Touched by CTCF

Analysis of a Multi-Functional Zinc Finger Protein

Suzanne van de Nobelen

The work presented in this thesis was performed at the department of Cell Biology at the Erasmus MC in Rotterdam.
Printing of this thesis was supported by the J.E. Jurriaanse stichting.

Cover design: Alexander Pino
Cover photography: http://www.istockphoto.com

Printed by: PrintPartners Ipskamp, Enschede

Touched by CTCF

Analysis of a multi-functional zinc finger protein

Aangeraakt door CTCF

Analyse van een multifunctioneel zinkvinger eiwit

Proefschrift

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

Prof.dr. S.W.J. Lamberts

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 7 mei 2008 om 15:45 uur

door

Suzanne van de Nobelen

geboren te Gorinchem

Promotiecommissie

Promoter:	Prof.dr. F.G. Grosveld
Overige leden:	Prof. dr. C.P. Verrijzer Dr.ir. D.N. Meijer Prof. dr. J.N.J. Philipsen
Copromoter:	Dr.ir. N.J. Galjart

"We've discovered the secret of life… the genetic information in any organism is carried by nucleic acid, usually by DNA"

Francis Crick, James Watson, Maurice Wilkins; Nobel Prize 1962

Contents

List of abbreviations

3C	Chromosome Conformation Capture
4C	Chromosome Conformation Capture on ChIP
APS	Amplification promoting sequence
ATP	Adenosine triphosphate
B	*Mus. Musculus* Domesticus (C57/Bl6)
bp	Base pairs
C	*Mus. Musculus* Castaneus
cDNA	Complementary desoxyribonucleic acid
ChIP	Chromatin Iimmunoprecipitation
CpG	Cytosine-guanine
CTCF	CCCTC-binding factor
CTCFL	CTCF-like
DMR	Differential methylated region
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
DNase I	Deoxyribonucleic acid I
DRB	5,6-dichloro-1β-D-ribofuranosyl benzimidazole
E	Embryonic day
ES	Embryonic stem
FACS	Fluorescence-activated cell sorting
FISH	Fluorescent in situ hybridization
FISH	Fluorescent in situ hybridization
GFP	Green fluorescent protein
GST	Glutathione S-transferase
HAT	Histone acetyltransferase
HDAC	Histone deacetyltransferase
HIS	Histidine
HMG	High mobility group
HMT	Histone methyltransferase
HS	Hypersensitive site
ICR	Imprinting control region
IGS	Intergenic sequence
IP	ImmunoPrecipitation
Kb	Kilo base pairs
kD	Kilo dalton
LCR	Locus control region
MEF	Mouse embryonic fibroblast
mRNA	messenger RNA
P	Postnatal day
PAC	P1 artificial chromosome
PCR	Polymerase chain reaction
PIC	Pre-initiation complex
PRMT	Protein arginine methyltransferase
RNA	Ribonucleic acid
RNAi	Ribonucleic acid interference
RPA194	RNA polymerase I large subunit of 194 kD
rRNA	ribosomal RNA
S/MAR	Scaffold/matrix associated region
TF	Transcription factor

Scope of the thesis

Multicellular organisms contain a complete set of genes in nearly all of their cells. However most cells are very different to each other and are able to form organs with distinct functions. The identity and survival of the cell is regulated by the activity of specific genes in time and space. Specific sets of genes encoding proteins become activated, whereas others are repressed. CTCF is a protein that mediates distinct processes of gene regulation, including transcription and the structural organization of the genome.

The aim of this thesis is to investigate the different functions of CTCF by a combined analysis of CTCF-interacting proteins and by deletion of CTCF *in vivo* and *in vitro*. Using these approaches we aimed to improve our understanding of the molecular mechanism underlying its functions.

Chapter 1 gives an overview of the information required to understand the foundations of studies presented and discussed in this thesis. It gives an introduction to gene regulation and how this process is influenced by chromatin modifications, nuclear organization and compartmentation. A specific nuclear compartment, the nucleolus, and its involvement in ribosomal RNA synthesis, are highlighted. Furthermore the characteristics of CTCF and its homolog CTCFL are described in detail.

Chapter 2 describes the generation and characterizations of a mouse line in which the *Ctcf* gene can be conditionally deleted. The chapter describes the phenotype of T cells lacking CTCF. The data show that CTCF controls cell cycle progression of β-selected T-cells. We propose that CTCF regulates cell size by controlling the activation of key metabolic processes in these rapidly proliferating and growing cells.

Chapter 3 describes experiments that were performed to identify proteins that interact with CTCF. A mouse line was generated that can be used to identify CTCF-interacting partners in a cell-type and developmental specific manner by making use of a biotinylation tagging approach. After identification of these factors by pull down assays coupled to mass spectrometry additional verification experiments were performed. These experiments provide insight in to what factors bind to CTCF and for which function they are important.

Chapter 4 describes the role of CTCF in ribosomal RNA transcription. We show that UBF, a factor involved in this process, is a novel binding partner of CTCF. We further show that CTCF controls transcription from one of the promoters present on the ribosomal DNA locus, namely the spacer promoter and that CTCF controls binding of UBF, RNA polymerase I, and other important proteins to the region near the spacer promoter. We propose that CTCF and UBF function together in enhancing transcription from the spacer promoter.

Chapter 5 describes a system that can used to study the regulation of expression of imprinted genes in the absence of CTCF. The data show that expression of some imprinted genes is changed in the absence of CTCF. Although the function of CTCF in regulating the imprinted genes *Igf2* and *H19* has been extensively characterized, our data indicate that CTCF may have other roles as well.

Chapter 6 gives an overall discussion of the work carried out in this thesis. Hypotheses on the implications of this work are presented and future directions for the work are put forward.

1

Introduction

Chapter 1 Introduction

Development of a single fertilized oocyte into a fully functional organism requires the highly regulated processes of cell division, growth and differentiation. All the information needed for the formation and functioning of an organism is present in its genetic material, called DNA (deoxyribonucleic acid). The functional diversity among cells that form different organs is, among others, the result of a variable read-out of this genetic information at different stages during development. This occurs in a process called transcription.

A functional unit of DNA controlling a discrete hereditary characteristic is called a gene. Most genes contain information to make proteins. The transfer of information from DNA to protein is achieved by "reading" the genetic DNA-code, which consists of four different bases arranged in a unique order, and "translating" it into a protein in a series of complex steps. A protein is composed of amino acids, which, like the bases in the DNA, are arranged in a specific order. The amino acid sequence determines how a protein will fold and what activities it will have. Each protein has a unique function in a cell. Examples of proteins are hormones, enzymes, and antibodies.

Some proteins function to regulate the "read out" of genetic information, a process termed transcription of DNA. The transcription of genes during development must be tightly regulated in order to produce, at the right moment and place, the multitude of different gene products (for example, proteins) for different cell types and tissues. Abnormal gene expression can lead to defects that eventually result in diseases such as cancer. In this chapter an overview will be provided about the organization of DNA in the nucleus, the regulation of transcription and the factors that play a role in these highly fascinating processes.

1.1 Transcription and organization of the chromatin template

1.1.1 Transcription

Genes are transcribed by RNA polymerases, enzymes that catalyze the synthesis of RNA from a DNA template. Three different RNA polymerases are present in eukaryotic cells. These enzymes share some common subunits and many structural features, but they transcribe different types of genes. RNA polymerase I transcribes the genes for ribosomal RNA (rRNA), which is incorporated into ribosomes, the megadalton factories acting in the cytoplasm to produce proteins. RNA polymerase II transcribes the protein encoding genes into pre-messenger-RNA (mRNA), while RNA polymerase III transcribes the genes for transfer RNA (tRNA) and various small RNAs. Ribosomes catalyze the translation of mRNA into proteins. The tRNA molecules carry the necessary amino acids towards the ribosome in order to make a protein from mRNA. Thus, it is at the ribosome that the products of all three RNA polymerases meet.

Transcription by the three polymerases begins with their binding at promoters, DNA elements that are found directly upstream of the transcription start site of genes and that are typically about 40-50 base pairs in length. RNA polymerases cannot bind DNA directly. Each polymerase needs a different set of general transcription factors (TFI's, TFII's, and TFIII's, respectively) to help it bind DNA (Reese, 2003). Promoters are involved in recruiting these TFs, and in stabilizing and initiating transcription by RNA polymerase. In the case of RNA polymerase

II a preinitiation complex (PIC) is formed consisting of most of the TFII subunits. In general TFIID is the first PIC component to bind to its recognition site (the so-called TATA-box) via its integral TATA-box binding protein (TBP). This provides the first signal for the start of transcription.

Of the general transcription factors, TFIIH, contains helicase activity which is thought to be required for access at the transcriptional start point. Then, upon phosphorylation of its carboxy-terminal domain (CTD), RNA polymerase II undergoes a conformational change, is released from the PIC and begins transcribing a gene. Once the polymerase has begun elongating the RNA transcript most of the general transcription factors are released from the DNA. The phosphorylated tail of RNA polymerase II recruits factors that are important for elongation and for mRNA processing (Buratowski, 2003). Transcription can be broadly divided into transcription initiation, elongation and termination.

1.1.2 Transcription factors

Besides the general transcription factors for RNA polymerases, there is a wide variety of other transcription factors, which are defined as proteins that can activate or repress the rate of transcription by the RNA polymerases. The importance of transcription factors in the regulation of gene expression is highlighted by the fact that approximately 6% of genes in the genome code for these factors, making this one of the largest classes of proteins (Babu et al., 2004).

Most transcription factors bind DNA directly and are classified into families based on structural characteristics of their DNA binding domain. Examples of such domains are the zinc-finger (ZF) motif, the helix-loop-helix (HLH) region and the POU- and homeo-domains. These regions form a specific motif that recognizes DNA (Harrison, 1991). However, DNA binding motifs can also be used for RNA recognition or for protein-protein interactions. Transcription factors interact with other proteins forming functional complexes that are important for the regulation of activation or repression of particular genes. They can, for instance, bind the basal transcription machinery, and recruit this complex to the correct transcription initiation site. Moreover they can bind proteins that influence the structure of chromatin (see paragraph 1.1.7), thereby influencing the accessibility of DNA and in this way influence gene expression (Cosma et al., 1999).

The specificity of DNA recognition and binding by a transcription factor can be very high. The smallest change in the DNA or the amino acid sequence of its DNA binding domain will modify the binding efficiency of a transcription factor. However, there is also a set of transcription factors that function primarily as DNA-bending proteins. Most of these proteins are characterized by a DNA-binding element called the HMG box, which is less specific in binding. These proteins are not thought to activate transcription by directly interacting with the transcription apparatus. Rather, they are able to bend the DNA so that activators and repressors can be brought into contact with the promoter (Falvo et al., 1995).

1.1.3 Chromatin

Humans have 2-3 times as many genes as a fruit fly. However, their whole genome is 23 times as big. If stretched out the entire human DNA would have a length of about 2 meters, yet it is stored in a nucleus with a diameter as small as 3-10 µm. In order to fit the complete DNA into such a small volume DNA is folded and compacted by proteins in a structure called chromatin that, during mitosis, is even further compacted into a chromosome (Kornberg, 1974). When the information that is embedded into the DNA needs to be read, a cell has to know which genes

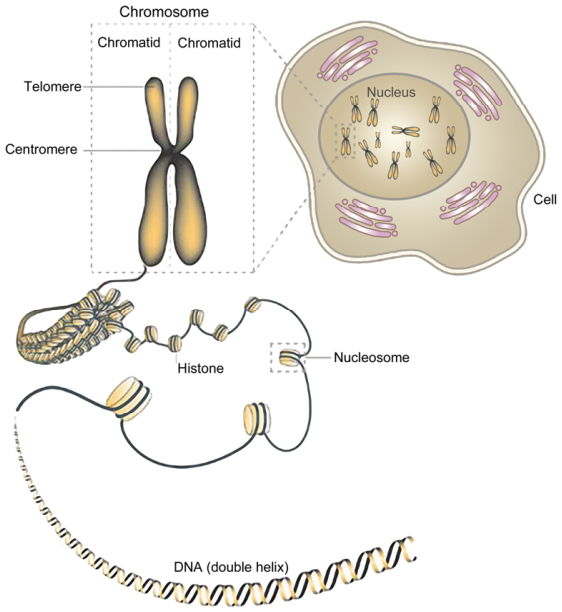

Figure 1.1 Organization of DNA in chromosomes
The organization of DNA in the nucleus of a eukaryotic cell. DNA is organized in the nucleus as chromosomes that compact the DNA more than 1000 fold. Unwrapping this compacted chromosome structure shows that the compacted state is a result of looping and higher order structures. The basis of these loops and structures is the (30 nm) chromatin fiber. This structure is the result of packaging of nucleosomes that consist of the naked DNA that is wrapped around histones. DNA the carrier of genetic information is present as two long entwined strands that forms a double helix structure.

need to be transcribed, and at what time. Here, the one-dimensional process of transcription (which involves reading a linear strand of DNA) is linked to the three-dimensional problem of DNA looping and compaction. Clearly, in order to understand transcription one also needs to understand the structure and composition of the highly compacted chromatin template.

In the first level of chromatin compaction, DNA is wrapped around a histone octamer, which consists of two copies each of histones H2A, H2B, H3 and H4. This so-called nucleosome is the fundamental packaging unit of chromatin. The nucleosome subsequently coils into what is termed a 30 nm fibre. Condensation into this structure is dependent on histone tails and is stabilized by H1 linker histones and other proteins (Fan et al., 2005). Further compaction into higher order chromatin structures is probably the result of self-association of these 30 nm fibers into a series of chromatin loops and coils (Bednar et al., 1998) (Figure 1.1) (see paragraph 1.2).

Chromatin can be broadly divided into two fractions: euchromatin, which is permissive for transcription, and heterochromatin, which is repressive. Heterochromatin itself occurs in two varieties, constitutive and facultative. DNA within constitutive heterochromatin is permanently silenced. Examples include centromeric regions and inactivated repetitive elements. In contrast, facultative heterochromatin is silenced only in certain contexts, for example throughout development or differentiation.

1.1.4 Gene regulatory regions

At the level of the DNA the proper spatial and temporal control of gene expression is achieved by the presence of *cis*-regulatory DNA sequences, which attract general and specific transcription factors. One example of a *cis*-acting element has already been mentioned above: the promoter. The simple model in which transcription is initiated from a single defined nucleotide position within a promoter is not correct, as genome wide characterization of core promoters revealed

that most genes have multiple promoters within which there are multiple start sites (Sandelin et al., 2007).

The TATA-box was thought to be important for transcription factor promoter binding, however it was shown that this element is present in only 10-15% of the human core promoters (Kim et al., 2005). Based on these observations it was proposed that the TATA-box is not a general promoter motif but that there is considerable diversity in core promoter structure and function. Examples of other DNA elements that contribute to core promoter activity and specificity include initiator (INR) and downstream core promoter elements (DPE). The latter regions were shown to often overlap with G/C rich sequences (CpG islands) (Yang et al., 2007).

Besides promoters, there are other *cis*-acting elements, including enhancers, silencers, locus control regions (LCR) and chromatin insulators. A common feature of these elements is that they can act at a distance, whereas a promoter cannot. Enhancers are elements capable of enhancing the basal transcription of a linked promoter. Enhancers were shown to differ in their strength of enhancement and span 200-1000 bp (Arnosti and Kulkarni, 2005). One of the first characterized enhancers was the SV40 repeat element. Transient expression experiments showed that this element could dramatically enhance transcription of a linked β-globin gene (Banerji et al., 1981). The increase in transcription was due to the recruitment of transcription factors, which enhanced the binding of RNA polymerase and thereby increased the level of transcription. Enhancers cannot only influence the level of transcription but are also involved in antagonizing gene silencing (Martin, 2001). Further studies on enhancer function have shown that they can act over long distances in an orientation independent manner. Since enhancers can be some distance away from a gene, they are thought to act by looping out the intervening DNA, forming protein interactions with promoter bound factors. Silencers act in a similar manner as enhancers, except that they are involved in repression of transcription.

The properties of a locus control region (LCR) are best understood in the context of experiments with the human β-globin locus and its linked LCR. When stably inserted into the genome of transgenic mice, it was shown that the LCR shielded the human β-globin locus (and other transgenic constructs) from position effects that are due to the random integration into the host genome. Besides this position-independent activity it was shown that the LCR has the ability to fully activate a linked gene in a tissue-specific and copy-number-dependent manner (Grosveld et al., 1987). Thus, an LCR differs from an enhancer in that it confers position-independent, copy-number dependent activity to a linked gene. Since the discovery of the β-globin LCR, other domains with LCR-like activity have been characterized in multiple loci, including the human growth hormone (GH) locus and in the Rad50 gene of the T-helper cell 2 (Th2) cytokine locus (Ho et al., 2006; Lee et al., 2003).

In contrast to enhancers, LCRs regulate transcription of linked genes not only if they are placed in euchromatic regions but also if they are present in heterochromatin regions (Dean, 2006). However the exact role of the LCR in the establishment or maintenance of an active chromatin environment is unclear and contradictory data exist between the human and mouse β-globin locus (Forrester et al., 1990; Schubeler et al., 2001). Deletion of the LCR from the mouse genome by homologous recombination tested its function at the endogenous location. β-globin expression was strongly reduced and a major effect on RNA polymerase II elongation at the β-globin gene could be observed (Bender et al., 2000; Epner et al., 1998) Besides its strong transcriptional enhancer activity the β-globin LCR also includes enhancer blocking/insulator abilities (Chung et al., 1993).

A consequence of DNA compaction is an increased resistance to enzymes that act upon the DNA, such as DNAse I, a non-sequence specific nuclease. In contrast, gene regulatory

regions have are generally more accessible to this enzyme, i.e. they are DNAse I "sensitive" or even "hypersensitive". This is because these regions are less occupied by histones and more by transcription and other factors (Szutorisz et al., 2005). The globin LCR, for example, contains multiple DNase I hypersensitive sites (HSS). Recently a genome wide analysis was performed in order to map ubiquitous and cell type specific regulatory elements by combining DNaseI hypersensitive site mapping, expression arrays and chromatin immunoprecipitation followed by hybridization to tiled arrays (ChIP-chip) (Xi et al., 2007). Using this approach it was shown that nearly all of the ubiquitous DNaseI hypersensitive sites correspond to either promoters, enhancers or insulator elements.

1.1.5 Chromatin insulators

Insulator elements are defined as elements that serve to restrict the action of regulatory elements to appropriate genomic targets. Insulators often demarcate independently regulated domains of eukaryotic genomes. They have been proposed to have enhancer-blocking and/or boundary functions. Enhancer blocking is the ability to inhibit enhancer-stimulated gene expression when the insulator is placed between an enhancer and promoter element. The boundary function is defined as the ability to counteract chromosomal position effects, for example preventing the spreading of heterochromatin. Although these functions should be seen as separable activities insulators may harbor both properties (Recillas-Targa et al., 2002).

Much of the early work that defined the properties of insulators was carried out in *Drosophila*. The first DNA sequences to be described as having properties of an insulator were the scs and scs' elements in *Drosophila* that were initially identified as marking the chromatin boundaries of a heat shock locus (Udvardy et al., 1985). When scs elements are placed on either side of a gene for eye color and this gene is introduced into *Drosophila*, the resulting flies all have similar eye color independent of the transgene's site of integration, an indication that scs has protected the reporter gene from both negative and positive endogenous influences or position effects (Kellum and Schedl, 1991). Moreover it was shown that the element could block enhancers from activating a promoter (Kellum and Schedl, 1992). Sequences that can act as insulators were subsequently identified and studied in multiple organisms (Bell et al., 2001; West et al., 2002). In fission yeast boundary elements have been isolated from the regions flanking the silent mating loci (Noma et al., 2001). In *Drosophila*, five different insulator binding proteins have been identified, Zw5, BEAF-32, GAGA factor, Su(Hw) and dCTCF while in vertebrates the only known insulator with enhancer-blocking abilities is CTCF. The observation that in *Drosophila* multiple insulator factors are present is likely due to its compact but largely euchromatic genome.

Evidence suggesting that insulators play a role in the regulation of higher-order chromatin structure has been provided, in part, by the analysis of the gypsy insulator. The gypsy insulator confers its activity through its associated protein complex consisting of Su(Hw), Mod(mdg4)2.2 and CP190. The ubiquitin ligase dTopors can also associate with this insulator complex and is important for gypsy insulator function because it mediates association of the complex to the nuclear lamina at the periphery of the nucleus (Capelson and Corces, 2005). In diploid cells Su(Hw), Mod(mdg4)2.2, and CP190 insulator proteins co localize at 20–25 large foci. These foci represent co localization of gypsy elements placed on different chromosomes. These complexes of gypsy insulators are called 'insulator bodies' (Gerasimova et al., 2000; Gerasimova and Corces, 1998). They are formed via association of multiple insulators and localize to the nuclear matrix (Nabirochkin et al., 1998; Pathak et al., 2007). The finding of insulator bodies implied a role

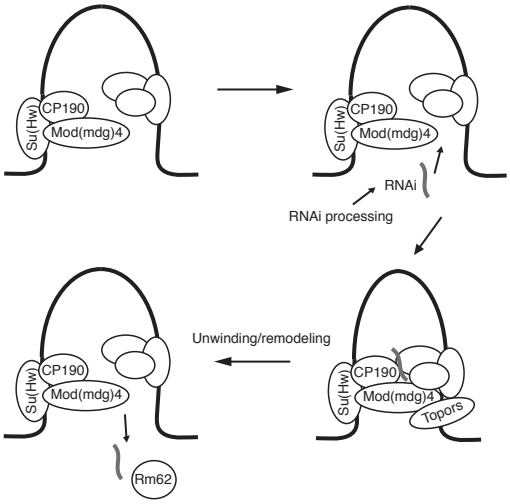

Figure 1.2 Loop formation stabilized by proteins and RNA bound to *gypsy* sites in *Drosophila*
For clarity, one set of interactions leading to the formation of a single loop is shown. Clusters of such sites form insulator bodies *in vivo.* The proteins recruited to the gypsy insulator include Su(Hw), CP190, Mod(mdg4)2.2 and Topors. Topors is an ubiquitin ligase, which associates with the nuclear lamina and the insulator complex and is required for insulator function. Loop formation and insulator activity also involve RNA and are dependent on members of the RNAi processing pathway. Loop formation can be interfered by the RNA-binding protein Rm26. The protein CP190 is additionally involved in insulator function at sites that do not contain Su(Hw) (Adapted from (Wallace and Felsenfeld, 2007))

for insulators in organizing chromatin, but did not prove a role for these elements in creating functionally independent chromatin domains. The discovery of gypsy insulator bypass helped to solidify this role. Using the enhancer-blocking assay, the gypsy insulator was shown to have the ability to block an enhancer from driving transcription of a promoter when placed in between two elements. Two independent groups found that placing two gypsy insulators between the enhancer and promoter restores enhancer-promoter communication (Cai and Shen, 2001; Muravyova et al., 2001). This implied that two insulators interacting to 'loop-out' specific sections of DNA might create independent chromatin domains.

Recently it was shown that RNA interference (RNAi), a conserved gene silencing mechanism that causes the degradation of specific RNA molecules or hinders the transcription of specific genes, is involved in regulating the gypsy insulator. It was shown that Insulator activity is decreased when Argonaute genes required for RNAi are mutated. An improvement of insulator function can be detected when the levels of the Rm62 RNA helicase, which is likely involved in unwinding or remodeling of RNA-insulator protein complexes, are reduced (Lei and Corces, 2006) (Figure 1.2).

1.1.6 Epigenetic modifications

Posttranslational modifications of histones, incorporation of histone variants and methylation of DNA can result in (heritable) changes in gene expression without changes in the actual genetic code. These so-called "epigenetic modifications" can alter nucleosome structure and accessibility of the DNA and thereby generate specific domains of chromatin in the genome (Jin et al., 2005). Specific epigenetically regulated processes include genomic imprinting and X chromosome inactivation. Genomic imprinting is a phenomenon whereby specific genes are expressed either from the allele inherited from the mother or from the allele inherited from the father. Imprints can be erased and re-established through each generation but not through cell division (Reik and Lewis, 2005).

Histones and histone modifications

The four 'core' histones (H2A, H2B, H3 and H4) are relatively similar in structure and are highly conserved throughout evolution. The nucleosome core is formed of two H2A-H2B dimers and a H3-H4 tetramer, forming two nearly symmetrical halves. This octamer of core histones represents a 108 kD protein complex around which 147 base pairs can be wrapped to form a nucleosome.

Each of the core histones has a long N-terminal amino acid tail, which extends out from the DNA-histone core. These histone tails are subject to several different types of covalent modifications, including acetylation, methylation, ubiquitination and sumoylation of lysines, and phosphorylation of serines (Kouzarides, 2007). Histone modifications control many aspects of chromatin structure. For example, these alterations regulate the accessibility of DNA for transcription factors, and create recognition signals to anchor protein complexes such as chromatin remodeling complexes (see paragraph 1.1.7).

Histone acetyl transferases (HATs) add acetyl groups to histone tails while histone deacetylases (HDACs) are involved in removing them. Modifications of the tails have little direct effect on the stability of an individual nucleosome but they seem to affect the stability of the 30-nm chromatin fiber and of higher-order structures. Acetylation of chromatin tails is generally linked to gene expression and histone deposition. By adding the acetyl group the positive charge of the lysine is removed making it more difficult to neutralize the charges on DNA as chromatin is compacted. Methylation of histone tails can affect gene expression depending on the residue that is methylated and the number of methylgroups that are present on the amino acid.

In general transcriptionally active regions are associated with methylation of lysine 4 of histone 3 (H3K4) and acetylation of histones H3 and H4. By contrast, inactive regions are marked by methylation of lysine 9 of histone 3 (H3K9) and histone hypoacetylation (Dillon and Festenstein, 2002) (Table 1.1). H3K9 methylation is mediated by the histone methyltransferase Suv39h and provides a high-affinity binding site for heterochromatin protein 1 (HP1). Suv39h/HP1 mediated repression is mainly associated with constitutive heterochromatin, but might also be involved in gene repression in euchromatic regions (Nielsen et al., 2001). Long term silencing

DNA modification	**Site of modification**	**Transcriptional role**
Methylated cytosine (meC)	CpG (but not CpG islands)	Repression
Histone post translational modification	**Site of modification**	**Transcriptional role**
Acetylated lysine (KAc)	H3(9,14,18,56), H4(5,8,13,16), H2A, H2B	Activation Repression
Phosphorylated serine/threonine (S/Tph)	H3(3,10,28), H2A, H2B	Activation
Methylated arginine (Rme)	H3(17,23), H4(3)	Activation
Methylated lysine (Kme)	H3(4,36,79) H3(9,27),H4(20)	Activation Repression
Ubiquitylated lysine (Kub)	H2B(123*/120#) H2A(119#)	Activation Repression
Sumoylated lysine (Ksu)	H2B(6/7), H2A(126)	Repression
Isomerized proline (Pisom)	H3(30-38)	Activation/Repression

Table 1.1 Chromatin modifications

DNA modifications and histone post-translational modifications and their role in transcriptional activation or repression.
* yeast: *S. Cerevisae*, # mammals (Adapted from (Li et al., 2007))

by polycomb group proteins is associated with H3K27 methylation and H2A ubiquitination (Cao et al., 2002; Wang et al., 2004), indicating a role for histone modifications in cellular memory and stable maintenance of cell fates throughout development. However, histone modifications are dynamic and their mode of inheritance through cell division remains to be defined.

Multiple other methyltransferases have been identified that are involved in methylating lysine residues. Also the group of histone demethylases that can remove methylated marks from lysines is growing rapidly. Histone modifying enzymes need specialized protein domains to bind to histones such as the chromodomain, WD40-repeat and tudor domain (Flanagan et al., 2005; Huang et al., 2006; Schuetz et al., 2006).

Recent observations suggest that in many cases there is not a strict division between active and repressive modification states. Modifications are not a simple code but transcription occurs against a mixture of complex modifications, which probably have several roles. The location of the modification and the combination with other modifications must be considered in order to understand its biological meaning (Berger, 2007). For example in embryonic stem cells there are unusual chromatin domains that include both H3K4 trimethylation and H3K27 trimethylation, a combination of what was thought to be positive and negative marks. These bivalent domains were shown to correlate with locations for genes encoding developmentally important transcription factors. These observations suggest that such developmental genes are transcriptionally silenced but poised for activation during differentiation (Bernstein et al., 2006). Recently it was shown that ubiquitination of H2A is the crucial modification involved in maintaining the poised state of this subset of genes (Stock et al., 2007).

Histone variants

Histones that are homologous to the standard histones but also have their own feature that is distinct from the major histones, are called variant histones. These variant histones can substitute for standard histones and modify nucleosomes to carry out specific functions. All the conventional histones, except H4, have a variant counterpart. Some variants like H3.3, H2A.X and H2A.Z are not considerably different from the main forms. Two variants, macroH2A1 and macroH2A2, have however a low level of homology with main histone forms (Henikoff et al., 2004).

Histone variants are distinguished from canonical core histones mainly by the fact that they are synthesized outside the S-phase and incorporated into chromatin in a DNA replication-independent manner (Kamakaka and Biggins, 2005). Variant histones can differ from canonical histones in their tails, fold domains, or in only a few key amino acid residues (Doyen et al., 2006; Henikoff and Ahmad, 2005).

Incorporation of histone variants can impact chromatin structure in various ways. H3.3 and H2A.Z, for example, exert an effect by affecting nucleosome stability. Promoters and enhancers of transcriptionally active genes and coding regions of highly expressed genes have nucleosomes that carry both H3.3 and H2A.Z and are therefore extremely sensitive to disruption (Jin and Felsenfeld, 2007). H2A.Z is mainly localized in promoter regions and boundaries and is not only associated with activation but also with repression (Jin et al., 2005). The centromere incorporates a specific H3 variant called CENP-A, which associates with centromeric heterochromatin (Smith, 2002). Centromere structure and function is blocked if H3 cannot be substituted for CENP-A in this specific region (Howman et al., 2000). How CENP-A is specifically loaded onto centromeres is not known. The histone variant macroH2A is concentrated on one of the two copies of the X chromosome present in female mammals, namely on the inactive X chromosome (Changolkar and Pehrson, 2006). Thereby this histone variant was suggested to be involved in silencing of one X chromosome in female mammals for reasons of dosage compensation (X inactivation). Later

it was shown that macroH2A is not important for silencing the X chromosome but is involved in fine-tuning the expression of specific genes (Changolkar et al., 2007). Another H2A variant H2A.X binds to DNA with double strand breaks and marks regions that undergo DNA repair (Redon et al., 2002).

Accessibility of *cis*-regulatory elements is not only characterized by nuclease sensitivity and nucleosomes depletion, but also by a high frequency of histone replacement (Mito et al., 2007). The rapid H3 replacement with H3.3 at *Drosophila* boundary-associated-regions suggests that constant replacement of nucleosomes serves to erase laterally spreading chromatin domains. Recently the epigenetic memory of an active gene state in the absence of transcription was shown to depend on H3.3 incorporation (Ng and Gurdon, 2008). This suggests that histone variants are important for epigenetic regulation and memory and might help to stabilize gene expression in normal development.

Histone variants can be exchanged by chaperones (like HIRA, FACT, CAF1) or by ATP-dependent nucleosome remodeling complexes like SWI/SNF and ISWI (Belotserkovskaya et al., 2003; Bruno et al., 2003). Also the cooperative binding of transcription factors and actively transcribing RNA polymerase II can mediate histone exchange (Kireeva et al., 2005; Lorch et al., 2001; Workman, 2006). How the complexes that facilitate incorporation of histone variants recognize where to perform their activity is not well known. Perhaps transcription factors or histone modifications provide positional clues. The functional interplay between variant histone incorporation and histone modifications is an important area of research that needs to be further explored.

DNA methylation

DNA methylation is a modification that is associated with stable gene silencing (X inactivation, genomic imprinting) through interference with transcription factor binding or through recruitment of repressors that specifically bind methylated DNA. DNA methylation is a more stably maintained chromatin mark than the reversible histone modifications mentioned above. In mammals, DNA methylation occurs upon cytosines within CpG dinucleotide residues. Between 60-70% of all CpGs present in the genome is methylated. CpGs are often grouped in clusters called 'CpG islands' that are present in the 5' regulatory regions of many genes. However cytosine methylation occurring at CpA and CpT dinucleotides is also reported in embryonic stem (ES) cells (Haines et al., 2001).

The enzymes that catalyze the methylation of CpG's are called DNA methyltransferases (Dnmt's). These enzymes both generate new methylation marks and maintain (and restore) existing DNA methylation patterns after cell division. Dnmt1 is a maintenance methyltransferase while Dnmt3b and 3a are *de novo* methyltransferases (Bird and Wolffe, 1999). After DNA replication DNA sequences are methylated on only one of the two DNA strands (hemimethylated). Dnmt1 recognizes hemimethylated CpG methylation patterns and methylates the opposite strand. In this way Dnmt1 stably transmits methylation patterns after each cell division. Although Dnmt1 can carry out *de novo* methylation *in vitro*, there is little evidence that it can do this *in vivo* by itself. Inactivation of Dnmt1 in mice causes global loss of methylation and bi-allelic expression or silencing of imprinted genes (Howell et al., 2001). DNA methylation is a primary imprint signal for imprinted gene clusters but between those clusters different mechanisms are used to establish and maintain this mark (Edwards and Ferguson-Smith, 2007). Dnmt3a and Dnmt3b are essential for *de novo* DNA methylation, mainly acting in embryonic stem cells and early postimplantation embryos (Okano et al., 1999). Knockout mouse models of Dnmt3a and Dnmt3L have shown that both proteins are necessary for appropriate establishment of maternal and paternal imprinting. Dnmt3b may participate in methylation of specific imprinted loci but is not absolutely essential

(Hata et al., 2002; Kaneda et al., 2004; Webster et al., 2005).

DNA methylation and histone modifications are able to influence each other (Fuks, 2005). For example, methyl-CpG binding domain proteins (MBDs) can recruit histone deacetylases (HDACs) and other chromatin remodeling proteins that can modify histones, thereby forming a compact inactive chromatin state (i.e. methylated DNA and modified histones). Moreover histone deacetylation and methylation at lysine 9 of H3 contribute to the establishment of DNA methylation patterns. Dnmt's and histone deacetylase (HDAC) can bind to each other and connect the two processes. The combination of the two modifications could be relevant to situations were a strong stable epigenetic state is crucial (Burgers et al., 2002). In embryonic stem cells that lack both Dnmt3a and 3b global levels of H3 acetylation and methylation are altered and the mobility of linker histones is decreased. These changes, however, do not have effects on compaction of chromatin but they do affect nuclear organization and nucleosome structure (Gilbert et al., 2007).

DNA methylation is involved in propagating the state of activity of particular genes. Beside DNA methylation another important system involved in memorizing gene expression patterns of developmentally important genes are the polycomb (PcG/Trx) group proteins (Bird, 2002). Polycomb and trithorax group of proteins work to maintain repressed or active transcriptional states respectively. The mechanism by which silencing or activation is transmitted between cell generations remains obscure.

1.1.7 Chromatin remodeling complexes

Modification of histone tails can result in decondensation of chromatin in the regions of promoters, but the modification cannot disrupt the structure of the nucleosome. This requires so-called "chromatin-remodeling complexes", which can displace nucleosomes from enhancer and promoter regions. Chromatin-remodeling complexes utilize energy of ATP hydrolysis to displace or exchange histones and in this way they change structure and position of nucleosomes temporarily (Saha et al., 2006; Workman, 2006). This results in DNA becoming less tightly bound to the histone core and being more accessible for transcription factors (Workman and Kingston, 1998).

Three chromatin-remodeling complexes are conserved from yeast to human: SWI/SNF, ISWI and a complex called Mi-2/CHD or NuRD (Eisen et al., 1995). The INO80/SWR chromatin-remodeling complex is present only in yeast (Eberharter and Becker, 2004; Narlikar et al., 2002). All of the ATP-dependent chromatin-remodeling complexes contain an ATPase subunit that belongs to the SNF2 super family of proteins. Remodeling enzymes contains specific histone binding domains so that they can be targeted to specific modified histones. For example, SWI/SNF family can bind acetylated histones with their bromo-domain, ISWI has a SANT domain to bind modified histones and the CHD complex binds methylated histones with its chromo-domain (Boyer et al., 2004; Flanagan et al., 2005; Hassan et al., 2002). While the SWI/SNF proteins disrupt DNA-histone contacts and release the histones and disassemble complete nucleosomes, the ISWI family member NURF merely slides the nucleosomes along the DNA (Hamiche et al., 1999; Langst et al., 1999).

The relationship between SWI/SNF complex and chromatin became apparent when the complex was purified and found to alter nucleosome structure in an ATP-dependent manner (Cote et al., 1994; Peterson and Herskowitz, 1992). Several mutations that suppressed SWI/SNF phenotypes appeared to correspond to genes encoding histones and other chromatin proteins

(Kruger et al., 1995; Recht and Osley, 1999). The SWI/SNF family contains, besides SWI/SNF, Brm and BRG1, also actin-related proteins. It has been suggested that these proteins link remodeling complexes to actin-binding proteins, to components of the nuclear matrix, or to chromatin itself (Chen and Shen, 2007).

The second group of ATP-dependent remodeling complexes contains the ISWI protein as the ATPase subunit. Using biochemical methods the most extensively studied members of this group (ACF, NURF and CHRAC) were originally purified from *Drosophila* extracts based on their ability to disrupt and generate regularly spaced nucleosomal arrays. The NuRD complex was shown to possess both chromatin remodeling and deacetylase activities (Tong et al., 1998; Xue et al., 1998). However, the question of how the chromatin-remodeling complexes and HDACs/HATs might act in concert to activate transcription still remains unanswered (Hogan and Varga-Weisz, 2007).

Remodeling complexes are recruited to promoters via interactions with sequence-specific transcription factors or through their association with RNA polymerase II. Not all genes require the function of a chromatin-remodeling complex for full activation. The strength of a particular promoter might play a role in its dependence on chromatin-modifying complexes. Thus, a weak promoter may require the complex for full activity, while a strong promoter may not. Another possibility is that remodeling complexes are only required for the transcription of promoters that possess positioned nucleosomes, which leave transcription factor sites unavailable for binding (Burns and Peterson, 1997).

Some important questions that remain to be addressed in this field include the extent of modifications that remodeling complexes exert on a given gene. It will be interesting to determine whether only the nucleosomes that are positioned on the promoter are remodeled or whether these modifications extend beyond the regulatory region, maybe also facilitating elongation by the RNA polymerase. Another important aspect is the effect that ATP-dependent remodeling complexes might have on higher-order chromatin structure (Vignali et al., 2000).

Besides recruiting transcription factors both enhancers and locus control regions (LCRs) can recruit complexes that carry out alterations in chromatin structure such as nucleosome remodeling and posttranslational modifications of histones. By altering the chromatin structure , the accessibility of DNA for transcription can be regulated. The order of recruitment of chromatin remodeling factors, transcriptional activators, general transcription factors and RNA polymerase II is not absolute an can vary depending on the gene that is activated (Agalioti et al., 2000).

1.2 Nuclear architecture

1.2.1 Chromatin organization in nuclear space

The subnuclear localization of chromatin is not random, specific genetic loci or whole chromosomes reside in specific locations within the nucleus (Cremer and Cremer, 2001). In interphase each chromosome tends to occupy a discrete and relatively small territory in the nucleus. Chromosomes are shown to have not only preferred positions within the nucleus but also with respect to each other (Parada and Misteli, 2002)(Figure 1.3). Changes in transcriptional activity are often coupled with changes in subnuclear localization of chromosomes (Chambeyron and Bickmore, 2004). Transcriptional regulatory elements such as LCRs, enhancers and insulators appear to act by

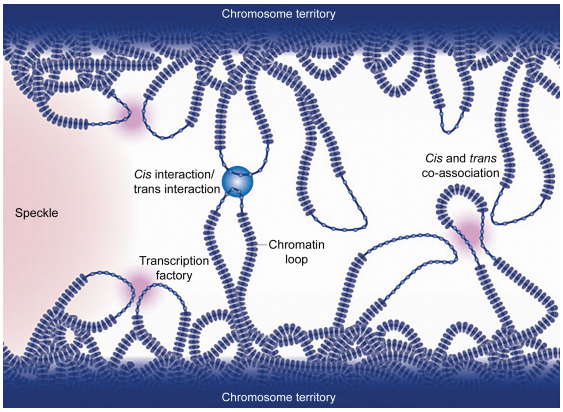

Figure 1.3 Nuclear organization of the genome
Colocalization of genes in the nucleus for expression or coregulation. Active genes on decondensed chromatin loops that extend outside territories can co localize both *in cis* and *in trans* at sites in the nucleus with local concentrations of RNA polymerase II and adjacent to splicing-factor enriched speckles. Interactions can also occur between regulatory element and/or gene loci and lead to co-regulation *in trans* (Adapted from (Fraser and Bickmore, 2007))

repositioning specific genetic loci to regions or loops with active and silent transcription (Ragoczy et al., 2003). Multiple active genes and gene clusters are often located together at places in the nucleus that have a high local concentration of the transcriptional and mRNA processing machinery. These sites are called “transcription factory” (Faro-Trindade and Cook, 2006). In a transcription factory individual genes that are widely separated on a chromosome, or even situated on different chromosomes, are brought in close proximity (Osborne et al., 2004).

Fluorescent in situ hybridization (FISH) and chromosome conformation capture (3C) are two main techniques used to understand how the genome is organized in the nucleus. In 3C, formaldehyde is used to cross-link DNA fragments that are close together in nuclear space. Cross-linked chromatin is subsequently digested with an excess of restriction enzyme. DNA ends are ligated under conditions that favor junctions between cross-linked DNA fragments, cross-links are reversed and the frequency of interaction between chromatin segments can be determined by semi-quantitative PCR (Dekker et al., 2002; Tolhuis et al., 2002). Using 3C, *cis*-regulatory elements of the β-globin locus were shown to form an erythroid specific spatial conformation (Palstra et al., 2003).

The recently developed chromosome conformation capture-on-chip (4C) technique can be used to identify not only intra- but also interchromosomal interactions. In contradiction to 3C, it allows for a high throughput genome wide analysis in an unbiased manner. In this way all DNA segments that physically interact with a DNA fragment of choice can be mapped. Using this technique it was demonstrated that both active and inactive genomic regions could interact over long distances with many loci in nuclear space. Active genes were shown to preferentially cluster with active genomic regions, but this is not absolute. Moreover 4C analysis shows that intrachromosomal interactions are favored over interchromosomal interactions (Simonis et al., 2006; Zhao et al., 2006).

For the β-globin locus it was determined that regions interacting with it in erythroid cells do not preferentially contain erythroid-specific genes (Simonis et al., 2006). Thus, one cannot predict locus interactions based on function. Interactions can also differ depending on species or cell-type. For example, the active α- and β-globin genes are located on different chromosomes and often come together in human, but not in mouse, erythroid cells. These data show that spatial proximity between these two loci is not conserved and therefore this interaction in humans is not expected to be functionally important for α- and β-globin gene regulation (Brown et al., 2006).

The T-helper cell 2 cytokine locus and interferon-γ genes were shown to contact each other (Spilianakis et al., 2005). Subsequently interchromosomal interactions were also detected

of the olfactory-receptor locus, *H19/Igf2* locus and HoxB1 locus (Lomvardas et al., 2006; Wurtele and Chartrand, 2006; Zhao et al., 2006). The relevance of these interactions is difficult to interpret since the characteristics of the associations were not always addressed in detail. In the case of the olfactory receptor locus, the initially reported chromosomal contact, proposed to be involved in regulating gene expression, were questioned by additional experiments (Fuss et al., 2007).

One question is whether the interactions detected by 4C originate from multiple dynamic interactions that occur transiently in almost all cells, or whether they reflect different interactions that occur from cell to cell and are just an average sampling of the population. Therefore, it will be important to clarify the cell-type, developmental and functional consequences of interchromosomal interaction. A combination with high-throughput FISH could be used to verify interchromosomal interactions.

1.2.2 Nuclear matrix

The nuclear matrix or scaffold has been defined as the insoluble non-DNA material left in the nucleus after a series of biochemical extraction steps. Transmission electron microscopy showed the nuclear matrix to be a network of protein fibers that extends throughout the nucleus and connects to the cytoskeleton at the nuclear envelope (Capco et al., 1982; Long et al., 1979). Some of the proteins that constitute the nuclear matrix were shown to bind specific DNA sequences called scaffold-associated- or matrix-associated regions (SAR/ MAR). These DNA sequences have been postulated to form the base of chromosomal loops and are implicated in attaching them to the nuclear envelope or other structures in the nucleus. MARs or SARs usually contain ATC sequences with AT-repeats (Izaurralde et al., 1989). Thus, the nuclear matrix might help to organize chromosomes, localize genes, and regulate gene expression and DNA replication by S/ MARs (Pemov et al., 1998). Experiments with isolated nuclear halos suggested that S/MARs are necessary but not sufficient for formation of chromatin loops (Heng et al., 2004; Iarovaia et al., 2005).

Several proteins such as lamins, nucleolar proteins, topoisomerase II and histone H1 were reported to be components of the nuclear matrix (Calikowski and Meier, 2006). Also proteins involved in DNA replication such as DNA polymerase, primases and PCNA have been classified as matrix-attached (Mika and Rost, 2005). A protein called SATB1 was originally identified as a protein that binds to MARs of the immunoglobulin heavy chain intronic enhancer (Dickinson et al., 1992). Later it was shown to be the first MAR-binding protein that is important for global regulation of expression of multiple genes (Alvarez et al., 2000). Several recent lines of evidence suggested a role for SATB1 in the formation of chromatin loops. A transcriptionally active chromatin structure is formed at the cytokine locus upon T-helper 2 (Th2) activation and subsequent induction of SATB1 expression. This structure is composed of numerous small and folded chromatin loops all anchored to SATB1 at their base. Moreover it was shown that upon knock-down of SATB1 not only chromatin loops are lost but that also the expression of interleukin genes present in the cytokine locus drops (Cai et al., 2006). The role of SATB1 in organizing higher-order chromatin-loop structures is directly coupled to its ability to tether MARs to the nuclear matrix associated PML bodies (Galande et al., 2007; Kumar et al., 2007).

Visualization in live cells of proteins that are supposed to form the nuclear matrix has never shown a structure of fibers. The existence of a matrix structure is therefore still controversial and, by some, claimed to be an artifact (Pederson, 2000). The degree in which gene expression depends on nonchromatin nuclear structure therefore remains an intriguing but unsolved issue.

1.2.3 Nuclear bodies

Although the nucleolus (see chapter 1.3) is the most prominent structure in the nucleus, several other nuclear bodies have been visualized and studied. These include the well-characterized Cajal bodies, nuclear speckles and promyelocytic leukemia nuclear bodies (PML) (Handwerger and Gall, 2006). In general, nuclear bodies organize the delivery and storage of essential RNAs and proteins that play a role in transcription, pre-mRNA biosynthesis and splicing (Figure 1.4). Moreover they are involved in sequestering and degradation of certain regulatory proteins (Zimber et al., 2004). Posttranscriptional modification of RNAs and of proteins is emerging as an important regulatory function of subnuclear organelles. These modifications can dramatically change the localization of molecules and thereby modify the structure and function of entire organelles. Despite the marked progress in characterization of these RNA and protein modifications, many questions remain about the enzymes and substrates that participate in these events.

Nuclear speckles contain high concentrations of pre-mRNA splicing factors and small nuclear ribonucleoprotein particles (snRNPs). In mammalian cells, the composition and intranuclear localization of speckles responds to changes in mRNA transcription and protein phosphorylation (Lamond and Spector, 2003). Speckles might serve not only as a reservoir of factors that participate in the co-transcriptional splicing of mRNA at the chromosomes, but also as a way-station for components that accompany mRNA on its journey to the nuclear pore and cytoplasm (Mermoud et al., 1994).

Cajal bodies are involved in processing of nuclear RNA, small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs) that are recycled or undergo their final modifications. They appear as a tangle of coiled threads and are characterized by the presence of the p80 coilin protein. Control of Cajal body number is mediated by phosphorylation of the C-terminus of Coilin, but how this mechanism operates is unknown (Hebert et al., 2002; Shpargel et al., 2003).

PML bodies are structures enriched for the promyeolocytic leukaemia RING-finger protein and are implicated in transcriptional regulation, apoptosis and DNA repair. Mice that lack the PML gene have impaired immune function, exhibit chromosome instability and are sensitive to carcinogens (Bernardi and Pandolfi, 2003; Wang et al., 1998). Upon DNA damage PML bodies disassemble and thereby enhance interactions between PML-associated factors involved in regulating DNA repair and apoptosis (Dellaire and Bazett-Jones, 2004; Dellaire and Bazett-Jones, 2007). The stability of PML bodies is not only affected by DNA damage but also by other forms of cellular stress and by changes in chromatin structure.

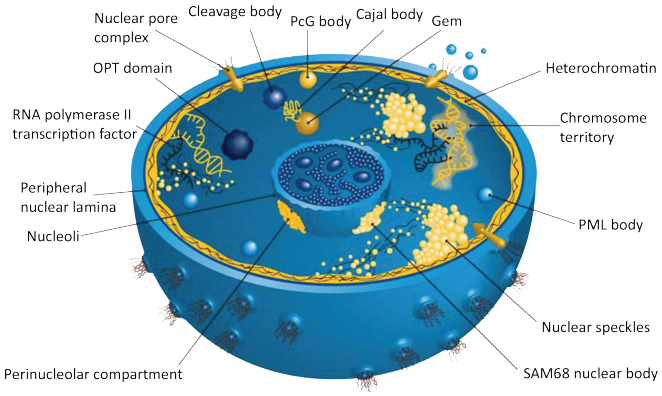

Figure 1.4 The nucleus and its sub-nuclear domains
Schematic overview of the nucleus and its different domains and bodies. Including the nucleolus, PML-bodies, Cajal bodies and speckles (Adapted from http://www.abcam.com).

Association with nuclear bodies is involved in organization of chromatin within the nucleus. For example, perinucleolar regions are involved in anchoring chromatin loops and in this way chromatin states can be maintained during replication by spatial separation in the nucleus (Zhang et al., 2007). PML bodies can associate with specific regions of high transcriptional activity in the genome and chromatin-modifying proteins can accumulate in them. It has been proposed that PML bodies functionally interact with chromatin and are important for the regulation of gene expression (Ching et al., 2005). Recently a direct link between PML, higher-order chromatin organization and gene regulation was demonstrated (Kumar et al., 2007).

1.3 The nucleolus

The nucleolus is a typically round granular body mainly composed of protein and RNA. Its primary role is to produce ribosomes, a complex that catalyzes the synthesis of proteins. However the nucleolus is also involved in signal recognition particle assembly, small RNA modification, telomerase maturation, trafficking of small RNAs, sensing of cellular stress and cell cycle regulation (Boisvert et al., 2007; Lo et al., 2006; Visintin and Amon, 2000). Nucleoli differ in size depending on the demand of the cell for rRNA (and, hence, for protein synthesis). They are large during periods of active protein synthesis, small in quiescent cells and completely invisible during cell division.

1.3.1 Nucleolar structure

Different substructures are present inside the nucleolus that can be visualized with an electron microscope. Three distinct regions have been identified: the fibrillar center (FC), the dense fibrillar component (DFC) and the granular component (GC). The spatial organization of these regions differs between cell types and organisms and is also dependent on nucleolar activity (Shaw and Doonan, 2005).

It is still debated in which of the different substructures transcription and processing of rRNA takes place. Several studies have shown that transcription occurs in the fibrillar regions of the nucleolus where RNA polymerase I is concentrated (Hozak et al., 1994). Using a pulse-chase procedure and an elongation inhibitor it was demonstrated that the ribosomal transcripts elongate in the cortex of the FC and then enter into the surrounding DFC (Cheutin et al., 2002).

During mitosis the nucleolus breaks down and the ribosomal DNA (rDNA) is arranged in a partially condensed form on the chromosome, which is recognizable, and has been termed the nucleolar organizer region (NOR). Inactive RNA polymerase I and UBF remain bound to most NORs during this stage of the cell cycle. Upon exit from mitosis these NORs fuse into one or more nucleoli. NORs lacking UBF and RNA polymerase I remain inactive and are not incorporated into nucleoli (Dousset et al., 2000; Roussel et al., 1996; Sullivan et al., 2001).

1.3.2 Ribosomal RNAs

Ribosomal RNAs (rRNAs) form the basic structure of the ribosome. There are four types of eukaryotic rRNA, each of which is present in one copy per ribosome. Three out of the four rRNAs (18S, 5.8S and 28S) are generated in the nucleolus (see below). The 5S rRNA is synthesized from a

separate cluster of genes by a different polymerase, RNA polymerase III. It is not known why this fourth RNA is transcribed separately.

rRNA synthesis is highly efficient and can account for 80% of total cellular RNA synthesis. The 18S, 5.8S and 28S rRNAs are generated from a large 45S rRNA precursor after multiple cleavage steps (Figure 1.5). The 45S precursor rRNA needs to be modified and processed by small nucleolar single strand RNAs (snoRNAs) before it can be correctly cleaved (Schneider et al., 2007). Both the synthesis and processing occur in the nucleolus and for correct ribosome assembly a tight coordination of rRNA transcription, processing and maturation is crucial.

