain the expression of male germ cell genes and is often located to promoters associated with loosely assembled histones **(Chapters 4 and 5)**.

Regulation of rDNA transcription by CTCF

The higher-order organization of chromatin is dynamic and specific chromatin regions can localize to various subnuclear compartments to influence genome configuration and transcription *(Brown et al., 2008; Capelson et al., 2010; Guelen et al., 2008; Nemeth et al., 2010; Peric-Hupkes et al., 2010; van Koningsbruggen et al., 2010)*. The nucleolus is a subnuclear compartment where rDNA is actively transcribed and ribosomal RNA is synthesized, processed and assembled with ribosomal proteins *(Andersen et al., 2002; Grandori et al., 2005; Hernandez-Verdun and Roussel, 2003)*. CTCF localizes throughout the nucleus, but seems to be excluded from the nucleolus **(Chapter 3)**, yet appears to associate with the borders of lamina associated domains (LADs) and the nucleolus *(Guelen et al., 2008; Yusufzai et al., 2004)*. rDNA is shifted according to its transcriptional activity between two subnucleolar compartments *(Nemeth et al., 2010; van Koningsbruggen et al., 2010)*. Within the nucleolus CTCF interacts at the borders with nucleophosmin *(Yusufzai et al., 2004)* and, as described in **chapter 2,** with UBF on the rDNA repeat. We found a specific CTCF binding site on the rDNA, at the spacer promoter, where it regulates the local epigenetic state of rDNA by stimulating the binding of UBF, H3K4me2 and H2A.Z. Additionally, CTCF enhances RNA polymerase I dependent transcription of non-coding RNA from the spacer promoter. All together, CTCF binding results in an open chromatin state near the spacer promoter, which indicates that CTCF binding maintains rDNA repeats poised for transcription. Our findings in mouse have been supported by a genome-wide analysis of human rDNA repeats that also

identified CTCF binding to the spacer promoter *(Zentner et al., 2011)*. Another proteomic approach identified CTCF interaction with RNA polymerase I and its binding to the spacer promoter. Here, CTCF positively regulates rRNA transcription in an RNA polymerase I-dependent fashion. Additionally, CTCF binding enhances histone acetylation across the rDNA locus to establish an open chromatin state *(Huang et al., 2013)*. This implies that CTCF acts as transcriptional activator on rDNA repeats by interacting with the RNA polymerase I transcription initiation complex at the spacer promoter. However, we could not show an effect on total rRNA levels when we deleted *Ctcf* in mouse embryonic fibroblasts **(Chapter 2)**.

The importance of CTCF's function with regard to the regulation of rDNA repeats was emphasized by the identification of a putative nucleolar localization sequence, RRGR motif, in the C-terminus of the protein *(Huang et al., 2013)*. However, in **chapter 3** we show that CTCF localizes throughout the nucleus but does not accumulate extensively inside the nucleolus. Beside UBF and RNA polymerase I, two subunits of the condensin complex, SMC2 and SMC4, were shown to interact with CTCF and to negatively regulate CTCF-mediated rRNA gene transcription *(Huang et al., 2013)*. Thus CTCF interacts with transcriptional activators to promote transcription but is also competed from the active rDNA repeats by transcriptional repressors. This suggests that there is a balance between activation and repression of CTCF-mediated transcription of rDNA repeats. Identification of other repressors and activators is required to gain better understanding on the regulation of the rDNA repeats.

CTCF co-localizes with cohesin on a subset of CTCF binding sites. Both factors mediate long-range interactions on a genome-wide scale. CTCF binding is related to the transcriptional activity of rDNA repeats and cohesin facilitates production of rRNA and thus enhances protein translation *(Bose et al., 2012)*. To gain insight how CTCF and/or cohesin organize the spatial organization of active and inactive rDNA repeats one should first determine whether both proteins are localized on either all rDNA repeats or only a subset (active or inactive). This could be analyzed by including rDNA repeats in the bioinformatic analysis of ChIP-Seq profiles of both proteins. DNA FISH experiment incombination with RNA FISH can be used to gain understanding into the spatial organization of the inactive and active rDNA locus. Both experiments can be done in a conditional *Ctcf* knock out mice background and/or in cells with a depletion of cohesin to examine whether the 3D configuration of rDNA repeats changes.

CTCF binding recognition and motif distribution

CTCF's binding to DNA is established by its ZF domain. All ZFs are highly conserved across species and must therefore be important for the functionality of the protein *(Moon et al., 2005; Pugacheva et al., 2006)*. ZFs 4-7 bind to the 20 bp core motif present in 80-99.5% of all CTCF binding sites, and therefore these ZFs are essential to establish CTCF binding *(Nakahashi et al., 2013; Renda et al., 2007; Rhee and Pugh, 2011)*. The remaining ZFs were proposed to be required for stabilizing CTCF binding to its target sites *(Nakahashi et al., 2013; Renda et al., 2007)*.

In order to understand the functional relevance of each ZF with regard to CTCF binding recognition we generated ES cells expressing CTCF ZF mutants in a *Ctcf* knockout background **(Chapter 3)**. Deletion of endogenous CTCF results in cellular lethality *(Heath et al., 2008)*. This can be prevented by substitution of endogenous CTCF with wild type GFP-CTCF *(Sleutels et al., 2012)*. In **chapter 3** we describe that mutant CTCF with distinct individual ZF deletions can also substitute for endogenous CTCF, allowing us to study ZF specific effects on DNA binding. However, because we delete ZFs our method results in shifted position of the remaining ZFs within the complete domain. This might influence CTCF binding to DNA. Indeed, we speculate that the major reason why GFP-CTCF-Δ8 fails to bind the bipartite CTCF consensus motif is because ZFs 9-11 are shifted and can not bind essential nucleotides. This is also the case for GFP-CTCF-Δ9 and GFP-CTCF-Δ10. Recently, an alternative approach was published, where ZF mutants were created with mutated key histidine residues that coordinate zinc binding

(Nakahashi et al., 2013). This method maintains the position of each ZF although, however, folding of the finger might be affected.

About 80% of the CTCF binding sites contain the 20 bp consensus core motif and 15-30% the bipartite motif *(Boyle et al., 2011; Chen et al., 2008; Kim et al., 2007; Nakahashi et al., 2013; Rhee and Pugh, 2011; Schmidt et al., 2012; Xie et al., 2007)*. It is therefore remarkable that CTCF zinc finger mutants are reduced in DNA binding on bipartite motif sites, since the core motif is present and ZF4-7 are not affected in these mutants. Thus, in addition to the shifted position of the ZFs, another environmental factor could be present that affect CTCF binding to core motif sites. It might be possible that CTCF protects bipartite motifs from binding of other factors. Reduction of CTCF binding in the zinc finger mutants might allow competition between CTCF and additional factors resulting in a reduction in DNA binding of CTCF zinc finger mutants to bipartite motifs.

In support of current understanding our rescue experiment suggested that ZFs 4-7 are essential for the functionality of CTCF. We were also not able to generate ES cells expressing GFP-CTCF-Δ2, or -3, suggesting that these fingers are also highly important for CTCF's function. It is possible that these two ZFs are required for binding to the core motif by recognizing sequences adjacent to it and/or are essential for protein-protein interactions involving important cellular processes. These hypotheses could be tested in ES cells with one remaining endogenous allele or through *in vitro* assays e.g. EMSA and immunoprecipitation experiments.

Although all ZFs of CTCF are highly conserved we were able to delete fingers 1 and 8-11 and demonstrated that these fingers are required for binding the bipartite CTCF recognition sequence. By establishing binding deficiencies of ES cells expressing CTCF mutants in these peripheral ZFs we identified the complete CTCF binding motif. Our data result in a model where ZFs 1-7 bind adjacent to each other, with finger 1 mainly binding to a nucleotide on 'position 44', 10 nucleotides downstream of the core sequence **(Chapter 3)**, indicating the 3'end of the binding sequence. The region bound by ZF1-7 is followed by an approximately 10 bp helical turn, which might be bridged by ZF8. After this ZFs 9-11 are able to bind to the upstream motif of the bipartite sequence. It is still unknown what the exact role of ZFs 8-11 is when CTCF is bound to sites containing only the core motif.

In **chapter 3** we show that CTCF sites containing the bipartite motif are excluded from transcription start sites and exons and preferentially located in intergenic and intronic regions. These sites are associated with the repressive chromatin mark H3K9me3 whereas CTCF is often located to open chromatin environments. As CTCF is one of the key players in the regulation of long-range interactions the question remains if these CTCF bipartite motif-containing sites are involved in 3D looping.

Functional analysis of CTCFL

Initial studies on the localization of CTCFL and CTCF in testis yielded contradictory results. First, both proteins were detected in a mutually exclusive pattern, with CTCFL being localized in primary spermatocytes, preceding CTCF expression in round spermatids *(Loukinov et al., 2002)*. Later reports identified CTCFL in gonocytes during embryonic development and in spermatogonia after birth *(Jelinic et al., 2006)*. *Ctcfl* expression was later also detected in round spermatids *(Kosaka-Suzuki et al., 2011; Suzuki et al., 2010)*.

In order to elucidate where CTCFL and CTCF are localized in the testis we used various mouse models and technologies **(Chapter 4)**. *Ctcfl* knock out, as well as *Ctcf*^{gfp} and *Ctcfl*^{gfp} knock-in mice were examined using immunohistochemistry, immunofluorescence staining and *ex-vivo* imaging with multiphoton confocal laser scanning microscopy. This analysis revealed that CTCF is expressed in all cell types of the seminiferous tubule, but that the amount of CTCF decreases drastically in spermiogenesis. CTCFL expression could only be detected in late spermatogonia and pre-leptotene spermatocytes. Our results oppose the theory that CTCF

and CTCFL are expressed in a mutually exclusive pattern *(Loukinov et al., 2002)*, since we showed that these proteins are expressed in the same cells during early spermatogenesis.

It has been hypothesized that CTCFL is involved in establishing DNA imprints during reprogramming *(Loukinov et al., 2002)*. Establishment of sex-specific imprints takes place at prospermatogonia stage (E15-16) *(Reik et al., 2001)*. This process is initiated prior to the expression of CTCFL, which indicates that CTCFL is most likely not involved in reprogramming. Furthermore, as described in **chapter 4**, *Ctcfl* knock out mice do not exhibit DNA methylation defects and do not show imprinted-related developmental defects.

In **chapter 4** we hypothesized that CTCFL acts as transcriptional activator, since many genes are down-regulated genes in *Ctcfl* knock out mice. Surprisingly, microarray experiments revealed only very few testis-related genes, including *Prss50* and *Gal3st1*. The absence of more deregulated testis-associated genes in these microarray experiments can be explained. CTCFL is only expressed in a small population of cells in the testis and the use of whole testis extracts will 'dilute' CTCFL-specific responses, which are subsequently not detected by microarray. The fact that *Prss50* and *Gal3st1* were found in this experimental set-up indicates that these genes are regulated by CTCFL. This was confirmed in **Chapter 5**.

To study CTCFL-dependent transcription regulation we isolated a pure CTCFL-expressing population by crossing *Ctcfl* mutant mice with *Stra8-gfp* transgenic mice as presented in **chapter 5**. STRA8 and CTCFL are expressed in the same cells during spermatogenesis *(Sleutels et al., 2012)*; therefore GFP driven by the Stra8 promoter can be used as a marker to isolate CTCFL-expressing cells via FACS. An RNA-Seq-based analysis confirmed our statement that CTCFL is a transcriptional activator, since the majority of mis-regulated genes are negatively affected. Next to *Stra8*, *Prss50* and *Gal3st1* we found additional testis-related genes such as *Dmrt1* and, surprisingly, a subset of genes involved in carcinogenesis. CTCFL is aberrantly expressed in cancer cell lines and different types of cancer. It belongs to the group of cancer germline genes (CG genes), which are germ cell-specific genes that are reactivated and aberrantly expressed in various cancer types *(Cheng et al., 2011)*. CTCFL stimulates the expression of CG genes NY-ESO-1 and several MAGE-A genes in cancer cell lines *(Bhan et al., 2011; Hong et al., 2005; Kang et al., 2007; Smith et al., 2009; Vatolin et al., 2005)*. Our RNA-Seq data from testis revealed new candidate genes that could enhance our understanding regarding CTCFL's function during cancer development.

Genome-wide analysis revealed that CTCFL binding sites are mainly located at promoters with loosely assembled histones **(Chapter 4)**. CTCFL enhances expression of several CG genes by binding to their promoters. A subset of CTCFL binding sites is also bound by CTCF and CTCFL is able to compete with CTCF on some of the 'shared' sites. CTCFL binds a similar 20 bp consensus sequence as CTCF, which is not so surprising given 71% identity of the ZF domains of these proteins. This begs the question how it is possible that there are both CTCF-only and CTCFL-only sites. For one, CTCFL ZFs 10 and 11 are less homologous to corresponding fingers of CTCF *(Loukinov et al., 2002)*, which suggests that CTCFL is not able to bind to CTCF's upstream motif described in **chapter 3**. Secondly, in **chapter 4** we also propose that nucleosome composition containing loosely assembled histones, e.g. H3.3 and H2A.Z, determines binding of CTCF and CTCFL.

It is unknown whether a subset of CTCFL-CTCF binding sites is involved in long-range interactions. If CTCFL is able to compete with CTCF on shared binding sites it would be interesting to examine whether CTCFL can interfere with CTCF-mediated loops to affect transcription. Can CTCFL mediate loops at all? The fact that in ES cells CTCFL binds to ~2000 sites to which CTCF does not bind but cohesin does *(Sleutels et al., 2012)*, indicates that CTCFL might also act in conjunction with cohesin. Whether CTCFL is involved in chromatin interactions *in vivo* can be tested using the *Ctcfl*$^{wt/wt}$*-Stra8-Gfp* and *Ctcfl*$^{del/del}$*-Stra8-Gfp* mice described in **chapter 5.** Purified cell populations from *Ctcfl* wild type and knock out mice can be used to perform chromatin conformation capture techniques such as 3C or 4C to examine whether the

presence or absence of CTCFL cause changes in chromatin structure.

In conclusion, the data presented in this thesis have enhanced our understanding of the biological functions of CTCF and CTCFL, and the interplay between these factors. The genome-wide approaches applied in our research have contributed enormously to this and provide future prospective and clues for research.