1.3.3 rRNA "gene" structure

Multiple copies of the 45S rRNA gene are present, divided over different chromosomes. Mammalian cells, for example, contain about 400 rRNA genes, which are organized in a head-to-tail fashion and spread over 5-8 chromosomes. Thus, each repeat contains about 50 copies. About half of these genes is transcribed at any given time. The 45S rRNA "coding units" are separated from each other by an intergenic spacer (IGS) that contains transcription termination signals on both ends, presumably to prevent "spill-over" of transcription. Besides terminators the IGS also contains enhancers, a spacer promoter, replication elements and an upstream control element (UCE) (Grummt, 2003; Russell and Zomerdijk, 2005). The IGS is in mammals about 30 kb in length, while in *Xenopus* and *Drosophila* this region is much smaller. However the relative position of regulatory elements that are present in the IGS is similar compared to mammals (Figure 1.6). The mammalian rRNA gene promoter contains a core element (-45 to +18 relative to the start site (+1)), which is essential for accurate transcription initiation. Furthermore the promoter contains an UCE (-156 to -107) that has a modulatory role in transcription initiation. Followed by a transcriptional terminator site (-170), the enhancer repeat and the spacer promoter (around -2 kb). The enhancers function by increasing transcription rates of active rRNA genes (Banditt et al.,

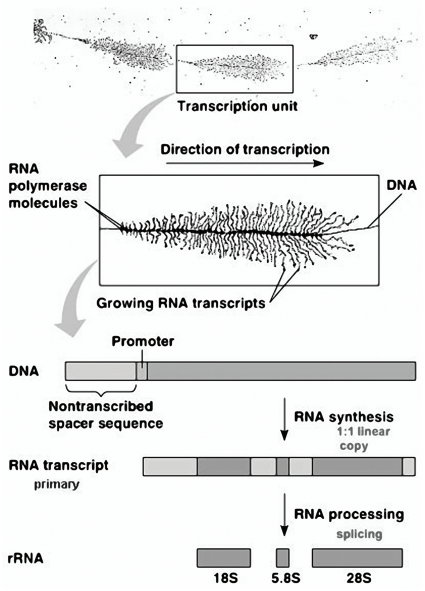

Figure 1.5 Organization of the rRNA genes
Each cell contains hundreds of copies of rDNA in the form of tandem repeats that are spread out on multiple chromosomes. At any given time, only a subset of rDNA repeats is actively being transcribed. Miller spreads (electron microscopic visualization of spread transcription units, RNA being synthesized from DNA templates) of transcriptionally active rDNA repeats demonstrate Christmas tree–like transcription units. The rRNA genes are present in a single transcription unit, transcribed by RNA polymerase I to yield a 45S precursor rRNA that is, in part, co-transcriptionally processed and modified by methylation and pseudo-uridinylation to produce the mature 18S, 5.8S and 28S rRNAs.

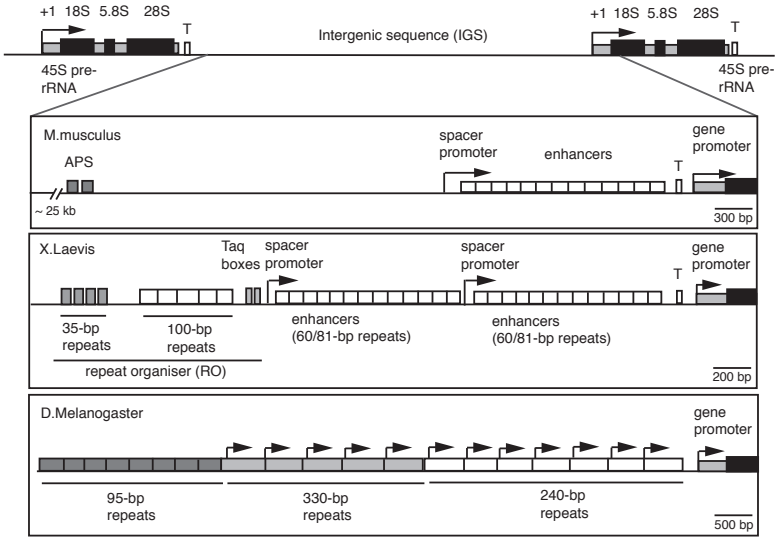

Figure 1.6 Intergenic sequence of the rDNA repeat
Overview of the organization of the rDNA repeat with indicated the main differences between species in the intergenic sequence (IGS). Organization of the mouse, *Xenopus* and *Drosophila* IGS with positions of spacer promoters, terminators and enhancers indicated. The relative position of these elements with respect to the transcriptional start is conserved. The *Xenopus* 100bp-repeats of the repeat organizer (RO) are reported to have enhancer-blocking abilities. *Drosophila* contains multiple spacer promoters that function as enhancers.

1999). More upstream (around -5 kb) the IGS contains sequences that are called amplification promoting sequences (APS).

The APS play a role in rDNA replication, they stimulate amplification of *cis*-linked plasmid DNA and coincide with nuclease hypersensitive sites (Langst et al., 1997). APS1 and 2 are methylated and located close to the zone of replication initiation (Wegner et al., 1989).

The enhancers and the spacer promoter are involved in enhancement of rRNA transcription (Paalman et al., 1995). Spacer promoter transcripts are terminated at the promoter proximal terminator, which is located between the enhancers and the UCE. This promoter proximal terminator is also involved in remodeling chromatin over the promoter (Langst et al., 1998). The spacer promoter contains only very little sequence homology with the gene promoter. The mechanism by which the spacer transcription might stimulate pre-rRNA transcription initiated from the gene promoter is unclear, particularly in light of recent data suggesting that spacer transcription participates also in NoRC-dependent rRNA gene silencing (Mayer et al., 2006).

1.3.4 rRNA transcription

EM studies suggested that a single 45S rRNA gene could be transcribed by a large number of RNA polymerase enzymes. Nascent 45S rRNAs form 'Christmas trees', the trunk representing the rDNA copy and each branch of the tree representing a transcript with a nascent ribosome forming at its tip. Transcription by RNA polymerase I is a comparatively simple process with a basic mechanism of regulation. Compared to RNA polymerase II only a few transcription factors are associated with RNA polymerase I. Because of this relative simplicity, RNA polymerase I can

optimally performed its function as polymerase with the highest synthesis rate (Grummt, 2003). Alterations in cell proliferation result immediately in a change in the transcriptional rate of rRNA genes.

'Basal' levels of transcription *in vitro* can be achieved in the presence of a pre-initiation complex (PIC) comprising only RNA polymerase I and selectivity factor 1 (SL1/TIF-IB) at the rDNA promoter. SL1 is a complex consisting of the TATA-box-binding protein (TBP) and its associated factors TAF110, TAF63 and TAF48 (Zomerdijk et al., 1994). These associated TAFs are involved in promoter- and polymerase recognition. SL1 recruits RNA polymerase I to the promoter via interaction of its TAFs with the RNA polymerase I associated factor RRN3/TIF-IA. RRN3 is tethered to RNA polymerase I by interacting with its core-subunit RPA43 and associated factor PAF67 (Miller et al., 2001; Peyroche et al., 2000). Activated transcription requires, in addition to RNA polymerase I and SL1, the upstream binding factor UBF (Bell et al., 1988). This is achieved by recruitment of RNA polymerase I by DNA bound UBF via an interaction with TIF-IB/SL1 and also by a direct interaction with the RNA polymerase I subunit PAF53 (Hanada et al., 1996). After recruitment of RNA polymerase I and its binding to the UBF/SL1 complex transcription starts (Figure 1.7).

If UBF cannot bind the UCE impaired PIC formation and repression of RNA polymerase I transcription is observed (Santoro and Grummt, 2001; Santoro et al., 2002). UBF is however not absolutely required for specific initiation on the promoter, but plays a role in activation of transcription. Addition of UBF to depleted extracts increases the efficiency of *in vitro* transcription in a dose dependent manner (Jantzen et al., 1990; Jantzen et al., 1992). Recently UBF was shown to activate transcription by stimulating promoter escape of the polymerase (Panov et al., 2006).

UBF binds to the enhancer region, core regions of the promoter and the UCE but does not have high sequence specificity and is also shown to bind weakly throughout the entire rDNA repeat (O'Sullivan et al., 2002). UBF can rapidly exchange from enhancer-bound to free UBF, so that a high UBF concentration can be maintained near the promoter (Putnam and Pikaard, 1992). Although UBF does not have high sequence specificity, it has a high affinity for bend or distorted DNA, and can bent linear DNA itself (Agresti and Bianchi, 2003). Moreover UBF functions in opening up chromatin to provide access to the DNA for transcription factors and RNA polymerase I. UBF dependent large-scale chromatin decondensation can be induced by its interaction with SL1 and does not involve common chromatin remodeling complexes (Chen et al., 2004). Rather UBF can form an open chromatin structure by competing with histone H1 binding and thereby displacing or dissociating it from a nucleosome core (Kermekchiev et al., 1997).

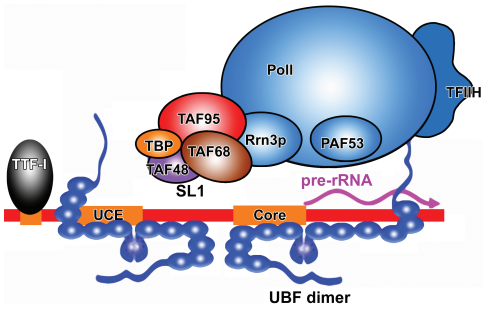

Figure 1.7 RNA polymerase I complex bound to the promoter region
Binding of the pre-initiation complex to the UCE and core region of the rRNA promoter. The UBF dimer bind via its first three HMG-boxes to the DNA and its C-terminal tail interacts with the PAF53 subunit of RNA polymerase I and with the SL-1/TIFIB complex. The TIFIB/SL-1 complex consists of TAF48, TAF68, TAF95 and TBP. SL1/TIFIB binds to RNA polymerase I via RRN3 and can bind to the UCE and core region of the rRNA promoter via its TAFs. In this way RNA polymerase I, which itself has no sequence specific DNA-binding activity is recruited to the DNA. The transcriptional termination factor TTF-I is bound to the promoter proximal terminator and can enhance as well as repress transcription from the gene promoter.

Elongation and termination of transcription

If RNA polymerase I transcription starts UBF and SL1 remain bound to the promoter poised to recruit the next RNA polymerase I complex and reinitiate transcription from the same promoter (Panov et al., 2001). The RNA polymerase I-associated factor TIF-IC can stimulate the overall rate of transcription elongation and suppress pausing of RNA polymerase I (Schnapp et al., 1994). If the process of elongation is interrupted by DNA damage, transcription-coupled repair proteins like TFIIH and CSB bind to RNA polymerase I and repair the damage (Bradsher et al., 2002).

A specific protein, called TTF-I, binds the transcription termination sites at the 5' and 3' ends of the rRNA gene. At the 3' end of rRNA genes TTF-I stops RNA polymerase I and dissociates it from the DNA with the help of the transcriptional release factor PTRF (Jansa and Grummt, 1999). After termination of transcription the released polymerase and other components can be reused to produce initiation-competent RNA polymerase I (Panov et al., 2001). TTF-I binding to the 5' end of the rRNA gene (the so-called promoter-proximal terminator site) serves to terminate transcripts originating from the spacer promoter (Langst et al., 1998).

Number of active rRNA genes

About half (i.e. 150-200) of all copies of mammalian rRNA genes are transcriptionally active. The ratio of active and inactive rRNA genes is stably propagated throughout the cell cycle and is independent of cellular rRNA synthetic activity (Conconi et al., 1989). Adjustments to the number of active copies are made only during development and differentiation (Haaf et al., 1991). rRNA transcription is therefore regulated by changing the rate of transcription initiation rather than by activating or silencing transcription units. Yeast is an exception, as it can actively regulate the number of active rRNA genes (Sandmeier et al., 2002). Because virtually all organisms contain a large number of inactive rRNA genes, this must have some kind of evolutionary advantage. However, what exactly the purpose is for maintaining such a large fraction of rRNA genes in a silenced state remains to be elucidated.

Re-establishment of the same number of active genes after mitosis is regulated by UBF and SL1, that remain bound to DNA, but are inactivated by phosphorylation in mitosis (Klein and Grummt, 1999; Voit and Grummt, 2001). TTF-I is also phosphorylated during mitosis but remains only weakly bound to the DNA in this stage of the cell cycle (Sirri et al., 1999). At first it was thought that RNA polymerase I also remains associated with DNA during mitosis. However, more detailed analysis using time-lapse imaging of single cells showed that during a small window at metaphase RNA polymerase I is not associated with the rDNA (Leung et al., 2004).

Silencing of ribosomal RNA transcription

In human and mouse the transcriptionally silent state of rDNA is established by the nucleolar remodeling complex NoRC. One component of this complex, TIP5, was identified in a yeast two hybrid screen for TTF-I interacting proteins. Subsequent analysis revealed that TIP5 is present together with the ATPase SNF2h and that these two subunits form the NoRC complex (Strohner et al., 2001). NoRC shares homology with the remodeling proteins CHRAC and WCRF that belong to the family of ISWI chromatin remodeling machines. NoRC is targeted to rDNA by interaction of TIP5 with TTFI that is bound to the promoter-proximal terminator (T0). The acetylation mark at lysine 16 of histone H4 (acH4K16) is required for TIP5 binding to DNA (Zhou and Grummt, 2005). Overexpression of NoRC reduces the size and number of nucleoli, impairs cell proliferation and represses RNA polymerase I transcription by increasing heterochromatic marks at the promoter (Li et al., 2005; Santoro et al., 2002).

NoRC can perform part of its silencing properties by recruiting the SIN3 corepressor complex that subsequently deacetylates histones near the rDNA promoter (Zhou et al., 2002). NoRC is able to establish heterochromatic features at the rRNA gene promoter not only by mediating histone H4 deacetylation but also by methylating histone H3 lysine 9 and by *de novo* DNA methylation (Santoro and Grummt, 2005). Methylation of a single CpG within the UCE of the mouse rDNA promoter can result in impairment of UBF binding, thereby preventing initiation complex formation (Santoro and Grummt, 2001). This result suggests that methylation is involved in inactivating rDNA genes and could play a role in the propagation of transcriptional silencing through cell division. Cells lacking the maintenance methyltransferase Dnmt1 that is recruited by NoRC show a complete disruption of nucleolar structure (Espada et al., 2007).

Another study shows that NoRC is able to silence rRNA transcription by positioning nucleosomes in such a way that the UCE and the core promoter element are separated not allowing cooperative binding of UBF and TIF-IB/SL1 (Li et al., 2006). This results in the inability to form a PIC needed for efficient rRNA transcription. Binding of NoRC to a small piece of RNA that covers the rDNA promoter and originates from the upstream spacer promoter is required for the association of NoRC with rDNA and formation of heterochromatin (Mayer et al., 2006). Taken together these results imply a major role for NoRC in the silencing of ribosomal RNA transcription.

Activation of ribosomal RNA transcription

The rate of rRNA gene transcription initiation is found to be mainly determined by RNA polymerase I promoter escape (Panov et al., 2001), that is, the rate-limiting step in rRNA transcription is the initiation of transcription itself and not the formation of the PIC. UBF was shown to be crucial for regulating promoter escape independent of the promoter-specific targeting of SL1 and RNA polymerase I during PIC assembly (Panov et al., 2006). It was suggested that changes in UBF protein amount during differentiation and quiescence affect the rRNA transcription rate (Glibetic et al., 1995; Poortinga et al., 2004). However, upon cell cycle-arrest UBF protein levels remain unaffected (Hannan et al., 2000). The question whether levels of UBF protein change throughout the cell cycle is a point of debate (Junera et al., 1997; Klein and Grummt, 1999). Furthermore the amount of UBF or its degree of phosphorylation is not necessarily correlated with its activity to effect rDNA transcriptional rates. Studies have led to the notion that UBF can be sequestered into an inactive complex, for example with Rb (Cavanaugh et al., 1995; Voit et al., 1997).

The promoter regions of active rRNA genes can be distinguished from inactive ones by the presence of histone H4 acetylation and H3K4 methylation marks, and by a different nucleosome position, which allows UBF and TIF-IB/SL1 binding to each other (Li et al., 2005). TTF-I binding to the rRNA gene promoter is not only required for repression but also for activation of RNA polymerase I transcription. The Cockayne syndrome B (CSB) protein, a member of the SWI/SNF2-like family of chromatin remodelers, binds to TTF-I and is involved in activating rDNA transcription. CSB is present in a complex with RNA polymerase I, TFIIH and TIF-IB/SL1 that is recruited to rDNA by TTF-I. Overexpression of CSB stimulates rDNA transcription, whereas in CSB-deficient cells rRNA synthesis is impaired (Bradsher et al., 2002). To establish an open chromatin structure at active rDNA, CSB binds to the remodeling complex WICH (Percipalle et al., 2006). Recently it was shown that CSB is directly involved in promoting transcription from the active rDNA promoter by recruiting histone methyltransferase G9a, which is able to methylate histone H3 on lysine 9 (H3K9me2) in the pre-rRNA coding region. This was suggested to facilitate RNA polymerase I transcriptional elongation (Yuan et al., 2007).

1.4 CTCF, a vertebrate chromatin organizer

1.4.1 Discovery and characterization of CTCF

CTCF (CCCTC binding factor) was identified independently by two labs: as NeP1, a factor involved in silencing of the chicken lysozyme gene (Baniahmad et al., 1990) and as CTCF, a factor binding to the *c-myc* promoter (Lobanenkov et al., 1990). Later studies showed that NeP1 and CTCF were one and the same protein (Burcin et al., 1997). Other studies then showed that CTCF is not only involved in transcriptional silencing but also in activation, both of the *c-myc* gene and of the amyloid β-protein precursor (*APP*β) gene (Klenova et al., 1993; Vostrov and Quitschke, 1997).

CTCF is highly conserved and can be traced back in evolution down to zebrafish (*Danio rerio*) and fruitfly (*Drosophila melanogaster*) (Moon et al., 2005; Pugacheva et al., 2006). Fungi, yeast, plants, or the nematode *Caenorhabditis elegans* do not contain a CTCF-like protein. CTCF contains two conserved protein domains, an eleven zinc finger domain and a proline-rich region containing an AT-hook domain (Figure 1.8).

C2H2-type (classical) zinc fingers were the first class of zinc-fingers characterized. This motif forms a short β-hairpin and an α-helix. Conserved amino acids are involved in the folding of the module. Most notably, two cysteine and histidine (C2H2) residues hold a single zinc atom in place in a tetrahedral array. This, in turn, allows specific amino acid residues from the zinc finger module, some of which are not conserved, to make highly specific contacts with DNA. C2H2 zinc fingers are the most common DNA-binding motifs found in eukaryotic transcription factors, and have also been identified in prokaryotes (Bouhouche et al., 2000). Zinc fingers are not only capable of making contact with DNA, this domain is also known to bind RNA and other protein motifs (Wolfe et al., 2000). The binding properties of a transcription factor largely depend on the binding specificity of individual zinc fingers, on the number of fingers and on the linker region between the zinc fingers (Iuchi, 2001).

It was initially proposed that only a few zinc fingers of CTCF are necessary to bind DNA (Quitschke et al., 2000). Subsequent studies showed that four zinc-fingers in the middle of the protein are sufficient for CTCF to bind to multiple different sites. Moreover it was shown that

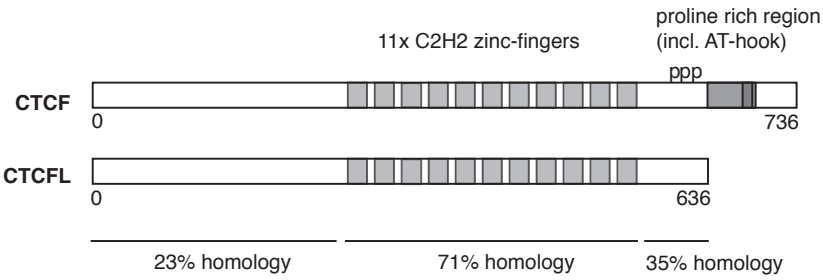

Figure 1.8 Schematic representation of mouse CTCF and CTCFL proteins
Mouse CTCF and CTCFL protein domains are shown. Including the eleven zinc-finger (ZF) region involved in protein-DNA and protein-protein interactions. CTCF contains a proline rich region with an AT-hook motif involved binding to AT-rich DNA sequences in the minor groove. CTCF and CTCFL are highly homologous in the eleven zinc-finger region, but in their N- and C-terminus the proteins differ from each other. On the top post-translational modifications are shown P:phosphorylation.

one zinc finger is responsible for recognizing the methylation status of the CTCF binding site in the *Igf2/H19* locus. The presence of a methyl group is interfering with the interaction of this zinc finger with DNA (Renda et al., 2007).

The AT-hook motif was first described in HMG-I(Y) group of proteins, where it was shown to direct binding to the minor groove of AT-rich DNA regions (Reeves and Nissen, 1990). AT-hook motifs, cooperate with other DNA-binding activities and play a role in facilitating changes in the structure of the DNA. It was proposed that the AT-hook plays a role in translocating proteins to MAR/SARs and positions chromatin fibers (Fujimoto et al., 2004). Interestingly AT-hook domains can be found in chromatin architectural factors and chromatin remodeling complexes such as Mdg4, ISWI, MeCP2, Swi/Snf2, and in the NoRC subunit TIP5 (Aravind and Landsman, 1998). The AT-hook domain in the CTCF protein is conserved down to zebrafish, however *Drosophila* CTCF is not reported to contain other domains than the eleven zinc-fingers. The exact role of this domain for CTCF functioning has not been addressed.

Although the *Ctcf* gene encodes only one protein isoform with a predicted molecular mass of 82 kD, CTCF is reported to be present in multiple different isoforms. The major isoform migrates at approximately 130 kD in SDS-PAGE (Klenova et al., 1993). It is unknown what causes this aberrant migration pattern of CTCF. Besides this isoform a protein of 70 kD is detected in many cell types and might represent a C-terminal truncation of CTCF (Klenova et al., 1997). Furthermore, poly(ADP-ribosyl)ation of CTCF results in a 180 kD form that is present in only very small amounts compared to the 130 kD protein (Yu et al., 2004). Moreover CTCF was reported to form dimers of ~200 kD (Yusufzai et al., 2004). Dimerization of the protein was suggested to be involved in stabilizing long range interactions between distant pairs of sites that bind CTCF (Ling et al., 2006; Splinter et al., 2006; Yusufzai et al., 2004).

1.4.2 Posttranslational modifications of CTCF

As mentioned above, CTCF can be posttranslationally modified. Both phosphorylation and poly(ADP-ribosyl)ation of the protein have been described (Klenova et al., 2001; Yu et al., 2004). These modifications might modulate CTCF function as they affect DNA binding. Casein kinase II is involved in phosphorylating four serines in a conserved part of the CTCF C-terminus. Substituting all serines resulted in enhanced repression of vertebrate *c-myc* promoters that contain CTCF binding sites (Klenova et al., 2001).

Poly(ADP-ribose) (PAR) is a negatively charged polymer whose polymerization onto acceptor proteins is catalyzed by a family of poly(ADP-ribose) polymerases (PARPs). The N-terminus of CTCF is the preferred *in vitro* target for this type of modification although the exact sites are not defined. PARP-1 was suggested to be involved in modifying these sites by interacting with CTCF (Yusufzai et al., 2004). Since PARP-1 can interact with nucleophosmin as well, and was shown to localize to the nucleolus, it was suggested that CTCF target sites are tethered to this subnuclear organelle and that *de novo* poly(ADP-ribosyl)ation of CTCF is essential for this localization (Meder et al., 2005). Indeed the poly(ADP-ribosyl)ated form of CTCF was predominantly found in nucleoli (Torrano et al., 2006).

When bound to the maternal *H19* ICR, CTCF appears to be poly(ADP-ribosyl)ated as well. Treatment of cells with 3-ABA, an inhibitor of PARPs, results in loss of *H19* ICR insulator function and bi-allelic expression of *Igf2* (Klenova and Ohlsson, 2005). This result has led to the notion that poly(ADP-ribosyl)ation of CTCF is involved in regulating its insulator properties. However, the poly(ADP-ribosyl)ation dependent insulator function of CTCF appears to be required on only 1%

of its binding sites (Yu et al., 2004). This indicates that this is not a general regulatory mechanism of CTCF insulator function. Indeed, besides the *H19/Igf2* ICR, no other poly(ADP-ribosyl)ation sensitive CTCF insulators have been described to date.

1.4.3 CTCF binding partners

To unravel the function of CTCF it is essential to know the proteins it interacts with. Logically, there have been many attempts to identify interaction partners of CTCF. Several techniques have been used, including GST-pull downs, yeast-two hybrid screens, and affinity chromatography procedures, including a two-step anti-Flag/anti-HA immunopurification protocol (Chernukhin et al., 2000; Defossez et al., 2005; Lutz et al., 2000b; Yusufzai et al., 2004). A short conclusion of these studies is that CTCF binding partners that have been identified to date are involved in multiple processes in the nucleus, reflecting the diversity of functions of CTCF (Figure 1.9).

CTCF was reported to bind to transcription factors, histone-modifying proteins and other regulatory proteins, including enzymes (Lutz et al., 2000b; Yu et al., 2004). Proteins that were reported to bind to the zinc-finger region of CTCF are Sin3, YB-1 and CHD8 (Chernukhin et al., 2000; Ishihara et al., 2006; Lutz et al., 2000b). Kaiso and RNA polymerase II were shown to bind to the C-terminus of CTCF, whereas Yy1 has affinity for the N-terminus (Chernukhin et al., 2007; Defossez et al., 2005). Proteins that were shown to bind *in vivo* to CTCF-dependent boundaries are CHD8, nucleophosmin and the largest subunit of RNA polymerase II. The interaction of CTCF with the latter is interesting in that CTCF has been implicated in blocking the transfer of RNA polymerase II between an enhancer and a gene (Zhao and Dean, 2004). The interactions of CTCF with nucleophosmin, CHD8 and Kaiso have been implicated in its insulator function. Nucleophosmin was suggested to tether the β-globin cHS4 insulator to the nucleolar periphery by binding to CTCF (Yusufzai et al., 2004). Using cell lines carrying a transgene with wild type or mutated CTCF sites it was shown that CTCF is required for localization of the transgene array to the nucleolus. Binding of CTCF insulators to the nucleolar periphery could be involved in spatially separating the communication between an enhancer and promoter. It remains to be determined whether the interaction with the nucleolar periphery is a general phenomenon and involved in tethering multiple CTCF dependent insulators (Yusufzai et al., 2004). It was suggested that the association of CTCF with the nuclear matrix also mediates its boundary function (Dunn et al., 2003).

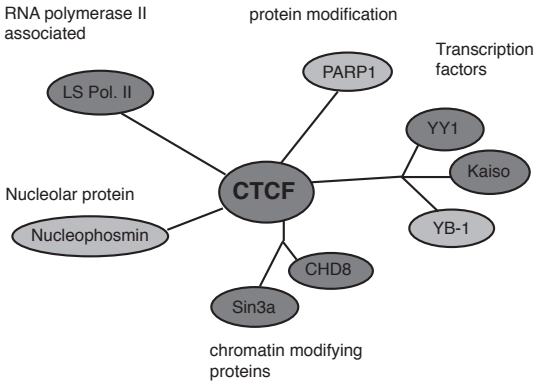

Figure 1.9 Published interacting protein partners of CTCF
Proteins reported to interact with CTCF are indicated and separated in groups regarding their function. Dark grey: proteins of which a direct interaction with CTCF using purified proteins is confirmed, light grey: proteins of which a direct interaction with CTCF is not confirmed.

CHD8, a Snf2-like chromodomain helicase, was reported to be present at multiple CTCF sites. Interestingly, knockdown of CHD8 disrupts enhancer blocking at the *H19* insulator, leading to biallelic *Igf2* expression without affecting CTCF binding at this site. Lack of CHD8 also results in a gain of DNA methylation and a decrease of acetylated histone H3 around the CTCF binding site at the *c-myc* insulator (Ishihara et al., 2006). CHD8 might affect epigenetic conditions near CTCF binding sites by bringing modifying enzymes and chromatin remodeling complexes together at CTCF dependent insulators.

The transcription factor Kaiso was shown to interact with CTCF and to bind close to the human HS5 β-globin insulator. A plasmid based enhancer-blocking assay showed that the presence of a Kaiso binding site reduces the enhancer-blocking activity of CTCF by half. These data indicate that the Kaiso-CTCF interaction negatively regulates CTCF insulator activity (Defossez et al., 2005).

In vitro GST-CTCF pull down assays showed CTCF binding to a histone modifying protein, the Sin3a/HDAC complex. It was shown that the ability of CTCF to retain histone deacetylase activity by binding to Sin3a correlates with the ability to repress gene activity (Lutz et al., 2000b). However, in another study using a different cell-type, CTCF was reported not to associate with HDACs (Dunn et al., 2003). The physiological significance of the CTCF-Sin3a/HDAC interaction therefore remains unclear.

The transcription factor Yy1, which contains four zinc fingers, interacts with CTCF. Together the two proteins function in activating *Tsix*, a transcript running antisense to the *Xist* gene. CTCF and Yy1 thereby appear to regulate X chromosome inactivation (Donohoe et al., 2007). Paired CTCF-Yy1 elements are present in a region that regulates X-chromosome counting and pairing. However it was recently shown that CTCF and not Yy1 is required for X-chromosome pairing (Xu et al., 2007). Furthermore, Yy1 binding elements are present in several imprinted loci suggesting that Yy1 may interact and bind next to CTCF not only at the X chromosome but also at imprinted domains (Kim et al., 2003; Kim et al., 2006). Interestingly, Yy1 was shown to regulate CTCF expression by binding to the *Ctcf* promoter and stimulating transcriptional activity of the gene (Klenova et al., 1998).

Another DNA binding protein, YB-1, interacts with CTCF and functions together in repression of the Myc oncogene, as shown in reporter assays (Chernukhin et al., 2000). Later it was reported that YB-1 also interacts with the serotonin transporter (5-HTT) and that CTCF can interfere with the ability of 5-HTT to support YB-1 directed gene expression *in vitro* (Klenova et al., 2004).

Recently *Drosophila* CTCF was shown to interact with the centrosomal protein 190 (CP190). dCTCF binding sites largely overlap with CP190 binding sites and binding of dCTCF to these sites requires CP190 in many cases (Mohan et al., 2007). The *Fab-8* insulator requires for example both dCTCF and CP190 to perform its enhancer-blocking activity. Mammals are not reported to contain a homolog of the *Drosophila* CP190 protein.

Several other proteins interacting with CTCF were identified by mass spectrometry after anti-Flag/anti-HA pull downs. These include the histone variant H2A.Z, the molecular chaperone Taf-1/Set, Topoisomerase II, Lamin A/C and Importin α3/ α1 (Yusufzai et al., 2004). These interactions should be interpreted with caution since experiments to confirm them were not shown.

1.4.4 The multiple roles of CTCF

***In vivo* analysis of CTCF**

Despite the enormous interest in CTCF only one study has appeared to date that describes its

in vivo function in mice. In this study, CTCF protein amounts were specifically knocked down in oocytes using a transgenic RNAi approach (Fedoriw et al., 2004). Reduction in CTCF levels resulted in the inability of zygotes to develop to blastocyst-stage embryos. These data demonstrate a requirement for CTCF during early developmental processes (Fedoriw et al., 2004). CTCF null mutations in *Drosophila* cause late pupal lethality and a homeotic phenotype showing a critical requirement for CTCF during early development in flies (Mohan et al., 2007). By making use of a conditional *Ctcf* knockout allele we have recently been able to examine the role of CTCF in T cell development (Heath et al submitted, see chapter 2 of this thesis).

CTCF, cell cycle and cancer

Most *in vitro* studies on CTCF have been done in immortal cell lines. It was shown that CTCF overexpression results in inhibition of proliferation without apoptosis. Viable CTCF overexpressing cells could be maintained for several days, but these cells did not divide. The cell cycle profile of CTCF overexpressing cells did not show a clear block in a particular stage of the cell cycle (Rasko et al., 2001).

B-cell receptor cross linking on immature B-cells results in cell-cycle arrest and induces repression of Myc and expression of p53, p21, p27 and CTCF. Conditional expression of CTCF has been shown to mimick this situation, resulting in repression of Myc and expression of p27, p21, p53 and $p19^{ARF}$ while a reduction in CTCF results in the opposite effect on those genes. These results showed CTCF to be a determinant of growth arrest and apoptosis signaling in immature B cells (Qi et al., 2003).

Furthermore CTCF has been shown to bind regulatory regions of several genes implicated in cell proliferation and cancer. For example, removal (or mutation) of CTCF binding sites in the *c-myc* promoter resulted in an increased Myc expression in transient reporter assays (Filippova et al., 1996; Lobanenkov et al., 1990), indicating that CTCF acts as a repressor of this important oncogene. However in another study mutant CTCF sites fused to a reporter gene showed a reduction in Myc expression, indicating that CTCF is an activator instead of a repressor of *c-myc* (Klenova et al., 1993). Another CTCF binding element was found upstream of the *c-myc* promoter. This site is located next to a MAR and is active in enhancer-blocking assays, suggesting not only a transcriptional role for CTCF but also a role in higher-order nuclear organization of the *c-myc* locus (Gombert et al., 2003). CTCF was also shown to bind the promoter of the gene encoding another important cell cycle regulator, i.e. the Retinoblastoma (Rb) tumor suppressor protein. CTCF binds the human *Rb* gene promoter in a methylation sensitive manner and ablation of the CTCF binding site induces silencing of reporter gene expression, suggesting that CTCF functions as an activator of Rb (De La Rosa-Velazquez et al., 2007).

RNA polymerase and most transcription factors cannot remain their binding to DNA during mitosis, since during this stage of the cell cycle chromosomes condense and are less accessible. However it was shown that CTCF does remain bound to mitotic chromosomes and that the CTCF-dependent chromatin loop at the *Igf2/H19* locus is maintained during mitosis. CTCF might therefore provide a novel form of epigenetic memory that is maintained throughout cell division (Burke et al., 2005). During mitosis CTCF not only binds to chromosomes but also associates with the centrosome, especially from metaphase to anaphase. At telophase, CTCF dissociates from the centrosome and localizes to the midbody (Zhang et al., 2004). In contrast, it was shown recently that the structure and binding of CTCF to the *c-myc* insulator is not preserved during mitosis (Komura et al., 2007). Indicating that this insulator element must be reassembled *de novo* with each new cell generation. This suggests that transfer of epigenetic information and reconstitution of domain structure after the completion of mitosis is not generally applicable to

all CTCF insulators.

The human CTCF gene is located in an area on chromosome 16q22 that is commonly deleted in sporadic breast and prostate tumors (Filippova et al., 1998). In breast-, prostate- and Wilms' tumors four different CTCF somatic missense mutations were observed that involved amino acids within the zinc finger domain. These mutations did not completely abrogate CTCF binding to the promoters of specific genes (i.e. *c-myc*, *p19^{ARF}*, *Igf2* and *Plk*) as analyzed by gel mobility shift assays. However expression of these genes was slightly altered in the tumors (Filippova et al., 2002). Besides mutations affecting CTCF binding, an increased expression of CTCF itself was also found in breast cancer cells, this was associated with resistance of these cells to apoptosis (Docquier et al., 2005). Furthermore, deletion of CTCF target sites in the *H19* ICR appears to be necessary, but is not sufficient, to cause Beckwith-Wiedemann syndrome and Wilms' tumors associated with misregulation of *Igf2* (Prawitt et al., 2005). In conclusion, extensive searches for CTCF defects in tumors have not revealed convincing evidence for a role of CTCF in cancer. On the other hand, CTCF does appear to be involved in regulating the cell cycle.

CTCF and the β-globin locus

The 5'HS4 element of the chicken β-globin locus, which is located at the 5' end of the locus, was the first element with insulating (i.e. enhancer blocking) properties to be described in vertebrates (Chung et al., 1993). Enhancer-blocking activities were subsequently found in the mouse and human β-globin locus (see below), as well as in the T-cell receptor αδ locus and in a matrix attachment 5'-boundary of the chicken lysozyme gene (Bell et al., 1999; Saitoh et al., 2000).

After narrowing down the enhancer-blocking activity to ~250 bp (Chung et al., 1997), the protein responsible was identified as CTCF (Bell et al., 1999). Subsequently, it was shown that enhancer blocking and another insulator property (i.e. a barrier function protecting against the spread of chromatin marks) are actually two separate activities at the 1.2-kb cHS4 element. Deletion of the CTCF binding site showed that CTCF is only responsible for enhancer-blocking activity at this site (Recillas-Targa et al., 2002). The transcription factor USF was identified as the protein important for boundary function at the cHS4 element. It was shown to function by recruiting histone modification complexes that are involved in H3K4 methylation and histone acetylation. In this way USF is likely to maintain the barrier by creating a local environment of active chromatin that counteract the propagation of condensed chromatin structures (Huang et al., 2007; West et al., 2004).

CTCF binding sites, their relative positions, and enhancer-blocking abilities are conserved between chicken, mouse and human. The CTCF binding sites are present at the 3' end of the locus (3'HS1), in the LCR (5'HS5/cHS4) and in mouse additional sites in the 5' end are reported (HS-62.5 and HS-85) (Bulger et al., 2003; Farrell et al., 2002) (Figure 1.10A). The role of CTCF and other factors in the regulation of β-globin gene expression and chromatin organization was defined by deleting the different binding sites in the locus. Deletion of the whole β-globin LCR that contains binding sites for multiple regulatory factors resulted in a large decrease in β-globin gene expression, but did not alter chromatin structure (Epner et al., 1998). Targeted deletion of the 5'HS5 showed only minimal effects on expression of globin genes, indicating that it does not function in maintaining the locus in an active conformation (Bender et al., 1998). Further studies on 5'HS5 function showed that its enhancer-blocking activity is restricted to embryonic erythroid cells (Wai et al., 2003). It was initially thought that the CTCF-bound HS-62.5 and 3'HS1 sites that flank the β-globin locus were involved in maintaining an open chromatin domain and were present to prevent regulatory effects of neighboring domains. HS-62.5 and 3'HS1 were shown to interact with the LCR to form a 'chromatin hub' involved in β-globin gene activation

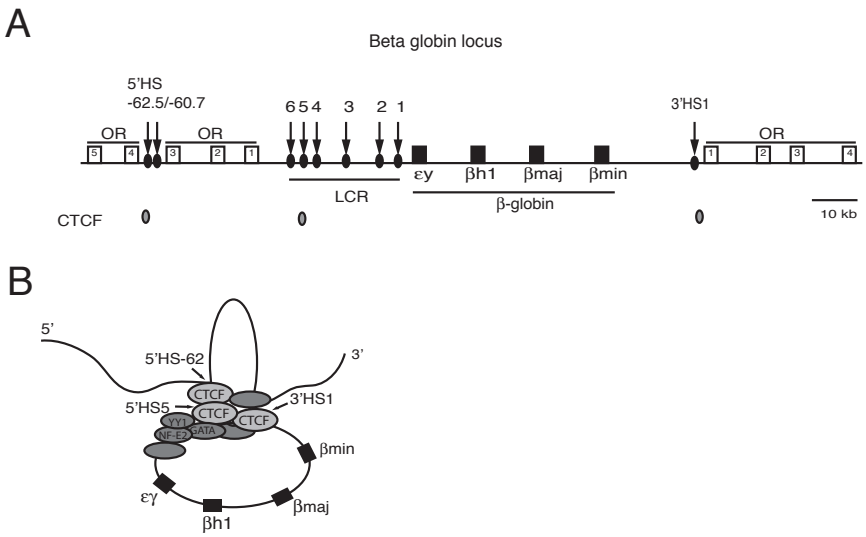

Figure 1.10 CTCF binding in the mouse β-globin locus
A. Schematic representation of the mouse β-globin locus with the position of CTCF binding sites indicated. Boxes above the horizontal line represent genes. The β-globin genes, aligned in the order of their developmental expression, are indicated by black boxes. Olfactory receptor genes (OR) are represented by white boxes and flank the locus. DNase I hypersensitive sites (HS) are indicated with arrows. **B.** Presentation of interactions that occur between regulatory DNA elements of the β-globin locus in erythroid progenitor cells. Several proteins that bind to the regulatory DNA elements are indicated. Contacts between distant CTCF-bindings sites (HS-62, 5'HS5, 3'HS1) in erythroid progenitor cells are shown. The β-globin genes are expressed at basal levels in these cells (Palstra et al., 2003; Splinter et al., 2006; Tolhuis et al., 2002).

specifically in erythroid cells (Palstra et al., 2003; Tolhuis et al., 2002)(Figure 1.10B). These results indicate that the formation of certain chromatin loops is important for regulating interactions between other regulatory elements and the β-globin genes. Both conditional deletion of CTCF as well as deletion of its 3'HS1 binding site were shown to be important for maintaining long-range interactions in the β-globin locus. However deletion of the CTCF binding sites HS-62.5 and 3'HS1 had no effect on transcription of the β-globin genes or the neighboring olfactory genes (Bender et al., 2006; Splinter et al., 2006). Absence of CTCF at 3'HS1 was shown to cause a local loss of histone H3 acetylation and gain of repressive H3K9/K27 di-methylation specifically at the CTCF binding site.

Combined these results imply that CTCF is involved in formation of a chromatin structure holding the LCR and the outside HSs together. Deleting individual CTCF binding sites apparently has no functional relevance to globin gene expression and the data indicate that CTCF does not function to prevent activation/silencing of neighboring regions. However, it is possible that in cells where specific CTCF binding sites have been deleted new sites are used to build specific chromatin hubs. Thus, CTCF could be involved in the formation of an alternative loop (with an increased size) and in this way the β-globin genes might still benefit fully from the presence of the LCR. One could call this "*in-cis*" redundancy of sites. Alternatively, *cis*-regulatory elements bound by other factors could substitute for the deleted regions.

CTCF, imprinting and X-inactivation

The Insulin-like growth factor 2 (*Igf2*) and *H19* genes are imprinted genes expressed only from the paternal or maternal allele, respectively. Identification of the region that shows allele-specific DNA methylation (DMR/ICR) and the subsequent deletion of this region suggested how imprinted expression of *Igf2* and *H19* might be established (Ferguson-Smith et al., 1993; Leighton et al., 1995). At first a model was proposed whereby *H19* and *Igf2* compete for a common set of enhancers downstream of *H19* under the influence of the ICR. However, the location of the enhancers appeared to be a critical determinant and thus an enhancer-blocking model was subsequently proposed (Webber et al., 1998). CTCF was identified as the protein that binds only to the maternal unmethylated ICR and was shown to function as an enhancer blocker and a crucial regulator of *Igf2/H19* imprinting (Bell et al., 1999; Hark et al., 2000) (Figure 1.11A).

Both maternal inheritance of mutated CTCF sites in the ICR as well as knock down of CTCF in oocytes resulted in loss of *Igf2* imprinting and a gain of methylation at the imprinting control region (ICR) (Fedoriw et al., 2004; Schoenherr et al., 2003). These results showed the requirement for CTCF to maintain the hypomethylated maternal ICR in somatic cells as well as during a critical period in the female germline when methylation marks are being set.

Besides the DMR/ICR that is located 2-4kb upstream of *H19* three other DMRs are present close to the *Igf2* gene. DMR0 is reported to be placenta-specific, DMR1 functions as a methylation sensitive silencer of *Igf2* in mesodermal tissues and DMR2 functions as a methylation sensitive activator of *Igf2* (Constancia et al., 2000; Feil et al., 1994). Using an approach whereby the ICR was tagged by knocking in a GAL4-binding motif (UAS) together with expression of a GAL4-myc fusion protein, interactions involving the ICR could be detected. The interactions were detected by performing a pull down with antibodies against the myc epitope. The results were confirmed using the 3C technique. It was shown that the DMRs in the *Igf2/H19* locus can interact with the ICR and partition imprinted *Igf2* and *H19* genes into parental-specific chromatin loops. On the maternal allele, the CTCF bound ICR interacts with DMR1 resulting in two chromatin domains, with *H19* in an active domain with its enhancers and *Igf2* in an inactive domain away from the enhancers. On the paternal allele, the methylated ICR was shown to associate with the methylated DMR2. Formation of the parental-specific chromatin loops was proposed to be dependent on the methylation status of the DMRs and differential CTCF binding (Murrell et al., 2004) (Figure 1.11B/C).

Using mice with mutated CTCF binding sites in the ICR it was confirmed that the maternal-specific interactions and chromatin loops depend on CTCF. A matrix attachment region (MAR3) located at the 3' end of the *Igf2* gene was also shown to participate with the ICR and DMR1 in the formation of a loop structure of the maternal allele. Moreover CTCF was shown to interact with the unmethylated maternal DMR1. CTCF binding to the maternal DMR1 was lost when the mutated ICR allele was inherited maternally. This suggests that CTCF is likely to be recruited to DMR1 through the physical interaction between the ICR and DMR1 (Kurukuti et al., 2006).

Examination of the chromatin loop detected on the maternal *Igf2/H19* allele during mitosis revealed that the structure is maintained throughout the cell cycle (Burke et al., 2005). It should be noted that the chromatin loops present in the *Igf2/H19* locus have never been examined in a quantitative manner or confirmed by DNA FISH so the frequency of specific interactions remains to be determined.

The crucial role for CTCF in maintaining correct imprinted expression of the *Igf2* and *H19* genes indicates that CTCF might perform a similar function in other imprinted genes. Imprinted gene clusters that contain parental specific methylation-sensitive CTCF binding sites in their DMRs are *Rasgrf 1*, *Kvlqt/P57Kip2*, *Meg3/Gtl2* and *Meg1/Grb10* (Fitzpatrick et al., 2007; Hikichi

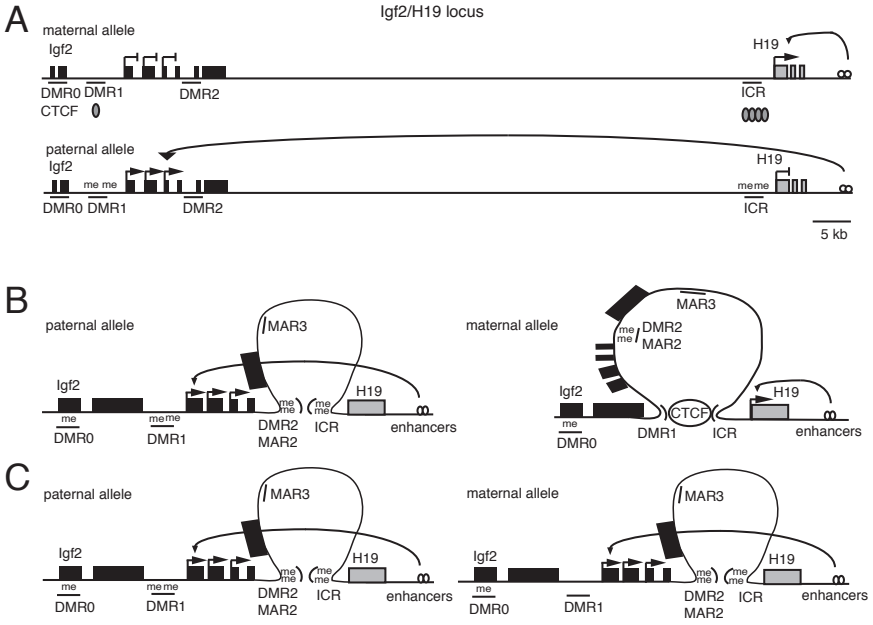

Figure 1.11 Action of CTCF at the mouse *Igf2/H19* locus
A. Schematic representation of the *Igf2/H19* locus. On the maternally transmitted allele, CTCF binds to sites in the imprinting control region (ICR) and prevents downstream enhancers from activating *Igf2* expression. The ICR of the paternal allele is methylated, CTCF does not bind and the enhancers are no longer blocked and activate the *Igf2* gene. B/C) representation of chromatin loop structures in the presence (B) and absence (C) of CTCF binding to the ICR upon deletion of its binding sites as analyzed by 3C (Kurukuti et al., 2006; Murrell et al., 2004). **B.** It is proposed that the conformation of the paternal allele allows contacts between *Igf2* promoters and the enhancers that are blocked on the maternal allele. The ICR on the maternal allele makes contact with the DMR1 located upstream of the *Igf2* promoters. It was suggested that the ICR-DMR1 interaction places the *Igf2* promoters in a repressive domain. On the paternal allele the ICR contacts another such site, DMR2, downstream of *Igf2*. This should make the *Igf2* promoters accessible to the enhancers located downstream of *H19*. **C.** Upon deletion of CTCF binding sites in the ICR. The ICR-DMR1 interaction is lost and resulting in access of the enhancers to the *Igf2* promoter. A gain of methylation is observed on the maternal ICR, which spreads over the *H19* gene and results in silencing. ICR: imprinting control region, DMR: differentially methylated region, MAR: matrix attachment region, Me: methylated site, small white circles: endodermal enhancers, light grey boxes: *H19* exons, black boxes: *Igf2* exons. Maternal and paternal alleles are indicated.

et al., 2003; Rosa et al., 2005; Yoon et al., 2005). It would be interesting to determine the role of CTCF in organizing chromatin structure and regulating gene expression in these loci.

In mice, X inactivation is imprinted in extra-embryonic tissues and occurs randomly in the embryo (Takagi and Sasaki, 1975). To start X-chromosome inactivation the two X-chromosomes need to communicate *in trans* through homologous pairing of the X-inactivation center (Xic). The pairing is necessary for counting and mutually exclusive choice (Lee, 2005; Xu et al., 2006). The Xic contains multiple regulatory elements including the *Xist* gene. Initiation of X inactivation depends on the coordinated expression of the *Xist* and *Tsix* genes, which overlap and are oriented antisense with respect to each other. The *Xist* RNA is exclusively expressed from, and coats, the inactive X chromosome (Xi). *Tsix* RNA is initially expressed on both X chromosomes and is down regulated on the Xi before inactivation, conversely *Tsix* expression persists longer on the active X chromosome (Xa) (Lee et al., 1999; Lee and Lu, 1999). Enhancers that are present in the *Xite* and

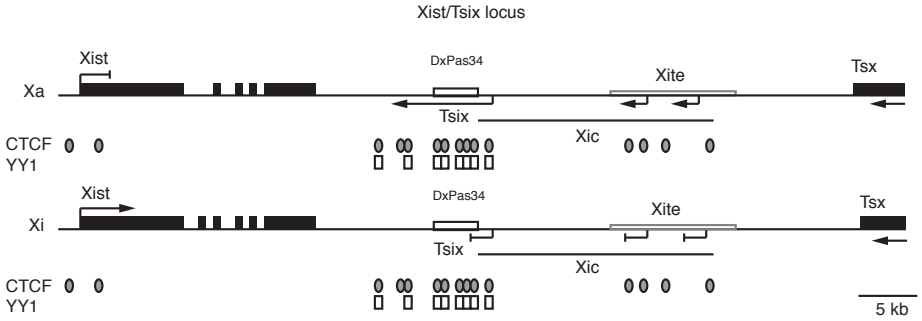

Figure 1.12 CTCF binding sites in the mouse *Xist/Tsix* locus
Schematic representation of the locus involved in mouse X chromosome inactivation. The regulatory elements implicated in counting and choice are shown. *Xist* and regulatory antisense *Tsix* transcripts are indicated, including their transcriptions status. Arrow indicates active transcription. X-inactivation intergenic transcription element (Xite), X inactivation centre (Xic) and DXPas34 element are indicated. Reported binding sites of CTCF and its interacting protein Yy1 in the *Xist/Tsix* locus are shown with small rounds (CTCF) or blocks (Yy1). Xa: active X chromosome, Xi: inactive X chromosome are indicated (Based on: (Donohoe et al., 2007; Navarro et al., 2006)).

DXPas34 elements developmentally regulate *Tsix*.

Multiple CTCF binding sites were identified in the DXPas34 element in *Tsix* and they were shown to possess enhancer-blocking activity *in vitro* (Chao et al., 2002). Differential methylation of these CTCF binding sites and of Xite correlate with X chromosome choice in mice (Boumil et al., 2006). Therefore CTCF was proposed to be a candidate factor involved in choice.

Using a bioinformatics approach additional CTCF binding sites were found in the *Tsix* gene that were frequently paired by Yy1 binding sites (Donohoe et al., 2007) (Figure 1.12). ChIPs confirmed binding of Yy1 to *Tsix* DNA and immunoprecipitations showed an interaction between Yy1 and CTCF. Furthermore it was shown that a deficiency of Yy1 leads to diminished *Tsix* levels and *Xist* up regulation. CTCF knockdown through RNA interference yielded an identical phenotype, indicating an important role for Yy1 and CTCF in transcriptional activation of *Tsix*. These results implied that a CTCF-Yy1 complex could potentially function as a key component in X chromosome pairing, counting or choice in both random and imprinted X chromosome inactivation (Donohoe et al., 2007).