References

Andersen, J.S., Lyon, C.E., Fox, A.H., Leung, A.K., Lam, Y.W., Steen, H., Mann, M., and Lamond, A.I. (2002). Directed proteomic analysis of the human nucleolus. Curr Biol *12*, 1-11.

Bhan, S., Negi, S.S., Shao, C., Glazer, C.A., Chuang, A., Gaykalova, D.A., Sun, W., Sidransky, D., Ha, P.K., and Califano, J.A. (2011). BORIS binding to the promoters of cancer testis antigens, MAGEA2, MAGEA3, and MAGEA4, is associated with their transcriptional activation in lung cancer. Clin Cancer Res *17*, 4267-4276.

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

Boyle, A.P., Song, L., Lee, B.K., London, D., Keefe, D., Birney, E., Iyer, V.R., Crawford, G.E., and Furey, T.S. (2011). High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. Genome Res *21*, 456-464.

Brown, C.R., Kennedy, C.J., Delmar, V.A., Forbes, D.J., and Silver, P.A. (2008). Global histone acetylation induces functional genomic reorganization at mammalian nuclear pore complexes. Genes Dev *22*, 627-639.

Capelson, M., Liang, Y., Schulte, R., Mair, W., Wagner, U., and Hetzer, M.W. (2010). Chromatin-bound nuclear pore components regulate gene expression in higher eukaryotes. Cell *140*, 372-383.

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*, 1106-1117.

Cheng, Y.H., Wong, E.W., and Cheng, C.Y. (2011). Cancer/testis (CT) antigens, carcinogenesis and spermatogenesis. Spermatogenesis *1*, 209-220.

Grandori, C., Gomez-Roman, N., Felton-Edkins, Z.A., Ngouenet, C., Galloway, D.A., Eisenman, R.N., and White, R.J. (2005). c-Myc binds to human ribosomal DNA and stimulates transcription of rRNA genes by RNA polymerase I. Nat Cell Biol *7*, 311-318.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W.*, et al.* (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature.

Heath, H., Ribeiro de Almeida, C., Sleutels, F., Dingjan, G., van de Nobelen, S., Jonkers, I., Ling, K.W., Gribnau, J., Renkawitz, R., Grosveld, F.*, et al.* (2008). CTCF regulates cell cycle progression of alphabeta T cells in the thymus. Embo J *27*, 2839-2850.

Hernandez-Verdun, D., and Roussel, P. (2003). Regulators of nucleolar functions. Prog Cell Cycle Res *5*, 301-308.

Hong, J.A., Kang, Y., Abdullaev, Z., Flanagan, P.T., Pack, S.D., Fischette, M.R., Adnani, M.T., Loukinov, D.I., Vatolin, S., Risinger, J.I.*, et al.* (2005). Reciprocal binding of CTCF and BORIS to the NY-ESO-1 promoter coincides with derepression of this cancer-testis gene in lung cancer cells. Cancer Res *65*, 7763-7774.

Huang, K., Jia, J., Wu, C., Yao, M., Li, M., Jin, J., Jiang, C., Cai, Y., Pei, D., Pan, G.*, et al.* (2013). Ribosomal RNA Gene Transcription Mediated by the Master Genome Regulator Protein CCCTC-binding Factor (CTCF) Is Negatively Regulated by the Condensin Complex. J Biol Chem *288*, 26067-26077.

Jelinic, P., Stehle, J.C., and Shaw, P. (2006). The testis-specific factor CTCFL cooperates with the protein methyltransferase PRMT7 in H19 imprinting control region methylation. PLoS Biol *4*, e355.

Kang, Y., Hong, J.A., Chen, G.A., Nguyen, D.M., and Schrump, D.S. (2007). Dynamic transcriptional regulatory complexes including BORIS, CTCF and Sp1 modulate NY-ESO-1 expression in lung cancer cells. Oncogene *26*, 4394-4403.

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*, 1231-1245.

Kosaka-Suzuki, N., Suzuki, T., Pugacheva, E.M., Vostrov, A.A., Morse, H.C., 3rd, Loukinov, D., and Lobanenkov, V. (2011). Transcription factor BORIS (Brother of the Regulator of Imprinted Sites) directly induces expression of a cancer-testis antigen, TSP50, through regulated binding of BORIS to the promoter. J Biol Chem *286*, 27378-27388.

Loukinov, D.I., Pugacheva, E., Vatolin, S., Pack, S.D., Moon, H., Chernukhin, I., Mannan, P., Larsson, E., Kanduri, C., Vostrov, A.A*., et al.* (2002). BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. Proc Natl Acad Sci U S A *99*, 6806-6811.

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

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*, 165-170.

Nakahashi, H., Kwon, K.R., Resch, W., Vian, L., Dose, M., Stavreva, D., Hakim, O., Pruett, N., Nelson, S., Yamane, A*., et al.* (2013). A Genome-wide Map of CTCF Multivalency Redefines the CTCF Code. Cell Rep *3*, 1678-1689.

Nemeth, A., Conesa, A., Santoyo-Lopez, J., Medina, I., Montaner, D., Peterfia, B., Solovei, I., Cremer, T., Dopazo, J., and Langst, G. (2010). Initial genomics of the human nucleolus. PLoS Genet *6*, e1000889.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W., Solovei, I., Brugman, W., Graf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M*., et al.* (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol Cell *38*, 603-613.

Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. Cell *137*, 1194-1211.

Pugacheva, E.M., Kwon, Y.W., Hukriede, N.A., Pack, S., Flanagan, P.T., Ahn, J.C., Park, J.A., Choi, K.S., Kim, K.W., Loukinov, D*., et al.* (2006). Cloning and characterization of zebrafish CTCF: Developmental expression patterns, regulation of the promoter region, and evolutionary aspects of gene organization. Gene *375*, 26-36.

Reik, W., Dean, W., and Walter, J. (2001). Epigenetic reprogramming in mammalian development. Science *293*, 1089-1093.

Renda, M., Baglivo, I., Burgess-Beusse, B., Esposito, S., Fattorusso, R., Felsenfeld, G., and Pedone, P.V. (2007). Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. J Biol Chem *282*, 33336-33345.

Rhee, H.S., and Pugh, B.F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell *147*, 1408-1419.

Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Goncalves, A., Kutter, C., Brown, G.D., Marshall, A., Flicek, P., and Odom, D.T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell *148*, 335-348.

Sleutels, F., Soochit, W., Bartkuhn, M., Heath, H., Dienstbach, S., Bergmaier, P., Franke, V., Rosa-Garrido, M., van de Nobelen, S., Caesar, L*., et al.* (2012). The male germ cell gene regulator CTCFL is functionally different from CTCF and binds CTCF-like consensus sites in a nucleosome composition-dependent manner. Epigenetics Chromatin *5*, 8.

Smith, I.M., Glazer, C.A., Mithani, S.K., Ochs, M.F., Sun, W., Bhan, S., Vostrov, A., Abdullaev, Z., Lobanenkov, V., Gray, A*., et al.* (2009). Coordinated activation of candidate proto-oncogenes and cancer testes antigens via promoter demethylation in head and neck cancer and lung cancer. PLoS One *4*, e4961.

Suzuki, T., Kosaka-Suzuki, N., Pack, S., Shin, D.M., Yoon, J., Abdullaev, Z., Pugacheva, E., Morse, H.C., 3rd, Loukinov, D., and Lobanenkov, V. (2010). Expression of a testis-specific form of Gal3st1 (CST), a gene essential for spermatogenesis, is regulated by the CTCF paralogous gene BORIS. Mol Cell Biol *30*, 2473-2484.

van Koningsbruggen, S., Gierlinski, M., Schofield, P., Martin, D., Barton, G.J., Ariyurek, Y., den Dunnen, J.T., and Lamond, A.I. (2010). High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. Mol Biol Cell *21*, 3735-3748.

Vatolin, S., Abdullaev, Z., Pack, S.D., Flanagan, P.T., Custer, M., Loukinov, D.I., Pugacheva, E., Hong, J.A., Morse, H., 3rd, Schrump, D.S.*, et al.* (2005). Conditional expression of the CTCF-paralogous transcriptional factor BORIS in normal cells results in demethylation and derepression of MAGE-A1 and reactivation of other cancer-testis genes. Cancer Res *65*, 7751-7762.

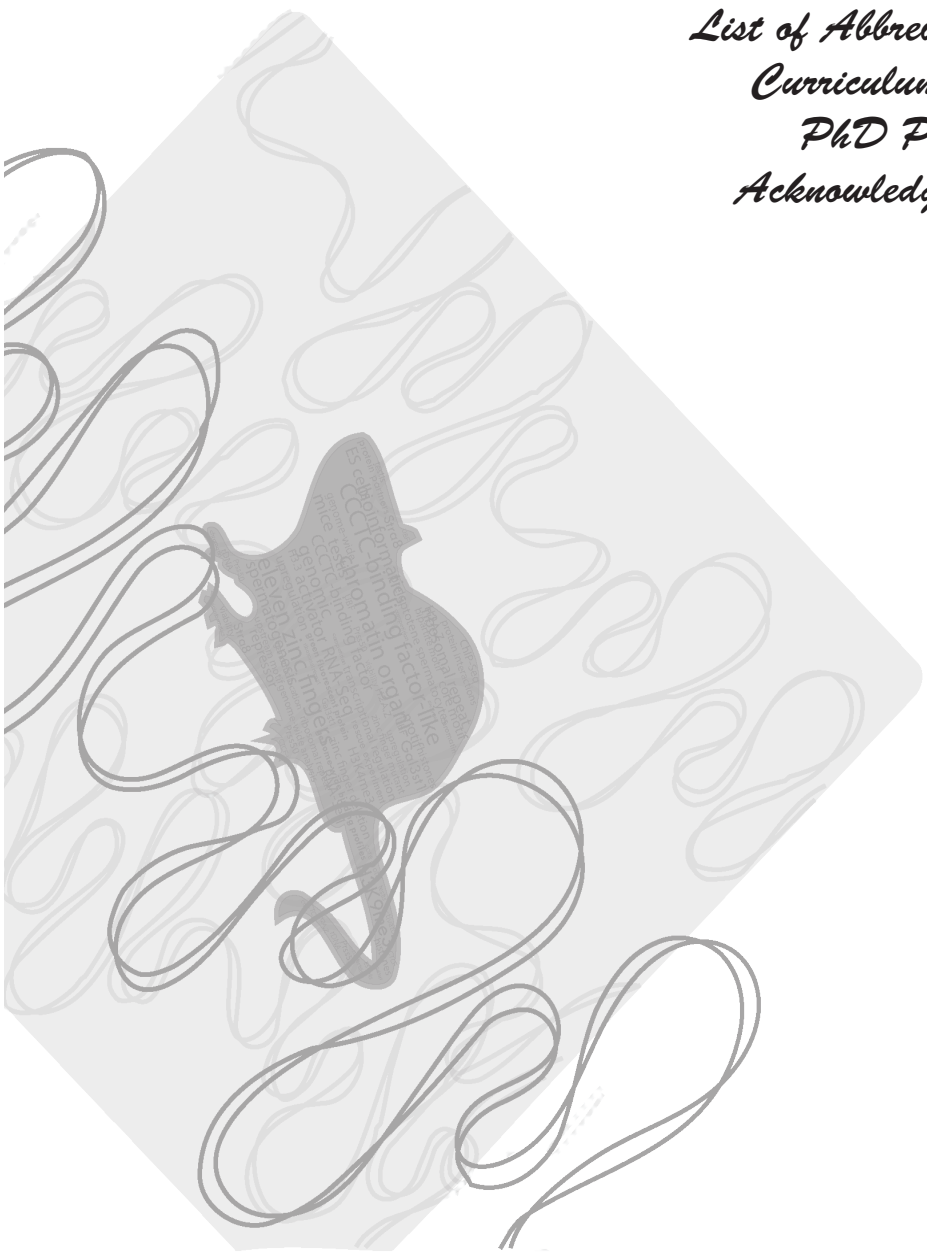
Xie, X., Mikkelsen, T.S., Gnirke, A., Lindblad-Toh, K., Kellis, M., and Lander, E.S. (2007). Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. Proc Natl Acad Sci U S A *104*, 7145-7150.

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*, 291-298.

Zentner, G.E., Saiakhova, A., Manaenkov, P., Adams, M.D., and Scacheri, P.C. (2011). Integrative genomic analysis of human ribosomal DNA. Nucleic Acids Res *39*, 4949-4960.

Chapter 7

Summary/Samenvatting
List of Abbreviations
Curriculum Vitae
PhD Portfolio
Acknowledgements



Summary

DNA is a long molecule that is stored inside the small volume of the nucleus by hierarchical wrapping and folding. DNA and the proteins that associate with it, which together are called chromatin, are organized to enable compaction on the one hand and to allow transcription, DNA repair and all kinds of other processes on the other. One of the key players in chromatin structure and spatial organization of the genome is CTCF (CCCTCF binding factor). CTCF is a multifunctional and highly conserved nuclear protein that is characterized by an eleven zinc finger (ZF) domain that is surrounded by N- and C-terminal regions. CTCFL (CCCTC binding factor like), whose function has been characterized to a lesser extent, is a less well conserved testis-specific paralogue of CTCF. In this thesis we examined the biological roles of CTCF and CTCFL using various techniques and approaches. We identified interacting proteins of CTCF(L), examined CTCF- and CTCFL dependent gene regulation, and analyzed the genome-wide distribution and intracellular localization of these proteins.

One of the CTCF- and CTCFL-interacting proteins that we identified is UBF (upstream binding factor). UBF is localized in the nucleolus and binds to the ribosomal DNA (rDNA) repeat to regulate the transcription of ribosomal RNAs (rRNAs), which are major components of ribosomes. CTCF is bound upstream of the unmethylated spacer promoter of the rDNA repeat. Its binding stimulates binding of RNA polymerase I and H2A.Z near the spacer promoter. This leads to enhanced transcription of non-coding RNA from the spacer promoter. Thus, CTCF affects RNA polymerase I-mediated events by regulating chromatin at the rDNA spacer promoter. CTCF may load UBF onto rDNA, thereby forming part of a network that maintains rDNA genes poised for transcription **(chapter 2)**.