Subsequently it was shown that homologous X-chromosome pairing requires CTCF binding and co-transcriptional activity of *Tsix* and *Xite*. By contrast, the CTCF-interacting partner Yy1 was shown not to be required for pairing (Xu et al., 2007). Whether CTCF could directly mediate pairing or collaborate with additional factors remains to be determined. It was speculated that together with *Tsix* and Xite RNAs CTCF might form a RNA-protein bridge holding the two X chromosomes together.

Interchromosomal interactions

CTCF was recently shown to mediate interchromosomal interactions. Using two different methods, networks of chromosomal interactions with the CTCF-bound maternal H19 ICR were identified (Ling et al., 2006; Zhao et al., 2006). In one approach a modification of the 3C technique, called associated chromosomal trap (ACT), was used (Ling et al., 2006). The procedure is similar to 3C except that the ligated fragments are cleaved by a frequently cutting second restriction enzyme followed by linker ligation and amplification with a linker primer and a specific

primer. Sequencing subsequently identifies the putative interacting DNA segments of unknown identity. Using this approach it was shown that the maternal *H19* ICR on chromosome 7, which binds CTCF because it is unmethylated, colocalizes with an intergenic sequence on chromosome 11. This intergenic sequence is located between the *Wsb1* and *Nf1* genes. ChIP experiments confirmed CTCF binding to the paternal allele of *Wsb1/Nf1*. If the full maternal ICR is deleted, or if CTCF levels are diminished, the interaction between the ICR and *Wsb1/Nf1* cannot be detected anymore and the two genes show an altered expression pattern (Ling et al., 2006). While the ACT approach identified only three different interactions with the ICR, a 4C approach yielded 114 unique sequences identified from mouse neonatal liver (Zhao et al., 2006). Technical differences might underlie the discrepancy between these two approaches.

In the 4C study the authors show that the maternal H19 ICR interacts *in trans* with many imprinted regions on other chromosomes (Zhao et al., 2006). For example, the maternal H19 ICR allele was shown to interact with the maternal DMR of the imprinted *Impact* gene. Three-dimensional DNA FISH confirmed that this interaction is lost in mice with a maternal inheritance of the ICR CTCF binding site mutant. In these mice a reduction in gene expression of the imprinted *Impact* gene can be observed compared to wild-type mice. The expression of its neighboring non-imprinted gene *Osbpl1a* was up regulated in neonatal livers from mice with a maternal inheritance of the mutated ICR CTCF binding site (Zhao et al., 2006). This study shows that imprinted domains can be functionally connected *in trans* via DMRs and in this way are able to regulate transcription *in trans*. Both studies provide proof for a central role of CTCF binding sites in a network of chromatin interactions. It is however a point of debate whether interchromosomal interactions represent true physical interactions between loci or simply reflect the close proximity of loci.

Discovery and functional analysis of *Drosophila* CTCF

Multiple different insulator binding proteins are present in *Drosophila,* these are Zw5, BEAF-32, GAGA factor and Su(Hw). None of these have an obvious counterpart in vertebrates. However, vertebrate CTCF was recently found to actually have a *Drosophila* counterpart (Moon et al., 2005).

Subsequent work showed that dCTCF is bound to several hundred loci on polytene chromosomes. Polytene chromosomes are formed in specialized cells that undergo repeated rounds of DNA replication without cell division. The banding patterns of these chromosomes provide easy visualization of transcriptionally active and inactive chromatin and the general chromatin structure. The banding pattern of dCTCF binding sites displayed no overlap with binding sites of other well known insulator proteins Su(Hw) and BEAF-32 (Mohan et al., 2007). The fact that dCTCF binds different targets indicates that it does not have a redundant function.

In *Drosophila* correct anterior-posterior patterning in the thorax- and abdominal segments is dependent on the precise expression of Hox genes of the Bithorax complex (BX-C) in specific parasegments. This is achieved by the subdivision of the regulatory regions of each of the three BX-C genes (Ubx, Abd-A and Abd-B) into distinct enhancer domains (Martin et al., 1995). There are at least nine distinct regulatory domains controlling the three genes of the BX-C in specific parasegments. Of the eight boundaries between those domains, three have been shown to have enhancer-blocking activity: Mcp, Fab-7 and Fab-8. At first dCTCF was shown to bind the Fab-8 boundary of BX-C, but more detailed analysis revealed dCTCF is bound to all but one (Fab-7) of these elements (Holohan et al., 2007). Fab-7 is bound by the boundary factor GAGA (Schweinsberg et al., 2004).

Strikingly, the binding pattern of a Su(Hw) cofactor called CP190 did show a significant overlap with that of dCTCF and the two proteins were subsequently shown to interact (Mohan EMBO J 2007). In CP190 mutant flies the total number of dCTCF labeled sites is reduced on the other hand the number of CP190 sites was not affected by dCTCF mutants (Mohan et al., 2007). Although it was subsequently reported by another group that in CP190 mutant flies the total number of dCTCF-bound sites is completely abolished and the number of CP190 sites is reduced in dCTCF mutants (Gerasimova et al., 2007). It is however not clear why there is a discrepancy between these two reports, but they indicate that dCTCF and CP190 are likely to cooperate with each other to bind to the DNA. Indeed, CTCF and CP190 were shown to occupy similar target sites in the BX-C. The Fab8 enhancer blocker was shown to require both CTCF and CP190 for its activity. The importance of CTCF binding to the BX-C elements was shown by a reduction in Abd-B expression in dCTCF mutants (Mohan et al., 2007).

Although dCTCF does not show an overlap in distribution with Mod(mdg4)2.2 on polytene chromosomes the proteins were reported to co localize in the nuclei of diploid cells. This supported the idea that dCTCF and *gypsy* are both involved in the formation of insulator bodies (Gerasimova et al., 2000). In the absence of CP190, both dCTCF and Mod(mdg4)2.2 fail to form insulator bodies and have an aberrant cellular localization. These results suggest that *gypsy* and dCTCF insulators share CP190 and cooperate in the formation of insulator bodies to regulate the organization of the chromatin fiber in the nucleus (Gerasimova et al., 2007).

The importance of dCTCF for fly development, enhancer-blocking function at specific sites, organization of chromatin structure and the shared binding site consensus points to conserved similarities between *Drosophila* and vertebrate CTCF.

1.4.5 Characteristics of binding sites

A first attempt to map CTCF binding sites in a genome-wide fashion was performed using ChIP-on-chip on fetal liver cells (Mukhopadhyay et al., 2004). However, only 200 new target sites were identified and it was suggested that more than two-third of the binding sites could not be identified in the mouse genome due to the fact that they mapped to heterochromatic regions. Most but not all of the newly identified binding sites were reported to be free of DNA methylation (Mukhopadhyay et al., 2004). As the chip used for this work contained only 2200 clones, the approach was inheritantly limited in scope.

More recent approaches have revealed a much more extensive, genome-wide distribution of CTCF binding sites, and allowed delineation of a characteristic CTCF consensus site (Barski et al., 2007; Kim et al., 2007; Xie et al., 2007). In one of the improved approaches a mixture of nine CTCF monoclonal antibodies was used for ChIP-on-chip analysis; detection was performed using high-resolution genome-tiling microarrays. About 14.000 CTCF binding sites were identified in the human genome. This result implies that CTCF binds on average once every 200 kb. Nearly half of the CTCF-binding sites are found in sequences between genes and likely define boundaries for gene clusters. An equivalent number of CTCF sites are located close to genes, in promoter regions (20%) or within genes (34%) (Kim et al., 2007). From these binding sites a twenty basepair conserved consensus sequence for CTCF binding was derived (Figure 1.13). Interestingly, a similar consensus was found independently by using a screen for regulatory motifs in conserved non-coding regions (Xie et al., 2007). The 20 basepair consensus site differs somewhat from a minimal CTCF DNA binding site that was determined to be twelve basepairs in length (Renda et al., 2007).

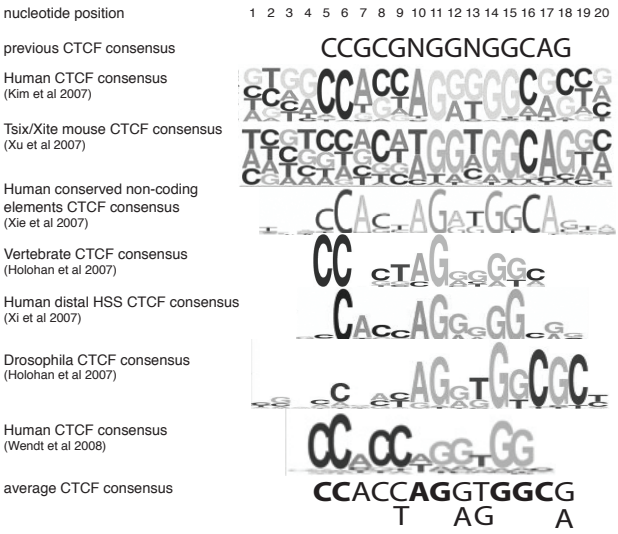

Figure 1.13 CTCF binding site consensus
The CTCF binding site consensus as adapted from different studies in human, mouse and *Drosophila*. Nucleotide position is shown on top. The height of each letter represents the relative frequency of nucleotides at different positions in the consensus. At the bottom a general consensus that represents an average of the different consensus sequences is indicated (Adapted from (Holohan et al., 2007; Kim et al., 2007; Renda et al., 2007; Wendt et al., 2008; Xie et al., 2007; Xu et al., 2007)).

Another study published recently reported on the genome-wide distribution of histone methylation- and CTCF-binding-profiles (Barski et al., 2007). In this approach ChIP was followed by adaptor ligation and direct sequencing using the Solexa 1G Genome Analyzer. In this study 20.000 CTCF binding sites were identified, and the nature of the different sites was analyzed in more detail. By linking histone modification profiles to gene expression data and RNA polymerase II and CTCF occupancy a genome-wide pattern of promoter and boundary characteristics was generated. CTCF-dependent insulators were found to be enriched for all three forms of H3K4 methylation, for H3K9me1 and for histone variant H2A.Z (Barski et al., 2007). CTCF is preferentially bound to ubiquitous and common DNaseI HS sites and not to cell-type specific HS sites (Xi et al., 2007). The conclusion from these experiments is that CTCF-dependent insulators play a major role in the genome. The question is whether the main function of CTCF is in chromatin folding, without necessarily having a positive or negative influence on gene expression.

Recently it was shown that half of the CTCF binding sites is associated with cohesin. Interestingly it was shown that CTCF is dispensable for cohesin loading onto DNA, but is needed to enrich and position cohesin at specific binding sites. Cohesin was shown to contribute to CTCF insulator function and it was suggested that cohesin is the crucial factor that performs insulating activities. How exactly cohesin contributes to transcription insulation remains to be determined, it is conceivable that cohesin physically connect different sites thereby creating DNA loops or blocks the spreading of chromatin remodeling or transcription factors on DNA (Parelho et al., 2008; Stedman et al., 2008; Wendt et al., 2008).

To date only a few CTCF binding sites that are flanked by thyroid hormone response elements (TREs) are reported. Examples include the silencer element of the chicken lysozyme gene and a region upstream of the human *c-myc* gene. Upon binding to the F1 element of the chicken lysozyme gene silencer CTCF can mediates repression by v-ERBA, thyroid hormone receptor, or retinoic acid receptor in a synergistic manner (Burcin et al., 1997). CTCF and thyroid hormone receptor (T3R) could potentially mediate this repression through recruitment of histone deacetylase activity by Sin3A (Lutz et al., 2000a).

Interestingly thyroid hormone is able to relieve the enhancer blocking activity of CTCF, when

CTCF and thyroid hormone receptor are bound to adjacent sites, even as CTCF remains bound to the TRE (Lutz et al., 2003). Loss of enhancer blocking activity after the addition of thyroid hormone is correlated with increased histone H4 acetylation at the CTCF binding site.

1.5 CTCFL, the testis-specific paralogue of CTCF

1.5.1 Discovery and initial characterization of CTCFL

CTCFL, or BORIS (Brother Of the Regulator of Imprinted Sites) was discovered in a database search for CTCF-like proteins (Loukinov et al., 2002). Like CTCF, CTCFL contains 11 zinc fingers. Most of these are of the C2H2-type, yet one matches an U1-type zinc finger, a C2H2-type-like module, which is present in matrin, U1 small nuclear ribonucleoprotein C, and other RNA-binding proteins (Dumortier et al., 1998). Mouse CTCFL is similar for about 71% to CTCF in the eleven zinc-finger region, but the N- and C-terminus of CTCF and CTCFL are very different (Figure 1.8). Furthermore, CTCFL does not contain the AT-hook domain that is present in the CTCF C-terminus. Thus, CTCF and CTCFL are alike, but only in the zinc finger domains. Consistently, both proteins can bind the same DNA probes *in vitro* (Loukinov et al., 2002).

In contrast to CTCF, CTCFL is expressed in a very restricted manner, i.e. the protein is found only in the male germ cells of the testis, in spermatogonia and spermatocytes, but not in the later stages of spermatogenesis (Loukinov et al., 2002). CTCFL is present in mammals, but absent in lower organisms. A comparison of mouse and human CTCFL reveals considerable divergence between these proteins. This is completely unlike CTCF, which is extremely well conserved. Perhaps the testis-specific expression of CTCFL allows for more mutations in the protein because of a reduced number of important interactions.

1.5.2 CTCFL and imprinting

It has been suggested that CTCF and CTCFL are expressed in a mutually exclusive manner (Loukinov et al., 2002). This excludes competition for binding sites and instead suggests that binding sites require a different protein depending on the differentiation stage. Although it was stated that the sequential expression of CTCFL and CTCF in the testis takes place at the moment of erasure and re-setting of DNA methylation marks during male germ-line development (Loukinov et al., 2002), this clearly is not correct, as this takes place earlier during gametogenesis (Davis et al., 2000; Ueda et al., 2000). Thus, the switch from CTCFL to CTCF expression is not linked with the initiation or targeting of *de novo* DNA methylation to paternal imprinting marks, like that of the *H19* ICR (Loukinov et al., 2002).

In a subsequent study it was reported that CTCFL expression could be detected much earlier in male germ-cell development, at 14.5 dpc, in pro-spermatogonia (Jelinic et al., 2006). Nuclear transfer analysis revealed that the erasure process of DNA methylation in primordial germ cells takes place around day 10.5 to day 11.5 (Lee et al., 2002). The paternal *H19* allele becomes hypermethylated in the ICR around 14.5 and 15.5 d.p.c. in pro-spermatogonia (Davis et al., 2000; Ueda et al., 2000). CTCFL is clearly not associated with erasure of methylation marks, a process that takes place earlier than expression of the protein. CTCFL expression, however, does coincide with DNA methylation and acquirement of allele specific imprints in the male germline.

Whether CTCFL plays a role in re-setting of the methylation imprint in the male germline remains to be determined. This should involve an association of the protein with factors like the *de novo* methyltransferases (Dnmts) that are responsible for setting methylation imprints. In order to identify proteins interacting with CTCFL a yeast two-hybrid screen was performed. PRMT7, a protein arginine methyltransferase was identified from this screen and further experiments confirmed its interaction with CTCFL (Jelinic et al., 2006). PRMT7 was shown to methylate histone H2A and H4 (Lee et al., 2005).

The interaction between CTCFL and PRMT7 appeared to be important for stimulation of PRMT7 activity and its interaction with histones. The question whether or not CTCFL and PRMT7 could participate in specific ICR methylation was addressed by co-injection of different expression constructs in *Xenopus* oocytes. These experiments showed that CTCFL and PRMT7 could contribute, in addition to Dnmt3a/3b, to methylation of the *H19* ICR. It was suggested that CTCFL-PRMT7 binding leaves a histone methylation mark on the ICR that is subsequently recognized by a complex that recruits Dnmt3a and Dnmt3b for *de novo* DNA methylation (Jelinic et al., 2006). However, the nature of this histone methylation mark that should recruit the Dnmts remains to be defined. Moreover Dnmts are not known to bind to PRMTs, although interactions could of course be modulated by another protein.

ChIP experiments showed that CTCFL binds to the *H19* ICR, both in young (15.5 dpc) and adult testis. It was not shown whether CTCFL binds in an allele-specific manner and whether CTCFL binding is sensitive to DNA methylation. To gain more insight in a possible role for CTCFL in re-setting methylation imprints it would be interesting to know whether CTCFL binds the ICR at the moment of the first methylation wave around 13.5-14.5 d.p.c.

1.5.3 CTCFL and cancer

Genome-wide DNA methylation patterns are frequently altered in cancer. A role for CTCFL in human cancer was proposed based on its potential function in setting up methylation marks. Indeed, besides its expression at certain stages of spermatogenesis CTCFL was also found to be expressed in certain tumors, including spermatocytic seminomas, in immortal cell lines, and several lung cancer cell lines (Looijenga et al., 2006; Risinger et al., 2005; Vatolin et al., 2005). High levels of CTCFL expression correlate with specific types of cancers for example, uterine cancers (Risinger et al., 2007). However in other cancer tissues and carcinoma cell lines, such as from the prostate and bladder, expression of CTCFL remained largely unchanged compared to normal tissues (Hoffmann et al., 2006).

Interestingly, DNA methylation itself appears to be an important mechanism in the control of CTCFL transcription. It was shown that increased DNA methylation and functional p53 negatively regulate expression from all three CTCFL promoters (Renaud et al., 2007). In another study CTCFL was dramatically induced in cancer cell lines upon treatment with the DNA methylation inhibitor 5-aza-2'-deoxycytidine (5-AzadC) (Hoffmann et al., 2006). Furthermore, it was reported that cancer-associated genome-wide demethylation results in aberrant expression of cancer/testis (CT) antigens that are encoded by genes that are normally expressed only in the human germ line (Old, 2001; Simpson et al., 2005). This led to the hypothesis that CTCFL could be itself a cancer/testis antigen or could be involved in regulating such genes. Derepression of CTCFL upon 5-AzadC treatment results indeed in expression of multiple cancer-testis antigens including NY-ESO-1 and MAGE-A1 in lung cancer cells (Hong et al., 2005; Vatolin et al., 2005). It was shown that overexpression of CTCFL is associated with changing methylation patterns of the

MAGE-A1 promoter and in derepression of the gene. This implies that CTCFL is directly involved in regulation of those genes. Both CTCF and CTCFL were shown to bind a region within the MAGE-A1 promoter in a methylation sensitive manner.

Gel shift and ChIP experiments in the NY-ESO-1 promoter also showed a binding site for CTCF/CTCFL. It was suggested that binding of this site by CTCF is associated with silencing of the NY-ESO-1 promoter whereas switching from CTCF to CTCFL on this site resulted in derepression of NY-ESO-1 (Hong et al., 2005). Later it was suggested that CTCFL and not CTCF could recruit Sp1 to mediate de-repression of NY-ESO-1 during pulmonary carcinogenesis (Kang et al., 2007). Thus, while in testis CTCF and CTCFL are not expressed in the same cells, in a variety of human cancers they are. In human fibroblasts that are negative for CTCFL, it was shown that CTCFL expression could be induced by a reduction in CTCF protein levels (Renaud et al., 2007). Despite these data the existence of a CTCFL/CTCF switch and its role in regulating DNA methylation, remains to be identified *in vivo*.

2

CTCF regulates cell cycle progression during β-selection in the thymus

Submitted

CTCF regulates cell cycle progression during β-selection in the thymus

Helen Heath[1+], Claudia Ribeiro de Almeida[2+#], Frank Sleutels[1], Gemma Dingjan[2], Suzanne van de Nobelen[1], Iris Jonkers[3], Kam-Wing Ling[2,] Joost Gribnau[3], Rainer Renkawitz[4], Frank Grosveld[1], Rudi W. Hendriks[2*#], and Niels Galjart[1*]

Department of Cell Biology and Genetics[1], Department of Immunology[2], and Department of Reproduction and Development[3], Erasmus MC, P.O. Box 1738, 3000 CA Rotterdam, The Netherlands; Institute for Genetics[4], Justus-Liebig-Universitaet Giessen, Heinrich-Buff-Ring 58-62, D-35392 Giessen, Germany. [+]Equal contributions, #Current address: Department of Pulmonary Medicine, Erasmus MC, *Corresponding authors

Summary

CTCF is a highly conserved protein, involved in imprinting, long-range chromatin interactions and transcription. To investigate its function *in vivo*, we generated mice with a conditional *Ctcf* knockout allele. Consistent with previous studies we found that CTCF was indispensable for early embryonic development. We subsequently inactivated CTCF specifically in early double-negative (DN) thymocytes, using Lck-Cre transgenic mice. This resulted in the accumulation of immature single positive (ISP) cells, which were smaller and inhibited in their cell cycle, and which contained highly elevated amounts of the cyclin-CDK-inhibitors p21 and p27 levels. *In vitro* studies provide evidence for a crucial role of CTCF in the activation and proper localization of key proteins, such as nucleophosmin, a CTCF-interacting factor. Using two independent approaches we show that CTCF levels increase with T cell size. Combined, our results show that CTCF controls cell cycle progression of β-selected T-cells. We propose that CTCF regulates cell size by controlling the activation of key metabolic processes in these rapidly proliferating and growing cells.

Introduction

The 11-zinc finger protein CCCTC-binding factor (CTCF) is a widely expressed and highly conserved transcriptional regulator implicated in many important processes in the nucleus (for reviews, see (Lewis and Murrell, 2004; Ohlsson et al., 2001). In line with this view, murine CTCF is essential for early embryonic development (Fedoriw et al., 2004). CTCF is the archetypal vertebrate protein that binds insulator sequences, DNA elements that have the ability to protect a gene from outside influences (Bell et al., 1999). Its methylation-sensitive interaction with the imprinting control region of the H19/insulin-like growth factor 2 (*Igf2*) genes indeed controls enhancer access (Bell and Felsenfeld, 2000; Fedoriw et al., 2004; Hark et al., 2000). CTCF-mediated insulator activity has been predicted at several other sites including the *DM1* locus and boundaries of domains that escape X-chromosome inactivation (Filippova et al., 2005; Filippova et al., 2001). We have shown that CTCF mediates long-range chromatin interactions and regulates local histone modifications in the β-globin locus (Splinter et al., 2006). Evidence has furthermore been presented for a role of CTCF in inter-chromosomal interactions between *Igf2* and other loci (Ling et al., 2006). During mitosis, CTCF remains bound to mitotic chromosomes, possibly facilitating reformation of higher order chromatin loops after mitosis (Burke et al., 2005). Combined these data suggest that CTCF is an essential organizer of imprinting, long-range chromatin interactions and transcription.

Genome-wide mapping of CTCF-binding sites revealed ~14,000 sites, whose distribution

correlated with genes but not with transcriptional start sites (Kim et al., 2007). Strikingly, the 20-bp consensus motif found in the majority of the sites is virtually identical to a consensus sequence LM2*, bound by CTCF and found in ~15,000 conserved non-coding elements in the human genome (Xie et al., 2007). High-resolution profiling of histone methylation in the human genome showed that CTCF marks boundaries of histone methylation domains (Barski et al., 2007) consistent with a role for CTCF as an insulator protein. CTCF may tether insulators to subnuclear sites together with nucleophosmin, a CTCF-interacting partner (Yusufzai et al., 2004). Very recently it was shown that CTCF binding sites on mammalian chromosomes largely overlap with those of cohesin (Parelho et al., 2008; Wendt et al., 2008). Interestingly, CTCF is required for cohesin localization to these sites, but cohesin might function as the actual insulator protein.

Genome-wide analyses also revealed CTCF-binding sites near genes displaying extensive alternative promoter usage, including protocadherin γ, the immuno-globulin λ light chain and the TCR α/δ and β chain loci. In mice, CTCF-dependent insulators were found downstream of the TCR α/δ and the immunoglobulin H chain loci (Garrett et al., 2005; Magdinier et al., 2004). Combined, these data suggest an important role for CTCF in lymphocytes, in particular in the regulation of gene transcription or recombination targeting in complex loci. To understand how CTCF regulates lymphocyte proliferation and differentiation *in vivo*, we generated mice with a conditional *Ctcf* allele (*Ctcf*$^{f/f}$), and deleted the gene in early T cell development. T cell progenitors differentiate in the thymus, where early double-negative (DN) precursors, expressing neither CD4 nor CD8 co-receptors, begin locus-specific recombination of their TCR loci (for review, see (Rothenberg and Taghon, 2005). Upon productive TCRβ gene rearrangement, the TCRβ chain associates with the invariant pTα chain on the cell surface. Cells that successfully pass this β-selection checkpoint enter the cell cycle and acquire CD4 and CD8 co-receptors to become double-positive (DP) thymocytes. Upon productive TCRα locus recombination in DP cells, TCRαβ is expressed on the cell surface. Positive selection results in the differentiation to CD4 and CD8 single positive (SP) cells, which express TCRαβ and recognize peptide antigens presented by MHC class II or class I molecules, respectively. Mature SP cells exit the thymus and circulate to the periphery as naive $CD4^+$ and $CD8^+$ T cells.

Here we show that CTCF acts as a critical regulator of cellular proliferation and differentiation following β-selection in the thymus. We demonstrate that CTCF is expressed at higher levels in immature single positive (ISP) cells, which normally form a relatively big and actively cycling subpopulation of T cells. In *Ctcf* knockout mice ISP cells are small and are blocked in their cell cycle. Interestingly, we detect increased expression of p2 and p27, two major cell cycle regulators, in *Ctcf* knockout ISP cells. Using *in vitro* cultures we show that CTCF is also required for the activation and localization of key factors in cell metabolism. Thus, CTCF regulates cell size and the cell cycle in β-selected T cells.

Results

Conditional deletion of the Ctcf gene in developing T lymphocytes

We generated a conditional *Ctcf* allele (*Ctcf*^{f}) by inserting *loxP*-sites upstream of exon 3 and downstream of exon 12 (Figure 2.1A). Normal amounts of CTCF were expressed in *Ctcf*$^{f/f}$ mice (data not shown). *Ctcf*$^{+/f}$ mice were crossed with mice expressing Cre recombinase ubiquitously (Sakai and Miyazaki, 1997). This caused removal of *Ctcf* exons 3-12 from *Ctcf*^{f}, yielding the *Ctcf*

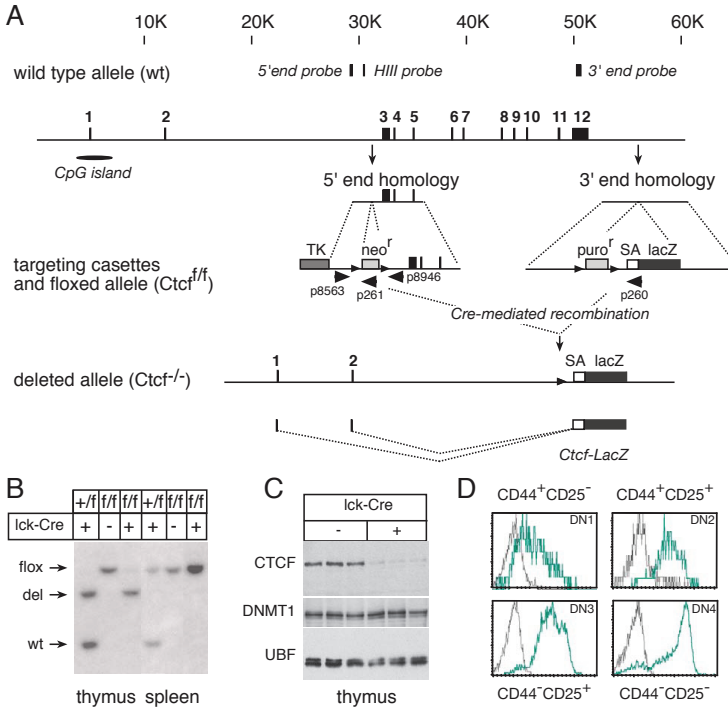

Figure 2.1 Conditional targeting of the mouse Ctcf gene

A. Murine *Ctcf* locus and gene targeting constructs. Exons of the *Ctcf* gene (solid boxes) are numbered, scale is in kilobase (K). Exon 1 is embedded in a CpG island. Exon 3 contains the start codon and exon 12 the stop codon. Southern blot probes are shown above the *Ctcf* gene. The two targeting constructs, with loxP sites (small triangles), flanking a PMC1-neomycin cassette (neo^r) or a PGK-puromycin cassette (puro^r), are shown with homologous regions. TK: thymidine kinase gene, SA-LacZ: Splice acceptor-lacZ cassette (Hoogenraad et al., 2002). PCR primers for genotyping (p8563, p8946, p260 and p261, large triangles) are indicated on targeting cassettes. Underneath the targeting constructs the deleted *Ctcf* gene is shown, that is generated after complete Cre-mediated recombination at the outermost loxP sites. Due to alternative splicing, the splice acceptor (SA) site, present at the 5'end of the reporter LacZ cassette, is spliced on to *Ctcf* exon 1 or 2, thereby generating a hybrid *Ctcf-lacZ* transcript. **B.** Southern blot analysis of Lck-Cre recombinase activity. Digested genomic DNA from thymus and spleen of mice of the indicated genotypes was analyzed by hybridization with the HIII-probe (see panel A). The positions of the wild type (WT), *Ctcf$^{f/f}$* (flox) and *Ctcf$^{-/-}$* (del) alleles are indicated (asterisk indicates a polymorphic WT allele from the FVB background). **C.** Western blot analysis of thymus. Total thymus lysates from Lck-Cre *Ctcf$^{f/f}$* and WT mice (+ indicates presence of Cre transgene; - indicates absence) were analyzed for CTCF, DNMT1 and UBF protein levels. **D.** Flow cytometric analysis of lacZ expression in CTCF conditionally deleted mice. LacZ expression was analyzed in conjunction with cell surface markers. The indicated cell populations were gated and lacZ expression data are displayed as histogram overlays of Lck-Cre *Ctcf$^{f/f}$* mice *(green)* on top of background signals in wild type mice *(black)*.

allele, in which a *Ctcf-lacZ* fusion transcript is expressed instead of *Ctcf* (Figure 2.1A). *Ctcf$^{+/-}$* mice appeared normal and were fertile, but we were unable to generate homozygous knockouts from *Ctcf$^{+/-}$* crosses (Table 2.1), consistent with an essential role for CTCF in early development (Fedoriw et al., 2004). Surprisingly, crosses among *Ctcf$^{+/-}$* mice and between wild type and *Ctcf$^{+/-}$* mice yielded more wild type animals than would be expected on the basis of Mendelian segregation (Table 2.1). These data suggest that CTCF is required in a dose-dependent manner.

In order to obtain a T-cell specific deletion of the *Ctcf* gene we crossed *Ctcf$^{f/f}$* mice with *Lck-Cre* transgenes, in which the Cre recombinase is driven by the proximal *Lck* promoter (Lee et al.,

	$Ctcf^{+/-}$ x $Ctcf^{+/-}$		
	genotype and number		
Age	wild type	$Ctcf^{+/-}$	$Ctcf^{-/-}$
E 9.5	13	14	0
E 3.5	10	7	0
adult	88	92	0
	$Ctcf^{+/-}$ x wild type		
adult	101	74	n.a

Table 2.1 Genotype of Ctcf+/- x Ctcf+/- offspring

2001; Wolfer et al., 2002). Southern blotting showed almost complete deletion of the *Ctcf* gene in thymus, while in spleen deletion was not evident (Figure 2.1B). These data reflect the specificity of the *Lck-Cre* transgene; they further indicate that *Ctcf* knockout T cells do not repopulate the spleen in large numbers. To evaluate the onset of $Ctcf^{f}$ gene deletion we analyzed lacZ expression in T cells using fluorescein-di-β-D-galactopyranoside (FDG) as a substrate. We found that deletion was almost complete from DN2 onwards (Figure 2.1D). Western blotting showed that in thymic nuclear extracts from *Lck-Cre* $Ctcf^{f/f}$ mice CTCF protein levels were reduced to ~8 % of control (Figure 2.1C). We conclude that ablation of the *Ctcf* gene results in an efficient depletion of the protein *in vivo*.

*Defective TCR*αβ *lineage development in Lck-Cre* $Ctcf^{f/f}$ *mice*

To examine the effects of a *Ctcf* deletion, thymocyte subpopulations in 6-8 week-old mice were analyzed by flow cytometry. *Lck-Cre* $Ctcf^{f/f}$ mice displayed reduced thymic cellularity, with a severe decrease in the proportions of DP and CD4 SP cells, and a concomitant increase in the proportions of DN and $CD4^{-}CD8^{+}$ cells (Figure 2.2A, B). $CD4^{-}CD8^{+}$ cells were mainly of the $CD3^{lo}CD69^{lo}$ type, indicative for ISP cells (see also below). αβ T cell development was partially arrested at the ISP to DP transition, causing accumulation of DN3, DN4 and ISP cells (Figure 2.2B). Heterozygous *Lck-Cre* $Ctcf^{+/f}$ mice also displayed a phenotype at the DP stage, showing that normal CTCF levels are important for proper T cell development. In these mice, thymic cellularity was modestly reduced and no accumulation of ISP cells was detected (Figure 2.2B). In agreement with impaired thymic SP cell production, the numbers of mature $CD4^{+}$ and $CD8^{+}$ T cells in spleen and lymph nodes of Lck-Cre $Ctcf^{f/f}$ and heterozygous Lck-Cre $Ctcf^{+/f}$ mice were significantly reduced (Figure 2.2A, B and unpublished data).

CTCF-deficiency had no adverse effect on γδ T cell development, since the number of $CD3^{+}TCR\gamma\delta^{+}$ thymocytes in Lck-Cre $Ctcf^{f/f}$ mice was ~2-fold higher than in wild type littermates (Figure 2.2C, D). The relative proportion of γδ T cells in the spleens of Lck-Cre $Ctcf^{f/f}$ mice was also markedly increased (Figure 2.2C), due to impaired αβ T cell production. In fact, in vitro culture of anti-CD3/CD28 antibody-stimulated peripheral T cell fractions from Lck-Cre $Ctcf^{f/f}$ mice resulted in a selective outgrowth of TCRγδ T cells (Supplementary Figure S2.1A). We could not detect CTCF protein in a mixed population of *in vitro* activated TCRαβ and TCRγδ T cells from Lck-*Cre* $Ctcf^{f/f}$ mice (Supplementary Figure S2.1B, C). We therefore conclude that CTCF is essential for TCR-mediated activation and proliferative expansion of TCRαβ but not of TCRγδ T cells.

The accumulation of CTCF-deficient ISP cells in Lck-Cre $Ctcf^{f/f}$ mice could result from a developmental arrest at the ISP stage or reflect defective up-regulation of CD4 expression in $CD3^{+}TCR^{+}$ cells (i.e. DP cells), similar to thymocytes deficient for the chromatin remodeler Mi-2β (Williams et al., 2004). To distinguish between these possibilities, we assessed expression of

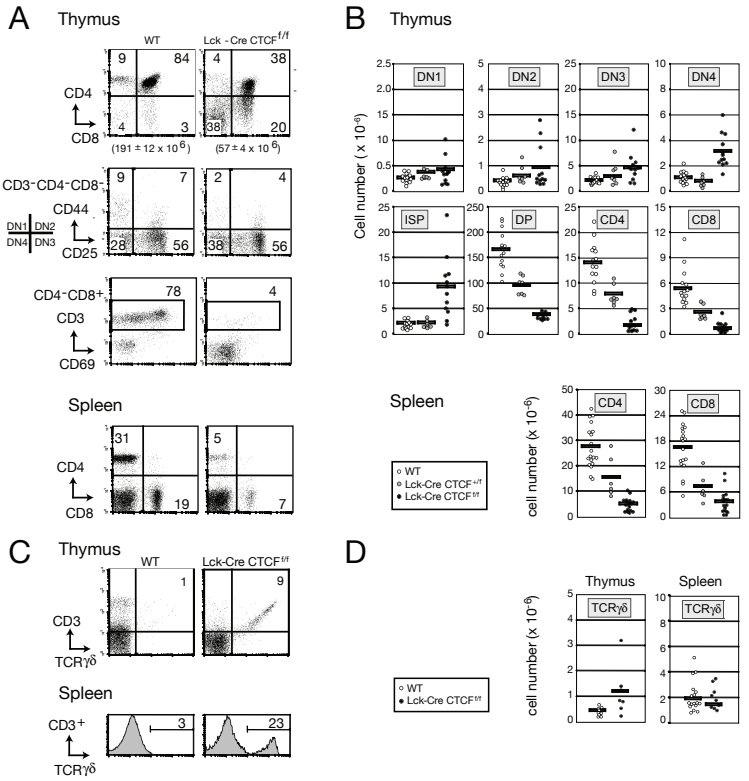

Figure 2.2 Defective TCRαβ lineage development in CTCF-deficient mice

A. Flow cytometric analyses of the indicated cell populations in thymus or spleen from the indicated mice. Expression profiles of surface markers are shown as dot plots and the percentages of cells within quadrants or gates are given. **B.** Absolute numbers of the indicated thymic and splenic T cell subpopulations. Each symbol represents one individual animal and lines indicate average values. Lck-Cre Ctcf$^{f/f}$ had increased numbers of DN3 ($p<0.01$), DN4 ($p=0.0002$) and ISP cells ($p<0.002$). In Lck-Cre Ctcf$^{f/f}$ mice and heterozygous Lck-Cre Ctcf$^{f/+}$ mice DP, CD4 SP and CD8 SP subsets in the thymus were significantly reduced ($p<0.0001$). CD4 and CD8 T cells in the spleen were significantly reduced in Lck-Cre Ctcf$^{f/f}$ mice ($p<0.00001$) and in heterozygous Lck-Cre Ctcf$^{f/+}$ mice ($p<0.01$). **C.** Flow cytometric analyses of total thymocytes and CD3$^+$ splenocytes. For the thymus, expression profiles of CD3 and TCRγδ surface markers are shown as dot plots and the percentages of CD3$^+$TCRγδ$^+$ cells are given. For the spleen, data are displayed as histograms and the percentages represent the fractions of CD3$^+$ cells that are TCRγδ$^+$. **D.** Absolute numbers of TCRγδ$^+$ T cells in thymus and spleen of the indicated mouse groups. TCRγδ$^+$ T cells were significantly increased in the thymus of Lck-Cre Ctcf$^{f/f}$ mice ($p<0.05$) Data shown are representative of 10-20 animals per group.

various cell surface markers. CD3 and TCRβ expression were very low in CTCF-deficient ISP cells, suggesting that these were true ISP cells (Figure 2.3A). The cells also expressed low levels of CD5, which is normally up-regulated on DP cells (Azzam et al., 1998), and of CD69, which is induced in a sub-fraction of DP cells, reflecting TCR-mediated activation (Bendelac et al., 1992). Expression of CD24 (HSA), which is normally high on DN and ISP cells and down-regulated at the ISP to DP transition (Williams et al., 2004), was reduced in CTCF-deficient cells throughout thymocyte differentiation (Figure 2.3A). Remarkably, differentiation of TCRβ$^+$ cells from *Lck-Cre Ctcf$^{f/f}$* mice was possible from the DN4 towards the ISP cell type, but cells were significantly smaller compared to wild type (Figure 2.3B, see also Figure 2.4). Combined our data show that a CTCF-deficiency

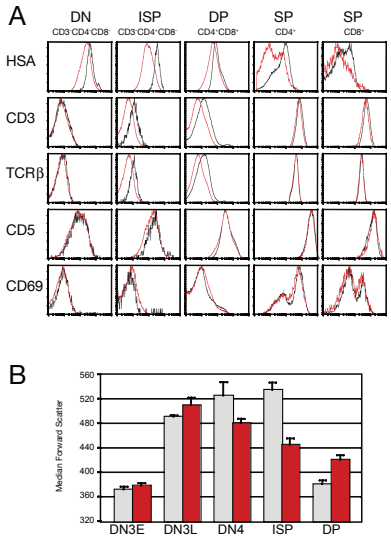

Figure 2.3 Characterization of CTCF-depleted cells in Lck-Cre Ctcf$^{f/f}$ mice

A. Flow cytometric analyses of HSA, CD3, TCR, CD5 and CD69 in the indicated thymocyte subpopulations, displayed as overlays of wild type mice (*black histograms*) and Lck-*Cre Ctcf$^{f/f}$* mice (*red histograms*). Data shown are representative of 5-8 mice per group. **B.** Flow cytometric analyses of intracellular TCRβ protein expression in the indicated thymic subsets from wild type (WT) and *Lck-Cre Ctcf$^{f/f}$* mice (KO). TCRβ/forward scatter (FSC) profiles are shown as dot plots and the percentages of TCRβ$^+$ cells are shown.

blocks β-selected T cells at the ISP stage. Lack of CTCF affects ISP cell size, suggesting that CTCF forms part of the regulatory network that monitors this fundamental property of cells.

Tcr rearrangements occur in Lck-Cre Ctcf$^{f/f}$ mice

As our findings indicated a specific role for CTCF at the ISP to DP transition, we focused our attention on possible molecular mechanisms underlying the hampered differentiation of these cells. We first analyzed TCR rearrangement, because of the many CTCF-binding sites found in the genes encoding the different receptors. *Tcr*β rearrangement is generally initiated and completed in DN3. This stage consists of early small cells that have not yet productively rearranged the *Tcr*β locus, and more mature large proliferating cells expressing TCRβ (Hoffman et al., 1996). We detected a significant population of large DN2 cells in *Lck-Cre Ctcf$^{f/f}$* mice that already contained intracellular TCRβ$^+$ (Figure 2.4). The proportion of large TCRβ$^+$ cells was still elevated in DN3 cells, thus the *Tcr*β locus can undergo functional V(D)J recombination in DN3 (and apparently even in DN2) cells that have deleted the *Ctcf* gene.

The severe reduction of DP cell numbers and low surface CD3/TCR expression on CTCF-deficient DP cells, together with the reported presence of CTCF-binding sites in the *Tcr* α gene locus (Garrett et al., 2005; Kim et al., 2007), suggest that defective TCRα V(D)J recombination may contribute to the arrest of CTCF-deficient thymocytes. We therefore crossed *Lck-Cre Ctcf$^{f/f}$*

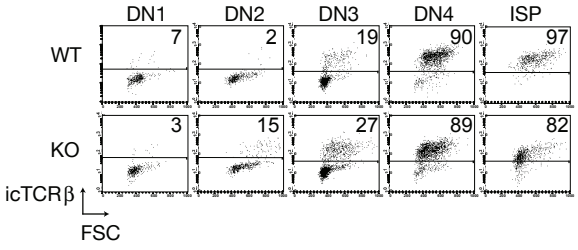

Figure 2.4 TCRβ rearrangements in Lck-Cre Ctcf$^{f/f}$ mice

Quantification of forward scatter values of the indicated thymocyte subpopulations in wild type (*gray bars*) and Lck-Cre *Ctcf$^{f/f}$* mice (*red bars*). Data are average values ± SEM from 5-8 mice per group.

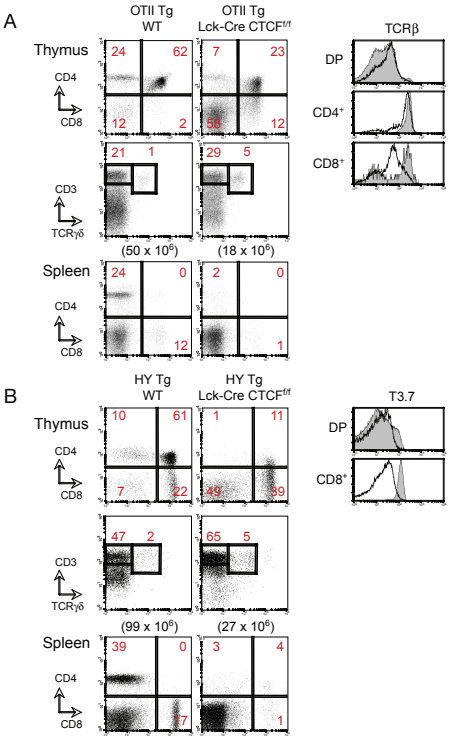

Figure 2.5 The arrest of CTCF-deficient thymocytes is independent of TCRα rearrangement

A. The OTII TCRαβ recognizes the $OVA_{323-339}$ peptide in the context of C57BL/6 MHC class II. OTII TCR transgenic thymocytes are positively selected towards the CD4 lineage {Barnden, 1998 #253}. **B.** The MHC class I-restricted HY TCRαβ recognizes a male-specific HY antigen peptide and in the C57BL/6 H-2^b class I female background {Kisielow, 1988 #221}; HY-specific thymocytes are positively selected towards the CD8 lineage. Providing Lck-*Cre Ctcf*$^{f/f}$ mice with a pre-rearranged TCRαβ transgene does not correct the developmental arrest of DP cells. Flow cytometric profiles of CD4/CD8 and CD3/TCRγδ in the indicated tissues are shown as dot plots; percentages of cells within quadrants or regions and total thymic cell numbers are given. The expression profiles of total TCRβ (A) or HY idiotype-specific T3.7 TCR (B) within the indicated cell populations are shown on the right as histogram overlays of TCR Tg Lck-Cre *Ctcf*$^{f/f}$ mice (*bold lines*) on top of profiles of TCR Tg wild-type littermates (*gray filled histograms*).

mice with transgenic mice expressing either the pre-rearranged OTII TCRαβ that is specific for albumin (Barnden et al., 1998), or the HY TCR, which is positively selected in HY Tg female mice (Kisielow et al., 1988). These molecules normally drive thymocytes into the CD4 or CD8 lineage, respectively. However, the impaired developmental progression of CTCF-deficient cells was not rescued (Figure 2.5). Rather, the presence of the OTII and HY TCR transgenes resulted in an even more severe arrest of T cell development in the thymus (Figure 2.5A and B, respectively), as in both cases we observed an almost complete absence of mature T cells in the spleen. These findings indicate that the developmental block in *Ctcf* knockout T cells is independent of TCRα rearrangement. Consistent with this, *in vivo* induction of DP cells by stimulation with anti-CD3ε antibodies, which mimics pre-TCR signaling (Azzam et al., 1998), was reduced in CTCF-deficient *Rag2*$^{-/-}$ DN cells, compared to CTCF-expressing *Rag2*$^{-/-}$ DN cells (Supplementary Figure S2.2).

Cell cycle arrest during TCRαβ lineage development in Lck-Cre Ctcf$^{f/f}$ *mice*

Interestingly, the CTCF-deficient ISP population from *lck-Cre Ctcf*$^{f/f}$ mice contain less cycling cells (29% ± 1 in S/G2/M phase), compared to wild type (53% ± 8; see Figure 2.6A for an example analysis from individual mice), indicating that CTCF is important for cell cycle progression in β-selected cells. To examine the underlying cause we analyzed mRNA expression patterns of important T cell factors using real-time PCR. We sorted wild type and CTCF-deficient T cells into DN, ISP, and DP fractions. In wild type cells *Ctcf* mRNA levels increased from the DN to ISP stage and then decreased again in DP cells (Figure 2.6B). *Ctcf* expression was severely reduced in the DN fraction from *lck-Cre Ctcf*$^{f/f}$ mice (Figure 2.6B). It should be noted that residual *Ctcf* mRNA is

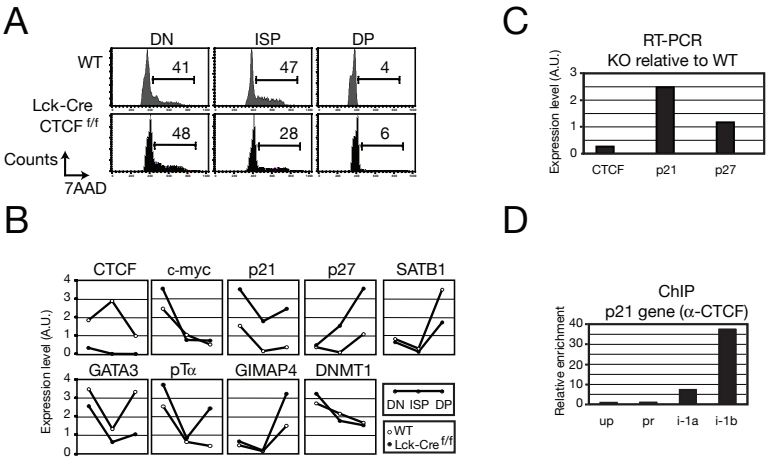

Figure 2.6 Impaired proliferation and differentiation of β-selected cells in Lck-Cre Ctcf$^{f/f}$ mice
A. Cell cycle status of DN, ISP and DP cells, using 7-AAD. **B.** Quantitative RT-PCR analysis in sorted DN, ISP and DP cell fractions from wild type (WT) and *Lck-Cre Ctcf$^{f/f}$* mice. The DP fraction also contained CD4 SP cells. **C.** Quantitative RT-PCR analysis in wild type and *Ctcf$^{f/f}$* MEFs, after treatment with Cre recombinase. Expression in Cre-treated *Ctcf$^{f/f}$* MEFs is shown relative to wild type. Although residual *Ctcf* mRNA is present in the MEFs *p21* expression is increased. **D.** Chromatin immunoprecipitation (ChIP) analysis in wild type and *Ctcf$^{f/f}$* MEFs, after treatment with Cre recombinase. ChIP analysis was performed with anti-CTCF antibodies on four regions in the *p21* gene. Potential CTCF binding sites in the mouse *p21* gene were chosen based on a genome-wide analysis in human cells {Barski, 2007 #207}. An identical binding pattern was observed in wild type MEFs, with relatively weak CTCF binding 2.3 kb upstream of the p21 promoter (up) and on the promoter (pr), and very strong binding on two adjacent regions within intron 1 (i-1a, i-1b).

still present in the DN pool because the *Ctcf* gene is deleted from DN2 onward (see Figure 2.1D). Thus, a small portion of DN cells still contains the *Ctcf* gene, and, consequently, *Ctcf* mRNA.

We next examined other T cell factors in the sorted cells. CTCF was reported to be a negative transcriptional regulator of c-Myc (Lobanenkov et al., 1990; Qi et al., 2003), and c-Myc is important for T cell development. In T cells we found that *c-Myc* was expressed in a different pattern than *Ctcf* in wild type cells and was not at all affected in CTCF-deficient cells (Figure 2.6B), showing that in T cells CTCF does not regulate expression of the *c-Myc* gene. GATA3 is critically involved in β-selection and development of CD4 SP cells (Pai et al., 2003), while SATB1 organizes cell type-specific nuclear architecture (Cai et al., 2006). Expression levels of these transcription factors mirrored the level of CTCF in wild type cells and were only somewhat reduced in CTCF-deficient cells (Figure 2.6B). These data suggest that neither GATA3 nor SATB1 is regulated by CTCF. PreTα expression was not affected in ISP cells and was up- rather than down-regulated in DP cells in the absence of CTCF (Figure 2.6B), indicating that this factor is not directly regulated by CTCF, but that its expression level is eventually strongly influenced by a *Ctcf* deletion. The expression of GIMAP4, which is induced by pre-TCR signaling and accelerates T-cell death (Schnell et al., 2006), was also increased in CTCF-deficient T cells (Figure 2.6B).

As ISP cells from *lck-Cre Ctcf$^{f/f}$* mice are blocked in the cell cycle, we next tested expression of two major cell cycle inhibitors, p21 and p27. In wild type cells, the expression of both factors mirrored that of the *Ctcf* gene, indicating that in ISP cells CTCF might act as a repressor of the *p21* and *p27* genes (Figure 2.6B). Strikingly, *Ctcf* knockout cells showed significantly increased *p21* and *p27* expression, consistent with a cell cycle arrest and defective proliferation of CTCF-

deficient β-selected cells at the ISP stage. The expression of *p21* appeared to be particularly tightly regulated by CTCF levels, i.e. in wild type cells p21 expression exactly mirrored that of CTCF and in cells from *lck-Cre Ctcf/f* mice the drop in *p21* was mild (~2-fold) in DN cells (where residual CTCF is still present), whereas in ISP cells *p21* expression was increased more than 50-fold (Figure 2.6B). Our data suggest that cell cycle progression in β-selected CTCF-deficient T cells is blocked due to the upregulation of the *p21* and *p27* genes.

To test whether p21 and p27 upregulation is a general consequence of a reduction of CTCF levels, we treated mouse embryonic fibroblasts (MEFs) from *Ctcf*$^{f/f}$ mice (Van de Nobelen et al, manuscript in preparation) with Cre recombinase (Splinter et al., 2006). Expression of *p21*, but not *p27*, was increased in the absence of CTCF (Figure 2.6C), consistent with a proliferation defect in CTCF-deleted MEFs (Van de Nobelen et al, manuscript in preparation). These data indicate that the *p21* gene is a target of CTCF in different cell types. A genome-wide CTCF-binding site analysis (Barski et al., 2007) suggested that the human *p21* gene contains four CTCF binding sites in the vicinity of its promoter, including two adjacent and very strong binding sites in intron 1. Chromatin immunoprecipiations (ChIP) in MEFs on the corresponding regions of the mouse *p21* gene revealed a virtually identical binding pattern of CTCF (Figure 2.6D). Our experiments were performed with different anti-CTCF antibodies than those used in human cells (Barski et al., 2007) verifying the CTCF binding pattern near the promoter of the *p21* gene. Thus, CTCF binding appears conserved across cell types and between species; we therefore propose that CTCF regulates *p21*.