In addition to binding to rDNA, CTCF binds to 25,000-50,000 sites in mouse and human genomes using its eleven ZF domain. It recognizes a 20 bp conserved consensus sequence (core motif). Additionally, it binds sites containing both the core motif and a 9 bp sequence upstream of the core motif (upstream motif). To determine how the different ZFs of CTCF contribute to binding specificity and how this relates to cellular function, we replaced the endogenous *Ctcf* gene in mouse embryonic stem (ES) cells with either wild type GFP-CTCF or with mutants in which individual ZF domains were deleted. ES cell lines with deletion of individual ZFs 2-7 could not be established suggesting that these ZFs are required for ES cell viability. By contrast, ZF1 and ZF 8-11 were dispensable. Based on ChIP-Sequencing analysis of GFP-CTCF- and GFP-CTCF-ZF-mutant-expressing ES cells we propose that ZF1-3 bind nine contiguous nucleotides immediately downstream of the CTCF core motif, which is bound by ZF4-7, and that ZF8-11 are required for binding a spacer sequence and the upstream motif. CTCF binding sites containing the core with upstream motif are specifically depleted from transcription start sites and exons, and are associated with the repressive chromatin mark H3K9me3. These sites are less well bound by GFP-CTCF-Δ8-11 mutants. Some of the genes with such CTCF binding sites show an altered expression. Combined, our data suggest that CTCF binds DNA throughout the genome to regulate essential nuclear processes, and that it binds near genes to regulate transcription locally **(chapter 3)**.

We also examined the function of CTCFL using mouse models and ES cells. Our data show that in the testis of adult male mice CTCFL is expressed in type B spermatogonia and pre-leptotene spermatocytes. Absence of CTCFL causes subfertility because of partially penetrant testis atrophy. Furthermore, CTCFL positively regulates expression of the germ cell-specific factors *Prss50*, *Stra8* and *Gal3st1*, suggesting that it acts as transcriptional activator. Genome-wide analysis of ES cells expressing CTCFL showed that the protein binds a 20 bp consensus sequence similar to that of CTCF and is able to compete with CTCF. However, only ~3,700 out of the ~ 5,700 CTCFL- and ~31,000 CTCF-binding sites overlap. Strikingly, CTCFL is mainly located on promoters with loosely assembled nucleosomes whereas CTCF binds to sites

surrounded by phased nucleosomes. This suggests that nucleosome composition specifies the genome-wide binding of CTCFL and CTCF. We propose that the transient expression of CTCFL in spermatogonia and pre-leptotene spermatocytes serves to occupy a subset of promoters and maintain the expression of male germ cell genes **(chapter 4)**.

The function of CTCFL in transcription regulation in the testis was further analyzed. The cellular expression pattern of CTCFL in the testis completely overlaps with that of STRA8. To examine the transcriptional function of CTCFL in more detail, we used a FACS-based approach to separate CTCFL-expressing from non-expressing testicular cells. We sorted GFP^+ (i.e. CTCFL-expressing) and GFP^- (i.e. not containing CTCFL) cell populations from a *Stra8-Gfp* transgenic mouse maintained in wild type or *Ctcfl* knock out backgrounds. RNA-Sequencing revealed highly specific enrichment of the *Stra8* mRNA in the GFP^+ fractions, validating our separation method. Many more genes were down-regulated than up-regulated genes in the GFP^+ *Ctcfl* knock out fraction, confirming the hypothesis that CTCFL is a transcriptional activator. Combined our data suggest that CTCFL activates transcription of a limited number of genes in the testis and that CTCF acts as repressor of these genes. This suggests that CTCFL competes with CTCF to maintain proper gene expression in the testis **(chapter 5)**.

In conclusion, this thesis contains studies that provide further insight into the biological functions of CTCF and CTCFL in terms of protein interaction, chromatin organization and transcriptional regulation.

Samenvatting

Het DNA is een lang molecuul, dat opgeslagen is in het kleine volume van de celkern doormiddel van hiërarchisch wikkelen en vouwen. Het DNA en de eiwitten die geassocieerd zijn met het DNA worden samen chromatine genoemd en zijn op zodanige manier georganiseerd om aan de ene kant compressie toe te staan en aan de andere kant transcriptie, DNA reparatie en overige cellulaire processen. Een van de belangrijkste factoren in het structureren van chromatine en de ruimtelijke organisatie van het genoom is CTCF (CCCTC-binding factor). CTCF is een multifunctioneel en zeer geconserveerd nucleair eiwit dat gekarakteriseerd wordt door een elf zink vinger (ZV) domein, die omgeven is door de N- en C-terminale regio's. CTCFL (CCCTCF-binding factor-like), van wie de functie op mindere mate gekarakteriseerd is, is een minder goed geconserveerd testis-specifiek paraloog van CTCF. In dit proefschrift bestuderen wij de biologische rollen van CTCF en CTCFL door gebruik te maken van verschillende technieken en benaderingen. We identificeren de interacterende eiwitten van CTCF(L), bestuderen CTCF- en CTCFL-afhankelijke gen regulatie, en analyseren de distributie over het gehele genoom en de intracellulaire lokalisatie van deze eiwitten.

Een van de CTCF- en CTCFL-interacterende eiwitten die wij geïdentificeerd hebben is UBF (upstream binding factor). UBF is gelokaliseerd in de nucleolus en bindt aan het ribosomaal DNA (rDNA) repeat om de transcriptie van ribosomaal RNAs (rRNAs), die de grootste component van de ribosomen zijn, te reguleren. CTCF bindt voor de ongemethyleerde spacer promoter van het rDNA repeat. Deze binding stimuleert de bindingen van RNA polymerase I en H2A.Z nabij de spacer promoter. Dit leidt tot verhoogde transcriptie van het niet-gecodeerde RNA van de spacer promoter. Dus, CTCF heeft een effect op RNA polymerase I gemedieerde gebeurtenissen door de regulatie van chromatine op de rDNA spacer promoter. CTCF kan UBF op het DNA zetten, waardoor het onderdeel wordt van een netwerk dat het rDNA toegankelijk houdt voor transcriptie **(hoofdstuk 2)**.

Naast het binden aan rDNA, bindt CTCF aan 25.000-50.000 plekken in het genoom van muis en mens gebruikmakend van het elf ZV domein. Het herkent een 20 bp geconserveerd consensus sequentie (kern motief/core motif). Bovendien bindt het ook plekken die het kern motief en een 9 bp sequentie voor het kern motief bevatten (voorop gelegen motief/upstream motif). Om te bepalen hoe de verschillenden ZVs van CTCF bijdragen aan de bindingscapaciteit en hoe dit relateert naar cellulaire functies, hebben wij het endogene *Ctcf* gen in muis embryonale stam (ES) cellen vervangen met wild type GFP-CTCF of met mutanten waarvan individuele zink vinger domeinen waren verwijderd. ES cellijnen met een deletie van individuele ZVs 2-7 kunnen niet bewerkstelligd worden, wat suggereert dat deze ZVs nodig zijn voor de vitaliteit van ES cellen. In tegenstelling tot ZV1 en ZV 8-11 die overbodig waren. Gebaseerd op ChIP-Seq analyses van GFP-CTCF en GFP-CTCF-ZV-mutant expresserende ES cellen stellen wij voor dat ZV 1-3 aan negen opeenvolgende nucleotide aan het einde van de kern motief binden. Het kern motief wordt gebonden door ZV 4-7, en ZV 8-11 zijn nodig voor de binding aan de tussenliggende sequentie en het voorop gelegen motief. CTCF bindingsplekken bevattende de kern en voorop gelegen motief zijn specifiek afwezig op transcriptie start plekken en exonen, en zijn geassocieerd met de repressieve chromatine marker H3K9me3. Deze plekken zijn minder goed gebonden door GFP-CTCF-Δ8-11 mutanten. Sommige genen met deze CTCF bindingsplek vertonen een andere expressie. Al met al suggereren onze data dat CTCF over het hele genoom bindt om essentiële nucleaire processen te reguleren, en dat het nabij genen bindt om lokaal transcriptie te reguleren **(hoofdstuk 3)**.

Wij hebben ook de functie van CTCFL bestudeerd met muis modellen en ES cellen. Onze data laten zien dat CTCFL in de testis van volwassen muizen in type B spermatogonia en pre-leptotene spermatocyten tot expressie komt. Afwezigheid van CTCFL zorgt voor subfertiliteit door een partiële penetrerende testis atrofie. Bovendien heeft CTCFL een positief

regulerend effect op de expressie van geslachtscel specifieke factoren *Prss50*, *Stra8* en *Gal3st1* wat suggereert dat het een transcriptionele activator is. Analyses van het hele genoom van CTCFL expresserende ES cellen laten zien dat het eiwit een 20 bp consensus sequentie bindt vergelijkbaar met dat van CTCF en dat is met CTCF kan competeren. Echter, alleen ~3.700 van de ~5700 CTCFL- en ~31.000 CTCF-bindingsplekken overlappen. Opvallend is dat CTCFL voornamelijk op promoters met los bindende nucleosomen bindt, terwijl CTCF op plekken bindt met gefaseerde nucleosomen. Dit suggereert dat nucleosoom compositie de bindingsplekken van CTCF en CTCFL specificeert. Wij stellen voor dat de tijdelijke expressie van CTCFL in spermatogonia en pre-leptotene spermatocyten als doel heeft om een aantal promoters te binden en de expressie van mannelijke geslachtscel genen te behouden.

De functie van CTCFL in de transcriptionele regulatie in de testis is verder geanalyseerd. Het cellulaire expressie patroon van CTCFL in de testis overlapt compleet met dat van STRA8. Om de transcriptionele functie van CTCFL in meer detail te bestuderen, maken wij gebruik van een FACS methode om de CTCFL-expresserende van de niet-expresserende testiculaire cellen te scheiden. We hebben GFP^+ (CTCFL-expresserend) en GFP^- (niet CTCFL-expresserend) cel populaties gesorteerd van een *Stra8-Gfp* transgene muis in een wild type of *Ctcfl* gedeleteerde muis. RNA-Sequencing laat zien dat er een hoge verrijking is van het *Stra8* mRNA in de GFP^+ fractie, wat onze methode valideerde. Meer genen waren omlaag dan omhoog gereguleerd in de GFP^+ *Ctcfl* gedeleteerde muis, wat de hypothesis dat CTCFL een transcriptionele activator is verder bevestigd. Al met al suggereren onze data dat CTCFL transcriptie van een aantal genen in de testis activeert en dat CTCF een repressor is van deze genen. Dit suggereert dat CTCFL met CTCF competeert om een juiste gen expressie in de testis te bewerkstelligen **(hoofdstuk 4)**.

Concluderend, dit proefschrift bevat studies die meer inzicht in de biologische functies van CTCF en CTCFL geven omtrent eiwit interacties, chromatine organisatie en transcriptie regulatie.

List of abbreviations

bp	base pairs
cDNA	Complementary desoxyribonucleic acid
CG gene	Cancer germ cell gene
ChIP	Chromatin immunoprecipitation
ChIP-Seq	ChIP-Sequencing
CT	Chromosome territory
CTCF	CCCTC-binding factor
CTCFL	CCCTC-binding factor like
DNA	Deoxyribonucleic acid
EM	Electron microscopic
ES	Embryonic stem cells
FACS	Fluorescence-activated cekk sorting
FDR	False discovery rate
FISH	Fluorescent in situ hybridization
FPKM	Fragments per kilobase of transcripts per Million mapped reads
FRAP	Fluorescence recovery after photobleaching
GFP	Green Fluorescence Protein
GST	Glutatione-S-transferase
H2A.Z	Histone H2A.Z
HAT	Histone acetyl transferase
HDAC	Histone deacetylase
HMG	High mobility group
HS	Hypersensitivity site
ICR	Imprinting control region
IGS	Intergenic spacer
Kb	Kilo base pairs
kDa	Kilo Dalton
LAD	Lamina associated domain
LCR	Locus control region
Mb	Mega base pairs
MEF	Mouse embryonic fibroblast
NAD	Nucleolus associated domain
PCR	Polymerase chain reaction
PGC	Primordial germ cell
qPCR	Quantitative PCR
rDNA	ribosomal DNA
RFP	Red fluorescence protein
RNA	Ribonucleic acid
RNAi	RNA interference
RNA-Seq	RNA-Sequencing
RPA194	RNA polymerase I large subunit of 194 kD
rRNA	ribosomal RNA
TAD	Topologically associated domain
TES	Transcription elongation site
TSS	Transcription start site
UC motif	Upstream core motif
WT	Wild type
ZF	Zinc Finger

Curriculum Vitae

Personal Information:

Name	**Widia Sabrina Wanita Soochit**
Date of birth	4 October 1986
Place of birth	Rotterdam, The Netherlands

Education

2009-2013	**PhD student** Department of Cell Biology, Erasmus MC, Rotterdam Promotor: Prof. dr. Frank Grosveld Co-promotor: Dr. ir. Niels Galjart
2005-2009	**Master of Science in Molecular Medicine** Erasmus MC, University Medical Center, Rotterdam
2008-2009	Department of Cell Biology, Erasmus MC, Rotterdam Master thesis: "DNA binding by CTCF, zinc fingers to the rescue", Supervisor: Dr. Frank Sleutels, Head of laboratory: Dr. ir. Niels Galjart
2008	Department of Internal Medicine, Erasmus MC, Rotterdam Internship: "Identification of osteoblast and adipocyte specific marker genes in differentiated human mesenchymal stem cells", Supervisor: Dr. Jeroen van de Peppel, Head of laboratory: Prof. dr. Hans van Leeuwen
2004-2008	**Medicine** Erasmus MC, University Medical Center, Rotterdam
2005-2008	Doctoral exam
2004-2005	Propaedeutic exam
1998-2004	**VWO Gymnasium** Rotterdams Montessori Lyceum, Rotterdam
2004	VWO exam Profile: Nature and Health with side subjects Art and Greek

List of publications

Suzanne van de Nobelen, Manuel Rosa-Garrido, Joerg Leers, Helen Heath, **Widia Soochit**, Linda Joosen, Iris Jonkers, Jeroen Demmers, Michael van der Reijden, Verónica Torrano, Frank Grosveld, M Dolores Delgado, Rainer Renkawitz, Niels Galjart and Frank Sleutels (2010)
"CTCF regulates the local epigenetic state of ribosomal repeats"
Epigenetics Chromatin 3,19