CTCF is required in actively dividing cells

CTCF-deficient ISP cells were small and, with some notable exceptions, many of the genes that we tested in the RT-PCR were downregulated. These results indicate that besides a cell cycle block other defects contribute to the ISP phenotype of *lck-Cre Ctcf*$^{f/f}$ mice. We performed several experiments to examine this in more detail. First, CTCF has been proposed to be required for the maintenance of methylation at the *Igf2/H19* locus (Schoenherr et al., 2003). Aberrant methylation of this locus, or of other loci, might cause defects in T cells. As CTCF deletion is highly efficient in T cells from *lck-Cre Ctcf*$^{f/f}$ mice and DNMT1, a maintenance methyltransferase with an important role in T cell development (Lee et al., 2001), is not significantly affected by deletion of CTCF in thymocytes (Figure 2.1C), we used these cells to examine DNA methylation in the *Igf2/H19* locus in the absence of CTCF. We found similar methylation of a CTCF-binding site in the *Igf2/H19* imprinted locus in wild type and CTCF-deleted thymocytes (Supplementary Figure S2.3A). We also examined the methylation status of the ribosomal DNA (rDNA) repeats in *Ctcf* knockout cells, as these repeats have been shown to be heavily methylated (Bird et al., 1981), and CTCF binds to a region of the rDNA repeat upstream of the transcription start site (Van de Nobelen et al, manuscript in preparation). We could not detect differences in rDNA methylation in the thymus (where CTCF is virtually absent) and spleen (where CTCF is not deleted) from *lck-Cre Ctcf*$^{f/f}$ mice (Supplementary Figure S2.3B). We conclude that a deletion of CTCF does not lead to aberrant methylation and that this cannot underlie the defects observed in *Ctcf*-negative ISP cells.

It is known that the paralogue of CTCF, named CTCF-L or BORIS, can bind the same DNA sequences as CTCF (Loukinov et al., 2002), and it has been suggested that CTCF-L can compete with CTCF and might antagonize its function (Vatolin et al., 2005). However, we did not detect *Ctcf-l* mRNA in T cells in the presence or absence of CTCF (Supplementary Figure S2.3C). Thus, the defects observed in T cells are not due to CTCF-L acting in a dominant-negative fashion in cells lacking CTCF.

Nucleophosmin is a multifunctional nuclear-cytoplasmic shuttling protein with an important role in cell division and, when mutated, in cancer (Grisendi et al., 2006). It has been proposed that insulator sequences are tethered to the peri-nucleolar rim by nucleophosmin, which interacts with CTCF and which binds insulators in a CTCF-dependent manner (Yusufzai et al., 2004). Importantly, the expression of nucleophosmin increases dramatically upon stimulation of T cells (Feuerstein et al., 1988). We examined nucleophosmin distribution in cultured T cells. These were derived from *Ctcf*$^{f/f}$ mice that were crossed with a CD4-Cre transgenic line (Lee et al., 2001; Wolfer et al., 2002), as much higher numbers of naive CTCF-deficient T cells could be obtained from these mice (Ribeiro de Almeida et al, manuscript in preparation). Consistent with other studies (Feuerstein et al., 1988) stimulation of wild type CD4$^+$ T cells with anti-CD3 and anti-CD28 caused a strong increase in the level of nucleophosmin after 1 day of culture, which resulted in the abundant accumulation of nucleophosmin at the perinucleolar rim (Figure 2.7A). These data indicate that nucleophosmin-mediated subnuclear tethering of insulators might be prominent in activated T cells. Strikingly, both the level as well as the distribution of nucleophosmin was severely affected in *Ctcf* knockout cells (Figure 2.7A), suggesting that CTCF controls nucleophosmin behaviour in activated T cells.

We could recently show that CTCF binds the ribosomal DNA (rDNA) repeat and regulates its epigenetic state (Van de Nobelen et al, manuscript in preparation). We therefore visualized the organization of rDNA repeats in CTCF-deficient T cells by fluorescent in situ hybridization (FISH) with an rDNA probe (Akhmanova et al., 2000). An example of such an experiment is shown in Supplementary Figure S2.4. Based on the FISH signals we counted the number of rDNA dots in CD4$^+$ and CD8$^+$ T cells, derived from *Ctcf* knockout (KO) and wild-type (WT) mice (Table 2.2). We found a very small shift towards a lower number of dots in *Ctcf* knockout cells. These

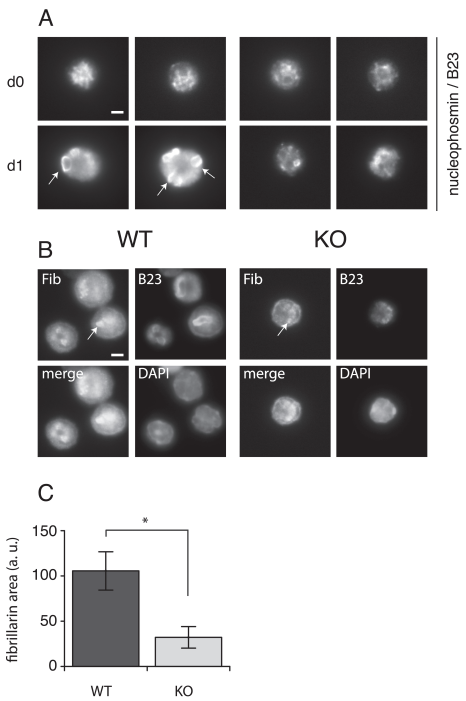

Figure 2.7 CTCF deletion affects nuclear organization in activated T cells

A, B. Immunofluorescent analysis of activated T cells. Cells were stimulated for 1 day and subsequently fixed and analyzed with antibodies against the indicated proteins. WT: T cells derived from wild type mice, KO: T cells derived from *CD4-Cre Ctcf*$^{f/f}$ mice. In panel (A) the arrows point to the intensely bright staining of nucleophosmin covering the perinucleolar rim. In panel (B) the arrows point to fibrillarin-positive nucleoli.

C. Quantification of fibrillarin-positive nucleolar signal. Cells were treated as above. Images were collected and a threshold was applied with Image J. The fibrillarin-positive surface area above the threshold was calculated and plotted (SEM is indicated). An unpaired t-test showed a significant ($P<0.05$) difference between WT and KO cells in fibrillarin-positive areas.

CD4+	number of rDNA dots (%)								
	0	1	2	3	4	5	6	7	Total
WT	0	2	17	20	32	22	5	1	100 (n=358)
KO	0	3	16	29	26	18	6	1	100 (n=368)

CD8+	number of rDNA dots (%)								
	0	1	2	3	4	5	6	7	Total
WT	1	5	9	27	26	21	7	3	100 (n=351)
KO	0	6	12	27	28	18	8	1	100 (n=354)

Table 2.2 Number of nucleoli in wild-type and CTCF-negative T cells

data indicate that in naive resting T cells nucleolar organization is not dramatically perturbed. However, the disorganized state of nucleophosmin (which normally abuts the nucleolus) in activated *Ctcf* knockout T cells suggested that CTCF might play a role in rRNA output upon T cell activation. Fibrillarin is an essential nucleolar enzyme involved in pre-rRNA processing (Newton et al., 2003) and an excellent tool to mark nucleolar activity. We therefore measured the intensity of fibrillarin staining in T cells from *Ctcf* knockout (KO) and wild-type (WT) mice after one day of *in vitro* activation. Whereas in wild type cells we detected increased fibrillarin staining, KO cells were unable to upregulate this marker of nucleolar activity (Figure 2.7C), suggesting that rRNA synthesis is hampered in activated T cells lacking CTCF. Thus, lack of CTCF causes a defective nuclear and nucleolar organization in fast growing T cells.

CTCF levels correlate with cell size

The mRNA expression data (Figure 2.4B) indicate that *Ctcf* is specifically upregulated in ISP cells. This result is consistent with lacZ staining data that indicate increased expression from the *Ctcf* gene promoter in ISP cells (not shown). However, both the RT-PCR and LacZ staining results reflect mRNA rather than protein levels of CTCF. To analyze the dynamic behaviour of CTCF *in vivo* we used a $Ctcf^{gfp}$ knock-in allele in which GFP-CTCF is expressed instead of CTCF (Heath et al, manuscript in preparation). We isolated T cells from mice carrying the *GFP-Ctcf* knock-in allele and used flow cytometry with markers for CD3, CD4 and CD8, as well as CD44, CD25, and TCRγδ, to subdivide cells from thymus and spleen and to identify GFP-CTCF levels in the different T cell compartments (Supplementary Figure S2.5). Interestingly, during αβT cell differentiation CTCF levels increased 2-3 times from the DN to the ISP stage (Supplementary Figure S2.5A). CTCF levels decreased again in DP cells (Supplementary Figure 2.5A). These data are in line with the mRNA expression pattern of *Ctcf*. In TCRαβ $CD4^+$ and $CD8^+$ T cells in the spleen, CTCF levels were identical to the levels in SP cells in the thymus. CTCF was somewhat up in γδT cells (Supplementary Figure S2.5B). Plotting GFP intensity versus T cell size (forward scatter) revealed a strict correlation between the amount of CTCF and cell size (data not shown). Collectively, these data indicate that CTCF levels specifically increase in large T cells with a high proliferative capacity. It is noteworthy that the major block in T cell differentiation in *lck-Cre* $Ctcf^{f/f}$ mice is observed in those cells that normally express high levels of CTCF.

Discussion

Here we report on the *in vivo* function of CTCF, a protein involved in chromatin organization and epigenetic regulation of gene expression. Consistent with previous studies (Fedoriw et al., 2004), we find that a knockout of *Ctcf* in early embryonic development is lethal. We also show that CTCF is essential for β-selected T cells. Somewhat surprisingly CTCF is neither required for γδ T cell development in mice, nor for proliferation of these cells *in vitro*. However, this result is similar to what has been reported for DNMT1 (Lee et al., 2001) and for the RNaseIII enzyme Dicer (Cobb et al., 2006). Apparently cell division, the regulation of chromatin structure, and gene expression in γδ T cells are quite different from αβ T cells, as they do not depend on proteins like CTCF, DNMT1 or Dicer, proteins that are absolutely essential in many other cell types.

It should be noted that we did not observe a more severe phenotype in CTCF-deficient female thymocytes compared to male cells suggesting that absence of CTCF does not cause deregulated expression of genes on the inactive X-chromosome. Also, the deletion of CTCF did not have an effect on the maintenance of methylation in the imprint control region of the *Igf2/H19* locus. Unlike other studies (Filippova et al., 2005; Schoenherr et al., 2003), our results therefore indicate that CTCF is not required to maintain X-inactivation and DNA methylation status of the *Igf2/H19* locus. Furthermore, within the T-cell lineage, we found that CTCF is not essential for V(D)J recombination at the TCRα or TCRβ loci. The developmental block of CTCF-deficient ISP into DP cells was not rescued when we crossed Lck-Cre *Ctcf*$^{f/f}$ mice with transgenic mice expressing a pre-rearranged αβ TCR. Therefore, the arrest of CTCF-deficient DP cells cannot result from impaired TCRα gene rearrangement. Because also the TCRγ and δ loci could undergo functional V(D)J recombination in the absence of CTCF, we conclude that the multiple CTCF-binding sites reported to be present in TCR loci (Barski et al., 2007; Garrett et al., 2005; Magdinier et al., 2004) are not essential for the process of V(D)J recombination.

Our findings show that CTCF is absolutely required for the efficient proliferation of β-selected cells, in particular for their maturation from ISP to DP cells, and for TCR up-regulation at the cell surface of DP cells. In line with the proliferative block, we detected a strongly increased expression of two major cell cycle inhibitors, p21 and p27. As T cells were directly isolated from mice, these data provide the first *in vivo* evidence for an important role of CTCF in cell cycle progression. Several lines of evidence suggest that the *p21* gene is a target of CTCF. First, ChIP analysis shows that CTCF binds the *p21* promoter region very efficiently. Thus, CTCF binding is conserved between man and mouse. Second, in wild type T cells that differentiate from the DN stage towards ISP and, later on, to DP cells, *p21* expression is inversely correlated with CTCF levels. These data indicate that CTCF functions to repress *p21* expression. As a major CTCF binding site is located within the *p21* gene, the repressive mechanism may simply be the blocking of RNA polymerase II elongation. Interestingly, CTCF was found to interact with the largest subunit of RNA Polymerase II (Chernukhin et al., 2007). Third, in knockout T cells the *p21* gene is massively induced, consistent with a repressive role of CTCF. Fourth, the level of *p21* induction appears to correlate with the reduction in the dose of CTCF. Thus, in DN cells, where some residual CTCF is still present, the induction of *p21* is less dramatic (~2-fold increase in knockout cells compared to wild type cells), than at the ISP stage (>50-fold induction). Fibroblasts lacking CTCF also upregulate *p21*, although the increase in knockout versus wild type cells is much less than in T cells. The latter results show that CTCF is not acting alone in its control of *p21* expression. Others have recently shown that *p21* is regulated by the differential localization of histone H2A.Z (Gevry et al., 2007). H2A.Z-binding sites were found to be enriched near CTCF-sites (Barski et al., 2007). Thus, it is tempting to speculate that CTCF and H2A.Z co-regulate *p21*.

Our results on *p21* regulation in T cells are completely opposite to those obtained in WEHI 231 B lymphoma cells, where conditional expression of CTCF resulted in the up- rather than the downregulation of *p21* and *p27*, while reduction of CTCF levels decreased (rather than increased) the expression of *p21* and *p27* (Qi et al., 2003). This could be due to the fact that the properties of the ISP thymocytes and WEHI 231 B cells are entirely different: whereas ISP thymocytes are highly proliferating as a result of pre-TCR stimulation, crosslinking of the B cell receptor on WEHI 231 immature B cells results in cell-cycle arrest and apoptosis (thereby providing a model for self-tolerance by clonal deletion). Thus, it may be possible that CTCF function is context-dependent. However, we note that our data were obtained *in vivo*, whereas the data in the WEHI 231 B lymphoma cells were obtained with stable transfectant clones selected for high expression of *CTCF* sense or antisense mRNA (Qi et al., 2003). Using our conditional *Ctcf*$^{f/f}$ mice in combination with existing B cell-specific Cre transgenes, we will be able to examine the interesting question whether CTCF has a different function in the cell cycle in B and T cells.

Despite the effects of a CTCF deficiency on mice and T cells, we found no evidence for an increased tumor incidence in heterozygous knockout animals, or for T lymphoid malignancies in CTCF-deficient T-cell lineages. These results argue against a role for CTCF as a crucial tumor suppressor (Klenova et al., 2002). Increased expression of p21 and p27 in CTCF-negative cells would explain why loss of CTCF does not induce tumours. However, we also show that CTCF controls nucleophosmin levels and behaviour. The nucleophosmin gene (NPM1) has been implicated in cancer pathogenesis, both as a putative proto-oncogene and tumor suppressor gene (Grisendi et al., 2006). NPM1 is frequently mutated in acute myeloid leukemia (AML), whereas deletion of the NPM1 gene is associated with myelodysplastic syndrome (MDS). Recently it was found that reduction in nucleophosmin in mice results in malignancies at higher frequency compared to wild type mice (Sportoletti et al., 2008). Thus, the potential role of CTCF in cancer merits a more detailed investigation.

Heterozygous *Ctcf* knockout mice are viable and fertile but are born in less than expected numbers. Thymocyte development is also affected when CTCF dosage is reduced by half. On the other hand γδ T cell express CTCF but do not require the protein for viability. Combined these results suggest that CTCF is required in a cell type- and dosage-dependent manner. Consistent with this idea we found, using two independent methods (RT-PCR and GFP-CTCF expression) that CTCF levels vary *in vivo* during T cell differentiation and that this is inversely correlated with *p21* expression. It was recently proposed that CTCF remains bound to most of its ~14,000 cognate binding sites irrespective of cell type (Kim et al., 2007). CTCF binding sites largely overlap with those of cohesin and deletion of CTCF affects cohesin binding (Parelho et al., 2008; Wendt et al., 2008). CTCF may work together with cohesin in setting up insulators (Parelho et al., 2008; Wendt et al., 2008). We show here that CTCF regulates nuclear organization by acting on nucleophosmin, which was proposed to tether insulators together with CTCF (Yusufzai et al., 2004). We also found that CTCF controls rRNA output in activated T cells. Accordingly, CTCF-negative ISP cells are smaller than their wild type counterparts. We therefore propose that CTCF function is more critical for actively proliferating and growing cell types than for non-dividing or resting cells. Indeed, we have found that deletion of CTCF does not affect nuclear organization in confluent fibroblasts (Van de Nobelen et al, manuscript in preparation) and in resting T cells (Ribeiro de Almeida et al, manuscript in preparation). It will be interesting to determine how CTCF performs its cell type-specific roles while remaining bound to cognate sites and how variations in the level of CTCF influence its dynamic behavior and function.

Materials and methods

Modified Ctcf alleles, mouse models and embryonic fibroblasts.

Human CTCF cDNA was used to screen a 129S6/SvevTac mouse PAC library (RPCI-21) (Osoegawa et al., 2000). PAC clones were used to isolate 6.7kb (for 5' end targeting) and 8kb (for 3'end targeting) EcoRI subclones. For 5'end targeting the 6.7kb EcoRI fragment was used to amplify 1360 bp of 5' end homology and 5340 bp of 3' end homology. The homologous arms were cloned into a vector containing the neomycin resistance gene flanked by loxP-sites (Hoogenraad et al., 2002). A viral thymidine kinase gene was inserted afterwards. For 3' end targeting we generated a SpeI-EcoRI subclone from the PAC DNA and used its unique BamHI site to insert a cassette containing the puromycin resistance gene flanked by loxP sites, followed by splice acceptor sequences and the bacterial β-galactosidase (lacZ) reporter (Hoogenraad et al., 2002). Relevant parts of the different constructs were verified by DNA sequencing.

Constructs were targeted into E14 embryonic stem (ES) cells as described (Hoogenraad et al., 2002). DNA from resistant ES cells was analyzed with external radiolabeled probes by Southern blotting. Confirmation of homologous recombination was performed using different 5' end and 3' end probes (Figure 2.1A, B) and a PCR-based assay for genotyping. $Ctcf^{f/f}$ mice were maintained on a C57BL/6 background. Mouse embryonic fibroblasts (MEFs) were isolated from $Ctcf^{f/f}$ mice using published procedures (Akhmanova et al., 2005). A more detailed characterization of the MEFs will be described elsewhere (Van de Nobelen et al, manuscript in preparation). Fibroblasts were treated with lentiviral Cre constructs as described (Splinter et al., 2006).

$Ctcf^{f/f}$ mice were bred to mice expressing chicken β-actin-Cre generating $Ctcf^{+/-}$ animals. T cell specific deletion of Ctcf was achieved by breeding to Lck-Cre (Lee et al., 2001), which were kindly provided by Dr. C. Wilson (University of Washington, Seattle, USA). Cre-specific primers were used for genotyping. $HY/Rag2^{-/-}$ (C57BL/10) mice were purchased from Taconic Europe A/S (Denmark). OT-II mice have been described (Barnden et al., 1998). Mice were bred and maintained in the Erasmus MC animal care facility under specific pathogen-free conditions and analyzed at 6-10 weeks of age. Experimental procedures were reviewed and approved by the Erasmus University committee of animal experiments.

DNA, RNA and protein analysis.

Genomic DNA was isolated, digested, and blotted onto Hybond N+ membranes (Amersham), and hybridized with radio-labeled probes. *Ctcf* probes are shown in Figure 2.1. Total RNA was prepared using RNA-Bee RNA isolation solvent (Tel-Test Inc.). RNA (0.5-1.0µg) was reverse transcribed (RT) with random and oligo-dT primers, in the presence of Superscript reverse transcriptase (Invitrogen). Real-time RT-PCR was performed as described (Splinter et al., 2006) with 100ng of each primer and 0.5 units of Platinum taq DNA polymerase (Invitrogen). Sybr-green (Sigma) was added to the reactions and PCR was performed on a DNA Engine Opticon PCR system (MJ Research Inc.) and Bio-Rad MyiQ iCycler single-color real-time PCR detection system. To confirm the specificity of the amplification products, samples were separated by standard agarose gel electrophoresis. Threshold levels were set and further analysis was performed using the SDS v1.9 software (Applied Biosystems). The obtained C_t values were normalized to the C_t value of Gapdh or β-actin. Each PCR was performed at least in triplicate. Primer sequences and PCR conditions used are available on request.

Nuclear extracts were prepared and analyzed by chromatin immunoprecipitation (Splinter et al., 2006), or western blot (Hoogenraad et al., 2002). Primary antibody incubation was done overnight at 4ºC in Tris-buffered saline (TBS), containing 5% (w/v) BSA and 0.15% (v/v) NP-40. Blots were incubated with secondary goat anti-rabbit or -mouse antibodies, coupled to horseradish peroxidase (GE Healthcare UK Ltd: 1:50000). Signal detection was performed using ECL (Amersham). Anti-CTCF (N3) and anti-fibrillarin antibodies were generated as described (Hoogenraad et al., 2000) using GST-linked chicken CTCF (amino acids 2-267) and mouse fibrillarin fusion proteins, were used in a 1:300 dilution. DNMT1 (Abcam) and UBF (Santa Cruz Biotechnology) mAbs were used 1:100. Western blots were scanned and the levels of protein were quantified using the gel macro function in ImageJ (Rasband, W.S., NIH, http://rsb.info.nih.gov/ij/). The

amount of CTCF was normalized to DNMT1 in the same sample.

Flow cytometric analyses.

Preparation of single-cell suspensions, FDG-loading, mAb incubations for four-color cytometry have been described (Hendriks et al., 1996). All mAbs were purchased from BD Biosciences (San Diego, CA). Samples were acquired on a FACSCalibur™ flow cytometer and data was analyzed using CellQuest™ software (BD Bioscience).

For cell cycle profiles of thymic subsets, cells were first stained for surface markers, fixed with 0.25% paraformaldehyde and permeabilized with 0.2% Tween 20. Next, 7-AAD was added to a final concentration of 15 µg/ml in PBS. Cell cycle status of T cell cultures was determined after fixing in ice-cold ethanol and subsequent staining in PBS, containing 0.02 mg/ml propidium iodide, 0.1% v/v Triton X-100 and 0.2 mg/ml RNAse. Doublet cells were excluded by measuring peak area and width.

FACS sorting of DN, ISP and DP cells was performed with a FACSVantage VE equipped with Diva Option and BD FACSdiva software (BD bioscience). The purity of fractions was >98%

In vitro T cell cultures

For in vitro T cell cultures from lck-Cre $Ctcf^{f/f}$ mice, either total T cells fractions (Figure 2.4A) were purified from lymph node by MACS depletion (using anti-B220, anti-NK1.1, anti-Ter119, anti-CD11b and anti-Gr-1 antibodies) or TCRαβ enriched T cell fractions (Figure 2.4B) were purified by MACS depletion using the same antibody mix supplemented with anti-TCRγδ antibodies (Figure 2.4B, C). Purity of obtained fractions was >98%, but the TCRαβ enriched T cell fractions still contained $TCR\gamma\delta^{low}$ T cells (<2% in WT and ~30% in CTCF KO mice). T cells were cultured at a concentration of 1 x 10^6 cells/ml in Iscove's modified Dulbecco's medium (IMDM) (Bio Whittaker, Walkersville, MD) containing 10% heat-inactivated FCS, $5x10^{-5}$M β-mercaptoethanol, 100 U/ml penicillin and 100 µg/ml streptomycin. Stimulation was with plate-bound anti-CD3 (145-2C11) and anti-CD28 (37.51) mAbs (coated at 10 µg/ml each at 4°C overnight for 7 days.

Fluorescent in situ hybridization (FISH) and immunofluorescence (IF)

FACS sorted naïve $CD62L^+$ $CD4^+$ T cells from *CD4-Cre* $Ctcf^{f/f}$ mice were cultured for 1 day (Ribeiro de Almeida et al, manuscript in preparation) and allowed to attach to glass slides for 30 min. Cells were fixed for 10 min with 4% PFA/PBS. For IF slides were used immediately. For DNA-FISH slides were stored in 70% EtOH until further use. IF procedures have been described (Hoogenraad et al., 2000). We used the anti-fibrillarin antibodies described above and anti-nucleophosmin antisera (Santa Cruz, sc-32256). As secondary antibodies we used FITC-conjugated goat antibodies against rabbit and mouse IgG (Nordic Laboratories) and Alexa 594-conjugated goat anti-rabbit antibody (Invitrogen) were used. Images were collected with a Leica DMRBE microscope equipped with a Hamamatsu ORCA ER camera, or with a Zeiss LSM510 confocal, as described (Akhmanova et al., 2005).

For DNA-FISH cells were pretreated by two PBS wash-steps followed by a permeabilization step of 4 min incubation in 0,1% pepsin in 0,01M HCl at 37 ˚C. Slides were washed once in PBS on ice and fixed again for 5 min in 4% PFA/PBS. Slides were washed twice in PBS and dehydrated. Denaturation was done for 2 min at 80 ˚C in denaturing solution (70% formamide; 2xSSC; 10 mM phosphate buffer, pH 7), after which the slides were cooled in 70% EtOH, dehydrated and hybridised as described (Gribnau et al., 2005). The rDNA probe (an 11.8 kb SalI fragment of a murine rDNA cosmid which contains non transcribed rDNA only (Akhmanova et al., 2000) was DIG labelled by nick translation (Roche). We used a Zeiss Axioplan 2 microscope for image acquisition and cell counting. Cells were counted without knowledge of the genotype, statistical significance was tested with the chi-square tool (Excel), with a p-value of 0.005.

Acknowledgements

We thank An Langeveld, Harmjan Kuiper, Janneke Samsom, Lisette van Berkel, Edwin de Haas, Tom Schoonerwille, Marjolein de Bruin and Jan Piet van Hamburg for their assistance. This work was supported

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

Supplementary data

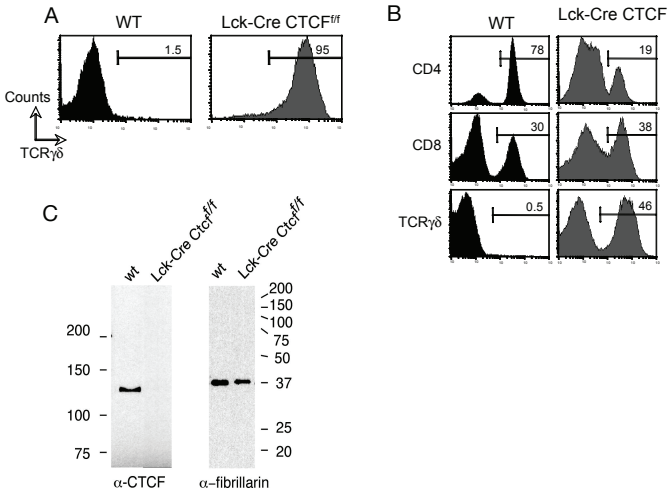

Figure S2.1 Selective advantage of TCRγδ cells in anti-CD3/CD28-stimulated cultures of Lck-Cre Ctcf$^{f/f}$ mice
A. Flow cytometric analysis of TCRγδ expression in T cell cultures from wild type (WT) and Lck-Cre *Ctcf$^{f/f}$* mice. Lymph node fractions were stimulated by anti-CD3/CD28 and cultured for 7 days. The percentages represent the fractions of TCRγδ$^+$ T cells. The proportions of γδ$^+$ T cells in the T-cell enriched cell suspensions before culture was <2% in WT and ~30% in Lck-Cre *Ctcf$^{f/f}$* mice (see also Figure 2C). **B.** Flow cytometric analysis of CD4, CD8 and TCRγδ expression in mixed T cell cultures from wild-type and Lck-*Cre CTCF$^{f/f}$* mice. Lymph node cell fractions were enriched for CD4 and CD8 cells and depleted for TCRγδ$^+$ T cells, stimulated by anti-CD3/CD28 and cultured for 7 days. **C.** Western blotting, showing the absence of CTCF protein in mixed T cell cultures from Lck-Cre *CTCF$^{f/f}$* mice. Fibrillarin was used as a loading control. Molecular weight markers are indicated in kD.

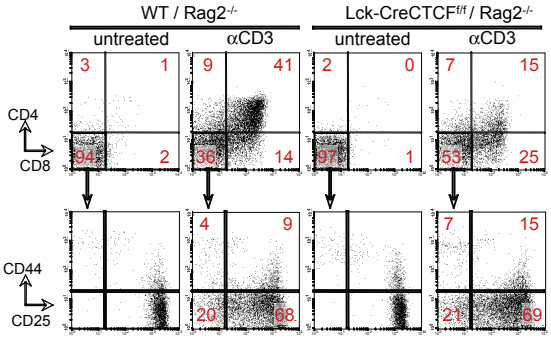

Figure S2.2 The arrest of CTCF-deficient thymocytes is independent of TCRα rearrangement
Flow cytometric analyses of the thymus of the indicated mice, which were either untreated or injected with 50 µg of rat anti-CD3 antibodies *in vivo*. CD4/CD8 expression profiles, 3d after injection, are shown as dot plots (*upper part*). DN cell populations were gated and analyzed for CD25 and CD44 (*lower part*). Data are shown as dot plots and the percentages of cells within the quadrants are given.

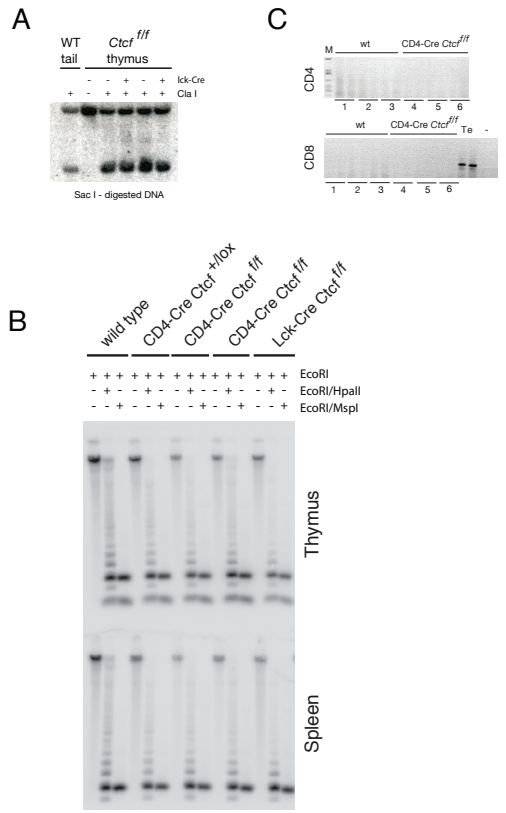

Figure S2.3 DNA methylation and CTCFL expression in T cells

A. DNA methylation analysis in the Imprinting Control Region (ICR) of the *Igf2/H19* locus. DNA was isolated from the thymus of *Ctcf*$^{f/f}$ mice, either not crossed (-, lanes 3 and 5) or crossed (+, lanes 4 and 6) with *Lck-Cre* transgenics. Samples were digested with SacI only (lane 2) or with both SacI and ClaI (other lanes). ClaI cuts within CTCF binding site 1 of the *Igf2/H19* ICR {Schoenherr, 2003 #158}. Normal tail DNA (lane 1) is shown as control. **B.** Methylation status of ribosomal DNA (rDNA) repeats. Southern blot analysis of genomic DNA from thymus and spleen of the indicated mice was digested with *Eco*RI (*lanes 1*), *Eco*RI and *Hpa*II (*lanes 2*), or (3) *Eco*RI and *Msp*I (*lanes 3*) and hybridized with the unstable 5' external transcribed spacer probe {Akhmanova, 2000 #227}. **C.** RT-PCR analysis for CTCF-L/BORIS expression in sorted naive peripheral CD62L$^+$ CD4$^+$ and CD8$^+$ T cell fractions from three wild-type (1-3) and three CD4-Cre *CTCF*$^{f/f}$ (4-6) mice. RNA was reverse transcribed, serially diluted, and used as a template for amplification. Amplification of testis cDNA samples (Te) and RNA samples without RT (-) were performed as controls. All samples were done in duplicate. Products were fractionated by gel electrophoresis and detected with ethidium bromide. M; molecular weight marker.

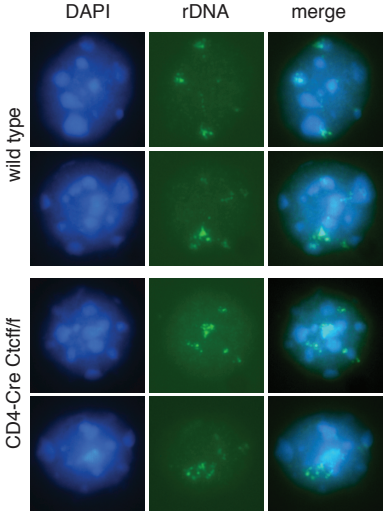

Figure S2.4 CTCF deletion does not affect nucleolar organization in resting T cells

FISH analysis of nucleolar organization in FACS-sorted naive CD62L$^+$ peripheral T cells from the indicated mice. Slides were hybridized with a DIG-labeled rDNA probe (green) containing non transcribed rDNA {Akhmanova, 2000 #227}. Cells were counterstained with DAPI (blue).

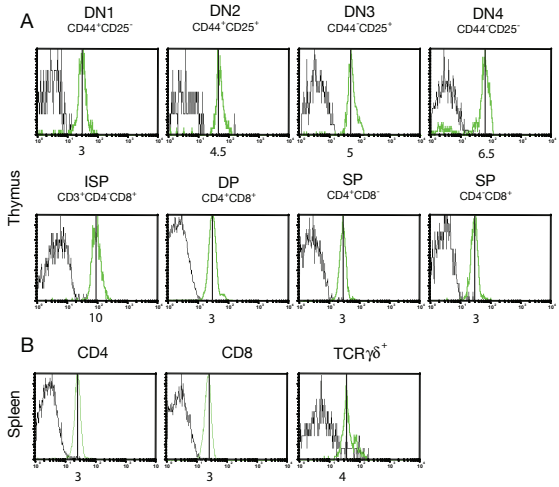

Figure S2.5 Flow cytometric analysis of GPF-CTCF protein expression

GPF-CTCF protein was analyzed, in conjunction with cell surface markers, in cell suspensions from thymus (**A**) and spleen (**B**) from mice carrying a green fluorescent protein *(GFP)-Ctcf* knock-in allele (*Ctcf*^{GFP}, H.H. *et al.,* manuscript in preparation, for targeting strategy see {Akhmanova, 2005 #205}. The indicated cell populations were gated and expression data are displayed as histogram overlays of GFP-CTCF *(green)* on top of background signals in wild type mice *(black)*.

3

Identification of CTCF protein partners using a proteomics approach

Work in progress

Identification of CTCF protein partners using a proteomics approach

Summary

CTCF binds to a wide variety of sites in the genome and is implicated in the regulation of gene expression by functioning as an activator, repressor or chromatin boundary factor. However, little is known about the molecular basis of CTCF function. Using a biotinylation tagging approach we isolated and characterized CTCF-interacting proteins in a developmental and cell type-specific manner. In addition to several known CTCF-interacting factors we describe novel proteins isolated from embryonic stem cells, thymus and lung. The identified protein partners were validated using different approaches. We have classified these proteins based on their biological functions. This analysis reveals potential roles of CTCF in the regulation of chromatin modifications, in cellular organization, and in transcription by RNA polymerases I and II. We also document tissue-specific transcriptional processes controlled by CTCF.

Introduction

CTCF is a ubiquitously expressed eleven zinc-finger transcription factor, which is able to bind to a wide variety of sites throughout the genome. Moreover the zinc-finger region is important for binding to different protein partners that are involved in modulating the multiple functions of CTCF. CTCF is a key player in the correct regulation of processes such as imprinted expression of *Igf2* and *H19* and X chromosome inactivation. A unique property of CTCF is its insulating capacity: it can form a boundary preventing the spread of heterochromatin, or, when placed in between elements, it can block an enhancer from activating a promoter. CTCF performs this unique function by forming chromatin loops and tethering sites to nuclear substructures (Wallace and Felsenfeld, 2007).

The mapping of CTCF binding sites has helped tremendously in interpreting its function (Kim et al., 2007). Also developmental and cell-type specific deletions of CTCF have revealed significant insight (Heath et al submitted). Another way to understand the multiple CTCF functions is by identification of interaction partners. In order to do this different approaches have already been used. Among them are yeast two hybrid, GST pull downs and Flag-HA two-step purification (Chernukhin et al., 2000; Defossez et al., 2005; Lutz et al., 2000; Yusufzai et al., 2004). Pull downs were performed with the full length CTCF protein as well as with the zinc-finger domain.

A number of CTCF-interacting partners have been identified to date. Of these, the SNF2-like chromodomain helicase CHD8 was shown to be important for CTCF insulator activity at the *H19* locus (Ishihara et al., 2006). Another interacting protein, called nucleophosmin, is involved in CTCF insulator function by tethering CTCF binding sites in the β-globin locus to the nucleolar periphery (Yusufzai et al., 2004). CTCF can also bind to YY1, a factor reported to act as an insulator by itself (Donohoe et al., 2007; Kim et al., 2006). Besides proteins involved in stimulating CTCF insulator activity, an interacting protein was identified that shows the opposite effect, i.e. *in vitro* enhancer blocking assay showed the transcription factor Kaiso to have a negative effect on CTCF insulator activity (Defossez et al., 2005). Finally, upon binding and modifying CTCF poly(ADP-ribose) polymerase 1 (PARP-1) changes the insulator activities of CTCF (Klenova and Ohlsson, 2005; Yusufzai et al., 2004).

Whereas a number of the proteins identified to interact with CTCF appear to be important for its unique chromatin insulating function, other interacting proteins are involved in modulating different activities of CTCF. For example, YB-1 and Sin3a are involved in regulating transcriptional repression by CTCF (Chernukhin et al., 2000; Lutz et al., 2000). CTCF interaction with the largest subunit of RNA polymerase II was suggested to be important for storage of this complex, or for initiation of transcription in promoter regions, or in pausing of transcriptional elongation (Chernukhin et al., 2007). Proteins suggested to interact with CTCF, but for which an interaction was not confirmed, are Topoisomerase II, Lamin A/C, Importin α3/α1, the histone chaperone Taf1-set and the variant histone H2A.Z (Yusufzai et al., 2004). Proteins that do not interact with CTCF have also been reported. These include the proto-oncogene Myc and the tumor-suppressors Rb1, P21 and P53. Regarding the involvement of CTCF in regulating RNA polymerase II some factors were excluded to interact with CTCF, such as TAF4 (TAFII 130), TBP and TFIIH. Moreover CTCF was shown not to interact with histone H2A or H3 or the thyroid hormone receptor TRα (Chernukhin et al., 2007; Chernukhin et al., 2000).

CTCFL is the testis-specific paralogue of CTCF (Loukinov et al., 2002), which is also expressed in some human tumors and tumor cell lines (Looijenga et al., 2006; Risinger et al., 2005; Vatolin et al., 2005). Recently a report was published describing unique interactions of this protein. CTCFL was reported to bind to the transcription factor SP1 and the histone methyltransferase Prmt7, neither of these proteins interact with CTCF (Chernukhin et al., 2007; Jelinic et al., 2006; Kang et al., 2007). Therefore, to date no common interacting proteins have been reported for CTCFL and CTCF.

Although a number of interaction partners of CTCF have been identified, we presume that the list is not complete. In order to gain more insight in the different CTCF-dependent molecular processes we aimed to identify more of its interaction partners. We used a biotin-tagging proteomics approach that involves the use of a small tag that can be biotinylated *in vivo* and that is able to bind with high affinity to streptavidin. In this way protein complexes can be identified in a single step high-affinity procedure. Proteins are subsequently identified by mass spectrometry. This strategy was previously shown to be highly efficient in isolating protein complexes from tagged factors present in the nucleus or in the cytoplasm (Lansbergen et al., 2006; Meier et al., 2006; Rodriguez et al., 2005). We have used this method to characterize new partners of CTCF. In chapter 4 of this thesis we describe UBF as a common interaction partner of CTCF and CTCFL. Here we describe the remainder of the proteins identified.

Results

Characterization of biotin-tagged CTCF

To determine the influence of a biotinylation tag on CTCF and CTCFL function, we tagged CTCF and CTCFL mouse cDNAs at the N- or C-terminus with a biotin-tag and transiently expressed each protein in HeLa cells (which express endogenous CTCF but which do not appear to express CTCFL), which were also transfected with the BirA biotin ligase. As shown in Figure 3.1A BirA biotinylates a specific lysine residue of the small biotin-tag. The biotinylation of tagged CTCF as well as CTCFL was visualized by streptavidin-HRP (Figure 3.1B). Because of the short length (23AA) of the biotin tag the difference in size between tagged and untagged proteins is minor and could hardly be visualized using antibodies against CTCF. Placing the biotin tag at the N- or C-terminus of CTCF or

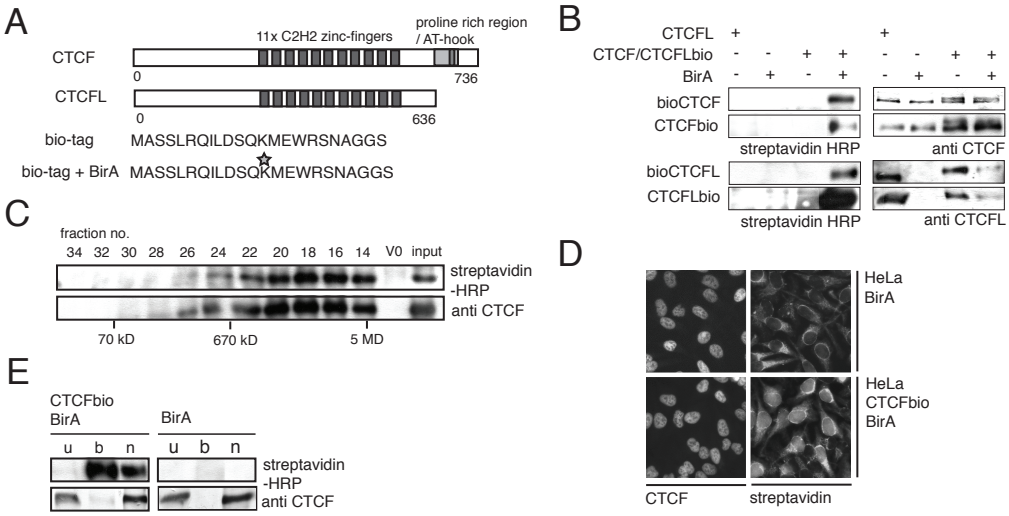

Figure 3.1 Biotinylation of CTCF and CTCFL

A. CTCF and CTCFL schematic overview of the proteins with domains indicated. Amount of amino acids is indicated below the schematic representation of each protein. The sequence of the 23-aa bio tag is shown. The star indicates the lysine residue that becomes biotinylated by the BirA biotin ligase. **B.** Transient biotinylation of N- and C-terminally tagged CTCF and CTCFL in HeLa cells. Western blot incubated with an anti-CTCF or anti-CTCFL antibody to detect endogenous and tagged CTCF or CTCFL protein (right). Western blot of the same extracts with streptavidin-HRP conjugate to detect biotinylated CTCF or CTCFL (left). Note that HeLa cells do not express endogenous CTCFL. **C.** Size-fractionation profiles by Superose 6 column of CTCF. Molecular mass markers are indicated below. V0: void volume, input : nuclear extract. Fractionation profile is shown of a MEL cell line stably expressing biotin-tagged CTCF and BirA (upper panel). The profile of tagged CTCF resembles that of untagged CTCF in stable MEL cell lines expressing only BirA (lower panel). **D.** Localization of stable expressed biotinylated CTCF and endogenous CTCF in HeLa cells. Immunofluorescent staining shows the expression of either endogenous CTCF or biotinylated CTCF in HeLA cell lines stably expressing CTCFbio and BirA or only BirA. Note the presence of endogenous biotinylated proteins present in the mitochondria. **E.** Streptavidin pull downs on MEL cells stably expressing biotinylated CTCF (left panel) or BirA only (right panel). U = unbound fraction; b = fraction bound to the beads after washes; n = nuclear extract used as input en represents 5% of bound material. Efficiency of CTCF biotinylation and binding to streptavidin beads is represented by this experiment.

CTCFL does not appear to affect the binding to streptavidin (Figure 3.1B).

Subsequently we performed mass spectrometry analysis on nuclear extracts of HeLa cells, transiently transfected with CTCF-bio and BirA. This failed to show CTCF itself, although western blot analysis did show that biotinylated CTCF was pulled down specifically. We therefore generated stable HeLa and MEL cell lines that overexpressed low amounts of CTCF. We isolated multiple CTCF expressing clones (unfortunately we could not identify any stable cell line expressing CTCFL). The clones expressing the highest level of CTCF-bio were used for further analysis.

It has been reported that CTCF is present in HeLa cells in high molecular weight complexes (up to about 400 kDa) (Yusufzai et al., 2004). This result was obtained with a 10-30% glycerol gradient. Using a Superose 6 gel filtration column we detected CTCF in complexes of >5 MDa in MEL cells (Figure 3.1C) as well as in ES cells (data not shown). CTCF-bio is also present in high-molecular weight complexes, indicating that tagged and endogenous CTCF behave in a similar manner in this assay.

We next tested whether biotinylated CTCF localized in a similar manner as CTCF. Using

streptavidin-based immunofluorescence, we could not recapitulate the typical speckled CTCF localization pattern in the nucleus of HeLa cells. Instead we detected a uniform distribution of CTCF-bio throughout the nucleus, indicating that in HeLa cells biotinylated CTCF might not mimic the behavior of endogenous CTCF (Figure 3.1D).

Streptavidin pull downs showed that in the clone expressing the highest level of biotinylated CTCF, this protein was still present in <10% of the endogenous CTCF (Figure 3.1E). An attempt to pull down CTCF-bio protein partners from this MEL cell clone was not successful. We presume that CTCF-bio expression levels in cells must be higher with respect to endogenous protein. Therefore, tagging of endogenous CTCF would be a preferred approach.

CTCF-bio expression from a knock-in allele

We generated a *Ctcf*$^{bio\text{-}neo}$ knock-in allele by targeting a construct with the biotinylation sequence (preceded by a TEV protease cleavage site) immediately upstream of the stop codon of CTCF, followed by a loxP-neo^r-loxP cassette, using homologous recombination in embryonic stem (ES) cells into the *Ctcf* gene (Figure 3.2A). It is important to remember that the neomycin resistance gene (neo^r) is still present in these ES cells, hence the name *Ctcf*$^{bio\text{-}neo}$. After selection targeted clones with a correct insertion of the construct were identified by PCR and Southern blot. The

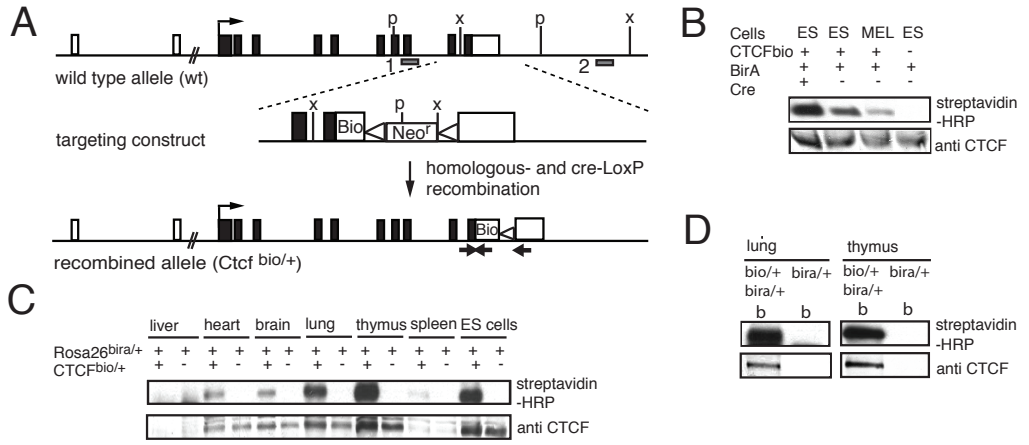

Figure 3.2 Generation of a CTCFbio knock-in mouse

A. Schematic overview of the targeting strategy used to generate a CTCFbio knock-in allele. The targeting constructs aims to remove the CTCF stop codon and place a biotin-tag (bio) at the complete end of the gene with in between a TEV protease cleavage site. The boxes represent the CTCF gene exons. Black boxes represent coding exons while white boxes indicate non-coding exons. Gray boxes represents the probes, probe 1 was used to select the correct targeted clones after XbaI digestion. Probe 2 was used to select the correct targeted clones after PstI digestion. Black arrows indicate primers used for genotyping. The triangles surrounding the neomycin (Neo) resistance cassette represent LoxP sites. **B.** Western blot showing expression levels of biotinylated CTCF in a stable MEL celline and in the different ES cell lines one containing the neomycin cassette that is placed in the antisense direction and one without the neomycin cassette. **C.** Western blot showing expression of biotinylated CTCF in nuclear extracts isolated from different adult mouse tissues. Incubation with streptavidin HRP and subsequent detection shows only biotinylated CTCF that is absent in the BirA only expressing tissues. CTCF antibody shows the presence of the protein in the samples and serves as a loading control. **D.** Streptavidin pull downs on biotinylated CTCF present in nuclear extracts isolated from lung and thymus of mice heterozygous for bira and CTCFbio. Biotinylated CTCF is specifically pulled down and not present in control bira lanes. b= bound fraction.

Ctcf$^{bio\text{-}neo/+}$ ES cell line with a correct karyotype was injected into blastocyst to generate *Ctcf*$^{bio\text{-}neo/+}$ knock-in mice.

A second targeting event was performed in *Ctcf*$^{bio\text{-}neo/+}$ ES cells by inserting the BirA biotin ligase into the *Rosa26* locus again using homologous recombination. This targeting event was verified by PCR. We showed that in this line biotinylated CTCF (CTCF-bio) was expressed and was readily detected. This ES cell line was used for the identification of CTCF interacting proteins together with a control cell line expressing only BirA (Exp 1; Table 3.2). Since the neomycin resistance gene was still present in the *Ctcf*$^{bio\text{-}neo/+}$*;Rosa26*$^{bira/+}$ cell line and transcribed antisense to CTCF-bio we argued this could inhibit CTCF-bio expression. In order to remove the neomycin resistance cassette we transfected ES cells with Cre recombinase. This yielded *Ctcf*$^{bio/+}$*;Rosa26*$^{bira/+}$ ES cells. Comparison between this ES cell line and the parent line indicated that in the latter expression of CTCF-bio was slightly reduced, however the expression levels were significantly higher compared to the stably expressing MEL cell lines generated before (Figure 3.2B).

Heterozygous *Ctcf*$^{bio\text{-}neo/+}$ knock-in mice were crossed with transgenic mice ubiquitously expressing Cre recombinase. This yielded *Ctcf*$^{bio/+}$ knock-in mice. These were in turn crossed with mice expressing the BirA biotin ligase from the ubiquitously active *Rosa26* locus (Driegen et al., 2005). The expression level of BirA was shown to be sufficient to biotinylate proteins *in*

A

Genotype Ctcf	Rosa 26	Age of death	Total number of mice born	Total number of breedings	Breedings without offspring
bio/bio	bira/bira	21	2	5	5
bio/bio	bira/+	16, 16, 34	8	4	4
bio/bio	+/+	30, 36	12	43	34

B

		Genotype parents (Ctcf ; Rosa 26)					
		bio/+ X bio/+ and no bira (n=29)		bio/+ X bio/+ and bira/+ or bira/bira (n=36)		bio/+ X bio/bio and bira/+ or bira/bira (n=25)	
Genotype offspring Ctcf	Rosa 26	%	Predicted %	%	Predicted %	%	Predicted %
+/+	+/+	24.1	25	11.1	7.6		
+/+	bira/+			5.5	12.5		
+/+	bira/bira			8.3	4.8		
bio/+	+/+	48.2	50	13.8	15.2	12	20
bio/+	bira/+			30.5	25	64	30
bio/+	bira/bira			8.3	9.7		
bio/bio	+/+	27.5	25	11.1	7.6	4	20
bio/bio	bira/+			8.3	12.5	20	30
bio/bio	bira/bira			2.7	4.8		

Table 3.1 Inheritance and properties of *Ctcf*$^{bio/+}$ alelle

A. Analysis of homozygous Ctcf$^{bio/+}$ mice. For each genotype as indicated the following properties are shown: number of mice found dead in the cage with an age (days after birth) below 70 days, total number of mice born, total number of breedings/crossing performed with mice of this genotype, total number of breedings/crossing with mice of this genotype that did not give any offspring. Breeding/crossing is defined as a period of 4 weeks in where the mouse (>56 days) is put together in a cage with a mouse of the opposite sex. It should be noted that no other mice than indicated were found dead below the age of 70 days. **B.** Inheritance of Ctcf$^{bio/+}$ alelle combined with or without inheritance of the Rosa26^{bira} allele. Offspring and genotype from different Ctcf-bio and Rosa26-birA crosses. Ctcf-bio allele of parents is indicated; bira genotype of parents was variable. n=number of mice, % = percentage of mice born with the indicated genotype, predicted % = predicted percentage of mice born with the indicated genotype based on mendelian ratio's.

vivo (de Boer et al., 2003; Driegen et al., 2005). Double heterozygous *Ctcf*$^{bio/+}$;*Rosa26*$^{bira/+}$ mice were analyzed for the expression of biotinylated CTCF in different tissues. All tissues examined showed CTCF-bio except for the liver, where, consistent with a recent report (Wendt et al., 2008) endogenous CTCF was not clearly detected either (Figure 3.2C). Expression of biotinylated CTCF can also be detected in the embryo (data not shown). Streptavidin pull downs on nuclear extracts from several tissues showed efficient purification of CTCF-bio (Figure 3.2D). These data show that CTCF-bio is produced *in vivo* in the mouse.

CTCF-bio is a functional protein

In order to test whether biotin tagged CTCF has the same properties as untagged CTCF its functionality was tested in different assays. First, we analyzed offspring number and genotypes from different crosses (Table 3.1). Although the numbers are too low to draw any firm conclusion we observed that mice with CTCF-bio were born and made it until adulthood without any obvious phenotype. However some homozygous *Ctcf*^{bio} pups died before the age of 36 days, while none of the mice with other genotypes were found dead at this young age (Table 3.1A). Furthermore homozygous *Ctcf*^{bio} mice are not efficient in giving offspring (Table 3.1B). Homozygous *Ctcf*^{bio} knock-in mice were born in normal numbers, compared to wild-type littermates. We therefore conclude that having CTCF-bio is compatible with viability.

This is in contrast to the N-terminally fused GFP-CTCF knock-in mice, which show early lethality at the homozygous level (Heath et al, unpublished observation). The phenotype observed in the GFP-CTCF knock-in mice is probably not due to a defective protein but rather to the low expression level of the fusion-protein compared to endogenous CTCF. In order to analyze whether this also holds true for biotin-tagged CTCF, expression levels of the protein were analyzed in thymus extracts from mice with different genotypes (Figure 3.3A). Expression levels of biotin-tagged CTCF as well as biotinylated CTCF were equal to untagged CTCF.

In order to examine binding properties of CTCF-bio we carried out a chromatin immunoprecipitation (ChIP) experiment on known CTCF target sites. ChIP was carried out with streptavidin beads using *Ctcf*$^{bio/+}$ ES cells. CTCF-bio binds both the β-globin 3'HS1 and the c-Myc insulator, indicating that biotinylated CTCF is able to bind DNA *in vivo* (Figure 3.3B). In order to compare the DNA binding efficiency we performed bandshifts using the 3'HS1 β-globin CTCF binding site as a probe. CTCF-bio binds this probe, albeit with an apparently reduced efficiently as compared to CTCF (Figure 3.3C). We asked whether biotinylation affects the nuclear distribution of CTCF-bio but in contrast to the HeLa cell system, we observed identical distribution of biotinylated CTCF and endogenous CTCF, indicating that the tagged protein behave similarly (Figure 3.3D). We conclude that biotinylated CTCF does reflect the nuclear distribution and DNA binding characteristics of endogenous CTCF.

Identification of CTCF-interacting partners - technical assessment and first evaluation

Multiple experiments were performed under varying conditions in order to identify CTCF-interacting partners. Because biotinylated CTCF is expressed at endogenous level a clear band representing CTCF-bio was not detected on coomassie stained gels of the different pull down experiments (Figure 3.4B/C/D). However, western blot analysis confirmed the presence of CTCF-bio in all experiments. We performed benzonase treatment because it reduced aspecific background, that is, protein-protein interactions that take place via DNA and/or RNA are eliminated. Benzonase is very efficient in the removal of RNA and DNA at low temperatures

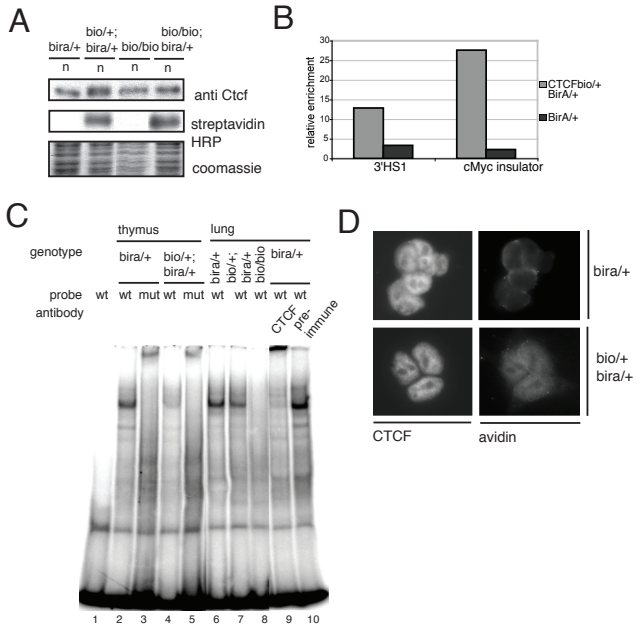

Figure 3.3 Biotinylated CTCF is a functional protein

A. Expression levels of tagged CTCF is mice with different genotypes are shown. A coomassie stained gel was used as a loading control and incubation of western blot with streptavidin HRP was used to detect biotinylated CTCF. n= nuclear extract. **B.** Chromatin IP (ChIP) performed on material from ES cells expressing BirA or biotinylated CTCF. ChIP is performed using streptavidin beads to show specific enrichment of biotinylated CTCF to its known binding sites the β-globin 3'HS1 and cMyc insulator compared to BirA control samples. **C.** Bandshifts on nuclear extracts isolated from lung and thymus tissues of homozygous CTCFbio knock-in (bio/bio) lane 8, heterozygous CTCFbio knock-in mice (bio/+;bira/+) lanes 4,5,7 or BirA (bira/+) control mice lanes 2,3,6,9,10. Genotypes are indicated above the gel. The probe used for the bandshift contains the β globin 3'HS1 CTCF binding site (wt) or a mutant form of this site that abrogates CTCF binding (mut). Lane 9 shows a supershift using CTCF antibody and lane 10 shows that pre-immune serum is not able to show a CTCF dependent shift. **D.** Immunofluorescent staining on ES cells expressing BirA or biotinylated CTCF. CTCF antibodies were used to show CTCF localization in these cellines and avidin was used to detect biotinylated CTCF is these cells.

(Figure 3.4A). The variable conditions further included different stringencies of washing and the use of different cell types. CTCF was identified by the mass spectrometer in all experiments, except for the experiment that contained high stringency washes (exp. 3). In total seven different experiments were carried out, five of which were done with nuclear extracts from ES cells, and two with extracts from thymus and lung. In each case proteins were identified by mass spectrometry after pull down (Table 3.2).

The complete list of identified proteins, obtained from the different experiments, including mascot scores and amounts of identified peptides, is shown in Appendix 1. Some proteins were specifically pulled down by CTCF-bio in one experiment but were present in the BirA control of another experiment (indicated in Appendix 1). We have included these proteins in the list, because the mass spectrometry is not quantitative and it could be that these proteins are enriched in the pull down of biotinylated CTCF. This is, for example, illustrated by the known CTCF interacting proteins PARP1, YB1 and nucleophosmin, that are identified in CTCF-bio pull downs but that are also found in some BirA control samples. Although the number of identified

Experiment number	Source	Benzonase treatment	Wash conditions	Proteolytic digestion	Mass spectrometer
1	ES cells CTCF$^{bio/+}$; Rosa26$^{bira/+}$	-	100 mM salt / 0,3% NP40	In-gel	Q-Tof
2	ES cells CTCF$^{bio/+}$; Rosa26$^{bira/+}$; Cre	-	100 mM salt / 0,3% NP40	In-gel	Q-Tof
3	ES cells CTCF$^{bio/+}$; Rosa26$^{bira/+}$; Cre	-	250 mM salt / 0,3% NP40	In-gel	Q-Tof
4	ES cells CTCF$^{bio/+}$; Rosa26$^{bira/+}$; Cre	+	100 mM salt / 0,3% NP40	In-gel	LTQ
5	ES cells CTCF$^{bio/+}$; Rosa26$^{bira/+}$; Cre	+	150 mM salt / 0,3% NP40	In solution	Orbitrap
6	Lung CTCF$^{bio/+}$; Rosa26$^{bira/+}$; Cre	-	100 mM salt / 0,3% NP40	In-gel	LTQ
7	Thymus CTCF$^{bio/+}$; Rosa26$^{bira/+}$; Cre	-	100 mM salt / 0,3% NP40	In-gel	Orbitrap

Table 3.2 Schematic overview of the different experiments performed to identify CTCF interacting proteins
In order to identify interacting proteins streptavidin pull down was performed on nuclear extracts from cells expressing biotinylated CTCF or birA only expressing negative control cells. Stringency conditions used for streptavidin pull downs were similar in all experiments (100 mM NaCl/0,3% NP40). Source of nuclear extracts used for pull downs, removal of DNA/RNA by benzonase treatment (+/-), stringency of wash conditions, the way trypsin digestion was performed and the mass spectrometer used to identify interacting peptides are indicated for each experiment.

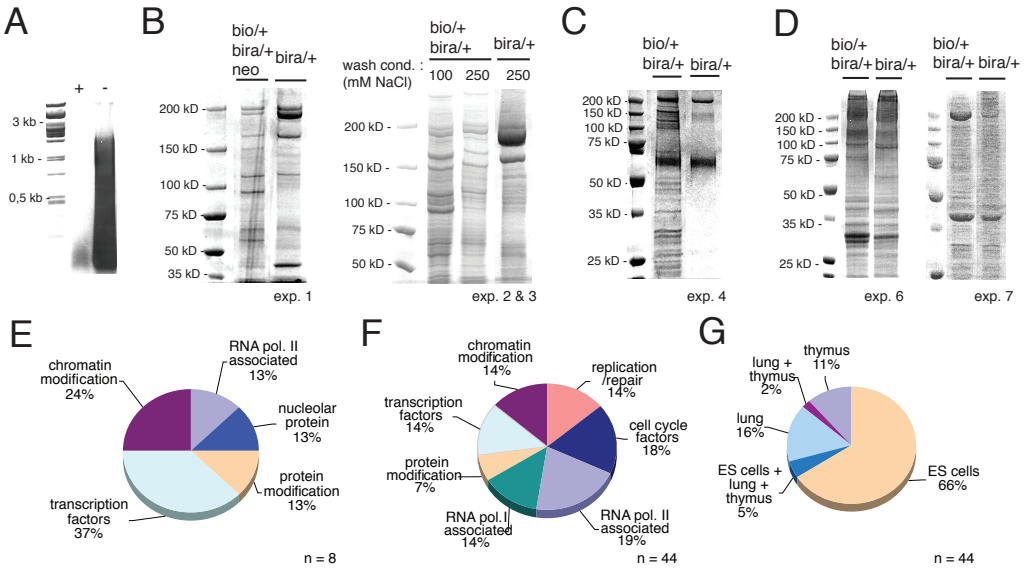

Figure 3.4 Identification of CTCF protein partners
A. Agarose gel showing DNA and RNA material present in nuclear extracts after 2h at 4°C incubation with benzonase (+) or without benzonase (-). **B.** Proteins eluted after direct binding to streptavidin beads of ~5 mg of nuclear extracts from biotinylated CTCF expressing ES cells. Coomassie stained gel used for mass spec. analysis showing nuclear proteins pulled down by biotinylated CTCF expressed in ES cells using different stringencies of washing (exp. 1, 2 and 3). **C.** Coomassie stained gel used for mass spec. analysis showing proteins pulled down by biotinylated CTCF after benzonase treatment

for 2h at 4°C to remove DNA and RNA (exp. 4). The clear band present at 60 kD represents benzonase. **D**. Coomassie stained gel used for mass spec. analysis. Streptavidin pull downs on nuclear extracts from the thymus or lungs of mice expressing biotinylated CTCF (bio/+;bira/+) or bira alone (bira/+). Left panel shows the material pulled down from the lung (exp.6) and right panel shows lung pull down material (exp. 7). Molecular weights are indicated. **E.** Published CTCF interacting proteins classified according to their most prominent function. Total number of proteins represented in the graph is 8. **F.** Classification by function of CTCF interacting proteins identified by mass spectrometry using extracts from ES-, thymus- and lung cells expressing biotinylated CTCF. All factors shown have a known function and were never found in BirA controls or reported to be background. Proteins are classified according to their most prominent function. Published interacting proteins that were identified by our analysis were not included. Total number of proteins represented in the graph is 44. **G**. Classification by cell type in which CTCF interacting proteins were identified by mass spectrometry. Proteins included are as in f). To identify interacting proteins five different experiments were performed on ES, one on thymus cell extracts and one on lung cell extracts. Total number of proteins represented in the graph is 44

peptides between sample and control gives some measure of amount, it cannot be taken as a truly quantative measurement.

Previously, splicing factors, ribosome biogenesis components (including snoRNAs and hnRNPs), cytoskeleton components (actin, lamin), metabolic enzymes, histones, Fibrillarin, YB1, Cdc5l, mRNA processing factors, and nuclear architecture proteins were identified from BirA control pull downs in MEL and therefore reported as background binding proteins (de Boer et al., 2003). Cytoplasmic factors pulled down by our analysis, like Caprin, G3bp, Dynein and Eif2 involved in transport, translation and degradation of mRNA are also likely to be background since CTCF is expressed only in the nucleus.

After exclusion of proteins that are likely to be background we sorted CTCF-interacting partners based on known biological function (Figure 3.4 and 3.5). Comparison of known interacting proteins (Figure 3.4E) with interacting proteins identified from CTCFbio pull downs (Figure 3.4F) revealed a novel pathways that CTCF might be involved in. We found, for example, CTCF-interacting partners functioning in replication, DNA repair, cell cycle regulation and RNA polymerase I transcription. The majority of CTCF-interacting proteins were identified from ES cells as the majority of experiments were performed in these cells. There are few common binding partners identified from ES cells, lung and thymus (Figure 3.4G). Interestingly, no transcription factors were identified in ES cells. The majority of the identified factors (68%) are identified in only one of the seven experiments (Figure 3.5B). This could be due to variability between experiments, to differences in the conditions used, cell type and/or cycle stage. Furthermore, different mass spectrometers were used, of which the sensitivity differs considerably. The number of CTCF peptides identified by the mass spectrometer was low compared to experiments using overexpressed proteins (Lansbergen et al., 2006; Rodriguez et al., 2005).

The coomassie stained gels of pulled down material show some prominent bands that are hardly present in the BirA control lanes (Figure 3.4B/C). These proteins either represent single factors that are present in large amounts or multiple interacting proteins with the same molecular weight. Some proteins migrate in gel at a different band size than expected from their predicted molecular weight. Antibody databases often show at what size proteins migrate, based on this information we predicted the identity of the prominent proteins observed on gel after CTCF pull downs (Appendix 1).

In nuclear extracts from lung (exp 6) a prominent band of ~28 kD and a less prominent band of ~75 kD were detected. The 75 kD band could represent Paf1, a member of the RNA polymerase II-associated Paf complex that was pulled down specifically from lung tissue material. The 28 kD band could represent either TAF15 or laminin binding protein. Furthermore, high molecular weight proteins (>200 kD) prominently seen in ES cells (exp 2 and 3) could be RIF1 or

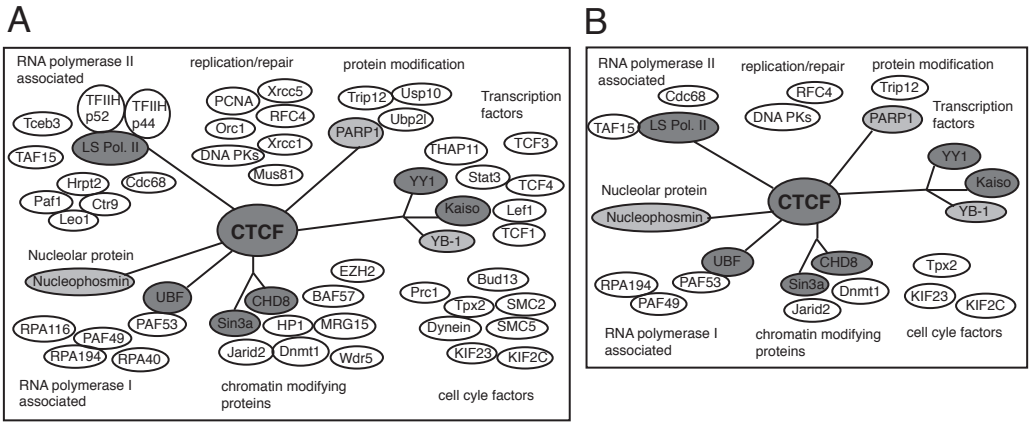

Figure 3.5 Classification of known and newly identified CTCF interacting proteins
A. Classification of CTCF interacting proteins by their most prominent function. All factors show were never found in BirA controls or reported to be background and have a known function. Dark grey: direct interaction and functional significance confirmed, light grey: interaction and functional significance confirmed but direct interaction is not confirmed, white: proteins identified from mass spec. analysis as copurifying with CTCF expect for PCNA/HP1. If boxes surrounding the protein names touch these factors are reported to interact with each other. **B.** Classification of CTCF interacting proteins (as described in a)) that were identified from more than one pull down experiment sorted by their most prominent function.

Ranbp2. The prominent 57 kD band observed in ES cells (exp 1,2,3) could be the ribonucleoprotein Dkc1 and the 90 kD band (exp 1,2,4) might represent Tpx2, DNA topisomerase I or Ssrp1.

Identified protein complexes that were pulled down by CTCF-bio include the thymus-specific TCF/LEF family, the RNA polymerase II associated complex Paf1 and the RNA polymerase I complex (Figure 3.5A). The thyroid hormone receptor interacting protein 12 (Trip12) is the only potential CTCF interacting protein identified from ES cells as well as from the lung and thymus. The function of Trip12 remains elusive, based on its protein domains it was suggested to be involved in ubiquitinylation and thyroid hormone regulation (Aravind, 2001). Two other proteins involved in ubiquitinylation were identified, Usp10 and Ubp2l. Interestingly the TFIID and RNA polymerase II associated factor TAF15 was identified from pull downs of both lung and thymus (Hoffmann and Roeder, 1996).

Many of the known interacting proteins were reported to be associated with the nuclear matrix. These include nucleophosmin, PARP, YY1 and the large subunit of RNA polymerase II. Proteins like PCNA, HDAC, Rb binding protein, Ddx5, DNA polymerase and Xrcc5 were also reported to be components of the nuclear matrix (Mika and Rost, 2005). Whether CTCF associates with the nuclear matrix should be evaluated carefully since nuclear matrix-bound proteins appear to be often pulled down non-specifically.

Verification of identified CTCF protein partners

In order to verify the mass spectrometry data we performed western blot analysis on selected partners (Figure 3.6A), co-immunoprecipitations and GST-based pull downs (Figure 3.7). An overview of CTCF interacting proteins identified by our analysis and the results of the validation

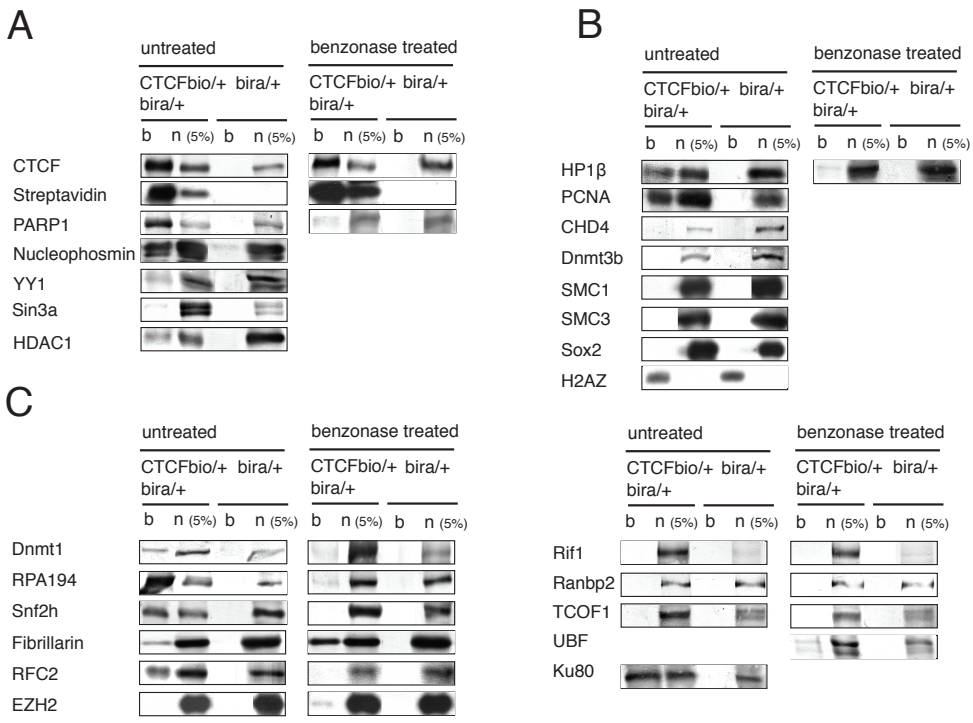

Figure 3.6 Verification of CTCF protein partners using streptavidin pull downs
A. Streptavidin pull downs on nuclear extracts from ES cells expressing biotinylated CTCF and from control BirA only expressing cellines. Biotinylated CTCF is detected by streptavidin-HRP and is absent from the BirA only transfected cells. Nuclear extract equivalent to 5% used in each pull-down or IP (b) was loaded as control for input material (n). **B.** Streptavidin pull-downs of some nuclear proteins that could potentially interact with CTCF. These nuclear proteins were not identified from the mass spec. data as interacting with CTCF. **C.** Confirmation of new CTCF interacting proteins identified by mass spec. using streptavidin pull downs on ES cells followed by western blot analysis

of interactions are shown in table 3.3. Two known interacting partners of CTCF (PARP1 and nucleophosmin) were confirmed by this analysis, validating our approach (note that both proteins were also identified in the BirA control sample, but they were enriched in CTCF-bio expressing cells). After benzonase treatment an interaction with PARP1 was still detected. Two other known CTCF-interacting proteins (YY1 and Sin3a) were not identified in our mass spectrometry analysis. However western blot analysis showed that minor amounts of YY1 and Sin3A were pulled down in ES cells by CTCF-bio. The Sin3A interacting histone deacetylases HDAC1 and HDAC2 were not found back in the mass spectrometry data but also showed an interaction on western blot (Figure 3.6A).

Western blot analysis showed that CTCF-bio interacts with HP1β and PCNA while these proteins were not found back in the mass spectrometry lists. HP1β is a component of both pericentric- and telomeric heterochromatin, while HP1α is detected in centromeric heterochromatin (Cheutin et al., 2003). HP1α was shown not to colocalise with CTCF, but HP1β shows a staining pattern more similar to that of CTCF (Burke et al., 2005). We could not show localization of CTCF with PCNA at replication foci (data not shown), leaving the question what the

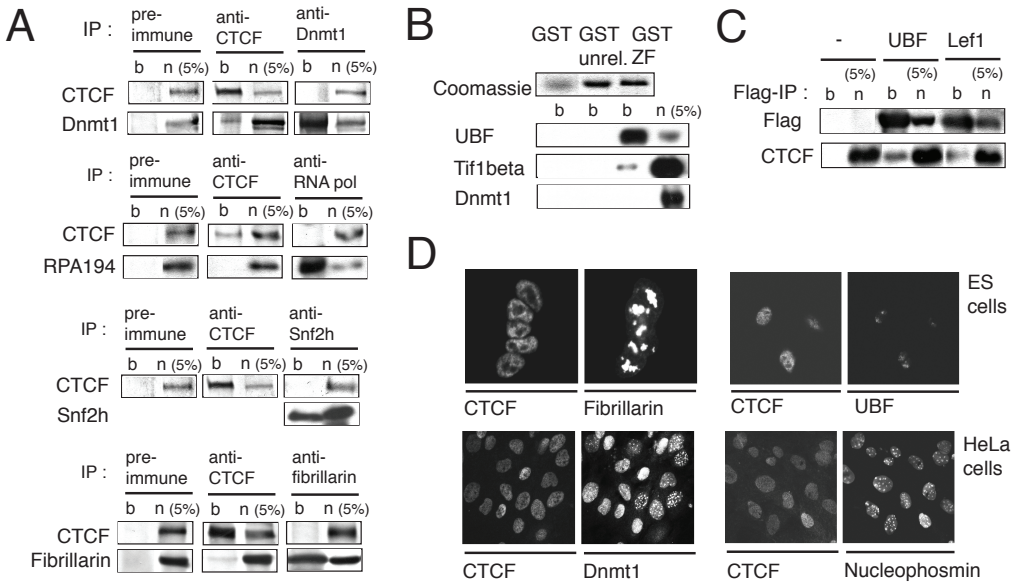

Figure 3.7 Verification of CTCF protein partners using immunoprecipitation and GST pull downs
A. Confirmation by immunoprecipitations (IP) of proteins identified as co-purifying with CTCF in ES cells. Antibody used for immunoprecipitation is indicated above the western blot, pre-immune serum is used as a negative control. Antibody used for western is indicated on the left side of the westerns. Note that all IPs were done in the absence of benzonase exept for the RNA pol. I IP. **B.** GST pull downs on ES cell nuclear extracts. GST alone and GST fusion proteins containing an unrelated protein and the zinc-fingers of CTCF were purified as shown on the coomassie stained gel. The westerns show the different interactions with the purified CTCF zinc-fingers after GST pull downs on ES cell nuclear extracts. **C.** Flag-immunoprecipitations performed in 293T cells overexpressing Flag-UBF or Flag-Lef1. Westerns are incubated with CTCF antibody to show an interaction with endogenous CTCF or with Flag-antibody to show the IP worked. Material that was left on the beads after washings is indicated with a 'b'. Input material of nuclear extracts used represents 5% of the material shown in the 'b' lane and is indicated with an 'n' . **D.** Immunostainings on Hela- and ES cells using different antibodies of CTCF and interacting proteins as indicated below.

function of this interaction could be. We tested several other proteins by western blot, including CHD4, Dnmt3b, Cohesin (SMC1/SMC3), Sox2 and H2A.Z but could not show binding to CTCF-bio (Figure 3.6B).

Western blot analysis confirmed the interaction of CTCF-bio with Dnmt1, RNA polymerase I, Snf2h, fibrillarin, RFC2, Ku80 and UBF. However an interaction with Rif1, Ranbp2 and TCOF1 could not be confirmed, despite identification of these proteins by mass spectrometry. Upon benzonase treatment interaction with Dnmt1, UBF and RNA polymerase I could still be detected. By contrast, interaction with Snf2h and RFC2 was lost (Figure 3.6C), indicating that the latter interactions are indirect. Surprisingly benzonase treatment actually enhanced the interaction of CTCF with fibrillarin and revealed an interaction with EZH2 (Figure 3.6C).

In order to further confirm an interaction of CTCF with Dnmt1, fibrillarin, RNA polymerase I, Snf2h and UBF co-immunoprecipitations were performed. Furthermore, using FLAG-tagged UBF and Lef1 we verified interaction of these proteins with CTCF (Figure 3.7C). A more detailed analysis of the interaction between CTCF and UBF is described in chapter 4. Using CTCF antibodies fibrillarin and Dnmt1 could be pulled down, but immunoprecipitations with fibrillarin and Dnmt1

Function	protein	Identified from celltype / tissue	times identified	remarks	verification CTCF i.a.	reported CTCF i.a.
Chromatin modifying enzymes	Dnmt1	ES cells	3		+	
	EZH2	ES cells	1		±	
	BAF57	thymus	1	SWI/SNF complex	n.d.	
	Snf2h	ES cells	1		+	
	MRG15	ES cells	1		n.d.	
	Jarid2	ES cells	2		n.d.	
	Wdr5	ES cells	1		n.d.	
	Dnmt3b	n.i.	0	not found in MS	-	
	HDAC1	n.i.	0	not found in MS	+	Lutz et al 2000
	HDAC2	n.i.	0	not found in MS	n.d.	Lutz et al 2000
	HP1	n.i.	0	not found in MS	+	
	CHD4	n.i.	0	not found in MS	-	
	Sin3a	n.i.	0	not found in MS	±	Lutz et al 2000
	Rbbp7*	ES cells	3		n.d.	
	BAF170*	thymus	1	SWI/SNF complex	n.d.	
	Tif1Beta*	ES cells	4		±	
RNA pol I associated factors	RPA194	ES cells	3	Pol I complex	+	
	RPA116	ES cells	1	Pol I complex	n.d.	
	RPA40	ES cells	1	Pol I complex	n.d.	
	PAF53	ES cells	3	Pol I complex	n.d.	
	PAF49	ES cells	2	Pol I complex	n.d.	
	UBF*	ES cells, lung, thymus	2	Pol I complex	+	
RNA pol II associated factors	Paf1	lung	1	Paf1 complex	n.d.	
	Leo1	lung	1	Paf1 complex	n.d.	
	HRPT2	lung	1	Paf1 complex	n.d.	
	Ctr9	lung	1	Paf1 complex	n.d.	
	Cdc68 (FACT)	ES cells	3		n.d.	
	TAF15	lung, thymus	2		n.d.	
	TFIIH p44	ES cells	1		n.d.	
	TFIIH p52	ES cells	1		n.d.	
	Tceb3	lung	1		n.d.	
	Ssrp1 (FACT)*	ES cells	1		n.d.	
Transcription factors	Stat3	lung	1		n.d.	
	Lef1	thymus	1	TCF/LEF complex	n.d.	
	TCF1	thymus	1	TCF/LEF complex	n.d.	
	TCF3	thymus	1	TCF/LEF complex	n.d.	
	TCF4	thymus	1	TCF/LEF complex	n.d.	
	THAP11	lung	1		n.d.	
	YY1	n.i.	0	not found in MS	+	Donohoe et al 2007
	Sox2	n.i.	0	not found in MS	-	
	YB-1*	ES cells, lung	3		n.d.	Chernukhin et al 2000
Replication/repair	RFC4	ES cells	2		n.d.	
	Orc1	ES cells	1		n.d.	
	DNA pks	ES cells	2		n.d.	
	Mus81	ES cells	1		n.d.	
	XRCC1	ES cells	1		n.d.	
	PCNA	n.i.	0	not found in MS	+	
	Ku80/XRCC5*	ES cells	1		+	
	RFC2*	ES cells	3		+	
	RFC1*	ES cells	3		n.d.	
	DNA pol delta*	ES cells	2		n.d.	
	DNA ligase*	ES cells	2		n.d.	
Cell cycle/cell organisation	Tpx2	ES cells	3		n.d.	
	KIF23	ES cells	2	motor proteins	n.d.	
	KIF2C	ES cells	2	motor proteins	n.d.	
	Dynein	ES cells	1	motor proteins	n.d.	
	p38-2G4	ES cells	1		n.d.	
	Prc1	ES cells	1		n.d.	
	Rif1	ES cells	3		-	
	Bud13	ES cells	1		n.d.	
	SMC2	ES cells	1		n.d.	
	SMC5	ES cells	1		n.d.	
	SMC1	n.i.	0	not found in MS	-	
	SMC3	n.i.	0	not found in MS	-	
	KIF1C*	ES cells	2		n.d.	
Protein modifying enzymes	Trip12	ES cells, lung, thymus	4		n.d.	
	Ubp2l	ES cells	1		n.d.	
	Usp10	ES cells	1		n.d.	
	Parp1*	ES cells, lung	1		+	Yu et al 2004
Nuclear architecture	Nucleophosmin*	ES cells, lung	3		+	Yusufzai et al 2004
	Tmpo*	ES cells	3		n.d.	
	Kpna2*	ES cells	2		n.d.	
	Ranbp2*	ES cells	2		-	

Table 3.3 Selection of CTCF interacting proteins identified by mass spectrometry classified according to their biological function

The cell type or tissue from which the specific protein is identified upon pulled downs for CTCF interacting proteins is indicated. The times the protein was identified in different experiments is shown. A number of these proteins have been validates by their interaction with CTCF using streptavidin pull downs. Note that this does not take into account an interaction via DNA. Proteins that were tested for an interaction with CTCF but were not found back in de mass spec. (MS) list are also indicated. Common background proteins (ribosomal proteins, splicing factors, metabolic proteins) were excluded from the list. Note that proteins indicated with a star (*) were identified at least once in the BirA control lane. n.d. = not done, n.i. = not identified, - = no interaction, + = interaction, ± = additional experiments needed to confirm this result.

antibodies could not bring down CTCF (Figure 3.7A). Immunoprecipitations did not show an interaction with Snf2h or RNA polymerase I. Using purified GST-tagged CTCF-zinc finger protein we did pull down significant amounts of UBF from ES cell extracts (see Chapter 4), but no Dnmt1 and only minor amounts of Tif1β were brought down, indicating that UBF can interact directly with the purified zinc-finger domain of CTCF but that Dnmt1 might bind another domain of CTCF (Figure 3.7B). Immunofluorescent stainings on Hela and ES cells show that CTCF shows some overlap in localization with Dnmt1 and nucleophosmin, however with UBF and fibrillarin co-localization is less clear (Figure 3.7D).

Discussion

Biotinylation method

The biochemical behavior of CTCF may be a complicating factor in a proteomics-based approach, as the protein is tightly bound to DNA (~70% of CTCF is immobile (Heath et al, unpublished observations)). This means that CTCF binds most of its interacting partners while attached to DNA, either using the DNA to adopt the right conformation, or as a co-factor for binding. Thus, even after vigorous washing of streptavidin beads we still observed multiple background bands. This contrasts to what is reported for other biotinylated proteins that were overexpressed (Lansbergen et al., 2006; Meier et al., 2006; Rodriguez et al., 2005). Upon benzonase treatment the background was reduced, but we also detected less specific interactions, as shown by western blots of known interacting proteins. Higher background signal might also be due to the expression levels in our study, as we used endogenously expressed CTCF-bio. This is reflected by the mass spectrometry results that show only minor amounts of CTCF peptides. We conclude that the relatively low expression level of tagged CTCF combined with its tight association with DNA (and perhaps its association with the insoluble nuclear fraction) makes it technically difficult to efficiently isolate intact protein complexes. Others have reported that CTCF is prone to aggregation and non-specific interactions (Klenova et al., 2002).

None of the proteins interacting with CTCF were highly enriched in binding, as compared to the input. This suggests that the interactions of CTCF with other proteins are often transient, or weak explaining why CTCF can interact with multiple factors. Despite the technical problems we were able to identify and verify several new (and established) CTCF-interacting proteins. For UBF we demonstrate that the interaction is direct (Chapter 4), by using purified proteins (Mackay et al., 2007). We have not done this for other interacting proteins. Bearing this in mind we discuss the potential physiological significance of several of the interactions, focusing on interacting proteins that belong to a complex of proteins or have a shared function.

HMG box proteins

Proteins containing high-mobility group (HMG) domains are segregated into two major groups. Members of one group are identified by the presence of more than one HMG domain. These proteins bind to DNA without sequence specificity. This group includes the HMG/UBF family. In contrast, members of the other group possess a single HMG domain and bind with high affinity to a specific DNA sequence. Members of this group generally resemble classic tissue-specific transcriptional regulators. This group includes members of the MATA/TCF/SOX family (Soullier et al., 1999).

Interestingly two proteins were identified from our pull downs that are closely linked to UBF, namely Ssrp1, a component of the FACT complex, and the SWI/SNF related protein BAF57. A CTCF-UBF interaction is discussed extensively in Chapter 4. Ssrp1 and BAF57 share 50-60% identity with the first HMG-box of UBF. Phylogenic and ancestral analysis revealed a related origin of these proteins, which is reflected in the nature of their DNA binding (unspecific) and which places these proteins in the HMG/UBF family (Wattler et al., 1999). This in contrast to other HMG-box containing proteins identified as interacting with CTCF, like Lef1 and TCF, that belong to the MATA/TCF/SOX family.

Using a phage display approach it was shown that HMGB1 could recognize several peptide motifs. A search through protein databases identified these peptide motifs in known interacting proteins of HMGB1, but also allowed to identify new potential candidates. Interestingly among them was CTCF, that contains a HMGB1-recognition motif in the proline-rich region present in its C-terminal region (Dintilhac and Bernues, 2002). Besides the fact that HMGB1, Ssrp1, UBF and BAF57 all belong to the family of HMG/UBF proteins that are able to recognize structural motifs of DNA, they contain structural properties themselves and are able to bend or kink DNA. However we have not yet confirmed an interaction between all of those proteins and CTCF. In Chapter 4 we show that the interactions of UBF with CTCF and CTCFL is mediated via the ZF domain of CTCF(L) and not via the C-terminus.

Taking into account the properties of CTCF to affect chromatin structure a link with the structural HMG family is very interesting. Like UBF, Lef1 plays a structural role by bending DNA in order to facilitate protein-protein interactions (Giese et al., 1992). For Lef1 and UBF we show a specific interaction with CTCF. Interestingly UBF2 was shown to associate with Lef1 and to stimulate RNA polymerase II dependent transcriptional activation by Lef1 from a Lef/TCF-responsive synthetic promoter (Grueneberg et al., 2003).

Regulation of RNA polymerase II transcription and elongation

The Paf1 complex is composed of Paf1, Ctr9, Hrpt2 and Leo1 and was originally isolated in yeast in association with the large subunit of RNA polymerase II. Paf1 complex components are abundant and co localize with RNA polymerase II on chromatin at promoters and in the coding regions of actively transcribed genes. Loss of Paf1 and Ctr9 results in a severe phenotype, affecting cell cycle regulation, protein synthesis and nucleic acid metabolism (Betz et al., 2002; Mueller and Jaehning, 2002). Deletion of Hrpt2 results in the dissociation of the remaining Paf1 complex members from chromatin and a significant reduction in binding of the complex to RNA polymerase II. In addition, loss of Paf1 complex components leads to a reduction in RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation (on Ser2) and in shortened poly(A) tails on most cellular transcripts, suggesting that the Paf1 complex facilitates linkage of transcription and pre-mRNA processing (Mueller et al., 2004; Penheiter et al., 2005; Sheldon et al., 2005).

The Paf1 complex is not only involved in elongation but also in histone methylation. Yeast mutants missing Paf1 and Ctr9 are defective in histone 3 lysin 4 (H3K4) methylation (Krogan et al., 2003). Paf1 is required for the recruitment of the yeast Set1 (COMPASS) methyltransferase to RNA polymerase II (Hampsey and Reinberg, 2003). Also the mammalian Paf1 complex is associated with a Set1-like histone methyltransferase complex, that methylates H3K4, this also appears to involve the MLL complex (Rozenblatt-Rosen et al., 2005). The Paf1 complex also mediates histone H3 methylation on lysines 36 and 79, thereby marking regions of active transcription (Hampsey and Reinberg, 2003). Recently it was demonstrated that elongation by RNA polymerase II through the nucleosomal barrier is dependent on the histone chaperone FACT (Facilitates Chromatin Transcription) and the recruitment of Paf1 and the H2B monoubiquitination machinery (Pavri et al., 2006). Paf1 plays an important role in controlling binding of histone methyltransferases and chromatin regulators (FACT, CHD1) with elongating RNA polymerase II (Warner et al., 2007). Two components of the FACT complex were identified from our mass spec analysis, Cdc68/Supt16h and the HMG-box containing Ssrp1. These components promote transcription elongation by nucleosome disassembly upon transcription (Orphanides et al., 1999).

All components of the Paf1 complex are specifically pulled down from lung nuclear extracts by CTCF. We do not exactly know why this complex was not identified from our ES cells or thymus experiments. Expression profiles based on ESTs (UniGene) show a higher abundance of Paf1 in lung than in thymus, and a more significant presence of this complex in adult than in ES cells/blastocysts. Moreover, the efficiency of the lung pull down appeared to be higher because more CTCF peptides were identified in these pull downs than in the thymus pull downs.

In light of a recently reported interaction of CTCF with the largest subunit of RNA polymerase II (Chernukhin et al., 2007) its potential interaction with the Paf1 complex and the TFIID component TAF15 is highly interesting. CTCF interacts with RNA polymerase both when the latter is associated with the initiation complex as well when it is elongating. Enrichment of RNA polymerase II binding to DNA, an indication for pausing, interference or release of the polymerase can be detected at some CTCF binding sites (Chernukhin et al., 2007; Filippova et al., 1996). Thus, CTCF might potentially pause or stall RNA polymerase II. Another possible function could be at the initiation of transcription, where promoter bound CTCF could function in gene activation by recruiting RNA polymerase II. Alternatively CTCF could regulate release and early elongation of promoter-proximal paused RNA polymerase II. This could explain its potential binding to the initiation component TAF15 and to factors that can increase the catalytic rate of RNA polymerase II, such as Tceb3/Elongin A. Components of the Paf1 complex are present at transcription initiation sites and can recruit transcription factors and modulate histone modifications. Interestingly, such a recruitment is reported for STAT3, a component that is found back in our pull downs on lung material, maybe via its interaction with the Paf component Ctr9 (Youn et al., 2007).

Boundary formation and chromatin modification

A chromatin barrier can act to maintain silenced chromatin domains and prevent their spreading by actively remodeling chromatin. In yeast barrier activity is modulated by histone modifiers, SWI/SNF chromatin remodelers, and by TFIID components (Oki et al., 2004). Recently CTCF was shown to be involved in organizing histone modification (Han et al., 2008; Splinter et al., 2006). Thus, barrier activity of CTCF may be linked to its ability to mediate histone modifications. Several chromatin-modifying factors are identified as interacting with CTCF (albeit unconfirmed), including BAF57, BAF170 and Snf2h, as well as a histone demethylase (Jarid2), histone methyltransferases (EZH2, Wdr5), a histone acetyltransferase (MRG15), and a DNA methyltransferase (Dnmt1).

Interestingly, a protein called USF1 that is involved in boundary formation at the β globin cHS4 has been reported to bind to Wdr5, CHD8 and Snf2h (Huang et al., 2007), which are all (potential) CTCF-interacting partners.

BAFs (BRG- or Brm-associated factors) are related to the SWI2/SNF2 group of ATP-dependent chromatin remodeling complexes. They include BAF47, BAF53, BAF57, BAF155 and BAF170. BAF57 is the only known DNA binding component of the BAF complex (Quinn et al., 1996; Wang et al., 1998). The Drosophila homolog of BAF57 is Bap111/Dalao no homolog of BAF57 has been found in yeast SWI/SNF complexes. The HMG domain of BAF57 is important since mutants lacking this domain are not functional. Thus, the DNA bending activity of the HMG domain is required for transcriptional regulation and chromatin remodeling by BAF57 (Chi et al., 2002; Papoulas et al., 2001). Mutations in BAF57 impair the function of the BAF complex in both silencing of the CD4 locus and activation of the CD8 locus. These loci express the co-receptor molecules CD4 and CD8 that cooperate with the T-cell receptor in antigen recognition on the surface of T-cells. BAF57 impairs CD4 silencing by direct binding to the CD4 silencer elements (Chi et al., 2002). BAF57 was identified as a potential CTCF interacting protein in the thymus. Interestingly, deletion of CTCF in the thymus result in significant effects on CD4- and CD8-positive T cells, the latter being less dependent of CTCF (Heath submitted).

Upon lymphocyte activation the BAF complex is rapidly targeted to chromatin via a phosphatidyl-inositol pathway (Zhao et al., 1998). The activity of SWI/SNF remodeling complexes can be modulated by inositol polyphosphates, although the biological significance of this regulation remains to be determined (Shen et al., 2003). Combined, these data suggest a role for BAF in rapid changes of chromatin structure, as for example upon antigenic activation of T lymphocytes. Interestingly stimulation of CTCF knockout T cells with inositol polyphosphates (PMA) partially but not fully restores the normal function of those cells (Heath submitted). Furthermore, CTCF has been shown to respond to insulin signaling, which acts via PI3-kinase (Gao et al., 2007). Thus, a CTCF-BAF interaction could be involved in quick transcriptional responses.

SNF2h belongs to the group of ISWI chromatin remodeling enzymes. In the mouse SNF2h can form two complexes, together with WSTF to form the WICH-complex, or with TIP5, p50 and p80 to form the nucleolar remodeling complex NoRC. In humans SNF2h can also bind to CHD4/Mi-2, HDAC's and the Cohesin complex. Interestingly human ACF (SNF2h, Acf1, WCRF) is reported to regulate chromatin folding into loop domains together with Sin3a and the transcriptional regulator SATB1 (Yasui et al., 2002). Snf2h was identified once in ES cell pull downs and it was verified to interact with CTCF. However, upon addition of benzonase the interaction was lost, indicating that it might be mediated via DNA. Recently it was shown that Snf2h does not bind to the CTCF binding sites in the *c-Myc* insulator and *Igf2/H19* ICR (Stedman et al., 2008).

Jarid2 is a member of the jumonji family involved in transcriptional repression and/or chromatin regulation, which plays an important role during development (Takeuchi et al., 2006). The jumonji group of proteins is characterized by a novel structural motif, the JmjC domain, which is implicated in histone demethylation. JMJD2 is involved in demethylation of H3K9/K36. Of the other family members, JMJD6 is involved in the demethylation of arginines 2 and 3 at histone H3, whereas JMJD3 is involved in demethylation of H3K27 (Whetstine et al., 2006) (Swigut and Wysocka, 2007). Jarid2 is identified in two separate ES cell pull downs, in which benzonase was not used. It would be interesting to confirm the interaction with CTCF.

Enhancer of zeste homologue 2 (EZH2) is a histone lysine methyl transferase, which is associated with transcriptional repression. It is a component of the polycomb repressive complexes PRC2 and PRC3. The PRC2 complex catalyses histone H3 K27 trimethylation, of which EZH2 is the catalytic subunit (Cao et al., 2002). EZH2 was shown to bind Dnmt1 and controls DNA

methylation thereby mechanistically linking the Polycomb group proteins to DNA methylation (Vire et al., 2006). We identified Dnmt1 and EZH2 in separate pull downs. EZH2 only binds to CTCF in the absence of DNA/RNA.

Wdr5 is a WD40 repeat protein that is part of MLL1, MLL2 and hSet1 histone H3K4 methyltransferase complexes. The yeast homolog of the human MLL complex is the Set1-containing complex COMPASS (Lee et al., 2007). Interestingly the CTCF interacting protein CHD8 is reported to be present in a complex containing Wdr5-MLL1-MOF (Dou et al., 2005). Wdr5 recognizes histone methyl-lysine with its WD40 and recruits H3K4 methyltransferases to K4-dimethylated histone H3. WDR5 is required for global and gene-specific tri methylation of H3K4 (Wysocka et al., 2005). Interestingly the MLL1 complex is linked to histone acetylation by its interaction with the H4K16 acetyltransferase MOF (Dou et al., 2005). It should however be noted that Wdr5 is identified only once in the performed ES cell pull downs with a significance around background.

MRG15 is a MOF related component of the NuA4/Tip60 histone acetyltransferase (HAT) complex that is involved in transcriptional activation of select genes principally by acetylation of nucleosomal histone H4 and H2A (Doyon et al., 2004; Garcia et al., 2007). Furthermore it can bind specifically to the methylated lysin 36 of histone H3 (Zhang et al., 2006). It should be noted that MRG15 is identified only once in the performed ES cell pull downs with significance around background.

The maintenance methyltransferase Dnmt1 is reported to be involved in imprinting and gene expression regulation its active mainly in S-phase when new methylation patterns have to be set on new daughter strands (Howell et al., 2001). Recently the RING- and SET finger domain containing Np95 was shown to be involved in loading Dnmt1 on newly replicated DNA (Sharif et al., 2007). Although CTCF is mainly reported to bind to unmethylated sites and was reported to be involved in keeping these sites methylation free it can also bind to methylated sites (Mukhopadhyay et al., 2004). At these methylated sites it could potentially cooperate with Dnmt1 to maintain DNA methylation.

T cell specific factors

All members of the of the HMG domain containing LEF/TCF family of transcription factors including LEF1 (lymphoid enhancer factor) and the T-cell factors TCF1, TCF3 and TCF4 were identified from CTCF pull downs on thymus material. In adult mammals, TCF1 is uniquely expressed in T lymphocytes, while LEF1 is expressed in T cells and early B cells (Travis et al., 1991). These factors play crucial roles in WNT/Wingless signaling, a signal transduction cascade that directs cell differentiation. LEF1 is redundant with TCF1 for correct development of T lymphocytes in the thymus. The Wnt mediated signaling pathway induces cytosolic β-catenin binding to TCF/LEF proteins within the nucleus, leading to the enhanced expression of the Wnt target genes. The β-catenin-TCF complexes are negatively regulated by the APC tumor suppressor protein, which phosphorylates β-catenin and, in turn, increases the degradation of cytosolic β-catenin and inhibits the transcriptional activity of the TCF/LEF proteins.

The *in vivo* functions of TCF1 and LEF1 have been explored by gene disruption experiments. *Tcf1* knockout mice are severely impaired in the generation of T cells, but are otherwise normal. *Lef1* mice lack hair, teeth and mammary glands and as a consequence die around birth (van Genderen et al., 1994). As deduced from direct analyses and from transplantation experiments, the *Lef1* mutation has no major effects on the immune system. In *Tcf1/Lef1* double knockout mice, development of T cells is completely abrogated, indicating that LEF1 can substitute for

TCF1 in T-cell differentiation. These mice show decreased expression of TCRα gene expression but normal TCRβ expression (Okamura et al., 1998; Verbeek et al., 1995).

LEF1 and TCF1 perform a major function in controlling expression of the T cell receptor (TCR) genes by binding to the enhancer region (Travis et al., 1991). The alpha enhancer is active in T cells and drives TCRα recombination in collaboration with a locus control region-like element located downstream of the Cα gene. CTCF was shown to bind in this region and can perform enhancer-blocking abilities at this site (Magdinier et al., 2004). This CTCF site is only 230 bp separated from the Lef1 binding site, but the two protein could potentially interact with each other because Lef1 is reported to bend DNA (Love et al., 1995). However recently CTCF was shown not to be critical for the locus control region activity. The region was shown to contain CTCF independent boundaries that have a more pronounced effect on LCR activity (Gomos-Klein et al., 2007). Also TCRα gene expression is not impaired in the absence of CTCF (Heath submitted). Leaving the question what the functional significance of a CTCF interaction with TCF/LEF could be. Potentially CTCF is involved in Wnt signaling by regulating binding to other LEF/TCF targets such as cyclin D1, E-cadherin, c-Myc or the oncogene Fra-1 (Mann et al., 1999; Tetsu and McCormick, 1999; van de Wetering et al., 2002).

Cellular organization

Structural Maintenance of Chromosomes (SMC) family proteins play critical roles in various nuclear events that require structural changes of chromosomes, including mitotic chromosome organization, DNA recombination and repair and global transcriptional repression. The SMC1-SMC3 heterodimer is part of the Cohesin complex, the SMC2-SMC4 heterodimer is part of Condensin, furthermore a SMC5-SMC6 heterodimer is also crucial for chromosome segregation. Our analysis identified SMC2 and SMC5 once in ES cell pull downs done in the absence of benzonase. Recently an interaction between CTCF and SMC3 was described, although similar studies on CTCF and Cohesin do not report an interaction between the two proteins (Parelho et al., 2008; Stedman et al., 2008; Wendt et al., 2008). However Cohesin was shown to mediate CTCF dependent transcriptional insulation. Likely an unknown protein factor mediates the positioning of Cohesin by CTCF.

SMC2 is a central component of the Condensin complex, a complex required for conversion of interphase chromatin into mitotic like condense chromosomes. SMC5 is required for DNA repair and segregation of repetitive regions (Torres-Rosell et al., 2005).

SMC's work together with Topoisomerase and are reported to be involved in higher-order chromosome organization. Furthermore it was suggested that Condensin might function as an intramolecular DNA cross-linker that folds a single DNA molecule, whereas Cohesin might act as an intermolecular DNA cross-linker that holds two different DNA segments together (Hirano, 1999). In addition to their essential contribution to mitotic chromosome condensation and segregation, the Condensin subunits play important functions at non-mitotic stages of the cell cycle (Hirano, 2002). Interestingly in *Drosophila* it was reported that Polycomb group proteins and Condensin subunits cooperate to maintain the silenced state of gene expression, possibly by assembling condensed heterochromatin-like structures (Lupo et al., 2001).

Some other proteins that are known to have prominent function during mitosis and cellular organization and were identified from our CTCF pull downs are RCC1-like, Tpx2, KIF23/MKLP1, Prc1, KIF2C/MCAK and Bud13/KAR9. Interestingly these proteins are all involved in organization, alignment and formation of the mitotic spindle. Mitotic spindle morphogenesis is a series of highly coordinated movements that lead to chromosome segregation and cytokinesis. The

mitotic spindle consists of bundles of microtubules, bundled by kinesins and nucleated at the centrosomes. During spindle assembly some of the spindle's microtubules attach to kinetochores that assemble on the centromere of the chromosomes.