Frank Sleutels, **Widia Soochit**, Marek Bartkuhn, Helen Heath, Sven Dienstbach, Philipp Bergmaier, Vedran Franke, Manuel Rosa-Garrido, Suzanne van de Nobelen, Lisa Caesar, Michael van der Reijden, Jan Christian Bryne, Wilfred van IJcken, J Anton Grootegoed, M Dolores Delagdo, Boris Lenhard, Rainer Renkawitz, Frank Grosveld and Niels Galjart (2012)
"The male germ cell gene regulator CTCFL is functionally different from CTCF and binds CTCFL-like consensus sites in a nucleosome composition-dependent manner"
Epigenetics Chromatin 5,8

Widia Soochit*, Frank Sleutels*, Marek Bartkuhn*, Michael van der Reijden, Frank Grosveld, Rainer Renkawitz, and Niels Galjart
"Reconstitution of wild type and mutant GFP-CTCF expression in embryonic stem cells lacking endogenous CTCF"
Manuscript in preparation
** these authors contributed equally*

PhD Portfolio

Name: Widia Sabrina Wanita Soochit
Department: Cell Biology, Erasmus Medical Center, Rotterdam
PhD period: 2009-2013
Promotor: Prof. dr. F.G. Grosveld
Co-promotor: Dr. ir. N.J. Galjart

General courses

- Radiation safety course 5B 2011
- Literature Course, Sex chromosomes 2010
- Safely working in the laboratory 2009

Specific courses

- Training omgaan met groepen voor tutoren 2012
- Photoshop and Illustrator CS5 Workshop 2011

Seminars and workshops

- Monday morning meeting presentations 2009-2013
- 2nd Winterschool of the Collaborative research Centre "Chromatin changes in differentiation and malignancies" Kleinwalsertal, Austria *(oral presentation)* 2012
- 1st Winterschool of the Collaborative research Centre "Chromatin changes in differentiation and malignancies" Kleinwalsertal, Austria *(oral presentation)* 2011
- 8th Winterschool of the International Graduiertenkolleg "Transcriptional control in developmental processes" Kleinwalsertal, Austria *(oral presentation)* 2010
- 7th Winterschool of the International Graduiertenkolleg "Transcriptional control in developmental processes" Kleinwalsertal, Austria *(oral presentation)* 2009
- 20th MGC PhD workshop, Luxembourg *(oral presentation)* 2013
- 18th MGC PhD workshop, Maastricht *(poster)* 2011
- 17th MGC PhD workshop, Cologne, Germany 2010
- MGC-Symposium, Leiden/Rotterdam 2009-2012

(Inter)national conferences

- 11th Dutch chromatin meeting, Rotterdam 2013
- Epigenetics and Chromatin, "Interactions and processes" Boston, U.S.A. *(poster)* 2013
- 10th Dutch chromatin meeting, Amsterdam 2012
- Cold Spring Harbor Meeting Asia, "Epigenetics, Chromatin and Transcription", Suzhou, China *(poster)* 2012
- Chromatin Symposium, "Chromatin changes in differentiation and malignancies" Giessen, Germany *(poster)* 2011
- 8th Dutch chromatin meeting, Leiden *(poster)* 2010

- Chromatin meeting “Chromatin and Epigenetics” Essen, Germany *(poster)* 2010
- EuTRACC 2nd Young Scientist Meeting, Dubrovnik, Croatia *(oral presentation)* 2010
- International Stem Cell Symposium, Amsterdam 2009

Supervising practicals and excursions, Tutoring

- Tutoraat 1e jaars Geneeskunde studenten 2012
- Junior Science Program high school students 2011

Acknowledgements

Gedurende mijn studie en promotie heb ik veel mensen leren kennen die mij op een of andere manier hebben geholpen. Ik wil graag iedereen hiervoor bedanken.

Als aller eerst wil ik mijn promotor prof. dr. **Frank Grosveld** bedanken voor de mogelijkheid om onderzoek te doen op zijn afdeling. Je commentaren tijdens de werkbesprekingen hebben me geleerd om kritisch naar mijn eigen werk te kijken.

Mijn co-promotor **Niels Galjart** wil ik graag bedanken met name voor de laatste ander half jaar toen ik echt onder jouw begeleiding viel en met jouw moest samenwerken. Je bent altijd zeer optimistisch (niet altijd even realistisch) en gaf mij veel vrijheid in het onderzoek. Bedankt voor de steun in de moeilijke periodes.

Als derde wil ik (mijn onofficieuze co-promotor) **Frank Sleutels** bedanken. Een groot deel wat ik van het labwerk, experimenten opzetten en daadwerkelijk onderzoek doen weet, heb ik van jouw geleerd. Ik kan me heel goed herinneren dat ik mijn aller eerste keer in het lab bij jou was en er vervolgens in geslaagd was om op de eerste dag je pipet te slopen! Sindsdien heb ik (gelukkig) veel geleerd en ik ben blij dat ik zowel mijn master stage als een deel van mijn PhD onder jouw begeleiding heb kunnen uitvoeren. Bedankt dat je ook de tijd wilde nemen om mijn proefschrift te corrigeren, de nodige tips en trics heb gegeven en in de grote commissie wilde plaatsnemen.

Hierbij wil ik de leden van de kleine commissie, prof.dr. **Joost Gribau**, prof.dr. **Bas van Steensel** en dr. **Willy Baarends**, bedanken voor de tijd om mijn proefschrift te becommentariëren. Het commentaar heeft zeker bijgedragen om er een beter stuk van te maken! Joost bedankt voor de commentaren en de discussiepunten en daarnaast ook bedankt voor het plaatsnemen in mijn eerstejaars commissie. Bas bedankt voor het commentaar en Willy heel erg bedankt voor het commentaar, met name op het onderdeel spermatogenese.

Naast de kleine commissie wil ik de leden van grote commissie, prof.dr. **Sjaak Phillipsen** en prof.dr. **Rainer Renkawitz** bedanken voor de tijd en moeite om hieraan deel te nemen. Sjaak bedankt voor het organiseren van de Kleinwalsertal meetings! Het was erg leuk om hier jaarlijks naar toe te gaan. Rainer, thank you for the nice collaboration and workdiscussions. I really appreciate that you are willing to come from Germany to join my PhD committee.

Natuurlijk wil ik het lab 10.30 bedanken voor alle leuke momenten! **Jeffrey**, ik ken weinig mensen die met zoveel passie en enthousiasme hun werk doen zoals jij. **Michael**, door de jaren heen heb ik veel surf/bodybuilding/dieet verhalen gehoord! Je bent heel behulpzaam en ik wens je veel succes. **Jessica**, we more or less started together on this crazy road. Thank you for your help, the nice chats and the friendship. One of the memorable things was the trip to America and our big struggle to order a decent ice tea ;). Good luck with your post doc and future career. **Umut**, it is great fun with you. Thank you for being my paranimf and most importantly for being a good friend. We had many nice chats, crazy Friday talks (with Jessica) and you gave me some nice writing tips. I wish you all the best with your career and family. **Kerstin**, thank you for your help, advice and discussion regarding CTCF. I wish you all the best with your career. **Kris**, it will take much more to convince me to pursue a career as neurosurgeon than only a monthy python sketch “Doctor, my brain hurts!”! Wish you and Athina all the best. **Dave**, bedankt voor alle leuke dingen, met name de movie nights samen met Linda en de CHAOS die jij af en toe brengt. Ik wens je veel succes in de toekomst. **Liu**, crazy camera guy thank you for the data analysis and I wish you all the best with your new career. **Linda**, veel succes met het behalen van je PhD. **Sreya**, Hritik Roshan is a big NO! Besides that it was great fun with you (especially the filmi discussion) and I wish you all the best with your career. **Luca**, my new office neighbor I wish you all the best with your PhD. (Master) students, **Lisa**, **Ralph**, **Alwin**, **Amr**, **Leila**, **Sophie** (bedankt voor je hulp met de tails), **Nuo** (crazy girl)

and **Lena** (thanks for designing my cover! It looks so good!) it was fun working with you guys. Although not member of the lab, **Martine**, thank you for the nice lab dinners!