RanGTP activates spindle assembly by activation and release of microtubule associated factors like Tpx2 and Numa (Kalab et al., 2002). RanGTP is generated at this location by the only known nucleotide exchange factor for Ran called RCC1, which remains bound to chromatin throughout mitosis (Moore et al., 2002). Parallel to the Ran-dependent spindle assembly promoting pathway is a second independent pathway exists in which Aurora-B play a role in regulating chromosome driven microtubule nucleation. Chromosome orientation and alignment within the mitotic spindle requires not only Aurora B but also the mitotic centromere-associated kinesin (MCAK) (Andrews et al., 2004). Proper spindle alignment is also ensured by multiple factors. The asymmetric loading of APC-related Kar9 onto spindle poles and microtubules is for example important in this (Liakopoulos et al., 2003). Localization of the central spindle depends on components of the spindle midzone including the microtubule bundling protein Prc1 and mitotic kinesin-like protein 1 (MKLP1) (Kurasawa et al., 2004).

Numa and Tpx2 are found in the interphase nucleus, presumably localized there by Ran/RCC1 and importins and they are preferentially activated in the vicinity of chromatin (Kahana and Cleveland, 2001). Besides its function in mitosis Tpx2 was shown to be involved in post mitotic nuclear assembly in *Xenopus* upon binding to Tmpo/Lap2 that localizes to the inner nuclear membrane (O'Brien and Wiese, 2006). LAP2 (lamina-associated polypeptide 2) plays a role in the regulation of nuclear architecture by binding lamin B1 and chromosomes. Moreover lamin B has been reported to be a structural component of the spindle matrix that promotes microtubule assembly and organization in mitosis. Depletion of lamin B resulted in defects in spindle assembly (Tsai et al., 2006). The periphery of the nucleus is suggested to provide a platform for sequestering transcription factors away from chromatin.

Potentially CTCF could function in chromosome condensation or have specific functions during mitosis. The identification of spindle components is interesting with regard to the localization of CTCF to the centrosome that is involved in the assembly and organization of the spindle during mitosis (Zhang et al., 2004). Moreover it was shown that during mitosis CTCF remains bound to mitotic chromosomes and is slightly enriched on centromeres (Burke et al., 2005). CTCF might play a role in regulating correct chromosome segregation during mitosis.

The identified factors localize to the nucleus in interphase, indicating that they might function in other processes outside mitosis. Interestingly two proteins only reported to function during mitosis, Cohesin and Aurora B, were reported to participate in gene regulation in post mitotic cells (Sabbattini et al., 2007; Wendt et al., 2008). This indicates that the function of CTCF binding to the mitotic factors could also be one outside mitosis. Potentially the factors participate in gene regulation or in the structural organization of chromatin together with CTCF. It would be interesting to further confirm an interaction of CTCF with the proteins identified and to investigate the functional meaning of this in more detail.

In conclusion CTCF interacts with multiple different proteins suggesting that it forms a platform that allows many functionalities to take place.

Materials en methods

Cell culture and generation of DNA constructs

Mouse cDNAs of CTCF (IMAGE 6825952) and CTCFL (Sleutels unpublished) were amplified using PCR with primers containing EcoRI and XmaI sites and cloned into the shuttling vector pGEMTeasy. To generate N-terminally bio tagged CTCF and CTCFL the cDNAs were cut out of this vector using EcoRI/XmaI and cloned into the C1-bio vector that contains the backbone of the pEGFP-C1 vector (Clontech) with the bio-tag cloned in place of the eGFP in AgeI/EcoRI sites. To generate C-terminally bio tagged CTCF and CTCFL the cDNA were cloned in a similar way to the N1-bio vector that contains the backbone of the pEGFP-N1 vector with the bio-tag cloned in place of the eGFP in EcoRI/XmaI sites. For BirA expression the pSCT-HaBirA vector was used. Transient transfections were performed using lipofectamine 2000 (Invitrogen) and Optimem transfection medium (GiboBRL). Cells were used for experiments 24 hours later or continued culturing in medium with 1 ug/ml puro to select single stable expressing clones. Flag-tagged UBF was a kind gift of Dr. R. Voit and flag-tagged Lef1 was generated by PCR amplification with primers containing NheI/XhoI sites on mouse cDNA (IMAGE 6401514) and cloned in NotI/NheI of the expression plasmid pCBA-Tag1-2xFlag. GST-tagged zinc-fingers of chicken CTCF present in pGEX4T1 were produced in BL21 E.Coli and purified using glutathione-Sepharose 4B (Amersham Biosciences). Hek293T and HeLa cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum. ES cells were cultured on 0,1% gelatin coated dishes in DMEM medium containing 10% FCS, BRL, LIF, NEAA, β-mercaptoethanol and split every two days.

Generation of CTCFbio knock-in mice

Oligo's containing bases to erase the CTCF stop-codon, TEV protease cleavage site followed by the bio-tag and a stop codon and a EcoRV or SacII overhang were ordered (Invitrogen) and annealed. The annealed fragment was cloned into SacII and EcoRV sites of pBleuscriptKS Lox-Neo-Lox. Afterwards 5'-homology and 3'-homology arms were PCR-ed from mouse 129 PAC Clones (Osoegawa et al., 2000) and cloned into the pBleuscript biotev-lox-neo-lox vector. Generating the final targeting construct that was confirmed by sequencing to be without any mutations and was subsequently linearized and used for homologous recombination to generate the $CTCF^{bio/+}$ allele. In this construct the neomycin resistance cassette was transcribed in the antisense direction under the PTK promoter. Constructs were targeted into 129 embryonic stem (ES) cells as described (Hoogenraad et al., 2002). Confirmation of homologous recombination was performed using different 5' end and 3' end external probes and a PCR-based assay for genotyping. Genomic DNA was isolated, digested, and blotted onto Hybond N+ membranes (Amersham) and hybridized with radioactive probes. Hybridization was performed in a rotating hybridizer at 65°C for 24h in ChurchHyb-mix (0,5 M Na2HPO4 pH 7.2, 7% SDS, 1 mM EDTA). Membranes were washed extensively with Church wash-buffer at 65°C (40 mM Na2HPO4 pH 7.2, 1% SDS). Hybridization signals were analyzed with a Phoshor Imager (Typhoon Amersham). Two targeted ES cell clones with the correct karyotype were chosen to inject into C57Bl/6 blastocysts. Chimeric males were mated to C57Bl/6 females to obtain germ line transmission of the modified allele. Mice with the modified allele were crossed to CAG-Cre transgenic mice to remove the neomycin resistance cassette. Injected CTCFbio ES cells were used for a second targeting event with CMV-Cre to remove the neomycin resistance cassette and afterwards using homologous recombination Rosa26-HABirA was inserted. Correct targeting and verification was done by southern blot analysis and PCR. Control BirA ES cellines were used as described (Driegen et al., 2005).
Genotyping was routinely performed by PCR using primers:
Rosa26 product (350 bp), Bira product (514 bp), 57°C annealing temperature.
Rosa26 F (265): GTGTAACTGTGGACAGAGGAG
Rosa26 F (266): GAACTTGATGTGTAGACCAGG
BirA_F (91): TTCAGACACTGCGTGACT
BirA_B (92): GGCTCCAATGACTATTTGC
For CTCFbio genotyping the following primers were used: CTCFbio (599 bp), wt (549 bp)

CTCFGB1: AGCAAAAGCAAAACCAGGTTA
CTCFGF14: AGGAGCCAGATGCCGAGCCTG
Primers used for probes:
Probe 1F: TCCTGCCTCTGTCCAGTCAGAGA
Probe 1B: GCAGATCACTGTGTGTTCAAGGC
Probe 2F: CGAATGCCACCTTTGACTCTACC
Probe 2B: AAGCCTCGTCCTTCCGAGCCT

Preparation of nuclear extracts and superose 6 size fractionation

Nuclear extracts were prepared as described (Andrews and Faller, 1991). Isolation of nuclear extracts from tissues was performed using mice of about 5 weeks old. Tissues were dissected cut into smaller pieces, after addition of small amounts of ice cold PBS further homogenizing was performed in a dounce homogenizer. Samples were centrifuged for 5 min 4°C at 1000 rpm and pellet of cells was used to isolate nuclear extracts. Size fractionation of protein complexes was done on an AKTA FPLC apparatus with a Superose 6 10/30 column (Amersham Biosciences). Fractions were precipitated with 100% trichloroacetic acid and analyzed by western immunoblotting as described. Molecular size standards were thyroglobulin (670 kD) and albumin (66 kD) (Amersham Biosciences)

GST pull downs, immunoprecipitation and immunoblotting

Prior to using isolated protein extracts for binding assays the salt and detergent concentrations of nuclear extracts samples as well as the protein concentration in the sample was adjusted to 100 mM NaCl/0.3%NP0 and 0,8-1,2 ug/ul nuclear protein. All binding reactions done for immunoprecipitations and pull downs (unless stated differently) were performed under low stringency conditions (100 mM NaCl, 0.3% NP40, 20 mM Hepes pH8, 0.2 mM EDTA, 10 mM MgCl2, 1x Complete protease inhibitor (Roche)). If indicated benzonase nuclease (Novagen) was added for 2h at 4°C to remove DNA and RNA. All pull downs were performed at least two times. For IPs nuclear extracts were pre-cleared at 4°C using Protein A sepharose beads. Immunoprecipitations were performed upon addition of antibody and samples (nuclear extracts in 100 mM NaCl/0.3% NP40) were rotated for 1 h at 4°C. Subsequently protein-A sepharose beads were added to the samples and incubation was continued for another 1h at 4°C while rotating. Flag-IPs were performed using the same protocol except binding was performed for 3h at 4°C upon addition of Anti-flag M2 agarose (Sigma). The beads were washed 7 times 5 min. each at 4°C in washing solution. Boiling in 1x Laemmli buffer eluted bound material. For immunoblot analysis samples were run on SDS polyacrylamide gels and blotted onto PVDF membranes (MilliPore) using a semi-dry blotting apparatus (Biorad). After antibody incubations of membranes signal detection was performed using ECL (Amersham).

Streptavidin pulldown and Mass spectrometry

For streptavidin ChIP preparation of cross linked chromatin ($2x10^7$ cells treated with 1% formaldehyde for 10 minutes at RT), sonication to 300-800 base pair fragments and immunoprecipitations were as described in the Upstate protocol (http://www.upstate.com). Exceptions for streptavidin ChIPs were as follows: streptavidin beads blocked for 1h RT in 0,2mg/ml sonicated salmon sperm DNA were used, elution was performed O/N at 65°C in elution buffer (0,1% NaHCO3, 1% SDS, 0,2M NaCl). For streptavidin pull downs: Dynabeads M-280 Streptavidin (Dynal) were blocked by 200 ng/ul chicken serum albumin (CSA) 1h at RT. Dynabeads were subsequently incubated with nuclear extracts at 4°C for 2 hours while rotating. After magnetic separation the beads were washed seven times in a buffer containing 100 mM NaCl, 20 mM Tris pH7.5, 0.3% NP40 and protease inhibitors (Complete; Roche). Subsequently the beads were resuspended and boiled in 1x Laemmli buffer. For mass spectrometry samples were run under keratin free conditions using a pre-cast 3-8% NuPAGE Tris-Acetate gel and stained with the colloidal blue staining kit (Invitrogen). Depending on the experiment mass spectrometric analysis was performed on a Q-Tof, LTQ or Orbitrap (see table 3.2). 1D SDS-PAGE gel lanes were typically cut into slices using an automatic gel slicer and subjected to in-gel reduction with dithiothreitol, alkylation with iodoacetamide and digestion with trypsin (Promega, sequencing grade), For Q-Tof analysis : NanoLC-MS/MS was performed on either a CapLC system (Waters,

Manchester, UK) coupled to a Q-ToF Ultima mass spectrometer (Waters, Manchester, UK), operating in positive mode and equipped with a Z-spray source, or an 1100 series capillary LC system (Agilent Technologies) coupled to an LTQ (Thermo) operating in positive mode and equipped with a nanospray source. Peptide mixtures were trapped on a Jupiter™ C18 reversed phase column (Phenomenex; column dimensions 1.5 cm × 100 µm, packed in-house) at a flow rate of 7 µl/min. Peptide separation was performed on Jupiter™ C18 reversed phase column (Phenomenex; column dimensions 15 cm × 50 µm, packed in-house). For LTQ or Orbitrap : NanoLC-MS/MS was performed on an 1100 series capillary LC system (Agilent Technologies) coupled to an LTQ ion trap mass spectrometer or an LTQ-Orbitrap mass spectrometer (Thermo) operating in positive mode and equipped with a nanospray source. Peptide mixtures were trapped on a ReproSil C18 reversed phase column (Dr Maisch GmbH; column dimensions 1.5 cm × 100 µm, packed in-house) at a flow rate of 8 µl/min. Peptide separation was performed on ReproSil C18 reversed phase column (Dr Maisch GmbH; column dimensions 15 cm × 50 µm, packed in-house) using a linear gradient from 0 to 80% B (A = 0.1 M acetic acid; B = 80% (v/v) acetonitrile, 0.1 M acetic acid) in 70 min and at a constant flow rate of 200 nl/min using a splitter. The column eluent was directly sprayed into the ESI source of the mass spectrometer. Mass spectra were acquired in continuum mode; fragmentation of the peptides was performed in data-dependent mode. Data analysis and protein identification was done by automatically creating peak lists from raw data files using the Mascot Distiller software (version 2.0; MatrixScience). The Mascot search algorithm (version 2.0, MatrixScience) was used for searching against the NCBI database (release date: 14/10/05 exp1; 21/10/05 exp2 and 3; 6/1/06 exp4; 5/3/06 exp5 18/6/06 exp6 and 7; taxonomy: *Mus musculus*). The peptide tolerance was typically set to 150 ppm and the fragment ion tolerance to 0.2 Da. Only doubly and triply charged peptides were searched for. A maximum number of 2 missed cleavages by trypsin were allowed and carbamidomethylated cysteine and oxidized methionine were set as fixed and variable modifications, respectively. The Mascot score cut-off value for a positive protein hit was set to 60. Individual peptide MS/MS spectra with Mowse scores below 40 were checked manually and either interpreted as valid identifications or discarded.

Immunofluorescence

Cells were fixed in 4% paraformaldehyde/PBS for 15 minutes at RT, permeabilised in 0,15% Triton X-100 in PBS, blocked in 1% bovine serum albumin in PBS and incubated with antibodies as described previously. For secondary antibodies, FITC-conjugated goat antibodies against rabbit and mouse IgG were purchased from Molecular probes and Alexa 594-conjugated goat anti-rabbit antibody from Nordic Laboratories. Conjugated streptavidin-Alexa594 and avidin Texas red were used to visualize biotinylated protein (Molecular probes). Images of cells were collected with a Leica DMRBE microscope.

Bandshifts

Protein extract was preincubated with bandshift buffer (10% Glycerol, 20 mM Hepes pH7.4, 20 mM KCl, 1 mM MgCl2, 5 mM DTT, 10 uM ZnCl2, 100 ug/ml BSA, 0.02% NP40) and 2-4 ug of salmon sperm DNA as a non-specific competitor. The reaction was incubated for 20 min. at the RT. Then upon addition of the ^{32}P end-labeled probe the binding reaction was performed for 20 min. RT. The complexes were analyzed by electrophoresis through a 5% acrylamide (37,5:1) 0.5x TBE non-denaturing gel at 8V/cm2 at 4°C. The sequences of the probe were as follows:

3'HS1-wt-s: CGGAAATCAGTGGAACACTTCTGCCCCCTACTGGTATGCAACAGG
3'HS1-wt-as: TCCTGTTGCATACCAGTAGGGGGCAGCCGTGTTCCACTGATTTCCG
3'HS1-mut-s: CGGAAATCAGTGGAACACTTCTGATATCTACTGGTATGCAACAGG
3'HS1-mut-as: TCCTGTTGCATACCAGTAGATATCAGAAGTGTTCCACTGATTTCCG

Antibodies

The following antibodies were used: anti-FLAG M2 monoclonal (Sigma), PARP1 (Alexis biochem / Santa Cruz), Sin3a (Santa Cruz / Upstate), CTCF (Upstate), Dnmt1 (Santa Cruz), YY1 (Santa Cruz), Snf2h (Abcam), nucleophosmin (Santa Cruz), HDAC1 (Santa Cruz), Dnmt3b, CHD4 (Santa Cruz), RFC2 (Abcam), fibrillarin (Heath unpublished), PCNA (mouse), Rif1 (Abcam), UBF (Santa Cruz), Tcof1 (Abnova Corp.), EZH2 (Cell

signaling), Tif1β (gift of Dr. Cammas), Sox2 (Santa Cruz), SMC1 (Bethyl labs.), SMC3 (Bethyl labs.), H2A.Z (Abcam), Ranbp2 (gift of Dr. Melchior), CTCFL (Sleutels unpublished), HP1 (Abcam), Ku80 (Abcam), anti-HIS (Qiagen). CTCF N3 rabbit polyclonal and RPA194 rabbit polyclonal were generated as described (Heath submitted) (Hoogstraten et al., 2002).

Acknowledgements

We like to thank Dr. R. Voit and Dr. I. Grummt for kindly providing us with Flag tagged UBF. Within the ErasmusMC we thank Siska Driegen en Dies Meijer for the pSCT-HaBirA and Rosa26-BirA construct and Jeroen Demmers and Karel Bezstarosti for the mass spectrometry analysis.

Identified protein	Mascot score	Molecular weight (kD)	Peptides total no.	Exp no.	BirA other exp	function /remarks
Abcf2	45,39		1, 1	1,2		GCN20, elongation factor 3
Arbp	93	34	3	4	6	Acidic ribosomal phosphoprotein P0
Atxnl2	73/149		2,3	2, 3		ataxin-2 like, rna splicing
BAF170	85	90/124	1	7	6	SWI/SNF-related RSC8, smarcc2
BAF57	80	47 (runs at 57)	1	7		SWI/SNF-related smarce1,nucleosome remodeling
BCAS 2	52	26	1	4	6,7	breast carcinoma amplified sequence 2
BTF	57	103	2	3	6,7	BCL2-associated transcription factor 1 isoform 2
Bud13	113		1	3		KAR9, positioning mitotic spindle
Capping protein	336	31	4	6	7	actin filament muscle Z-line, beta
Caprin1	103/105/75		2,1,1	1,2,3		cell cycle associated protein 1, GPI-anchored membrane protein
Cdc5l	210/47		5,2	2,3		pre-mRNA splicing
Cdc68	57/228/52	140	1,5,2	1,2,3		FACT complex subunit supt16h, Fact140
CRFG	79/75	74	1, 1	2,6	7	G protein-binding protein GTPbp4
CTCF	69/102/50/369/475/56	84 (runs at 130)	2,2,1,4,6,1	1,2,4,5,6,7		CCCTC binding transcription factor
Ctr9	128	108	2	6		Paf1 complex, SH2-domain bp1
Ddx5	276		5	3,4,5	7	DEAD box polypeptide 5
Dido3	69/68	248	1,3	2,3	6,7	Death inducer-obliterator3, centrosome associated
Dkc1	45/102/137/64	58	1,6,2,1	1,2,3,5	6,7	Dyskerin, H/ACA ribonucleoprotein complex
DNA heli. II	102	61	1	2	1,6,7	helicase
DNA lig. III-beta	75/44	103 (runs at 97)	2,1	2,3	6,7	ligase
DNA PKs	140/105	469	3	2,3		DNA-dependent protein kinase
DNA pol beta	59		1	2	6	DNA polymerase beta
DNA pol. delta ip 3	146/81	46	3,2	2,3	6,7	DNA polymerase delta
Dnmt 1	96/201/42	183 (runs at 180)	3,5,2	1,2,3		DNA methyltransferase
Dync1h1	64	53	1	6		dynein motor protein
EBNA1 bp 2	214/117/46		4,3,1	2,3,4	6,7	Epstein Barr virus nuclear antigen
EIF-4A-I	227	45	3	6	7	Eukaryotic initiation factor 4A-I,RNA helicase
Eif2s2	268	38	5	2		translation initiation factor IF-2
ERCC3/XPB	54		1	3	6,7	excision repair cross-complementing, DNA repair

EZH2	77	90 (runs at 90)	1	4		enhancer of zeste homolog 2
Fibrillarin	260/141/213/142	34 (runs at 34)	7,4,5,3	2,3,4,5	6,7	snRNP component rRNA processing
G3bp1	72/137		2,5	1,2		ras-GTPase-activating protein SH3-domain binding protein
Grwd1 protein	74	25	1	2	6	glutamate rich WD repeat, ribosome biogenesis
Histone H1d	92/101/172	22	2,2,5	2,3,4	6	Histone
Hnrpa3	87/271/53	37	2,7,2	1,2,4,5	6,7	(hnRNP) heterogeneous nuclear ribonucleoprotein
HRPT2	369	60 (runs at 34, 65)	6	6		parafibromin,cdc73, paf1 complex, tumor suppressor
IL- enhancer bp 2	88/85	43	2,2	2,4	6,7	interleukin enhancer binding protein
IL- enhancer bp 3	40/108	36	1,1	2,5	6,7	interleukin enhancer binding protein
Jarid2	66/74	139 (runs at 140,200,80)	1,1	2,3		Jumonji, histone modifications, transcription regulation
KIF1C	54/24		1	3,5	7	Kinesin motor protein
KIF22	37		1	3		Kinesin motor protein (OBP)
KIF23	65/324	56/110	1,6	2,3		Kinesin motor protein (MKLP1)
KIF2C	63/62		1, 1	1, 2		Kinesin motor protein (MCAK)
Kpna2	125/74		2,1	2,3	7	importin alpha2, nuclear import
KSRP	139	36 (runs at 36)	2	1		KH-type splicing regulatory protein
Ku80/XRCC5	54		2	1	7	Ku autoantigen
Laminin-binding protein	98	32	1	6		basal lamina component
Lef1	63	44 (runs at 55)	1	7		wnt1 signaling,lymphoid enhancer binding factor
Leo1	93	75 (runs at 110)	1	6		Paf1/RNA pol. II component
MACF	39	61	1	3		microtubule-actin crosslinking factor
Mago-nashi homolog	47	17	1	6	7	proliferation associated
Mdn1	126/63		4,1	2,3		MIDAS-containing protein, nuclear chaperone, dynein related
Mina	46	53	1	2		myc induced nuclear antigen
MLF1 i.p.	72	46	2	3	6	Myeloid leukemia factor, nucleolus, CENP-50
MRG15	42	41	1	2		MORF-related gene 15 isoform 1, chromatin remodeling
Mus81	50	62	2	4		endonuclease
NCBP80	81/ 68	80	1,1	2,3		similar to 80 kDa nuclear cap binding protein
NHP2 like	73	14	1	5	6	snoRNP, HMG like

NOP56	366/405/	64	8,7	2,3,5	6,7	nucleolar protein
N-pac	78/92/78	60	1,1,1	2,3,4	6,7	Cytokine-like nuclear factor
Nucleolar RNA hel. II/Gu	543	94	14	3,4	1,6,7	helicase
Nucleolin (C23)	491/179	76	11,3	2,3	1,5,6	nucleolar protein, nuclear matrix, histone chaperone
Nucleophosmin (B23)	54/71/198	33	7,2	2,4,5	6,7	snoRNP, nuclear matrix. Pre-rRNA processing
Nucleostemin	85/61		1,1	2,4		nucleolar GTP-binding protein stem cell enriched
NuMa	48		1	5	6,7	nuclear mitotic apparatus protein 1
Orc1	55/293	95	1,6	2,3		origin recognition complex, subunit 1
p38-2G4	46		1	2		cell cycle,proliferation-associated protein 2G4
p62 ras-GAP	89/76	48	1, 1	2,6	7	splicing factor
Paf1	296	60 (runs at 85/73)	5	6		RNA polymerase II associated factor, elongation
Pantetheinase pre-cursor	95		1	5		mrna processing
PARP1	197	114	3	4	1,2,3,6,7	Poly (ADP-ribose) polymerase family, member 1
Peter pan homolog	71	52	2	3	7	splicing
PNK	82/88		2	2	1	polynucleotide kinase 3'- phosphatase
Prc1	54		2	3		protein regulator of cytokinesis 1,MAP65/ASE1
PRP19	96/174/41/115		2,4,1,3	1,2,3,4	6	SNEV nuclear matrix, pre-mRNA processing
PRP40	52/67		2,1	1,2		pre-mRNA processing factor
Pspc1	107/229		3,3	1,2		paraspeckle
PTB-1	84 /138		1, 2	2,4	1,7	polypyrimidine tract-binding protein
PTB-associated factor	174		2	3	7	splicing, U5snRNP associated kinase
Rac GTPase-ap1	48		1	3		
RANbp 2	151 /292	358 (runs at 360)	3,8	2,3	7	RAN binding protein 2 , nuclear pore component
Rbbp7	73/53/45	19	1,1,1	2,3,6	1,7	retinoblastoma binding protein 7, RbAP46 chromatin remodeling
Rbm14	116	66	2	2		synaptotagmin interaction, RNA binding motif 14
Rbm15	40	104	1	2	6	nuclear pore complex
Rbm28	52	26	1	4	6,7	RNA binding protein
RCC2 (1-like)	207/41		3	2,3	7	nuclear transport, chromosomal organization
Rent1	83	124	1	2		regulator of nonsense transcripts 1

Rex2	65	83	1	5		reduced expression 2 guanine nucleotide bp
RFC1	89/58		2,1	2,3,4	7	activator 1 large subunit (RFC1) 145 kD subunit
RFC2	74/114/76	39	2,4,2	1,2,3	7	loads PCNA onto DNA
RFC4	44/37		1,1	2, 4		replication factor
RIF1	42/234/195	265 (runs at 265)	1,9,11	1,2,3		DNA damage respons/telomere
RNA pol I (116,1-2)	38	128	1	1		RPA116, RPA135
RNA pol I (194,1-4)	130/210/294	194	4,6,8	1,2,3		RNA polymerase I largest subunit
RNA pol I (40,1-1)	127	40	1	3		RNA polymerase I small subunit
RNA pol I (PAF49)	165/213		4,6	2,3		RNA polymerase I associated factor 49
RNA pol I (PAF53)	115/108/62	44/49/54	2,1,2	1,2,3		RNA polymerase I associated factor 53
RRS1	78/66	41	1,1	2,3	1,7	ribosome biogenesis regulator homolog
SAP155	155/108/60	146	3,2,2	1,2,3	6	pre-mRNA processing factor, splicing factor 3b
Sipa-1	212		3	7	6	Signal-induced proliferation-associated protein 1
SMC2	53	137 (runs at 140)	1	3		condensin
SMC5	161	129 (runs at 145)	3	2		condensin
Smu-1	107	85	2	2	6,7	suppressor of mec-8 and unc-52 homolog
Snf2H	109	120	2	2		SWI/SNF related, matrix associated subfamily a 5
Son	52	226	1	2		cell proliferation protein truncated isoform
Ssrp1	131	81 (runs at 90)	2	2	6	HMGbox Ssrp1/T160 FACT component
STAR	76		1	6	7	KH domain containing, signal transduction associated 1
Stat3	86	85 (runs at 80-85)	1	6		signal transducers and activators of transcription 3
TAF15	48/163	62 (runs at 32)	1,2	6,7		RNA polymerase II, TBP-associated factor (TFIID)
TAR DNA bp1	130	45	1	3		TDP43 RNA/DNA binding, transcription, splicing
Tceb3	107		2	6		transcription elongation factor B (SIII), polypeptide 3, elonginA
TCF-1	63	22-55 (runs at 50)	1	7		T-cell-specific transcription factor 1
TCF-3	63	75	1	7		T-cell-specific transcription factor 3
TCF-4	63	60 (runs at 60)	1	7		HMG-box transcription factor
TFIIH subunits	174/88	44/52	3,1	2,3		general transcriptionfactor, p44 and p52 subunits
THAP11	74	34	1	6		DNA binding domain THAP domain containing 11
Thrap3 (TRAP150)	85/135/80	150 (runs at 150)	2,4,2	2,3,5	6,7	thyroid hormone receptor associated protein 3

Thymopoietin (Tmpo)	154/122/180	75 (runs at 83)	4,4,4	1,2,3	7	LAP2 nuclear architecture lamin binding
TIF1 beta	207/180/285/316	100	1,4,5,6	2,3,4,5	1,7	co-repressor / HP1 binding
TPX2	86/205/443	86 (runs at 92)	2,5,7	1,2,3		microtubule-associated protein homolog
Treacle (Tcof1)	76/40/156	134	2,1,4	1,2,3	6,7	pre-rRNA methylation;rDNA transcription
Trip 12	118/168/56/57	124/67	3,3,1,1	2,3,6,7		thyroid hormone receptor interactor 12, ubiquitin ligase
TSEP-1	88	35	1	6		TSH receptor suppressor element-binding protein-1
U2AF	144/53	28	3, 1	2,4	6,7	splicing factor
UBF	110/85	94/97 (runs at 94/97)	2, 2	1,2	6,7	upstream binding transcription factor, RNA polymerase I
UBP2L	46/105		1,2	2,3		Ubiquitin associated (Ubap21/Nice4)
Usp10	47		1	1		ubiquitin specific peptidase
Wdr18	277		5	5		WD-repeat domain nucleolar
Wdr5	47	36 (runs at 36)	1	2		recognition of histone modifications, MLL1 complex
Wdr61a	73		1	5	7	repeat domain 61 isoform a
Wdr72.5	38	37	1	2		Bub3, mitotic checkpoint
Wdr74	37	43	1	2,3	7	contains WD40 domain
XPMC2	57		1	6	7	prevents mitotic catastrophe 2 homolog
XRCC1	49	69	1	1		replication, DNA repair
YB1	143/60/88	36 (runs at 36)	3, 1,2	4,5,6	1,2,3	Y box-binding protein

Appendix 1 Overview of mass spectrometry results of CTCFbio pull downs

Identified proteins, Mascot score, total amount of identified peptides, experiment number, molecular weight (kD) are indicated. The table shows proteins identified in pull downs with streptavidin beads from extracts of ES cells (exp. 1-5), P38 lungs (exp.6), P38 Thymus (exp.7) expressing biotin-tagged CTCF and biotin ligase BirA in a heterozygous manner. Conditions used are as indicated in table 1. The lists are corrected for background proteins, which were identified in control pull-downs of the same experiment using extracts expressing only BirA. If proteins was identified from BirA only expressing controls in another experiment this is indicated with the experiment number. For each identified protein, the lists are filtered for duplicates and show only the hits with the highest Mascot score and most identified peptides. The most prominent function of the protein and/or alternative names are indicated in the remarks/function column.

4

CTCF defines the local epigenetic state of the rDNA repeat

Submitted

CTCF defines the local epigenetic state of the ribosomal DNA repeat

Suzanne van de Nobelen[1], Jorg Leers[2], Manuel Rosa-Garrido[3], Helen Heath[1], Anna Akhmanova[1], Verónica Torrano[3], Frank Grosveld[1], Dolores Delgado[3], Rainer Renkawitz[2], Frank Sleutels[1], and Niels Galjart[1].

Department of Cell Biology and Genetics[1], Erasmus MC, The Netherlands; Institute for Genetics[2], Justus-Liebig-Universitaet Giessen, Heinrich-Buff-Ring 58-62, D-35392 Giessen, Germany, Departamento de Biología Molecular[3], Facultad de Medicina, Universidad de Cantabria, Avenida Cardenal H. Oria s/n, 39011-Santander, Spain.

Summary

CTCF is a highly conserved zinc finger protein, which organizes chromatin domains and RNA polymerase II-mediated gene transcription by binding to approximately 15,000 sites in the human and mouse genomes. Here we show that UBF, a key regulator of RNA polymerase I, directly interacts with CTCF and its testis-specific paralogue CTCFL. We identify conserved CTCF-binding sites immediately upstream of the spacer promoter of the ribosomal DNA (rDNA) repeat. CTCF is required for the accumulation of RNA polymerase I, variant histone H2A.Z and UBF near the rDNA spacer promoter, and regulates the transcription of non-coding RNAs from this promoter. Interestingly, these transcripts accumulate strongly after mitosis, marking a fraction of all nucleoli. We conclude that RNA polymerase I-mediated transcription at spacer- and gene-promoters can be independently regulated and suggest that a CTCF-UBF interaction regulates spacer promoter transcription. Interestingly, CTCF organizes chromatin at the rDNA spacer promoter similar to RNA polymerase II-promoters. Such a local epigenetic control may maintain the reservoir of rDNA genes that are poised for transcription.

Introduction

CTCF is a conserved and ubiquitously expressed factor, which was shown to both activate and repress transcription, and to organize epigenetically controlled chromatin domains (Ohlsson et al, 2001). It is generally thought that CTCF acts as an insulator or enhancer-blocker: its binding to DNA prevents the spreading of heterochromatin from chromatin domains separated by CTCF or inhibits the illegitimate interactions between regulatory elements on adjacent chromatin domains, respectively (Wallace & Felsenfeld, 2007).

CTCF function has been studied in detail for a number of genes. In the imprinted *Igf2/H19* locus, for example, CTCF was shown to bind the imprint control region (ICR) when this domain is unmethylated, i.e. when the locus is inherited via the mother (Hark et al, 2000). CTCF binding blocks the interaction between enhancers downstream of the *H19* gene with elements near the *Igf2* promoter. Consequently, *Igf2* expression is inhibited and *H19* is expressed. Importantly, DNA methylation on CpG-residues in the ICR, as occurs in paternally inherited alleles, abolishes CTCF binding, allowing long range interactions between the *H19* enhancer and *Igf2* promoter and leading to the expression of *Igf2* instead of *H19*. These data establish CTCF as a methylation-sensitive enhancer blocker. CTCF-mediated boundaries have also been described in the *Xist/Tsix* locus involved in X-chromosome inactivation (Chao et al, 2002). In this case CTCF would act as an insulator, creating a barrier against the spread of heterochromatin. A third example where CTCF function has been studied in detail is the β-*globin* gene locus. In chicken CTCF acts as an

insulator (Chung et al, 1993), whereas in mouse CTCF mediates the formation of a so-called active chromatin hub, a chromatin state that coincides with high expression of the β-*globin* genes (Palstra et al, 2003). CTCF deletion causes a local alteration in histone modification and the results indicate that CTCF might function to maintain the globin hub in a conformation "poised" for activation (Splinter et al, 2006).

Recently, genome-wide studies have revealed a multitude of CTCF binding sites (Barski et al, 2007; Kim et al, 2007; Xie et al, 2007). For example, in the human genome ~15.000 sites are present, whose distribution over chromosomes correlates with gene density (Kim et al, 2007). The CTCF binding sites that were identified with anti-CTCF antibodies turned out to be virtually identical to a group of motifs (called LM2*), which were discovered in a bio-informatics screen (Xie et al, 2007). These motifs are found within conserved non-coding elements (CNEs), regulatory sequences which do not code for proteins but which are conserved between species. Furthermore, comparison between CTCF binding sites, RNA polymerase II binding sites, transcription start sites (TSSs) and DNAse I hypersensitive sites (HSs) revealed that CTCF is located somewhat upstream of an HS which in turn precedes a TSS (Boyle et al, 2008). CTCF also colocalizes with the variant histone H2A.Z. These data emphasize the central role played by CTCF as a general organizer of RNA polymerase II-mediated transcription.

CTCF binds DNA through an 11-zinc finger (ZF) domain. Different combinations of individual ZFs were originally proposed to modulate binding specificity (Quitschke et al, 2000), however, more recent studies have shown that ZF4-8 are sufficient to bind to a number of physiologically relevant sites (Renda et al, 2007). Interestingly, a testis-specific paralogue of CTCF has been characterized, called CTCFL or BORIS (Brother Of the Regulator of Imprinted Sites), that is highly similar to CTCF in the ZF domain, and that has overlapping DNA binding specificity (Loukinov et al, 2002). CTCF and CTCFL share very little similarity outside their ZF region. To date no common interaction partners of CTCF and CTCFL have been reported.

Although in most localization studies CTCF is not detected inside the nucleolus (Zhang et al, 2004), some cell lines do show an enrichment of CTCF within this compartment, and it has been suggested that CTCF inhibits nucleolar transcription through a poly(ADP-ribosyl)ation-dependent mechanism (Torrano et al, 2006). The nucleolus is the site where 18S, 5.8S and 28S ribosomal RNAs (rRNAs) are synthesized and processed and, together with 5S rRNA, assembled into ribosomes (Grummt, 2003; Shaw & Doonan, 2005). In all organisms the genes for 18S, 5.8S and 28 S rRNA are repeated many times and organized in a head-to-tail fashion in clusters spread over several chromosomes. In mouse and man each ribosomal DNA (rDNA) repeat spans ~44 kb and approximately half of the repeats are inactive (Conconi et al, 1989). Although it is known that methylation plays a role in rDNA inactivation, it is still unknown how exactly the active and inactive genes are marked and segregated.

Transcription of the rDNA repeats occurs by RNA polymerase I and is tightly coordinated with cellular metabolism and cell proliferation. Transcription occurs at two promoters: transcription from the gene promoter gives rise to a ~13 kb (or 45S) ribosomal precursor RNA (pre-rRNA), that is processed into the three mature rRNAs in a complex manner. Efficient transcription from the ribosomal gene promoter requires a multiprotein complex including UBF (upstream binding factor), SL1 (selectivity factor 1) and RNA polymerase I (Bell et al, 1988; Comai et al, 1992; Jantzen et al, 1990). UBF enhances transcription by regulating initiation of transcription after formation of the RNA polymerase I pre-initiation complex (Panov et al, 2006). Another promoter in the rDNA repeat, the spacer promoter, is located in the intergenic spacer (IGS) ~2 kb upstream of the gene promoter. Transcription from this promoter gives rise to intergenic or non-coding RNA (ncRNA), which is thought to serve a regulatory function (Kuhn & Grummt, 1987; Mayer et al,

2006; Paalman et al, 1995).

In order to better understand the function of CTCF we performed a screen for CTCF-interacting proteins. Here we report that UBF, a key regulator of RNA polymerase I-mediated transcription, directly interacts with CTCF and CTCFL. We characterize conserved CTCF binding sites immediately upstream of the ribosomal spacer promoter. Our data show that CTCF controls the binding of RNA polymerase I, UBF and H2A.Z and acts as a regulator of spacer promoter transcription. We suggest that CTCF does this by organizing chromatin near the spacer promoter and that it functions to maintain rDNA repeats in a state poised for activation.

Results

Characterization of biotinylated CTCF

In order to identify CTCF-binding partners we used a biotinylation-tagging and proteomics approach (Figure 4.1A) that was shown to be successful in the identification of interacting proteins for other transcription factors (Rodriguez et al, 2005). As CTCF levels are critical for cells (Heath et al, submitted), 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, prior to the stopcodon of CTCF (Figure 4.1B). This tag is biotinylated upon addition of the bacterial biotin ligase enzyme, BirA (de Boer et al, 2003). A TEV-protease cleavage site was placed in between CTCF and the biotin tag (not indicated in Figure 4.1B). Southern blot and PCR analysis identified homologous recombination events (Figure 4.1C). The resulting allele was termed $Ctcf^{biof}$.

One ES cell line was injected into blastocysts in order to generate knock-in mice carrying a $Ctcf^{biof}$ allele. In these mice CTCF-interacting proteins can be identified *in vivo* in a developmental and tissue specific manner, provided BirA is co-expressed. In addition, the $Ctcf^{biof}$ ES cells were transfected with a plasmid expressing Cre recombinase to remove the neomycin resistance gene, thereby generating the $Ctcf^{bio}$ allele (Figure 4.1B). Furthermore, by using homologous recombination the BirA biotin ligase was placed into the *Rosa26* locus (data not shown). Genotyping and verification of these targeting events was done by PCR (Figure 4.1D). This yielded an ES cell line expressing normal CTCF (from a wild type allele) and biotinylated CTCF (from the $Ctcf^{bio}$ allele). As the biotin tag was placed at the C-terminus of CTCF, the fusion protein was called CTCF-bio.

CTCF-bio cannot be distinguished from untagged CTCF using anti-CTCF antibodies because the bio-tag does not cause a major difference in migration behavior in SDS-PAGE (Figure 4.1E, upper panel). However, CTCF-bio can be clearly detected using streptavidin-based methods (Figure 4.1E, lower panel). Our results indicate that CTCF-bio and CTCF are expressed at similar levels. A pull down on ES cell extracts using streptavidin-coupled magnetic beads showed highly efficient binding of CTCF-bio to the beads (Figure 4.1F). Size fractionation experiments showed that CTCF and CTCF-bio are present in high molecular weight complexes in ES cells (Figure 4.1G). Furthermore, CTCF-bio binds known CTCF target sites such as the *c-Myc* insulator and the β-*globin* 3'HS1 (Figure 4.1H). Combined these data indicate that CTCF-bio is a functional protein.

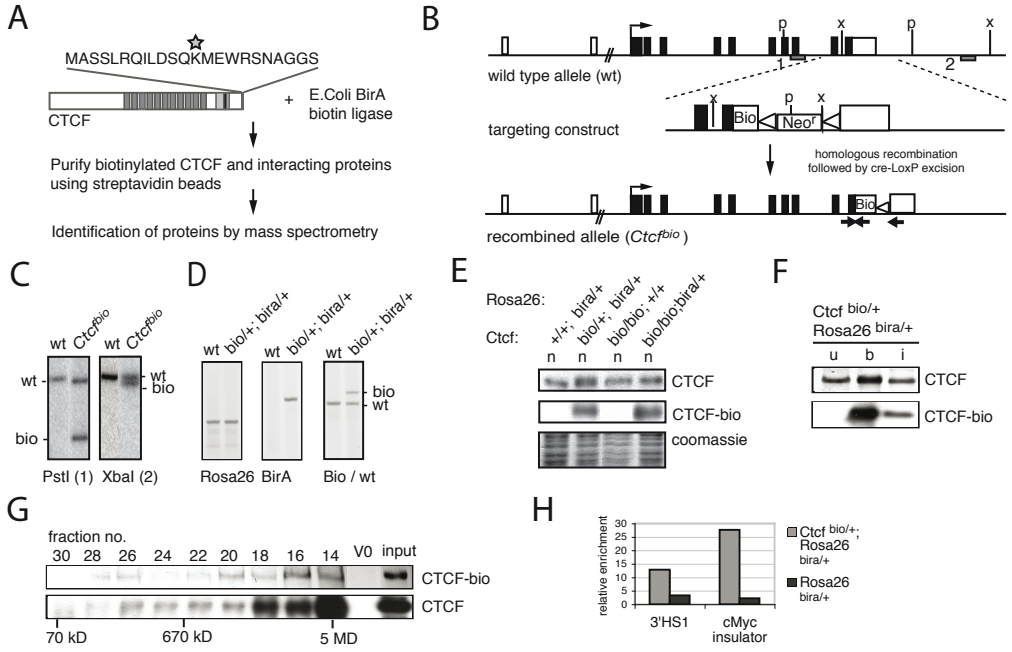

Figure 4.1 Characterization of biotinylated CTCF

A. Biotinylation-based strategy to identify CTCF binding partners. CTCF (grey boxes in middle: zinc-finger region, gray box towards the C-terminus: proline-rich region with AT-hook) is tagged at its C-terminus with a 23 amino acid sequence (indicated in single letter code), that can be biotinylated on a lysine residue (star) upon addition of the E.Coli BirA biotin ligase. Using streptavidin beads biotinylated CTCF (CTCF-bio) and its interacting proteins are pulled down. Interacting proteins were identified by mass spectrometry. **B.** Murine CTCF locus and gene targeting construct. Exons of the The *Ctcf* gene is indicated with a solid line with the exons as boxes. Exon 3 contains the start codon (indicated with arrow) and exon 12 the stop codon. Solid boxes indicate coding parts of the exon, open boxes non coding parts. Southern blot probes (1 and 2) are indicated below the gene. X: XbaI, P: PstI. The targeting construct, with biotin-tag (BIO) and loxP sites (small triangles), flanking a PMC1-neomycin cassette (Neo^r) are shown together with the homologous arms (1,6 kb at 5' end, and 3,8 kb at 3' end). After homologous recombination in ES cells and Cre-mediated excision of sequences in between loxP sites the *Ctcf^{bio}* knock-in allele is obtained. **C.** PstI- or XbaI-digested genomic DNA from ES cells was analyzed by Southern blotting using probe 1 or 2. Targeted (bio) and non-targetted (wt) alleles are indicated. **D.** PCR analysis of genomic tail DNA after targeting of *Ctcf^{bio}* knock-in ES cells with Cre recombinase (to remove the Neo cassette) and Rosa26-BirA (to biotinylate the biotin tag). Genotypes are shown above the lanes. **E.** Nuclear extracts from thymus of mice with indicated genotypes were analyzed by western blot with the indicated reagents. Coomassie staining shows equals loading of the lanes. **F.** CTCF-bio is specifically pulled down with streptavidin beads. Nuclear extracts from ES cells expressing CTCF-bio and BirA were used. The input (n) nuclear extract lane represents 5% of the material used for streptavidin pull down. b: material bound to beads, u: unbound fraction (6 %). **G.** Size-fractionation of CTCF and CTCF-bio. Nuclear extract were fractioned using a superose 6 column. Molecular mass markers are indicated at the bottom. V0 : void volume, input : nuclear extract (5%). **H.** Chromatin immunoprecipitations. CTCF-bio was precipitated with streptavidin-coupled magnetic beads from formaldehude fixed nuclei. In the extract from *Ctcf^{bio}/Rosa26^{BirA}* mice CTCF-bio is present, whereas in extracts from *Rosa26^{BirA}* mice CTCF-bio is absent. Binding is shown to two known CTCF binding sites, i.e. the β-globin 3'HS1 and c-Myc insulator.

UBF interacts directly with CTCF and CTCFL

CTCF-bio was purified from ES cell nuclear extracts under mild conditions using streptavidin-coupled magnetic beads (Figure 4.2A). Proteins co-purifying with CTCF-bio were identified

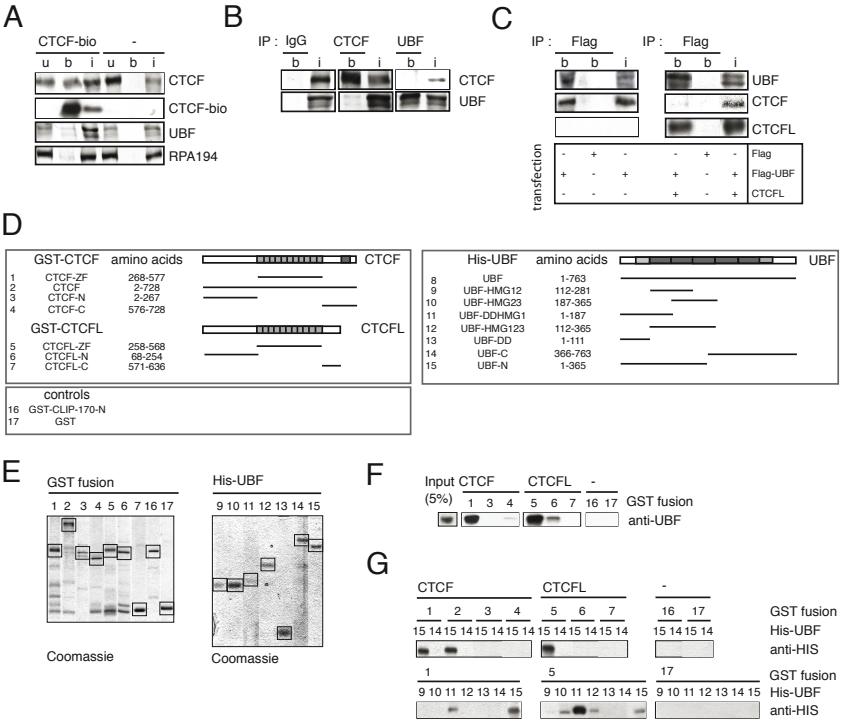

Figure 4.2 Interaction of CTCF and CTCFL with UBF

A. CTCF interacts with UBF and RNA polymerase I in ES cells. Nuclear extracts from ES cells expressing only BirA (negative control) and ES cells expressing CTCF-bio were used for streptavidin pull downs. Extracts were treated with benzonase for 2 hours at 4°C prior to pull down. Western blot were incubated with the indicated antibodies (CTCF-bio was detected with streptavidin-coupled HRP). RPA194: the large subunit of RNA polymerase I. UBF is detected as a doublet consisting of UBF1 and UBF2. U: unbound fraction (6%), b: bound fraction, i: input (5%). **B.** Endogenous CTCF can pull down UBF. Immunoprecipitation (IP) and detection were carried out with the indicated antibodies. C: control (rabbit IgG). B: bound fraction, i: input (5%). **C.** UBF interacts with CTCF and CTCFL. IPs were done with anti-FLAG antibodies from HEK293T cells transfected with FLAG-UBF. To detect CTCFL cells were also cotransfected with CTCFL (right hand panel). Proteins were analyzed with the indicated antibodies.**D.** Schematic representation of the GST and HIS fusion proteins used in this study. **E.** Purified fusion proteins shown on a Coomassie-stained gel. **F.** GST pull down assays of CTCF and CTCFL with nuclear protein extract from ES cells. Binding was performed under low salt conditions washing was done under more stringent conditions. Western blot of pull down material was incubated with antibody against UBF. **G.** GST pull downs of purified GST proteins incubated with UBF N- and C-terminal halfs tagged with HIS (upper panel) and smaller His-UBF fusion proteins (lower panel). Western blot of pull down material was incubated with antibody against His.

by mass spectrometry, classified by BLAST searches and compared to control pull downs. We identified known protein partners of CTCF such as YB-1, Parp1 and nucleophosmin (data not shown), validating our approach.

The mass spectrometry analysis revealed several proteins involved in RNA polymerase I-mediated transcription, including UBF. We also detected proteins that form a complex with UBF, such as the large subunit of RNA polymerase I (RPA194) and its associated factor PAF53 (Hanada et al, 1996; Panov et al, 2006). Moreover the 40 kD and 135 kD subunit of RNA polymerase I (RPA40, RPA135) and the polymerase associated factor PAF49 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.

Streptavidin pull downs followed by western blot analysis confirmed the interaction of UBF and the large subunit of RNA polymerase I with CTCF (Figure 4.2A). Nuclear extracts were treated with benzonase, to exclude the possibility that CTCF interacts with these proteins via DNA and/or RNA. Co-immunoprecipitation (co-IP) with CTCF antibodies indicated that an interaction exists between CTCF and UBF at the endogenous level (Figure 4.2B). We also showed an interaction of CTCF-bio and UBF in thymic extracts (Supplementary Figure S4.1B). However, a co-IP with UBF antiserum did not bring down CTCF (Figure 4.2B). These data suggest that the interaction between CTCF is not very strong, consistent with the fact that UBF is mainly detected in the nucleolus (Jantzen et al, 1990) whereas CTCF is mainly detected in the nucleus (Zhang et al, 2004).

As CTCF and CTCFL are 71% 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 (Figure 4.2C) and overexpressed CTCFL (Figure 4.2C). Interestingly, a diminished interaction between CTCF and UBF was detected in cells expressing CTCFL (Figure 4.2C). These results reveal UBF as the first common interaction partner of CTCF and CTCFL. They also indicate that CTCF and CTCFL compete for binding to UBF.

We next purified bacterially produced GST-CTCF and -CTCFL and his-tagged UBF (for scheme, see Figure 4.2D). Equal amounts of purified GST fusion proteins were first incubated with nuclear extracts from ES-cells, and analyzed for UBF binding by western blot. The data demonstrate that GST-tagged CTCF and CTCFL are able to specifically pull down UBF (Figure 4.2E), verifying results with CTCF-bio. As expected, the interaction with UBF occurs mainly via the ZF domains of CTCF and CTCFL (Figure 4.2E). Subsequently GST-CTCF(L) and his-UBF (both full length as well as truncated fusion proteins, see Figure 4.2D) were incubated with each other. These experiments reveal that the CTCF- and CTCFL-ZF domains interact with UBF HMG-box 1 and the dimerization domain (Figure 4.2F, G). His-tagged proteins containing either the dimerization domain of UBF or HMG boxes 1 and 2 bind less well to CTCF(L) (Figure 4.2G), indicating that for an efficient interaction both regions are necessary.