Marek, I would like to thank you for the data analysis. It was nice working with you, especially with the CTCF zinc finger work and my apologies if I have bother you too much with my emails and questions. Good luck in the future. **Phillip**, thank you for the Kleinwalsertal CTCFL discussions. **Reinier**, bedankt voor het vele FACS sorten! Het heeft mij enorm geholpen met de vervolg experimenten. **Jos**, bedankt voor het isoleren van de testis en het maken van de testis extracten. **Wilfred**, **Mirjam**, **Selia** en **Christel** bedankt voor de sample prepping en het sequencen! **Marja**, bedankt voor de hulp met de IHC. **Ralph** en **Robert-Jan**, heel erg bedankt voor de hulp en adviezen om 3C experimenten op te zetten. **Adriaan**, **Maarten**, **Gert-Jan** en met name **Tsion** bedankt voor de hulp en adviezen met de confocal and spinning disk. Tsion die overnacht time-lapse filmpjes hebben we wel mooi gemaakt. Nu nog de analyse pfffff! **Petros** and **Tobias**, thank you for the contribution during the workdiscussions.

Thanks to all the colleagues from the 6th, 7th, 9th and 10th floor for the nice chats, fun and help. It would be difficult to mention everyone, but I will mention some of you. The members from our neighboring lab, **Robbert** (altijd vrolijk en behulpzaam), **Ana** en **Marjon** (altijd in voor een praatje), **Kim** (veel succes met je IPs), **Heleen** (altijd vrolijk), **Joshua** (good luck with finishing). **Fanny** (the usual phrase during the weekend "You again!"), **Alvin** (altijd lachen met jou), **Maureen** (testis adviezen), **Siska** (altijd behulpzaam), **Linde** (verhaal van de elfjes in ijsland is me toch bij gebleven), **Thomas** (the guy with the best dance moves), **Xiao** (good luck with your new job), **Romana** (altijd grappig) en **Lab 663** (voor de lekkere barbecue). Party committee members, **Celine**, **Friedemann** (sorry I confess I'm not able to learn you any sranang tongo;)), **Frederica** (no nonsense girl) and **Fabrizia** (great dance moves;)) thank you for all the fun and your help to organize all the nice things. If I didn't mention you here, blame it on the chaos in my head to finish this thesis! Thank you all.

Hierbij wil ik ook alle vrienden en familie leden bedanken voor alle support en leuke tijden. **Lalini**, bedankt voor de vele praatjes en dat ik je paranimf kon zijn. Ik heb behoorlijk om je moeten lachen. Bedankt voor je vriendschap! Veel succes met je co-schap, je carriere (als MDL arts of was het nou huisvrouw?) en natuurlijk met...huisje…boompje…tja dat beestje zie ik toch niet echt voor me. Uiteraard wil ik de **chikengunya's** bedanken voor alle plezier die we samen hebben beleefd en de support die we elkaar hebben gegeven in minder goede tijden. Ondanks onze drukke agenda's weten we wel tijd voor elkaar te maken en ik hoop dat we ook in de toekomst zo met elkaar kunnen blijven omgaan. **Natasja** (Natiiii), heel erg bedankt dat je me paranimf wilde zijn, je support, luisterend oor en natuurlijk voor je vriendschap! We zijn samen aan dit avontuur begonnen en het einde is bijna inzicht! Even doorzetten en op naar de co-schappen! Ga we nou wel of niet naar India!?!? ;) **Roqzana** (Roqzie/doksie/moksie…ow ja deze niet he), bedankt voor je steun, begrip en zachtaardigheid. Ik wens je heel veel succes met je carriere als jeugdarts en veel geluk met Ackbar. **Fatma** en **Erman**, jullie heel veel geluk samen toegewenst en met jullie carriere als huisarts and dermatoloog. **Busra**, power girl veel succes met je opleiding tot neuroloog. **Remy**, filmfreak, grappenmaker en onze aanstaande MDLer ik hoop dat je je doel gaat bereiken. Ik wens je veel geluk met Johan. De volgende keer dat ik kom eten iets minder gember gebruiken please :p! En ow jaaaaaaa **Saphiraaaaaaa**, je zit in Groningen maar we horen het meest van jou! Het is heel grappig/leuk om te zien hoe jij ontzettend blij kan worden van dode mensen en vol enthousiasme dit met ons wilt delen. Veel succes met je opleiding tot patholoog en een slumber party in Groningen moeten we nog doen.

Van de hele familie wil ik toch **Wikram mamoe** en **Usha mamie** bedanken voor alles! Jullie betekenen veel voor mij en staan altijd paraat zowel in goede als in slechte tijden. **Ranoe**, voor alle leuke uitstapjes die we hebben gemaakt en die we in de toekomst nog gaan maken ;). Ik wens jou en Bas veel geluk met de kleine.

Kishan, we hebben goede en helaas ook hele slechte tijden meegemaakt. We hebben bijna alles samen gedaan (geneeskunde, master, promoveren). Ik heb veel geleerd van jouw kijk op zaken. Heel erg bedankt voor alle hulp en steun die je me hebt gegeven. Ik hoop dat we de slechte tijden achter ons kunnen laten en aan een betere toekomst kunnen werken. Veel succes met het afronden van je PhD en verdere carriere.

Kishen, ik vind het heel tof van je dat je me paranimf wilt zijn of zoals jij dat zegt paranimfer (want dat moet cooler klinken!?). Af en toe heb ik mijn kleine broertje wel nodig ;). Bedankt voor alle gekkigheid die we samen hebben gedaan. Je steun in moeilijke tijden heb ik zeer gewaardeerd en ook je lichte opvattingen over sommige dingen (boeiend, waar maak je je druk over! ;)). Veel succes met je opleiding en met de zoektocht “wat wil ik nou precies doen”.

Papa en **mama**, ik heb niet genoeg woorden om jullie te bedanken voor de onvoorwaardelijke steun, vertrouwen en adviezen die jullie me door de jaren hebben meegegeven. Jullie hebben me altijd geleerd dat je op je eigen benen moet staan en je eigen weg moet kiezen. Kijk niet teveel naar een ander en doe vooral dat gene waarin je goed in bent en waarin jij gelukkig van wordt. Dat advies heb ik zoveel mogelijk proberen te volgen, wat niet altijd even makkelijk is. Dank je wel voor alles!

Liefs,

Widia