CTCF binds upstream of the rDNA spacer promoter

Chromatin immunoprecipitation (ChIP) analysis has shown that UBF binds throughout the whole rDNA repeat with relatively low sequence specificity (Mayer et al, 2006; O'Sullivan et al, 2002). We tested whether CTCF also binds to this regulatory region of the rDNA locus, as this would provide a functional explanation for its selective binding to UBF. We first examined primary mouse embryonic fibroblasts (MEFs) derived from wild type mice. ChIP experiments recapitulated the binding pattern of UBF, showing binding at the promoter and spacer promoter regions and a mild enrichment at the enhancers (Figure 4.3A left panel, blue line). By contrast, ChIP of CTCF revealed highly specific accumulation immediately upstream of the rDNA spacer promoter (Figure 4.3A, left panel, red line). CTCF binding coincided with RNA polymerase I (Figure 4.3A, left panel, green line). Strong RNA polymerase I binding to the spacer promoter relative to the gene promoter has been shown previously (Mayer et al, 2006). Using a ChIP approach we also showed CTCF and CTCF-bio binding to the rDNA spacer promoter region in extracts of adult thymus (Supplementary Figure S4.1A).

To examine the functional significance of CTCF binding to the rDNA spacer promoter we generated a system to deplete CTCF *in vitro*. MEFs were isolated from mice homozygous for a

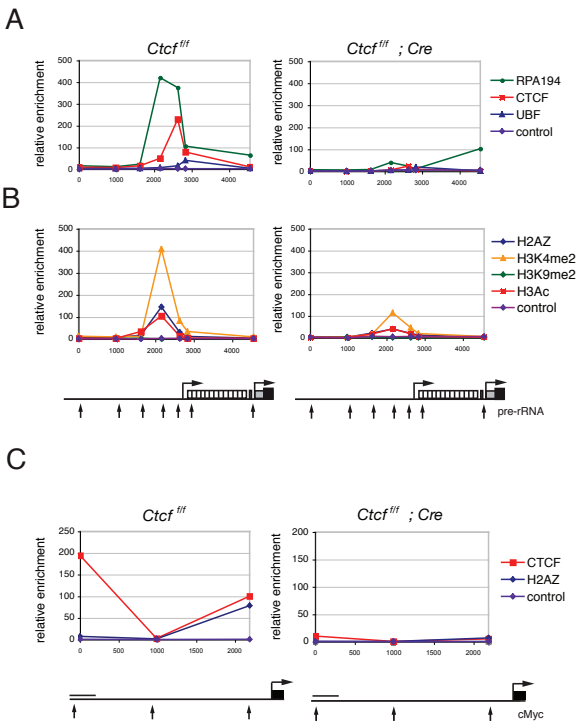

Figure 4.3 Chromatin organization by CTCF
A., B. Chromatin immunoprecipitations (ChIPs) on primary MEFs carrying the conditional *Ctcf* knock-out allele. MEFs were either infected (right hand panels) or not infected (left hand panels) with a lentivirus expressing Cre recombinase. Deletion of CTCF was measured by RT-PCR and was always found to be higher than 80 %. Nuclei were fixed with 1% formaldehyde, and protein-DNA complexes were immunoprecipitated with antibodies against the indicated proteins (RPA194, CTCF, and UBF in A, variant and modified histones in B). Primer sets, located in the 5 kb regulatory upstream of the gene promoter, are indicated with arrows. The location of spacer promoter, enhancer repeats and gene promoter are shown. Control: ChIP with rabbit IgG. **C.** Chromatin immunoprecipitations (ChIPs) on primary MEFs carrying the conditional *Ctcf* knock-out allele. MEFs were either infected (right hand panel) or not infected (left hand panel) with a lentivirus expressing Cre recombinase. ChIP on CTCF and H2A.Z was performed in the regulatory region upstream of the c-Myc transcriptional start site.

Ctcf conditional knockout allele (*Ctcf*^{f}, Heath et al, submitted) and the *Ctcf* gene was deleted by infecting confluent MEFs with a replication-deficient lenti-virus expressing Cre recombinase (Splinter et al, 2006). After four days of culturing only very low levels of *Ctcf* mRNA were detected by qRT-PCR (data not shown), and antibody staining revealed that only a small proportion of MEFs still expressed detectable levels of CTCF (data not shown). CTCF binding to the rDNA spacer promoter was virtually undetectable in cells treated with virus (Figure 4.3A, right panel, red line). These data demonstrate the specificity of the anti-CTCF antibodies in the ChIP experiments. In the absence of CTCF, binding of UBF and RNA polymerase I was severely reduced (Figure 4.3A, right panel, blue and green line, respectively). Remarkably, only binding at the spacer promoter was affected, i.e. the absence of CTCF did not significantly perturb RNA polymerase I binding to the gene promoter. Thus, CTCF exerts a local influence.

CTCF maintains specific histone marks at the spacer promoter

CTCF binding close to the *c-Myc* gene and in the *Igf2/H19* locus is associated with high levels of histone H3 acetylation (Gombert et al, 2003; Pedone et al, 1999). We have shown that CTCF regulates local histone modifications such as histone H3 acetylation and H3K9/K27 dimethylation (Splinter et al, 2006). Recently, CTCF boundaries were found to be highly enriched for the histone variant H2A.Z, for all three states of H3K4 methylation, and for H3K9 monomethylation, but not for H3K9 di- or trimethylation (Barski et al, 2007). H2A.Z (Htz in yeast) was reported to act as a boundary that prevents the spreading of silent heterochromatin into flanking euchromatin regions (Meneghini et al, 2003). H2A.Z was also found to be enriched at promoter regions and represents

a characteristic of active genes in human and chicken (Barski et al, 2007; Bruce et al, 2005). We therefore 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, and low levels of H3K9 dimethylation just upstream of the CTCF binding site (Figure 4.3B, left panel). In the absence of CTCF H2A.Z, H3K4 dimethylation and H3 acetylation (i.e. marks of "active" chromatin and of insulator sites) were clearly downregulated (Figure 4.3B, right panel). Combined the data show that CTCF not only regulates local histone modifications at the spacer promoter but also the presence of a histone variant.

Enrichment of H2A.Z at CTCF binding sites appears to be a general phenomenon (Barski et al, 2007; Boyle et al, 2008). Since we found that CTCF is responsible for depositing H2A.Z at the rDNA spacer-promoter, we tested whether this also occurs with H2A.Z sites near RNA polymerase II-dependent genes. We found that in the absence of CTCF H2A.Z is lost from the *c-Myc* promoter (Figure 4.3C). This results implies that CTCF can mediate deposition of this histone variant close to RNA polymerase I and II promoters. The constant levels of H3K9 dimethylation and histone H3 in the rDNA locus (not shown) indicated that the observed loss of histone modifications and of H2A.Z were not caused by a local loss of nucleosomes in the absence of CTCF.

As the effect of a CTCF deletion was most pronounced on H2A.Z and RNA polymerase I binding, we analyzed the intracellular distribution of these proteins by immunofluorescence. In non-treated MEFs CTCF partially colocalized with H2A.Z (Supplementary Figure S4.2A), consistent with genome-wide ChIP results (Barski et al, 2007). Although signals were more intense in the nucleus both proteins were also detected in the nucleolus. Absence of CTCF did not cause obvious differences in the distribution of H2A.Z and RNA polymerase I (Supplementary Figure S4.3). A reduction in the local binding of RNA polymerase I without affecting global protein levels was also reported in cells where Cockayne syndrome B (CSB) was knocked down (Yuan et al, 2007). UBF and other nucleolar markers (fibrillarin, nucleophosmin), as well as the maintenance methyltransferase Dnmt1, did not show obviously different staining patterns in cells with or without CTCF (data not shown). Immunofluorescence results were confirmed by RT-PCR (data not shown). These results indicate that local changes in DNA binding by specific proteins, that are seen in the absence of CTCF, are not the result of global changes in the levels of these proteins.

Identification of a conserved CTCF binding site in the rDNA locus

The ChIP experiments suggested the presence of a CTCF binding site near the spacer promoter of the mouse rDNA locus. We searched for potential binding sites within the mouse rDNA locus using our own algorithm. One site (R30), which conforms to the CTCF consensus sequence published recently (Kim et al, 2007), was found in the spacer promoter area (Figure 4.4A). A probe (also called R30) was designed and tested in bandshift analysis, using nuclear extracts of non-transfected cells and of cells overexpressing CTCF. The known chicken lysozyme F1 site (Arnold et al, 1996) was used as control. We detected binding of endogenous CTCF as well as bacterially purified GST-CTCF-ZF to the R30 probe (Figure 4.4B, lanes 6-10 and 11-13). Competition experiments indicated that CTCF binds the known FI probe less efficiently than R30, (Figure 4.4C, lane 3). These data demonstrate that CTCF binds R30 through its ZF domain. Previous studies have shown that the mouse, rat and hamster rDNA repeats share significant sequence similarity in the spacer promoter region of the IGS (Tower et al, 1989). Rat and hamster rDNA also contain the CTCF binding site (Supplementary Figure S4.3A). Based on alignment information we mutated three residues within R30 and performed bandshifts with normal and mutant R30 probes. As

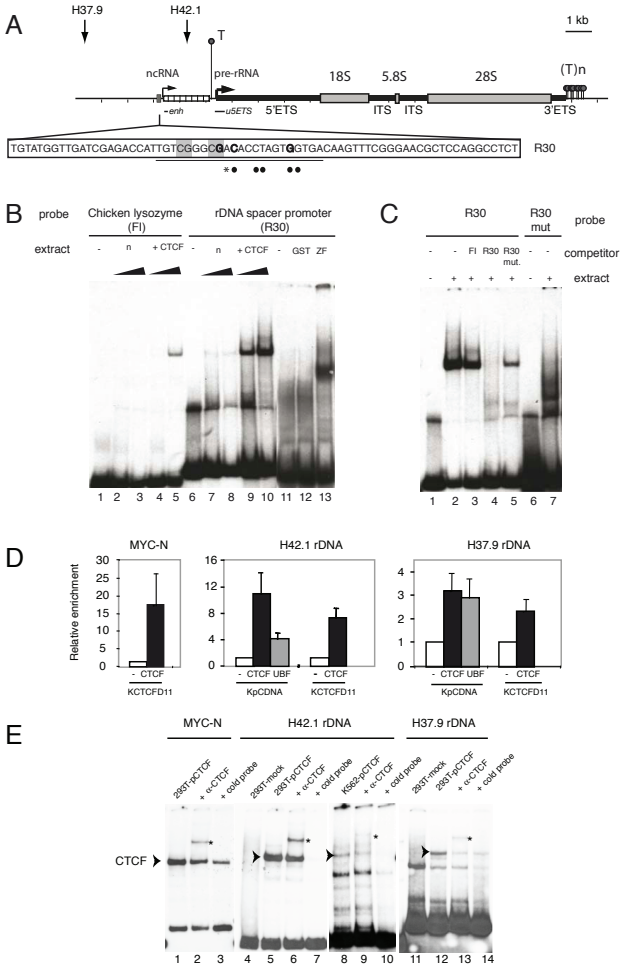

Figure 4.4 CTCF binds near the rDNA spacer promoter

A. Schematic representation of the mouse rDNA repeat. Transcription initiation from the spacer-promoter (yielding ncRNA) and the gene promoter (yielding pre-rRNA) is indicated by right pointing arrows. The 45S pre-rRNA is divided into 5'- and 3'-external transcribed spacer (ETS), internal transcribed spacers (ITS), and 18S, 5.8S and 28S rRNA genes. Positions of the CTCF consensus site (grey box), enhancer repeats (white boxes), TTFI binding sites (T) and probes used for RNA FISH (*enh* and *u5ETS*) are shown. The approximate position of the CTCF binding sites (H37.9 and H42.1) in the IGS of the human rDNA repeat are indicated above the mouse rDNA repeat by downward arrows. The sequence of the R30 oligo that contains the consensus site (underlined) is shown below the rDNA repeat. CpG di-nucleotides are indicated in grey, highly conserved CTCF consensus site residues are indicated by a dot (the asterisk indicates deviation between consensus site prediction and real residue). Nucleotides in bold were mutated (R30mut). **B., C.** EMSA analysis to verify binding of CTCF to mouse rDNA. Bandshifts were performed using a known CTCF binding site probe (F1, chicken lysozyme) as control, the spacer promoter probe (R30), which is close to the PCR fragment used for ChIP, and the mutated R30 site (R30mut). Extracts of HEK293T cells (increasing amounts are indicated), either non-transfected or transfected with CTCF (+CTCF), were incubated with the indicated probes. Competitor (non labelled probe) was added in 300-fold excess. Bandshifts were also done with purified bacterial proteins (GST only and zinc-fingers of CTCF tagged with GST). **D.** CTCF interacts with human rDNA *in vivo*. ChIP analysis with CTCF and UBF antibodies showing CTCF binding to the IGS of the rDNA repeat. Chromatin was prepared from K562 cells stably transfected with CTCF (KCTCFD11) and the empty vector (KpCDNA). Relative enrichment was quantified by real-time PCR with the indicated primer sets. A known CTCF target site (MYC-N) was used as positive control for ChIP. Data were normalized against the enrichment for the negative control

MYC-H.1 (Gombert et al, 2003). The value for the amount of PCR product present from ChIP assay without antibody was set as 1 (white bars). Error bars represent s.e.m. of five to seven independent experiments for CTCF and two for UBF. **E.** CTCF interacts with human rDNA *in vitro*. EMSA analysis with nuclear extracts from 293T cells or K562 cells transfected with CTCF or mock transfected. P^{32}-labelled 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. Arrowheads indicate binding of CTCF; asterisks indicate supershift bands that appear after incubation with anti-CTCF antibody.

shown in Figure 4.4C CTCF binds less efficiently to mutant R30 (lanes 5 and 7). Thus, we have identified a novel CTCF binding site in the mouse rDNA repeat that serves to assemble important RNA polymerase I associated factors. The site is conserved in rat and hamster rDNA repeats. Interestingly the CTCF-binding site includes two CpG residues, indicating it can be methylated *in vivo*.

The IGS of the human rDNA repeat is completely divergent in sequence from the mouse IGS (Supplementary Figure S4.4) and the presence of a spacer promoter has to our knowledge not been described. Despite this divergence we identified two potential CTCF binding sites in the rDNA repeat 0.9 kb and 5.1 kb upstream of the ribosomal gene promoter (called H42.1 and H37.9, respectively, see Figure 4.4). ChIP analysis revealed occupancy of CTCF in the rDNA at H42.1 and H37.9 sites (Figure 4.4D). The MYC-N probe was used as positive control for CTCF binding (Ohlsson et al, 2001). We also found binding of UBF to these rDNA regions, as described (Mais et al, 2005). Binding of CTCF was more prominent in the region near H42.1 than near H37.9 (Figure 4.4D). ChIP results were confirmed *in vitro* by EMSA analysis (Figure 4.4E). Nuclear extracts from cells transfected with CTCF showed stronger binding to H42.1 and H37.9 rDNA probes (Figure 4.4E, lanes 5, 8 and 12) compared to extracts from mock-transfected cells (Figure 4.4E, lanes 4 and 11). The specificity of the binding was shown by competition with unlabeled probes (Figure 4.4E lanes 7, 10 and 14) and by supershifts using anti-CTCF antibody (Figure 4.4E lanes 6, 9 and 13). Incubation with an anti-actin antibody, used as a negative control, did not produce supershifts (data not shown). Altogether these results demonstrate that CTCF associates at very similar positions within the IGS of mouse and human rDNA. Why CTCF binds twice in human and only once in mouse rDNA is unclear, but it might be linked to the presence of a highly repetitive region of Alu repeats in the human rDNA (Supplementary Figure S4.4).

CTCF regulates 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 could not consistently detect an effect on the ratio of pre-rRNA (45S) compared to *Gapdh* mRNA in CTCF depleted MEFs (Figure 4.5A). Furthermore, the ratio of mature 18S rRNA versus *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 fibroblasts.

Using nuclear run-on analysis we investigated transcription from spacer and gene promoters in the presence and absence of CTCF. In non-treated *Ctcf*$^{f/f}$ MEFs we observed higher transcription from the spacer promoter than from the gene promoter (Figure 4.5B). This result is consistent with published data (Espada et al, 2007; Kuhn & Grummt, 1987). Deletion of CTCF significantly reduced transcription from the spacer promoter but did not affect transcription from the ribosomal gene promoter (Figure 4.5B). These results suggest that CTCF regulates transcription from the spacer promoter and that it does so independently of transcription from the gene promoter.

3T3L1 cells can be differentiated into adipocytes, which results in the repression of rRNA

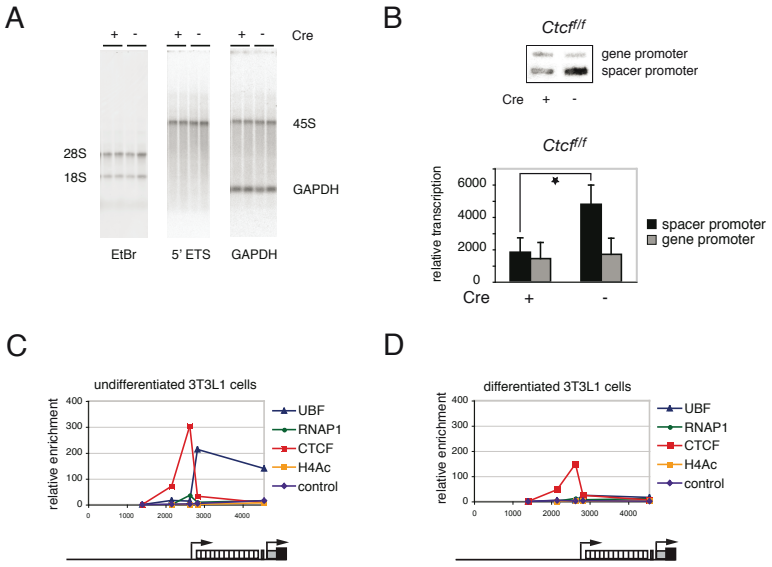

Figure 4.5 CTCF regulates spacer promoter transcription
A. Northern blot analysis. Total RNA was isolated from *Ctcf*$^{f/f}$ primary MEFs either treated (+) and or not treated (-) with Cre (2 samples are shown of each genotype, >10 independent samples were analyzed per genotype). RT-PCR confirmed *Ctcf* deletion (>90 %, not shown) in Cre-treated samples. RNA was blotted and analyzed sequentially with a probe against the 5' ETS region (Akhmanova et al, 2000), which clearly reveal the 45S pre-rRNA precursor, and a probe against *Gapdh*. As the same blot was used, the 45S rRNA is still observed in the second hybridization. **B.** Spacer promoter transcription is repressed in the absence of CTCF. A nuclear run-on analysis was performed for RNA polymerase I-mediatedtranscription (α-amanitin added at 100 ug/ml) on the spacer- and gene promoters in *Ctcf*$^{f/f}$ primary MEFs either treated (+) and or not treated (-) with Cre. In the upper panel the result of a typical run-on experiment is shown, the graph represents the average of three independent experiments (*: P< 0.02, student t-test). **C.,D.** Chromatin immunoprecipitations (ChIPs) on undifferentiated (C) and differentiated (D) 3T3L1 cells. Nuclei were fixed with 1% formaldehyde, and protein-DNA complexes were immunoprecipitated with antibodies against the indicated proteins (RPA194, CTCF, UBF and histone H4 acetylation).

transcription by more than half (Li et al, 2006). Increased heterochromatin features at the rDNA promoter, such as DNA methylation and histone H4 hypoacetylation, accompany this repression (Li et al, 2006; Santoro et al, 2002). Using this cell system we investigated CTCF, UBF and RNA polymerase I binding on the rDNA repeat. ChIP analysis revealed binding of all three proteins at the spacer promoter (Figure 4.5C). The binding patterns were indistinguishable from MEFs, demonstrating that binding at the spacer promoter is similar in different cell types. As reported before (Li et al, 2006) UBF and RNA polymerase I binding to the rDNA repeat were reduced in differentiated 3T3L1 cells (Figure 4.5C, D). Interestingly, also CTCF binding was reduced by half in differentiated compared to undifferentiated 3T3L1 cells (Figure 4.5C, D). The fact that turning down transcription and increasing methylation in 3T3L1 cells affects CTCF binding suggests that CTCF might mark the active ribosomal genes in a methylation-sensitive manner.

Localization of spacer-promoter-derived transcripts

Transcripts arising from the spacer promoter are highly unstable, i.e. they are quickly processed into much smaller products (Mayer et al, 2006). The 3' end product of the spacer-promoter-

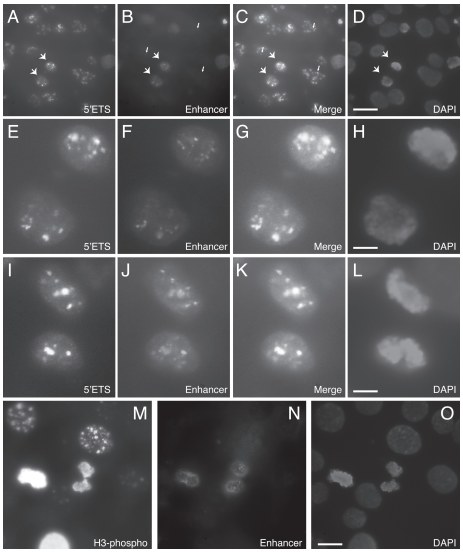

Figure 4.6 Localization of transcripts from the spacer-promoter

A.-L. RNA FISH on MES cells. Proliferating MES cells were fixed and hybridized with a digoxygenin-labelled (green) unstable 5' ETS probe (Akhmanova et al, 2000), which marks all transcribing nucleoli, and with a biotin-labelled (red) oligonucleotide against the enhancer region. Cells were counterstained with DAPI. Two cells that have just undergone mitosis are indicated by large arrows. Small arrows indicate enhancer signal in nucleoli of cells in interphase. **M.-O.** Combined RNA-immuno FISH on MES cells. Proliferating MES cells were fixed and hybridized with a biotin-labelled oligonucleotide against the enhancer region (red) and subsequently incubated with antibodies against phosphorylated histone H3, a cell cycle marker. Cells were counterstained with DAPI.

derived transcript binds the Nucleolar Remodelling Complex (NoRC) with high affinity, and is required for its activity (Mayer et al, 2006). To examine the intracellular distribution of spacer-promoter-derived transcripts we performed RNA FISH using a biotinylated oligonucleotide probe against the 5' end region of the spacer-promoter-derived transcript and a digoxygenin-labelled probe covering the the 5' external transcribed spacer (ETS), which marks all nucleoli (Akhmanova et al, 2000). We analyzed spacer-promoter-derived transcripts in rapidly proliferating MES cells. The advantage of using this cell type is that all phases of the cell cycle are well represented, whereas confluent MEFs accumulate in G1.

Spacer-promoter-derived transcripts were most readily detected in cells with compacted nuclei (Figure 4.6A-C, large arrows). Signal was often detected in two nearby nuclei, indicative of cells exiting mitosis. In cells with larger nuclei we occasionally observed a weak spacer-promoter-derived signal (Figure 4.6A-C, small arrows). The staining patterns of spacer-promoter-derived and 5'ETS probes were overlapping but not identical (Figure 4.6E-L), suggesting a different spatial coordination. These data indicate that spacer promoter transcription is highly regulated during the cell cycle and underscore the conclusion that transcription from the spacer promoter can occur independently of the gene promoter.

Histone H3 is specifically phosphorylated during mitosis, whereas upon exit of mitosis a global dephosphorylation of H3 can be observed (Hans & Dimitrov, 2001). We performed immuno-FISH experiments with the biotinylated probe against spacer-promoter-derived transcripts, in combination with antibodies against phosphorylated H3, to examine at which stage of the cell cycle spacer promoter transcription is most active. Our analysis revealed that both cells in metaphase and telophase contained spacer-promoter-derived transcripts (Figure 4.6M-O). These data indicate that transcription from the spacer promoter is most active at the end of mitosis.

Discussion

Here we identify UBF as the first common interaction partner of CTCF and CTCFL. This emphasizes a role for both proteins in the proper chromatin organization of the rDNA repeats. We demonstrate a direct interaction between the ZF domains of CTCF and CTCFL and HMG box1 and the dimerization domain of UBF, explaining the observation that CTCF and CTCFL compete for binding to UBF. CTCF and CTCFL may act antagonistically in cultured cells (Vatolin et al, 2005). It remains to be determined whether this is the case in rRNA transcription.

We did not observe a clear accumulation of CTCF in nucleoli of MEFs, even though we showed an important role for CTCF in ribosomal spacer promoter transcription. This is not uncommon, for example, c-Myc is also hardly found in the nucleolus, but is very important for rDNA transcription (Arabi et al, 2005). If CTCF is only bound to non-methylated rDNA and one CTCF molecule is present per mouse rDNA repeat, then with about half of the rDNA repeats being completely methylated (Brock & Bird, 1997) we expect ~200 CTCF molecules in the nucleolus. This small number explains why nucleolar CTCF is difficult to detect by immunofluorescence procedures in MEFs. UBF on the other hand is abundantly present in the nucleolus, binding with ~100 molecules per active rDNA repeat. Thus in MEFs only a small fraction of CTCF interacts with UBF. This observation and the fact that UBF is highly dynamic (Chen & Huang, 2001) makes it likely that the interaction between CTCF and UBF is transient.

HMG-boxes are involved in protein-protein interactions and DNA binding. The dimerization domain of UBF is involved in the formation of hetero- and homodimers (O'Mahony et al, 1992). A combination of the dimerization domain and the first three HMG-boxes of UBF is important for the formation of small ~175bp DNA loop structures, called enhancesomes (Stefanovsky et al, 2001), which bring UBF-bound enhancers in close contact to the gene promoter (Bazett-Jones et al, 1994; Sullivan & McStay, 1998). The fact that CTCF interacts with the first HMG box of UBF argues against cooperative binding of these proteins to DNA, since HMG-box 1 is absolutely necessary for UBF binding to DNA (Putnam et al, 1994). Instead, the UBF-CTCF interaction might be important for the recruitment of RNA polymerase I to the spacer promoter. Direct binding by CTCF to components of the RNA polymerase I complex, as suggested by our mass spec data, might aid in this recruitment. Alternatively, CTCF may load UBF on rDNA. UBF, as part of the architectural HMG box protein family, might change the topology of the spacer promoter, thereby facilitating binding of other factors (Bazett-Jones et al, 1994; Chen et al, 2004).

The importance of CTCF binding near the mouse spacer promoter is underscored by the observation that CTCF binding sites are present at similar positions in rat, hamster and human rDNA repeats, despite the fact that the IGS's of these different species are not conserved. ChIP analysis shows enrichment of CTCF on two binding sites in K562 cells indicating that CTCF might have additional regulatory functions in the human rDNA repeat. Interestingly, a prominent nucleolar localization of CTCF was described in K562 cells, which correlates with poly(ADP-ribosylat)ion and growth arrest of cells (Torrano et al, 2006). Post-translational modifications may alter the function of CTCF and influence its interactions. In future studies we will address this issue in more detail. Previously, the *Xenopus laevis* rDNA repeat was reported to contain multiple weak CTCF binding sites near its spacer promoter (Bell et al, 1999). Although a physiological significance for rDNA transcription was not investigated the result is consistent with our data.

In differentiated 3T3L1 cells reduced binding of CTCF correlates with decreased RNA polymerase I binding at the spacer promoter. In MEFs CTCF is essential for the highly localized binding of RNA polymerase I and H2A.Z. In the absence of CTCF not only these factors and UBF bind much less efficiently, also histone modifications associated with gene activity are reduced

and spacer promoter transcription is downregulated. This is the first example of a role for CTCF in maintaining promoter activity by regulating the presence of specific histone marks and RNA polymerase components.

A subset of the 15,000 CTCF binding sites in the human genome (Barski et al, 2007) have been shown to be located immediately upstream of RNA polymerase II promoters (Boyle et al, 2008). As depicted in Figure 4.7A, there is a striking similarity in the organization of these promoters and of the RNA polymerase I-regulated spacer promoter. First, in both types of promoters CTCF binds ~200 bp upstream of the TSS. Second, in both promoters a DNaseI HS has been described (Boyle et al, 2008; Langst et al, 1997), which is located in between the TSS and CTCF binding site. Furthermore, 200-300 basepairs upstream of the CTCF binding site enrichment of H2A.Z and of H3K4me2 (and H3K4me3) is detected in both types of promoters. In the case of RNA polymerase II promoters, these modifications were shown to mark active or "poised" promoters. H2A.Z was suggested to prevent spreading of heterochromatin into, for example, promoter regions (Fan et al, 2004; Meneghini et al, 2003; Raisner & Madhani, 2006). We show here that CTCF is required for H2A.Z accumulation at genes transcribed by RNA polymerase I and II. We also demonstrate a critical role for CTCF in spacer promoter transcription. We therefore propose that when CTCF is bound immediately upstream of a promoter, it is essential for the local organization of chromatin near that promoter and is required for proper transcription. Interestingly, it was reported recently that the transcription factors p53 and c-Myc provide positioning cues that direct the location of H2A.Z-containing nucleosomes (Gevry et al, 2007). We hypothesize that CTCF also regulates transcription of genes by positioning H2A.Z within chromatin.

The biological function of the spacer promoter and the ncRNA transcript that is generated

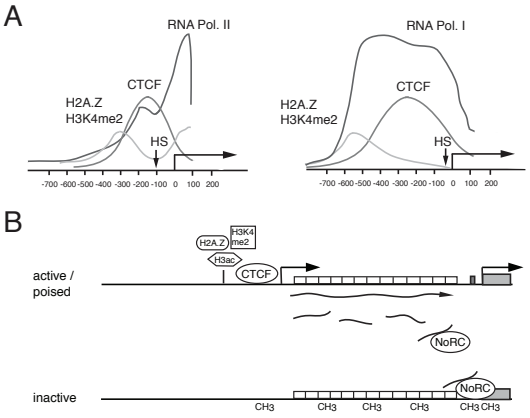

Figure 4.7 Models for CTCF function in the rDNA locus

A. Conserved function for CTCF as a local organizer of chromatin. The left hand panel depicts the local chromatin environment around CTCF-positive RNA polymerase II promoters, data are based on (Boyle et al, 2008). The graph represents an average of multiple active promoters analyzed. Position relative to the transcriptional start (TSS, indicated by right pointing arrow) is indicated in base pairs. Peak height is indicative of relative enrichment. HS: DNaseI hypersensitive site. The right hand panel represents results of our ChIP analysis in the mouse rDNA locus (note that the accuracy of this representation is lower than the actual experiment). The position of the HS is based on (Langst et al, 1997). Positions are relative to the transcription initiation site of the spacer promoter. **B.** A balance between active and inactive rDNA repeats. The upper cartoon represents an active (or potentially active, or "poised") rDNA repeat. CTCF is bound upstream of the spacer promoter, thereby regulating the deposition of other critical factors, enhancing transcription of ncRNA, and maintaining the fraction of active repeats. ncRNAs are processed and the 3'end product is bound by NoRC, which can then act "in trans" to maintain the balance of inactive rDNA repeats (lower cartoon).

from it are still poorly understood. Early experiments suggested that the spacer promoter, as well as the enhancer region that is often located in between spacer and gene promoters, act together to stimulate pre-rRNA transcription (De Winter & Moss, 1986; Grimaldi et al, 1990; Tower et al, 1989). More recent experiments have shown that ncRNAs generated from the spacer promoter are unstable; transcripts are rapidly processed and degraded and only the 3' end (~150 nt) of the transcript, that matches the rDNA gene promoter, is bound to NoRC (Mayer et al, 2006). Surprisingly, in this context the spacer-promoter transcript functions in rDNA silencing instead of activation (Mayer et al, 2006). This appears to be at odds with earlier results. Our data uncouple the regulation of spacer promoter and gene promoter transcription both in terms of cell cycle control and in terms of regulatory factors. For example, we show that in rapidly dividing cells transcription of the spacer promoter peaks immediately after mitosis. CTCF remains bound to mitotic chromosomes (Burke et al, 2005), and it could therefore aid in the upregulation of spacer promoter transcription at the end of mitosis. During cell division specific chromatin states must be inherited from one generation to the next. We speculate that the increased expression of ncRNAs from the spacer promoter at the end of mitosis is required to regulate or mark the fraction of active and inactive rDNA repeats.

We propose a model that incorporates our results into other published data (Figure 4.7B). There is a link between inactive rDNA repeats and DNA (hyper)methylation (Grummt, 2007). As CTCF binding is sensitive to methylation (Renda et al, 2007) and CTCF binding sites on the rDNA contain CpG residues, methylation of these sites may regulate CTCF binding and CTCF may therefore act as a methylation-sensitive binary switch that marks a subset of active or "poised" ribosomal genes. CTCF binding allows the set-up of an appropriate chromatin environment for transcription from the spacer promoter. Consistently, under conditions of repressed pre-rRNA transcription in differentiated 3T3L1 cells (Li et al, 2006) CTCF binding to the spacer promoter is reduced. Transcription from the spacer promoter is stimulated by CTCF. By generating spacer promoter transcripts CTCF is indirectly "feeding" NoRC with its 3'end degradation product. The association of NoRC to this short RNA molecule is required for NoRC binding to chromatin and for heterochromatin formation. This results in establishment and maintainance of inactive rDNA repeats by NoRC (Santoro et al, 2002). We propose that CTCF and NoRC are involved in balancing active and inactive rRNA genes and that the ncRNAs coming off one CTCF-driven spacer promoter act "in trans" together with NoRC to maintain inactive rDNA repeats. An important role in this balance is necessarily played by processing factors that degrade the spacer transcript and generate the NoRC substrate. Recent data implicate such factors in the regulation of rRNA transcription (Prieto & McStay, 2007).

Materials and methods

Antibodies

CTCF N3 rabbit polyclonal and RPA194 rabbit polyclonal were generated as described (Heath et al submitted) (Hoogstraten et al, 2002). The CTCFL polyclonal antibodies are described elsewhere (Sleutels et al, manuscript in preparation). Anti-Histone H2A.Z (ab4174), anti-dimethyl-Histone H3 (Lys4) (ab7766), anti-Histone H3 (ab1791) antibodies were from Abcam. Anti-acetyl Histone H3 (06-599), anti-acetyl Histone H4 (06-866), anti-dimethyl-Histone H3 (Lys9) (07-441) antibodies were from Upstate. Anti-UBF (sc-13125) antibody is from Santa Cruz. Streptavidin-HRP (RPN1231VS) and secondary HRP labeled anti-mouse (NA931VS) and anti-rabbit antibodies (NA934V) are from Amersham. Anti-HIS antibody is from Qiagen.

Cell lines, transfections and lentiviral transduction

To generate the *Ctcf*^{biof} knock-in allele, a CTCF-TEV-bio in-frame fusion DNA was generated by PCR. The biotinylation sequence has been published (de Boer et al, 2003), the TEV (Tobacco Etch Virus) protease cleavage site of seven amino acids was based on published data (Dougherty & Parks, 1989). The loxP-Neo-loxP vector (Neo: neomycin resistance gene) has been described (Hoogenraad et al, 2002). The 5'-end and 3'-end homology arms from the *Ctcf* gene were generated by PCR using mouse 129 PAC clones (Osoegawa et al, 2000) as template. A schematic representation of the knock-in construct is shown in Fig. 1A. Targetting into IB10 129 embryonic stem (ES) cells was performed as described (Hoogenraad et al, 2002). ES cell DNA was analyzed by Southern blot using radiolabeled probes outside of the region of homology (see Fig. 1A). For confirmation of homologous recombination we used different 5' end and 3' end probes as well as a PCR-based genotyping assay (for primer sequences, see Supplementary Table S4.2).

Ctcf$^{bio/f}$ ES cells were transfected with CMV-Cre to remove the neomycin resistance cassette. These cells were used in a second round of homologous recombination to target the Rosa26 locus with a construct encoding HA-tagged BirA (Driegen et al, 2005). Verification of homologous recombined clones was in this case done by PCR only (for primer sequences, see Table S3). Control BirA-positive ES cell lines used in the different experiments, have been described previously (Driegen et al, 2005).

3T3L1 cells were purchased from the ATCC (CL-173). The culturing and differentiation of 3T3L1 cells was done as described (Li et al, 2006). 293T cells were cultured as described (Splinter et al, 2006). The *Ctcf*$^{f/f}$ primary mouse embryonic fibroblasts (MEFs) were isolated as described (Akhmanova et al, 2005) from E13.5 embryos derived from conditional *Ctcf*$^{f/f}$ knockout mice (Heath et al, manuscript in preparation).

Transient transfections in 293T cells with Flag-UBF and pcDNA3-CTCFL were done using Lipofectamine^{TM}2000 (Invitrogen) or Optimem (GiboBRL). Cells were analyzed 24 hours after transfection. Cre-lentivirus production and transduction of confluent primary CTCF cKO MEFs was done as described (Splinter et al, 2006), with the exception that cells were split and diluted two-fold 24h 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 subline derived from K562 myeloid leukemia cells, which is stably transfected with a constitutive CTCF expression vector and which overexpresses moderate CTCF levels (2-3-fold) compared to cells transfected with the empty vector (KpCDNA subline) (Torrano et al, 2005). For EMSA experiments, 293T cells or K562 cells were transfected with 8 µg of pcDNA3-CTCF expression vector (Torrano et al 2005) using Lipofectamine^{TM}2000 (Invitrogen).

Affinity chromatography and size fractionation

Nuclear extracts were prepared as described (Splinter et al, 2006). Salt concentration in the extract was adjusted to 100 mM NaCl. Unless stated differently all immunoprecipitation (IP) and pull down reactions were performed in 100 mM NaCl, 0.3% NP40, 20 mM Hepes pH8, 0.2 mM EDTA, 10 mM MgCl2, protease inhibitors. Benzonase (Novagen) was sometimes added to remove DNA and RNA. Streptavidin pulldowns were done as described (Rodriguez et al, 2005) with the exception that wash buffer and binding buffer were the same as the IP buffer indicated here. For IPs nuclear extracts were pre-cleared at 4°C using Protein A sepharose beads. Washes were done at 4°C in wash buffer (100 mM NaCl, 20 mM Tris pH7.5, 0.3% NP40 and protease inhibitors (Complete; Roche)). IPs were performed by adding antibodies to the samples and incubating for 1h 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 UBF fusion proteins were generated by PCR using mouse UBF cDNA from a FLAG-tagged UBF construct (a kind gift of Dr. I. Grummt). Primers contained NheI and BamHI sites for subcloning into the pET28a vector. HIS-tagged proteins were purified using Ni-NTA (Qiagen) and eluted from the beads with 200 mM imidazole in 20 mM Hepes (pH 7.5), 100mM KCl, 10 mM beta mercaptoethanol, 10% glycerol (v/v). Imidazole was removed by dialysis. GST-tagged fusions of mouse CTCF and CTCFL were amplified using mouse CTCF (IMAGE 6825952) and CTCFL (Sleutels, unpublished) cDNAs as templates. cDNAs were

cloned into pGEX-3X and purified using glutathione-Sepharose 4B (Amersham Biosciences). Chicken CTCF GST tagged fusion proteins were described previously (Lutz et al, 2000). Pull downs were performed in binding buffer (20 mM Tris-HCl pH 8, 100 mM NaCl, 0.05% Triton X-100) for 2 h at 4°C. Washes were done in high salt wash buffer (20 mM Tris-HCl pH8, 400 mM NaCl, 0.05% Triton X-100). GST pull downs on ES cell nuclear extracts were done using the binding and washing conditions as described in the IP section.

Size fractionation of protein complexes was done on an AKTA FPLC apparatus with a Superose 6 10/30 column (Amersham Biosciences). Fractions were precipitated with 100% trichloroacetic acid and analyzed by western blotting as described (Lansbergen et al, 2006). Molecular size standards were thyroglobulin (670 kD) and albumin (66 kD) (Amersham Biosciences).

SDS-PAGE, western blot analysis and mass spectrometry

Bound proteins were eluted from beads by boiling in sample buffer (1x Laemmli buffer). For western blot analysis samples were electrophoresed on SDS polyacrylamide gels and blotted onto PVDF membranes (MilliPore) using a semi-dry blotting apparatus (Biorad). Signal detection was performed using ECL (Amersham).

For mass spectrometry samples were treated and analyzed as described (Wilm et al, 1996). Data analysis and protein identification was done as reported (Lansbergen et al, 2006). The Mascot search algorithm (version 2.0, MatrixScience) was used for searching against the NCBI database (release date: 14/10/05 exp1; 21/10/05 exp2 and 3; 6/1/06 exp4; 5/3/06 exp5; taxonomy: *Mus musculus*). The Mascot score cut-off value for a positive protein hit was set to 60. Individual peptide MS/MS spectra with Mowse scores below 40 were checked manually and either interpreted as valid identifications or discarded.

Chromatin immunoprecipitation

We used mouse (Acc. Number: BK000964), human (Acc. Number: U13369), rat (Acc. Number: X04084) and hamster (Acc. Number: DQ235090) rDNA sequences for alignments, PCR and ChIP experiments and for probe generation.

Preparation of crosslinked chromatin ($2x10^7$ cells treated with 1% formaldehyde for 10 minutes at RT), sonication to 300-800 base pair fragments and immunoprecipitations were as described in the Upstate protocol (http://www.upstate.com). At least two independent chromatin immunoprecipitations (ChIPs) were carried out per experiment. For streptavidin ChIPs exceptions to the protocol were as follows: streptavidin beads blocked for 1h RT in 0,2mg/ml sonicated salmon sperm DNA were used, elution was performed O/N at 65°C in elution buffer (0,1% NaHCO3, 1% SDS, 0,2M NaCl). Quantitative real-time PCR (Opticon I, MJ Research and MyiQ ,Biorad) was performed using SYBR Green (Sigma), Platinum Taq DNA Polymerase (Invitrogen) and 100ng of each primer under the following cycling conditions: 95°C for 3 min., 40 cycles of 10s at 95°C, 30s at 60°C, 15s 72°C (during which measurements are taken). Enrichment for a specific DNA sequence was calculated using the comparative Ct method, and normalized to *amylase.* PCR products were all smaller than 150 bp. Sequences of primers are shown in Supplementary Table S4.3.

For ChIP analysis with nuclei derived from human cell lines 5×10^7 cells were fixed in 1% formaldehyde, lysed and sonicated (4 bursts). The ChIP was performed by using Dynabeads-Protein G (Dynal Biotech) coupled to anti-CTCF () and anti-UBF (Santa Cruz Biotechnology F9). Dynabeads were incubated with lysates for 4 h at 4ºC and washed consecutively with Low Salt, High Salt and LiCl Immune Complex Wash Buffers (Upstate). Chromatin was eluted with 200 µl of elution buffer (Upstate), decrosslinked for 8h at 65ºC and purified through Qiaquick columns (Qiagen). Real-time PCR of immunoprecipitated DNA was performed with primers shown in Table S5. MYC-N amplicon was used as positive control (Ohlsson et al, 2001) and MYC-H.1 amplicon as negative control (Gombert et al, 2003) for CTCF binding sites.

Electrophoretic mobility shift analysis (EMSA)

The EMSA, or bandshift analysis, was performed using extracts from mouse and human cells. For the first protein extract was preincubated with bandshift buffer (10% Glycerol, 20 mM Hepes pH7.4, 20 mM KCl, 1 mM MgCl2, 5 mM DTT, 10 µM ZnCl2, 100 µg/ml BSA, 0.02% NP40) and 2-4 µg of salmon sperm DNA as a non-specific competitor. The reaction was incubated for 20 min. at the room temperature. Upon addition

of the ^{32}P end-labeled probe the binding reaction was performed for 20 min. Complexes were analyzed by electrophoresis through a 5% acrylamide (37,5:1) 0.5x TBE non-denaturing gel at 8V/cm2 at 4°C. When specified 300 fold excess of unlabeled probe or specific competitor was added at the same time as the probe.

For human cells nuclear extracts from 293T cells or K562 cells were prepared as described. EMSA was carried out using P^{32}-labelled PCR fragments (MYC-N, H42.1 rDNA and H37.9 rDNA, see primer sequences above) and 10 µg of protein from nuclear extracts. For supershift experiments, 1 µl of anti-CTCF mouse monoclonal (BD Biosciences) or anti-actin (Santa Cruz Biotechnology sc-1616, used as non-specific antibody) was added to the binding reaction prior to the radiolabeled probe. Sequences of primers using for the EMSA are shown in Supplementary Tables S4.1 and S4.5.

Northern blot analysis

Total RNA was isolated using RNA-Bee RNA isolation solvent (Tel-Test Inc.). Approximately 6 mg of total RNA was size separated by gel electrophoresis and blotted onto Hybond N+ membrane (Amersham). Probes were radioactively labeled by PCR. Blots were exposed to PhophorImager screens (Molecular Dynamics) to quantify results. Sequences of primers are shown in Supplementary Table S4.4.

Fluorescent in situ hybridization (FISH) and immunofluorescence analysis

Fluorescent in situ hybridization (FISH) experiments were carried out on MES cells as described previously (Akhmanova et al, 2000). The unstable 5'ETS probe has been described (Akhmanova et al, 2000), the enhancer oligonucleotide was ordered from Eurogentec and was prelabelled with biotin-UTP.

For immunofluorescent staining cells were fixed in 4% paraformaldehyde/PBS for 15 minutes at RT, permeabilised in 0,15% Triton X-100 in PBS, blocked in 1% bovine serum albumin in PBS and incubated with antibodies as described previously. We used FITC-conjugated goat antibodies against rabbit and mouse IgG were purchased from Molecular probes and Alexa 594-conjugated goat anti-rabbit antibody from Nordic Laboratories as secondary antibodies. Images of cells were collected with a Leica DMRBE microscope equipped with a Hamamatsu ORCA ER camera, or with a Zeiss LSM510 confocal, as described (Lansbergen et al, 2006).

Nuclear run-on

Cells were collected and washed twice with cold PBS. The cells were lysed in nuclear isolation buffer (10 mM Tris pH7.5; 10 mM NaCl, 10 mM MgCl2, 0.5% NP40). The nuclei were spun at 2000 rpm and resuspended in storage buffer (50 mM Tris pH8.5, 0.1 mM EDTA, 5 mM MgCl2, 40% glycerol) 1.10^6 nuclei (50 µl) were pre-incubated for 20 min. on ice with 100 ug/ml α-aminitin. Nuclei were than mixed with 50 ul 2x reaction buffer (300 mM KCl; 5 mM MgCl2; 10 mM Tris PH7.5; 5 mM DTT; 20U RNA guard 0.5 mM of each ATP, UTP, GTP and 100 µCi of alpha-32P CTP (800 Ci/mmol, 10 mCi/ml, Amersham). The labeling reaction was performed for 30 min. at 30°C. The reaction was stopped on ice by adding 1ml of RNA-Bee and total RNA was extracted as indicated above. Using a slot blot hybridization system with Hybond-N+ nylon membranes (Amersham Pharmacia Biotech), 5 µg of DNA PCR fragments was hybridized with ± 2x10^5 c.p.m. of labeled RNA. Hybridization and detection wash done as described in the ES cell targeting section. Incubation was performed in 2 ml of Churchhyb-mix (0,5 M Na2HPO4 pH 7.2, 7% SDS, 1 mM EDTA) in a rotating hybridizer at 65°C for 24h. Membranes were washed extensively at 65°C with Church wash-buffer (40 mM Na2HPO4 pH 7.2, 1% SDS). Hybridization signals were quantified with a Phoshor Imager (Typhoon Amersham) using Imagequant software. Signal was corrected for the amount of CTGs in the probe. Sequences of primers are shown in Supplementary Table S4.4.

Acknowledgements

We would like to thank Dr. I. Grummt for kindly providing us with FLAG-tagged UBF. We thank Dr. D.N. Meijer for the Rosa26-BirA construct and Dr J. Demmers and K. Bezstarosti for performing the mass

spectrometry analysis. This research was supported by the Dutch Cancer Society (KWF) and the European Science Foundation EUROCORES Programme EuroDYNA (ERAS-CT-2003-980409).

Supplementary data

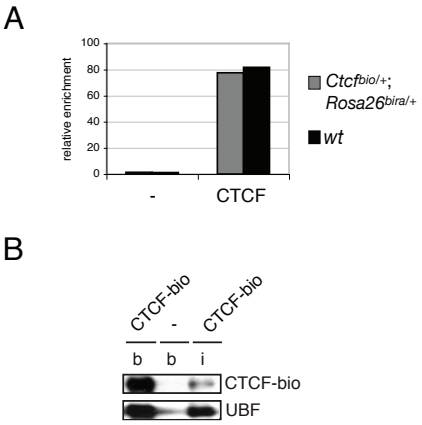

Figure S4.1 Characterization of CTCF and CTCF-bio in thymic extracts
A. rDNA spacer promoter ChIP analysis. Extracts of adult thymus from wild type and $Ctcf^{bio/+}$; $Rosa26^{bira/+}$ mice were analyzed for CTCF and CTCF-bio binding to the rDNA spcaer promoter using anti-CTCF antibodies or a control serum (-). **B.** Western blot analysis. Streptavidin pull downs on thymus nuclear extracts isolated from mice expressing biotinylated CTCF (CTCF-bio) or non-tagged CTCF (-).

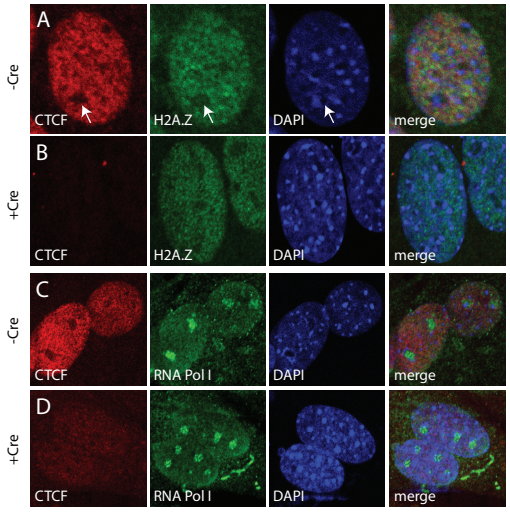

Figure S4.2 Immunofluorescence analysis in primary MEFs
Primary MEFs carrying the conditional *Ctcf*^{f} knock-out allele were either infected (+ Cre) or not infected (- Cre) with a lentivirus expressing Cre recombinase. Cells were fixed and incubated with antibodies against the indicated proteins. Cells were counterstained with DAPI (blue). The right hand panels show a merge.

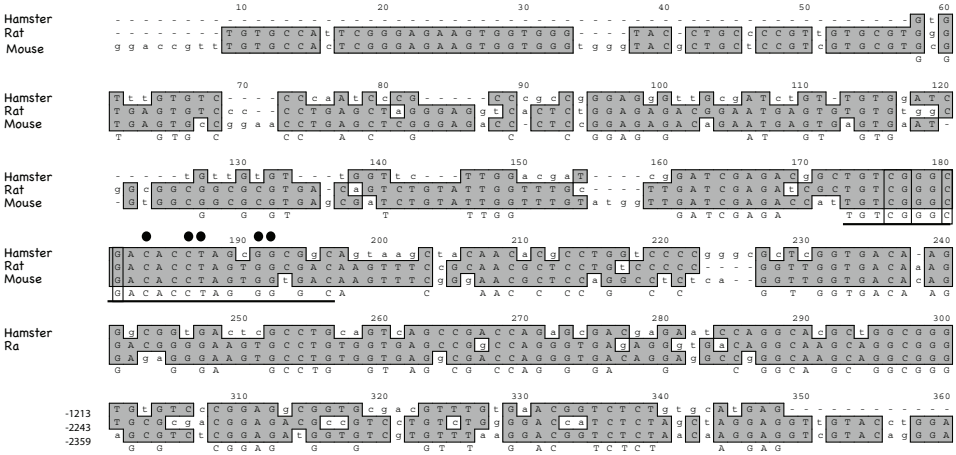

Figure S4.3 Comparison of mouse, rat and hamster rDNA repeat regions

A. Comparison of nucleotide sequences of the mouse, rat and hamster rDNA repeats. Only the regions around the spacer promoter are indicated. The CTCF consensus site (Kim et al, 2007) is underlined. CpG di-nucleotides are boxed, highly conserved CTCF consensus site residues are indicated by a dot (the asterisk indicates deviation between consensus site prediction and real residue).

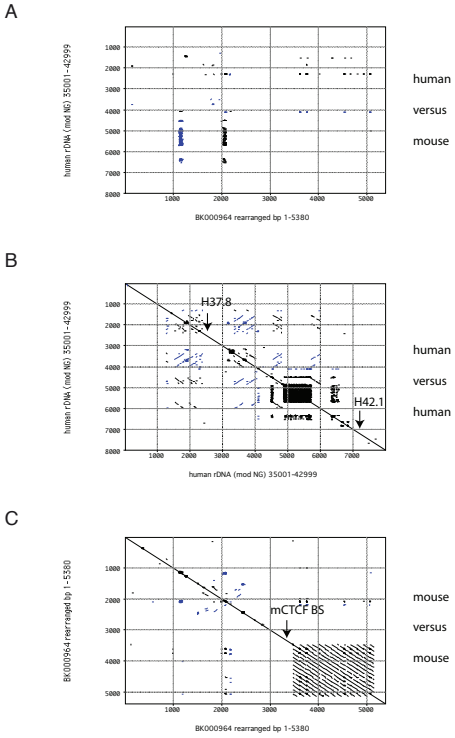

Figure S4.4 Comparison of mouse and human rDNA repeat regions

A. Matrix plot comparison of nucleotide sequences of mouse and human rDNA repeats in the region upstream of the gene promoter. No significant similarity is detected. **B.** Matrix plot comparison of the nucleotide sequence of the human rDNA repeat in the region upstream of the gene promoter. Notice the presence of a highly repetitive Alu element ~2.5 kb upstream of the gene promoter. Remarkably CTCF binding sites (H37.8 and H42.1) surround this repeat. It is unknown where the spacer promoter is localized in the human rDNA. **C.** Matrix plot comparison of the nucleotide sequence of the mouse rDNA repeat in the region upstream of the gene promoter. The CTCF binding site is indicated. It precedes the enhancer repeat region. One could imagine that the different local environment of human and mouse rDNA repeats has caused a difference in CTCF binding.

Table S4.1. Primers used for bandshifts

name	sequence (5' to 3')
R30s	TGTATGGTTGATCGAGACCATTGTCGGGCGACACCTAGTGGTGACAAGTTTCGGGAACGCTCCAGGC CTCT
R30as	AGAGGCCTGGAGCGTTCCCGAAACTTGTCACCACTAGGTGTCGCCCGACAATGGTCTCGATCAACCA TACA
R30mut-s	TGTATGGTTGATCGAGACCATTGTCGGGCAATACCTAGTAGTGACAAGTTTCGGGAACGCTCCAGGCC TCT
F1 _F[1)]	CTAGATGAAGAAATTGAGACCTCTACTGGATAGCTATGGTATTTACGTGTCTA
F1_B[1]	AGCTTAGACACGTAAATACCATAGCTATCCAGTAGAGGTCTCAATTTCTTCAT

1) From the chicken lysozyme gene (Burcin et al, 1997).

Table S4.2 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).

Table S4.3 Primers used for 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
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

Table S4.4 Primers used for northern blot and nuclear run-on

name	sequence (5' to 3')
Northern	
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

Table S4.5 Primers used for northern blot and nuclear run-on

	Primers 5'-3'	Human rDNA coordinate site	Amplicon size (bp)
MYC-N	ACAAGGAGGTGGCTGGAAAC TTCCCCTCCTGGCTTTTAGT		181
MYC-H.1	CAACGCAACACAGGATATGG TTCCCCTCCTGGCTTTTAGT		108
H42.1 rDNA	GCTTCTCGACTCACGGTTTC CCGAGAGCACGATCTCAAA	42012-42031 42117-42135	124
H37.9 rDNA	CCCTGGTCGATTAGTTGTGG GTGCTCCCTTCCTCTGTGAG	37818-37837 37997-38016	199

5

A system to study the role of CTCF at the imprinted Igf2/H19 locus

Work in progress

A system to study the role of CTCF at the imprinted Igf2/H19 locus

Summary

The conserved transcriptional regulator CTCF is involved in many essential processes in the cell including the transcriptional regulation of imprinted genes. Imprinted genes are expressed in a parental specific manner. The best-characterized locus is that containing the paternally expressed *Igf2* gene and maternally expressed *H19* gene. CTCF binds to the unmethylated maternal imprinting control region (ICR) of this locus and thereby functions as a blocking factor that prevent enhancers from activating the *Igf2* promoter on this allele.

We aimed to define the exact role of CTCF in regulating the imprinted expression of *Igf2* and *H19*. Therefore a cellular system was set up that is able to discriminate between the parental alleles and can be used to efficiently delete *Ctcf in vitro*. Using this approach it was shown that deletion of *Ctcf* results in a loss of both *Igf2* and *H19* expression and results in a change in expression level of several other imprinted genes. However a gain of DNA methylation on the ICR in the absence of CTCF could not be observed. These results are in contradiction to previous reports.

Based on recent reports that show CTCF binding to other sites in the locus we suggest that CTCF plays a yet undefined role in regulating expression of *Igf2* and *H19*. We hypothesize that CTCF might be involved in enhancing the expression of both *Igf2* and *H19* by transferring activity from the enhancers to the promoters or by maintaining the locus in an open chromatin state. Our system provides the basis to address this question in more detail and can moreover be used to analyze the role of CTCF in regulating the expression of other imprinted genes.

Introduction

The role of CTCF in regulating the imprinted *Igf2/H19* locus has been studied in detail within the last eight years. The imprinting control region (ICR) is the major element involved in regulating the paternal specific transcription of the *Igf2* gene and the maternal specific transcription of *H19* (Ferguson-Smith et al, 1993; Leighton et al, 1995a). The ICR is located 2-4kb upstream of *H19* and is methylated on the paternal allele. CTCF was shown to bind to the unmethylated maternal ICR and is crucial in controlling imprinted *Igf2* and *H19* expression (Bell et al, 1999; Hark et al, 2000; Kanduri et al, 2000).

The enhancer-blocking model was introduced in order to explain the imprinted expression of *Igf2* and *H19* (Webber et al, 1998). This model states that CTCF, while bound to the maternal ICR, prevents the enhancers downstream of *H19* from activating the *Igf2* promoter on this allele. On the paternal allele CTCF cannot bind due to DNA methylation allowing the enhancers to activate *Igf2*. Spreading of methylation over the paternal *H19* promoter keeps the gene inactive on this allele.

Several studies were performed to test this model. A mouse line was generated in which the CTCF binding sites of the ICR were mutated. This resulted in loss of CTCF binding to this site. When the mutated sites are maternally inherited a gain of methylation on the ICR could be detected, coinciding with loss of *Igf2* and *H19* imprinting. Loss of imprinting is reflected by biallelic expression of *Igf2* and repression of *H19* (Schoenherr et al, 2003). These results showed CTCF to

be a critical determinant in regulating imprinted expression of both *Igf2* and *H19*. Additional evidence for the essential role of CTCF in regulating the imprinted expression of *Igf2* and *H19* came from transgenic RNA interference (RNAi) experiments. Knocking down CTCF in oocytes confirmed a critical requirement for CTCF in maintaining the unmethylated maternal ICR and thereby controlling imprinting of *Igf2* and *H19* (Fedoriw et al, 2004). These experiments showed that CTCF functions not only as an enhancer-blocker but also maintains the unmethylated state of the maternal ICR throughout development.

In order to address the role of the ICR in regulating and maintaining the imprinting of *Igf2* and *H19* in more detail, additional studies were performed. Using an approach whereby the ICR was tagged by insertion of a GAL4-binding motif (UAS) together with expression of a GAL4-myc fusion protein, interactions involving the ICR were analyzed. ICR dependent interactions with other regions in the *Igf2/H19* locus were detected by performing a pull down with antibodies against the myc epitope (Murrell et al, 2004). Using this approach it was shown that on the maternal allele the unmethylated ICR interacts with the differentially methylated region (DMR1) that is present just upstream of the major *Igf2* promoters. It was proposed that this interaction places *H19* in an active domain with the enhancers and *Igf2* in an inactive domain. On the paternal allele the methylated ICR interacts with DMR2 located in the *Igf2* gene body. This places the *Igf2* promoters in an active chromatin domain that is located close to the enhancers. The formation of different chromatin loops on the parental alleles provides a model to explain how allele specific expression of *Igf2* and *H19* is regulated by the ICR (Murrell et al, 2004).

Subsequently an extensive 3C analysis on the whole locus was performed to address the role of CTCF in the formation of chromatin loops on the maternal *Igf2/H19* allele. This study confirmed the ICR-DMR1 interaction on the maternal allele. Moreover this interaction was shown to be lost in mice containing mutations in the CTCF binding sites in the ICR. These data together with a confirmation that CTCF binds to the unmethylated maternal DMR1 showed CTCF to be important for the formation of the maternal ICR-DMR1 chromatin loop. Moreover the enhancers needed for transcriptional activation of the *Igf2* promoter on the paternal allele were confirmed to be in close proximity to the promoter (Kurukuti et al, 2006).

Recently the CTCF-bound maternal ICR was shown not only to interact with regions within ~ 200 kb but also with regions on different chromosomes (Ling et al, 2006; Zhao et al, 2006). A co-localization was reported between the maternal ICR on chromosome 7 and an intergenic sequence on chromosome 11, located between the *Wsb1* and *Nf1* genes. Deletion of CTCF binding sites in the ICR and knockdown of CTCF using RNAi showed the association of these regions as well as the expression of *Wsb1/Nf1* genes to depend on the CTCF bound ICR (Ling et al, 2006). Moreover the maternal *H19* ICR interacts *in trans* with many imprinted regions on other chromosomes (Zhao et al, 2006). For example, the maternal *H19* ICR allele was shown to interact with the maternal DMR of the imprinted *Impact* gene. Three-dimensional DNA FISH confirmed that this interaction is lost in mice with a maternal inheritance of the ICR CTCF binding site mutant. In these mice a change in gene expression of the imprinted *Impact* gene and its neighboring non-imprinted gene *Osbpl1a* can be observed compared to wild-type mice (Zhao et al, 2006). This study shows that imprinted domains can be functionally connected *in trans* via DMRs and in this way are able to regulate transcription *in trans*. Both studies provide proof for a central role of CTCF binding sites in a network of chromatin interactions.

Here we describe an *in vitro* system that is suitable to study the role of CTCF in regulating *Igf2* and *H19* gene expression in more detail. By making use of *M.m. domesticus* (C57/Bl6) conditional CTCF knockout mice with distal chromosome 7 in a *M.m. castaneus* background we are able to distinguish the parental origin of both *Igf2* and *H19* genes. In this way we are able to determine

the role of CTCF in this locus in an allele specific manner. While setting up this system several papers were published describing the role of CTCF in regulating the chromatin conformation of the *Igf2/H19* locus and its role in regulating interchromosomal interactions associated with this locus (Kurukuti et al, 2006; Ling et al, 2006; Zhao et al, 2006). However these studies exclusively address the role of CTCF upon binding to the ICR. Our analysis suggests additional roles for CTCF beside the ones reported so far. CTCF is likely to be involved in fine-tuning *H19* and *Igf2* expression in different tissues and in regulating the expression of other imprinted genes.

Results

Selection of a cell system to study Igf2/H19 expression in the absence of CTCF

Igf2 and *H19* expression is reported to be at the highest level just after birth with almost undetectable levels in adult mice (Weber et al, 2001). Nearly all studies done on the *Igf2/H19* locus are performed on samples isolated from neonatal livers. In order to study the *Igf2/H19* locus in the absence of CTCF this would be the tissue of choice. Liver specific recombination was reported by using transgenic mice expressing the Cre recombinase under the Albumin promoter (Postic & Magnuson, 2000). However the recombination efficiency of this line is reported to be only 60% one week after birth and to be complete at six weeks after birth. Indicating that this mouse line is not suitable for deletion of *Ctcf* specifically in the neonatal liver where *Igf2* and *H19* are expressed.

In order to select an appropriate cell system for *in vitro* deletion of *Ctcf* and subsequent study of the *Igf2/H19* locus, expression levels were checked in different cell types. We confirm expression of *Igf2* mRNA to be high in embryonic and postnatal liver and low in adult liver. The non-coding RNA transcript of *H19* is less abundant than *Igf2* but present in embryonic and postnatal liver and not in adult liver (Figure 5.1A).

The majority of the cells present in the liver represent erythrocytes and hepatocytes. The abundance of these cell types depends on the developmental stage, in the embryo the liver consists mainly of erythrocytes and erythroid progenitors while at later stages hepatocytes represent a larger population of the cells present in the liver. In order to define in which cell type of the embryonic and postnatal liver *Igf2* and *H19* are expressed adherent cells isolated from neonatal livers were cultured and analyzed. This population of cells contains mainly hepatocytes but consists also of supporting Kupfer cells and epithelial cells. The $p53^{-/-}$ derived 1/11 cells were used as a model system to analyze *Igf2* and *H19* expression in erythroid progenitors and erythrocytes. 1/11 cells can be expanded indefinitely *in vitro* and upon exposure to stimuli such as erythropoietin, they undergo terminal differentiation into enucleated erythrocytes in a synchronized manner (Dolznig et al, 2001; von Lindern et al, 2001). Analysis of those cell types shows *Igf2* and *H19* to be expressed at low levels in the erythroid lineage and at more easily detectable levels in hepatocytes. In addition to hepatocytes, mouse embryonic fibroblasts (MEFs) also show clearly detectable levels of *Igf2* and *H19* expression, indicating that these cell types can be used to study the expression of both genes in the absence of CTCF (Figure 5.1B).

Expression in the thymus was also analyzed since this tissue can be used to delete *Ctcf in vivo* (Heath submitted). The postnatal thymus shows detectable levels of *Igf2* and *H19* expression however in the adult thymus expression is very low. A transgene expressing the Cre recombinase under the *Lck* promoter is widely used for LoxP recombination specifically in the thymus. However

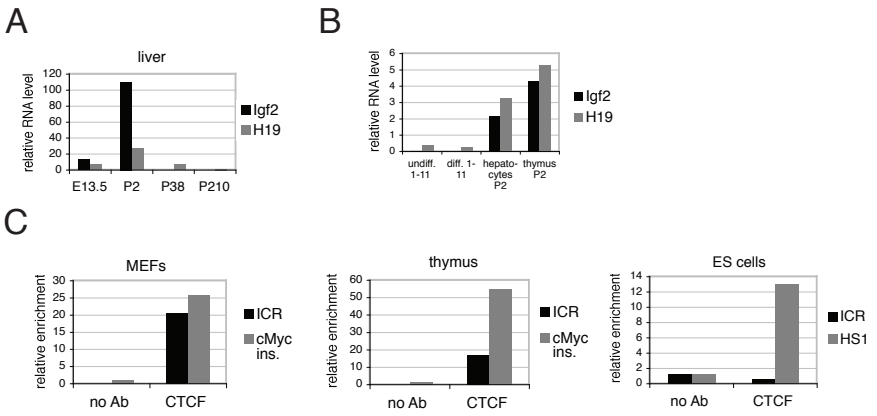

Figure 5.1 Selection of a cell system to study the role of CTCF in *Igf2/H19* regulation
A. Relative *Igf2* and *H19* mRNA levels as determined by quantitative RT-PCR on livers taken at different stages during developmental. Embryonic days (E) and days after birth (P) are indicated. **B.** Relative *Igf2* and *H19* mRNA levels as determined by quantitative RT-PCR on undifferentiated (undiff.) 1-11 cells, differentiated (diff.) 1-11 cells, adherent cells isolated from postnatal day two liver (hepatocytes) and whole thymus. **C**. Chromatin immunoprecipitation (ChIP) using an antibody against CTCF. Plotted are relative enrichments measured for CTCF binding sites in the *Igf2/H19* locus (ICR), the β-globin locus (HS1) and cMyc insulator (cMyc ins.) in (from left to right) primary MEFs, adult thymus and ES cells.

it should be noted that *Lck* promoter is active at different stages of thymocyte maturation in newborn mice compared to adult mice (Molina et al, 1998).

In addition to analyzing the expression levels of *Igf2* and *H19* in different cell types we also analyzed CTCF binding to the ICR in several cell types and tissues. We confirmed CTCF binding to the ICR in MEFs and thymus. Binding of CTCF to the ICR is undetectable in embryonic stem (ES) cells (Figure 5.1C).

The lenti-viral Cre system as a tool to delete Ctcf in vitro

In a first attempt to set up a system to delete *Ctcf in vitro*, *Ctcf* conditional knockout mice were crossed to tamoxifen inducible Cre mice. However *in vitro* addition of tamoxifen to MEFs isolated from those mice was not sufficient to delete the *Ctcf* gene with high efficiency (Sleutels unpublished). As an alternative to this system a replication deficient lentivirus expressing the Cre recombinase was generated. Retroviral based gene transfer vectors such as lentiviruses provide effective means for stable and high-level delivery, integration and expression of exogenous genes in primary dividing and non-dividing cells (Delenda, 2004; Miyoshi et al, 1998; Naldini et al, 1996).

MEFs and hepatocytes were isolated from homozygous (lox/lox) and heterozygous (lox/+) *Ctcf* conditional knockout mice and transduced with lentivirus expressing Cre or GFP. Fine tuning the conditions such as virus concentration and days of culturing led to the most efficient protocol to obtain a population of cells deleted for *Ctcf*. In the final protocol the cells were grown to almost full confluency and transduced with highly concentrated lentivirus expressing Cre, split in two the next day and left for another three days (Figure 5.2A).

The absence of CTCF protein upon Cre deletion was visualized by immunostaining with CTCF antibody on paraformaldehyde (PFA) fixed cells (Figure 5.2B). The efficiency of deletion can

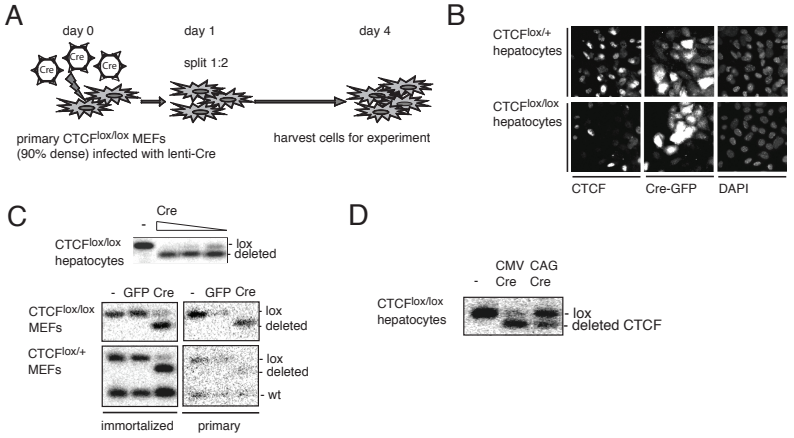

Figure 5.2 *In vitro* deletion of CTCF using the lentiviral Cre system
A. Schematic overview of the strategy used to delete CTCF *in vitro*. Confluent primary CTCF conditional knock-out MEFs are transduced with lentivirus expressing the Cre recombinase at day 0, the next day cells are split 1:2 and left in the dish for another 3 days. At day 4 cells show in general a 80-90% deletion of CTCF at the mRNA level and are used for experiments. **B.** Immunostaining with CTCF antibody on homozygous (lox/lox) and heterozygous (lox/+) CTCF conditional knockout hepatocytes transduced with lentivirus expressing Cre-GFP. **C.** Southern blot analysis showing deletion of CTCF conditional knock-out alleles in homozygous (lox/lox) and heterozygous (lox/+) MEFs and hepatocytes upon addition of Cre lentivirus (Cre) but not after addition of GFP expressing virus (GFP). D) Southern blot analysis showing deletion of CTCF conditional knockout alleles in homozygous (lox/lox) hepatocytes upon addition of similar amounts of lentivirus expressing Cre under the CMV or CAG promoter.

be determined quantitatively on the DNA level by southern blot analysis or by RT-PCR to analyze mRNA levels. A clear deletion of the *Ctcf* gene can be detected on southern blot in both MEFs and hepatocytes upon transduction with Cre but not with GFP expressing virus (Figure 5.2C).

If the Cre recombinase is expressed under a cytomegalovirus (CMV) promoter it appears to be more efficient in deleting the CTCF gene in hepatocytes than when it is expressed under the cytomegalovirus enhancer/chicken β-actin (CAG) promoter (Figure 5.2D). Moreover we found that CMV-Cre-ires-GFP lentivirus is less efficient in deleting the CTCF gene than CMV-Cre lentivirus. We have not measured the titer (m.o.i.) of the different viruses nor have we analyzed expression levels of the Cre recombinase so we do not know which of the two factors is causing the difference in deletion efficiency. We found that a high concentration of virus particles is needed to infect all cells. This becomes especially important when trying to transduce primary cells. Deleting CTCF from immortalized conditional knockout MEFs is more efficient than deleting the gene from primary MEFs upon addition of the same amount of lenti-viral Cre (data not shown).

Cells deleted for *Ctcf* can be maintained for multiple days without dividing or going into apoptosis (data not shown). A similar phenotype is reported for cells overexpressing CTCF (Rasko et al, 2001). For this reason the population of cells that we want to delete for *Ctcf* upon expression of Cre should be confluent in order not to be overgrown by non-deleted cells. Populations of confluent cells deleted for *Ctcf* and wild-type cells that were used for experiments are in a similar non-cycling state (data not shown). This indicates that the changes we observe between *Ctcf* deleted and wild-type cells are not likely to be caused by difference in cell cycle or cellular proliferation.

Expression of imprinted genes in the absence of CTCF

The gain of maternal *Igf2* expression and coinciding loss of maternal *H19* expression, as seen in ICR CTCF binding site mutant mice, is not consistently reported upon knockdown of CTCF using RNAi. The expression levels of *Igf2* and *H19* in the total absence of CTCF were not reported in studies using RNAi to knockdown CTCF in oocytes (Fedoriw et al, 2004). Knockdown of CTCF in bone marrow fibroblast cell lines results in biallelic *Igf2* expression and no change in *H19* expression compared to wild-type cells (Ling et al, 2006). Recently depletion of CTCF in synchronized (G1 or G2) HeLa cells does show an upregulation of *Igf2* expression and a loss of *H19* expression (Wendt et al, 2008).

We analyzed the expression of *Igf2* and *H19* in the absence of CTCF in both primary hepatocytes and MEFs. In heterozygous samples that show a reduction in amount of CTCF mRNA by half, *Igf2* and *H19* expression levels were not affected (Figure 5.3A). Also infection of cells with GFP expressing lenti-virus does not change the expression of *Igf2* and *H19*. In cells with a reduction in *Ctcf* mRNA levels to 5-10% compared to wild type both *Igf2* and *H19* expression levels were downregulated (Figure 5.3A). In hepatocytes this effect was more pronounced than in MEFs (Figure 5.3B). In hepatocytes *Igf2* and *H19* mRNAs are downregulated ~80% and ~90% and in MEFs ~40% and ~30% respectively (Figure 5.3A/B). This is in contradiction to many reports showing *Igf2* to be upregulated and *H19* to be downregulated upon deletion of CTCF binding in the ICR. In order to further clarify these results it would be interesting to examine the imprinting status of both genes in the absence of CTCF using an RNase protection assay.

None of the other imprinted genes analyzed showed any difference in expression levels in

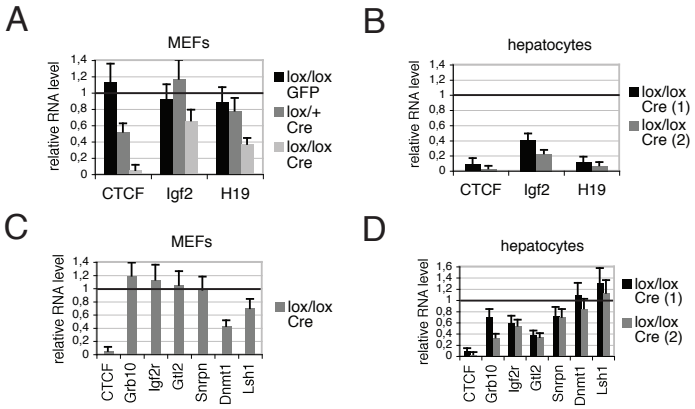

Figure 5.3 Expression of imprinted genes in the absence of CTCF

A. Relative mRNA levels as determined by quantitative RT-PCR on conditional CTCF knockout primary MEFs. Homozygous (lox/lox), heterozygous (lox/+) cells transduced with GFP or Cre expressing lenti-virus are shown. The expression levels are corrected for non-virus treated samples that are set at 1. **B.** Relative mRNA levels as determined by quantitative RT-PCR on homozygous (lox/lox) conditional CTCF knockout primary hepatocytes transduced with different amounts (1 and 2) of Cre expressing lenti-virus are shown. The expression levels are corrected for non-virus treated samples that are set at 1. C) Relative mRNA levels as determined by quantitative RT-PCR on homozygous (lox/lox) conditional CTCF knockout primary MEFs transduced with Cre expressing lenti-virus are shown. The expression levels are corrected for non-virus treated samples that are set at 1. D) Relative mRNA levels as determined by quantitative RT-PCR on homozygous (lox/lox) conditional CTCF knock-out hepatocytes transduced with different amounts (1 and 2) Cre expressing lenti-virus are shown. The expression levels are corrected for non-virus treated samples that are set at 1.

the absence of CTCF in primary MEFs, while Dnmt1 and Lsh1 both involved in DNA methylation do show a drop in expression (Howell et al, 2001; Myant & Stancheva, 2008) (Figure 5.3C). In contrast, in hepatocytes imprinted genes are downregulated in the absence of CTCF, while Dnmt1 and Lsh1 expression remained unchanged. The *Igf2* receptor, *Grb10* and *Gtl2* show a reduction in their expression level by half. *Snrpn* shows a less dramatic downregulation in expression (Figure 5.3D). Interestingly CTCF was reported to bind the *Grb10* promoter in a methylation sensitive (and allele specific) manner (Hikichi et al, 2003). CTCF can bind to a maternally unmethylated DMR close to the promoter of the non-coding *Gtl2* gene (Rosa et al, 2005). This CTCF binding site, as well as its allele specific methylation pattern was shown to be conserved in human, mouse and sheep (Paulsen et al, 2001; Takada et al, 2002).

Igf2/H19 parental allele discrimination

In order to analyze the *Igf2/H19* locus in the absence of CTCF in an allele specific manner *M.m. domesticus* C57/Bl6 conditional knockout mice were crossed to the *M.m. castaneus* mice strain. Several single nucleotide polymorfisms (SNPs) in the *Igf2/H19* locus were mapped or identified by sequencing parts of the locus in the two different strains. The SNPs can result in restriction fragment length polymorfisms (RFLPs) or simple sequence length polymorfisms (SSLPs) that can be used to discriminate between the two strains by making use of PCR (Figure 5.4A/B). These SSLPs and RFLPs were used to genotype the C57/Bl6 conditional CTCF knock-out mouse line with distal chromosome 7 containing the imprinted *Igf2/H19* and *Kvlqt1* locus in a *M.m. castaneus* background (Figure 5.4A).

Upon crossing this line to full C57/Bl6 mice the parental alleles of those mice can discriminate between maternal and paternal alleles. We were able to confirm the imprinting status of both *Igf2* and *H19* in postnatal liver as well as in MEFs by RNase protection (Figure 5.4C/D). The C/B postnatal liver sample showed very low expression of *H19* from the paternal allele.

DNA methylation of the Igf2/H19 ICR remains unchanged in the absence of CTCF

The ability to discriminate between the parental alleles can moreover be used to analyze the methylation status of the ICR in an allele specific manner. By making use of an MboI and SacI RFLP present in the ICR of *M.m. castaneus* but not in *M.m. domesticus* (C57/Bl6) we can discriminate the parental origin of the first CTCF binding site in the ICR (Schoenherr et al, 2003; Szabo et al, 2004). Complete digestion of SacI and MboI is shown by southern blot on single digests with these enzymes. Combined digestion of SacI with MboI or of ClaI with HhaI digestion respectively can identify methylated or unmethylated alleles, since both enzymes are sensitive to methylation and cut only unmethylated DNA (Figure 5.5A). Using these enzymes to digest genomic DNA isolated from mice containing the *Igf2* and *H19* genes in a mixed (*M.m. castaneus* / *M.m. domesticus*) background we are able to confirm the paternal ICR specific methylation. Mice with a mixed background containing either a *M.m. domesticus* (C/B) or a *M.m. castaneus* (B/C) paternal allele show this allele to be methylated at the ICR and the maternal allele to be unmethylated (Figure 5.5B). As an alternative bisulphite sequencing can be used to perform more detailed analysis of DNA methylation patterns.

Methylation analysis without allele discrimination shows a 60% unmethylated ICR in immortalized MEFs and a remarkably methylated ICR in ES cells (Figure 5.5C). The full methylation of the ICR in ES cells fits with the observation that CTCF binding is absent at this site and shows this cell type to be unsuitable to study *Igf2/H19* imprinted regulation (Figure 5.1C). Differentiation

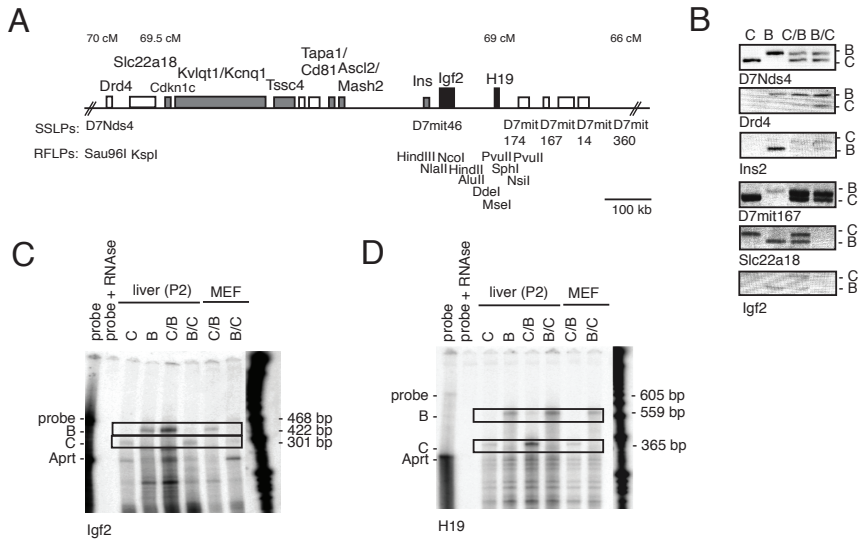

Figure 5.4 *Igf2* and *H19* parental allele discrimination
A. Schematic representation of distal chromosome 7 with indicated the different SSLPs and RFLPs between *M.m. domesticus* and *M.m. castaneus*. Filled boxes represent the imprinted genes present on this part of the chromosome and open boxes represent non-imprinted genes. **B.** Strain background of mouse distal chromosome 7 analyzed by gel electrophoresis after PCR for SSLPs or PCR followed by digestion for RFLPs. The different genotypes are represented on top with (C) representing pure *M.m. castaneus* , (B) C57/Bl6 *M.m. domesticus,* (C/B) maternal *M.m. castaneus* allele and paternal C57/Bl6 *M.m. domesticus* allele, (B/C) maternal C57/Bl6 *M.m.* domesticus allele and paternal *M.m. castaneus* allele. **C.** Parental allele specific analysis of *Igf2* RNA transcripts in MEFs and P2 liver as determined by RNase protection. The different genotypes are represented on top with (C) representing pure *M.m. castaneus* , (B) pure C57/Bl6 *M.m. domesticus,* (C/B) maternal *M.m. castaneus* allele and paternal C57/Bl6 *M.m. domesticus* allele, (B/C) maternal C57/Bl6 *M.m.* domesticus allele and paternal *M.m. castaneus* allele. Aprt was used as a loading control. **D**. Parental allele specific analysis of *H19* RNA transcripts in MEFs and P2 liver as determined by RNase protection. The different genotypes are represented on top with (C) representing pure *M.m. castaneus* , (B) pure C57/Bl6 *M.m. domesticus,* (C/B) maternal *M.m. castaneus* allele and paternal C57/Bl6 *M.m. domesticus* allele, (B/C) maternal C57/Bl6 *M.m.* domesticus allele and paternal *M.m. castaneus* allele. *Aprt* was used as a loading control.

of ES cells with retinoic acid (RA) shows a very modest loss of methylation in the ICR (Figure 5.5C). This is likely to result in a gain of CTCF binding to the ICR.

Surprisingly we were not able to detect a gain of methylation in the ICR upon deletion of CTCF in primary MEFs (Figure 5.5D). Methylation analysis of the ICR upon *in vivo* deletion of CTCF in T-lymphocytes confirmed this result (Sleutels and Heath unpublished). No difference in methylation of the ICR could be detected between wild type and CTCF deleted samples. This is in contradiction to results reported by different groups that show a gain of methylation at the maternal ICR in somatic cells upon mutation of its binding sites or in oocytes in the absence of CTCF (Schoenherr et al, 2003; Szabo et al, 2004).

In order to analyse the role of CTCF in the spatial organization of the *Igf2/H19* locus 3C analysis was set up. After cross-linking DNA fragments that are close together in nuclear space, digestion with an excess of restriction enzyme is performed in order to reduce the size of the fragments for efficient ligation. Preliminary experiments to optimize the 3C procedure to analyze the chromatin conformation of the *Igf2/H19* locus in MEFs resulted in low EcoRI digestion

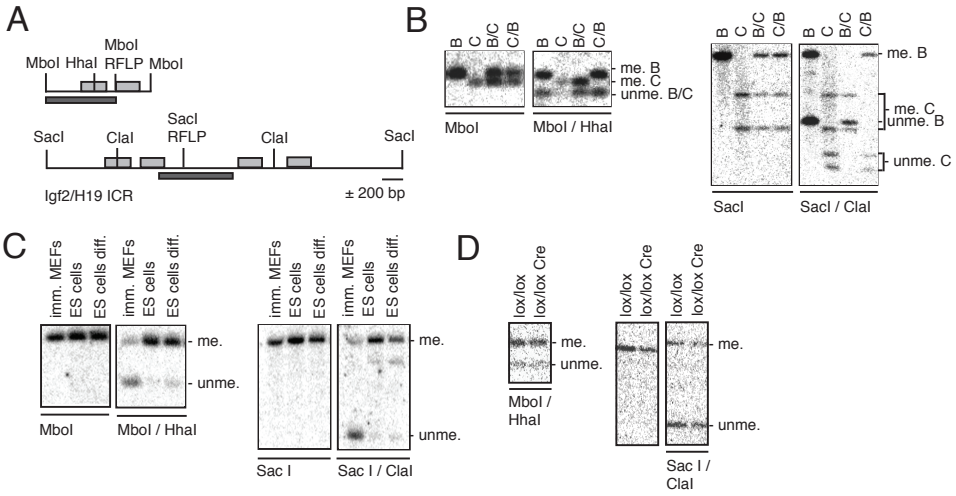

Figure 5.5 DNA methylation of the *Igf2/H19* locus
A. Schematic representation of part of the *Igf2/H19* ICR showing the position of restriction enzyme digestion sites used for DNA methylation analysis. Gray boxes represent CTCF binding sites in the ICR, dark boxes represent probes used to detect specific fragments by southern blot. The indicated RFLPs are present in *M.m. castaneus* and not in C57/Bl6 *M.m. domesticus.* **B.** Parental allele specific DNA methylation analysis of the *Igf2/H19* ICR. Southern blot analysis on genomic DNA isolated from MEFs. ClaI and HhaI are unable to cut methylated DNA. The different genotypes are represented on top with (C) representing pure *M.m. castaneus* , (B) pure C57/Bl6 *M.m. domesticus,* (C/B) maternal *M.m. castaneus* allele and paternal C57/Bl6 *M.m. domesticus* allele, (B/C) maternal C57/Bl6 *M.m.* domesticus allele and paternal *M.m. castaneus* allele. **C.** Methylation analysis of the Igf2/H19 ICR in different cell types. Southern blot analysis on genomic DNA isolated from immortalized (imm.) MEFs, ES cells and ES cells differentiated for 7 days with retinoic acid. **D.** Methylation analysis of the *Igf2/H19* ICR in conditional CTCF knockout MEFs (lox/lox) untreated or treated with Cre expressing lentivirus. Southern blot analysis on genomic DNA isolated from primary MEFs.

efficiencies (data not shown). When using the standard protocol for 3C analysis a clumping of nuclei was observed before starting the digestion step. This resulted in a digestion efficiency of about 50% as measured by PCR. At least 60-70% but preferntially 80% or more of the DNA should be digested before continuing with the ligation step (Simonis et al, 2007). In order to get a more homogenous suspension and thereby achieve higher digestion efficiency the volume of restriction buffer during digestion was increased six times compared to the standard protocol. This resulted in an improvement of digestion efficiency. The 3C analysis to analyse the role of CTCF in the spatial organization of the *Igf2/H19* locus is in progress.

Discussion

CTCF is involved in maintenance of allele specific methylation at and around its binding site on the ICR (Engel et al, 2006; Fedoriw et al, 2004; Schoenherr et al, 2003; Szabo et al, 2004). Elimination of all four CTCF binding sites in the ICR results in an acquirement of methylation on the maternal mutant ICR. It was suggested that CTCF binding prevents *de novo* methylation on the maternal allele (Rand et al, 2004). Surprisingly we do not observe a difference in DNA methylation in the absence of CTCF. This may be explained by the inability of the majority of cells

lacking CTCF to undergo replication, as shown by the low number of BrdU incorporation in these cells (Sleutels unpublished). Another explanation for the observed contradictory result could be that passage through the germline is needed in order to detect a gain of methylation on the maternal ICR in the absence of CTCF. In mice containing CTCF mutated ICR binding sites a gain of methylation at this site was not observed in the oocyte but only at later stages of development, during postimplantation (Schoenherr et al, 2003). This might be explained by a modulation in CTCF binding by cofactors at different stages of development.

We show CTCF binding to the ICR and parental allele specific expression of *Igf2* and *H19* in both hepatocytes ad primary MEFs. This indicates that these cell types are suitable to study imprinting of the *Igf2/H19* or *Kvlqt1* locus in more detail. It should however be noted that both cell types do not represent a homogenous population of cells making the study of chromatin conformations more difficult. For example the *Igf2/H19* locus was reported to contain multiple cell type- and embryonic lineage specific regulatory elements. MEFs represent a population of cells derived from endoderm as well as mesoderm.

The reduction in expression of imprinted genes in the absence of CTCF in hepatocytes but not in MEFs is quite striking. The mRNA levels of factors involved in DNA methylation show an opposite expression pattern and are reduced in MEFs and not in hepatocytes. Lsh1 has been implicated in regulation of specific imprinted genes, such as the *Cdkn1c/p57Kip2* gene. In the absence of Lsh1 the silenced paternal *Cdkn1c/p57Kip2* is reactivated, while expression of *Igf2*, *H19* and *Igf2r* remains unchanged (Fan et al, 2005). However we still have to test expression of *Cdkn1c/p57Kip2* in the absence of CTCF.

The decrease in expression of not only *H19* but also of *Igf2* in the absence of CTCF was unexpected since it does not match with existing data that show upregulation of *Igf2* in the absence of CTCF binding to the ICR. It would be interesting to check the parental allele specific expression of Igf2 in the absence of CTCF. *Igf2* could be biallelicly expressed but at very low levels. Multiple tissue- and allele specific elements were reported to be present in the locus suggesting additional roles besides the ICR in regulation of *Igf2* and *H19* expression (Constancia et al, 2000; Davies et al, 2002; Drewell et al, 2000; Lopes et al, 2003; Murrell et al, 2001; Schoenfelder et al, 2007). These elements could be directly or indirectly regulated by CTCF. Additional unknown stimulatory roles for CTCF in regulating the expression of both genes can explain the loss of both *Igf2* and *H19* expression in the absence of CTCF.

Interestingly recent genome wide mapping of CTCF binding sites in human mesoderm derived $CD4^+$ T-cells confirms binding of CTCF to the ICR and the mesoderm specific silencer, DMR1 (Barski et al, 2007) (Constancia et al, 2000). Moreover this analysis shows binding to DMR0 just upstream of the placental specific *Igf2* P0 promoter and to a region in between the *Igf2* and *H19* genes (Barski et al, 2007)(Figure 5.6A). This intergenic region (known as: A6A4 or CCD) is highly sensitive to nuclease on both parental alleles and conserved between mouse and human (Koide et al, 1994). Furthermore the region is unmethylated on both parental alleles suggesting CTCF could bind this region on the maternal as well as paternal allele. The CCD was reported to function as an enhancer for *Igf2* both in specific tissues where the gene is imprinted (i.e., somites, tongue, eye) and where *Igf2* is biallelically expressed such as the CNS (Charalambous et al, 2004; Jones et al, 2001; Ward et al, 1997). The same region was also reported to function as a skeletal muscle specific repressor of *Igf2* expression (Ainscough et al, 2000). Activity of this region in mouse is not reported in blood from which the $CD4^+$ T-cells used for the CTCF binding site analysis in human were isolated. Leaving aside the question what the function of CTCF binding to this area could be, it may function as a regulator of the HUC mesodermal enhancers that are present in between the CCD and ICR (Drewell et al, 2002). CTCF binding to the DMR1 is not restricted

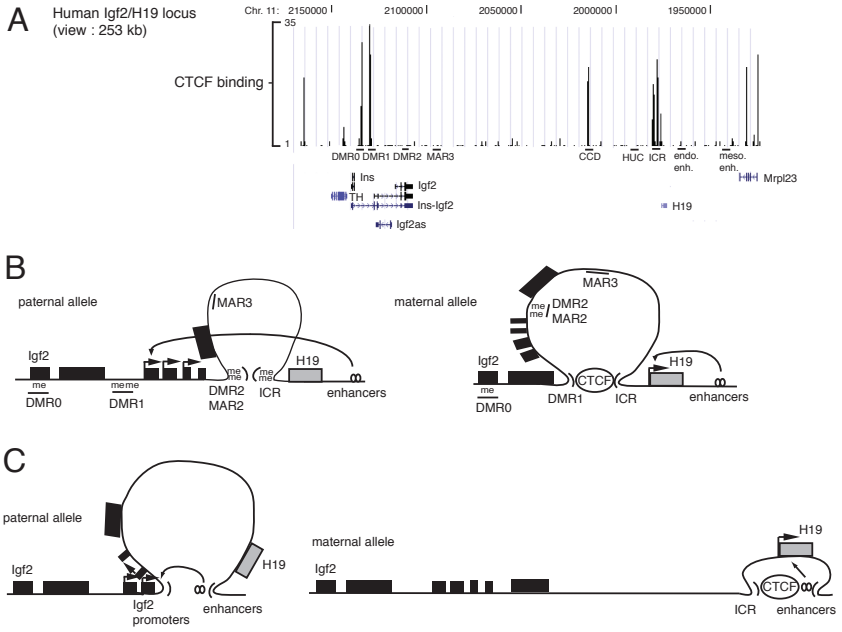

Figure 5.6 CTCF binding in the *Igf2/H19* locus and models that explain its role in regulating this locus
A. Schematic representation of the *Igf2/H19* locus and CTCF binding sites present in this locus. Vertical lines indicate potential CTCF binding sites. The height of these lines is indicative of the confidence CTCF will bind at that position. CTCF binding sites in the *Igf2/H19* locus as analyzed using human $CD4^+$ T cells (Barski et al, 2007). UCSC gene predictions based on refseq., uniprot, genbank, and comparitive genomics. Relative position of regulatory elements is indicated. Endo. enh.: endodermal enhancers, Meso. enh.: mesodermal enhancers, DMR: differentially methylated region, ICR: imprinting control region, HUC: mesodermal enhancers, CCD: conserved HSS. **B.** Model of allele specific chromatin loops in the *Igf2/H19* locus as analyzed by 3C (Murrell et al, 2004). It is proposed that the conformation of the paternal allele allows contacts between *Igf2* promoters and the enhancers that are blocked on the maternal allele. The ICR on the maternal allele makes contact with the DMR1 located upstream of the *Igf2* promoters. It was suggested that the ICR-DMR1 interaction places the *Igf2* promoters in a repressive domain. On the paternal allele the ICR contacts another such site, DMR2, downstream of *Igf2*. This should make the *Igf2* promoters accessible to the enhancers located downstream of *H19*. It should be noted that CTCF was reported to bind the maternal DMR1 (Kurukuti et al, 2006). ICR: imprinting control region, DMR: differentially methylated region, MAR: matrix attachment region, Me: methylated site, small white circles: endodermal enhancers, light grey boxes: *H19* exons, black boxes: *Igf2* exons. Maternal and paternal alleles are indicated. C) Model proposed by Yoon et al MCB 2007. The authors report an interaction between the enhancers and ICR on the maternal allele bringing the enhancers into close contact to the *H19* promoter. On the paternal allele an interaction between the enhancers and the *Igf2* promoter is reported bringing the enhancers in close contact to the *Igf2* promoters and not the *H19* promoter. The authors report predominant binding of CTCF to the maternal enhancers but also binding to the paternal enhancers.

to mesoderm derived tissues were this region function as a silencer also binding is reported in neonatal liver (Kurukuti et al, 2006).

Mapping of CTCF binding sites in human $CD4^+$ T-cells shows not only binding of CTCF within the *Igf2/H19* locus but also in regions directly flanking it. This separates the locus from the nearby (~200 kb distance) *Kvlqt* imprinted locus. CTCF is likely to form boundaries upon binding to these locus-flanking sites and thereby separate these imprinted domains into individually regulated clusters. Alternatively upon binding to these locus-flanking sites CTCF functions in maintaining the whole locus in an active and open chromatin state (Barski et al, 2007).

A similar downregulation in both *Igf2* and *H19* expression as we observe in the absence of CTCF is reported upon deletion of the endodermal enhancers. Deletion of these enhancers results in a ~90% reduction of *H19* and *Igf2* transcripts in liver, which is primarily composed of endodermal cells. In gut, kidney and lung where both endodermal and mesodermal cells are present a decline of ~70-75% in detected transcripts is observed (Leighton et al, 1995b). These results are similar to the *H19* and *Igf2* mRNA levels we measured in endodermal derived hepatocytes and combined endo- and mesodermal derived MEFs. Interestingly it was recently shown by ChIP that CTCF is able to bind the endodermal- and mesodermal enhancers and *Igf2* promoter in fetal and postnatal mouse tissues (Yoon et al, 2007). However binding of CTCF to enhancers was not reported in mesoderm derived adult human $CD4^+$ T cells (Barski et al, 2007). This indicates that CTCF binding to different sites in the *Igf2/H19* locus depends on the tissue and developmental stage. Binding of CTCF to the enhancers was reported to be biallelic although binding to the maternal enhancers was more pronounced (Yoon et al, 2007).

It would be very interesting to verify the binding of CTCF to the CCD, DMR1, locus-flanking sites and enhancers in different postnatal and adult tissues in an allele specific manner. This then allows to determine whether CTCF has additional functions in regulating *Igf2* and *H19* expression than reported so far.

The data as presented by Murrell et al. and Kurukuti et al. that show allele specific chromatin loops in the *Igf2/H19* locus can be questioned at some points. The paternal ICR-DMR2 interaction is apparently not as solid as the maternal ICR-DMR1 interaction since it was described only by Murrell *et al.* (Figure 5.6B). Kurukuti *et al.* on the other hand describe another interaction on the paternal allele, between the enhancers and *Igf2* promoters. However an interaction between the enhancers and *H19* promoter is detected on both parental alleles. The resolution of the restriction fragments that were used in this study is not high enough to analyze interactions between closely located fragments such as the enhancer-*H19* promoter interactions. Furthermore the quantitative manner is which the analysis was done can be criticized. Recent data from another lab show indeed some disagreement with the data of Murrell et al. and Kurukuti et al. (Yoon et al, 2007).

Yoon et al. generated a mouse line whereby the ICR was inserted upstream of the non-imprinted *Afp* gene, separating the enhancers from the promoter. In this way the functioning of the ICR was analyzed in a context independent manner. It was shown that upon insertion the ICR keeps its characteristics such as paternal methylation and absence of methylation and CTCF binding on the maternal allele (Park et al, 2004; Yoon et al, 2007). Maternal inheritance of the ICR insertion prevents association of the enhancers with the *Afp* promoter, resulting in a loss of *Afp* expression from the maternal allele. This result together with experiments comparing chromosomes with and without the ICR show that the ability of the ICR to regulate gene expression and to organize chromosome conformation is entirely context independent. Locus specific interactions such as the ICR-DMR1 interaction were suggested to be dispensable for enhancer-blocking activity (Yoon et al, 2007)(Figure 5.6C). Instead it was shown that the maternal ICR associates with the enhancers. Implicating a model whereby active insulators associate with active promoters and enhancers. This is in contradiction to the results of Kurukuti et al. that showed the enhancer-ICR*/H19* promoter interaction to be biallelic. It would be very interesting to confirm an involvement of CTCF in associating with promoters and enhancers by making use of the conditional CTCF knockout cells with a *M.m. castaneus* allele. If CTCF besides functioning as an insulator on the ICR has an additional role in binding to the enhancers on both alleles and keeping them active or transferring their activity than this could explain the reduction in *Igf2* and *H19* expression observed in the absence of CTCF.

The distal end of mouse chromosome 7 contains not only the *Igf2/H19* imprinted locus but

also a cluster of imprinted genes regulated by the KvDMR1 present in the *Kvlqt1* gene. The distal chromosome 7 imprinted region is well conserved during evolution and deregulation of the genes present in this region results in the overgrowth- and tumor-associated Beckwith–Wiedemann syndrome. The opposite functions of the two domains on embryo growth could have facilitated the acquisition and/or conservation of imprinting in this cluster. However the two clusters were shown to be imprinted by separate mechanisms demonstrating that each of the two domains of the cluster contains the *cis*-acting elements required for the regulation of its own genes. Deletion of the *H19* ICR was shown not to affect imprinting of *Kvlqt1*, *Mash2* or *Cdkn1c/p57(Kip2)* (Caspary et al, 1998).

All the genes located in the *Kvlqt* locus are required for normal development and are silenced by a KvDMR1 dependent mechanism on the paternal allele (Oh et al, 2008). Interestingly CTCF binding sites were identified in the unmethylated paternal KvDMR1 and a potential role of CTCF was suggested to be silencing of *Cdkn1c* (Fitzpatrick et al, 2007; Shin et al, 2008).

It will be interesting to analyze the role of CTCF in regulating the imprinting within the *Igf2/H19* and *Kvlqt* loci in different cell types in more detail by making use of the lenti-viral Cre system and conditional CTCF knockout mice with distal chromosome 7 in a *M.m. castaneus* background

Materials and methods

Cell culturing and immunofluorescent staining

Culturing and differentiation of 1/11 cells was done as described previously (Kooren et al, 2007). Immortalized and primary MEFs were cultured in Dulbecco's modified Eagle's medium (DMEM) and ham's F10 supplemented with 10% fetal calf serum and 1 mM β-mercaptoethanol. ES cells were cultured on 0,1% gelatin coated dishes in DMEM medium containing 10% FCS, 50% BRL, 10^3 U/ml LIF, 1:200 NEAA, 1 mM β-mercaptoethanol and cells were split every two days. In order to generate a primary hepatocyte culture, livers were isolated from postnatal day 2 (P2) embryos, resuspended by repeated pippeting and digestion with shaking for 30 min at 37°C in liver digest medium (GibcoBRL). Cells were cultured in William's E medium with glutamax containing 10% FCS, 1% insulin-transferrin-selenium (ITS) (GibcoBRL) and 10 nM dexamethason (Sigma) and grown on dishes coated with 0,1% gelatin. Cells were washed several times to get rid of debris and floating hematopoietic cells. Cells were used for experiments about 5 to 7 days after isolation. E13.5 embryos were used to isolate MEFs. Head and blood organs were removed and a piece of tail was used to isolate genomic DNA for genotyping. The remaining embryo was finely minced using scissors and incubated for 15min. 37°C in Trypsin/EDTA. Culture medium was added and dissociated tissue was transferred to a cell culture dish and cultured at 37°C with 5% CO_2, 3% O_2. Next day cells were washed and when confluent cells were frozen. To generate immortalized MEFs, cells were cultured for serial passages until spontaneous immortalization accured. Primary MEFs were used below passage five. For immunofluoresent staining cells were fixed in 4% paraformaldehyde/PBS for 15 minutes at RT, permeabilised in 0,15% Triton X-100 in PBS, blocked in 1% bovine serum albumin in PBS and incubated with antibodies as described previously. Antibodies used: anti-CTCF (Heath submitted), Alexa 594-conjugated goat anti-rabbit (Nordic Lab.). Images of cells were collected with a Leica DMRBE microscope.

RNA analysis

Total RNA was isolated using Trizol (Invitrogen) reagent according to the manufacturers instructions. RNase protection analysis (RPA) was performed with the RPAIII kit (Ambion). The probes for RPA were: Igf2RPA2-12 (T7): a 422 bp fragment protecting exon 6 of Igf2 and covering a G to A SNP between *M.m. domesticus* and *M.m.castaneus.* The probe was generated by PCR on BACs RP23-50N22 (244 kb) or RP23-209O22 (194 kb) using the following oligo's:

Igf2RPA2_F: CACACTAAGATCTCTCTGCTCCAC and Igf2RPA2_B: GGTTCTGGGATCCAAGTCTTAGC
H19RPA5-2 (T7) 554 bp probe in part of exon 5 covering G to T and G to A SNP between *M.m. domesticus* and *M.m.castaneus*, SpeI linearized template The probe was generated by PCR on BACs RP23-50N22 (244 kb) or RP23-209O22 (194 kb) using the following oligo's: H19RPA2_F: AAGATGGGAGAGCTGGAGGAG and H19RPA2_B: CAGCCGCTTCTTCTCACCAG. APRT (T7) probe was described previously (Sleutels & Barlow, 2001).

Southern blotting and DNA methylation analysis

Genomic DNA was isolated by the sodium dodecyl sulfate-proteinase K procedure (Sleutels & Barlow, 2001). DNA was digested, and blotted onto Hybond N+ membranes (Amersham) and hybridized with radioactive probes. Hybridization was performed in a rotating hybridizer at 65°C for 24h in ChurchHyb-mix (0,5 M Na2HPO4 pH 7.2, 7% SDS, 1 mM EDTA). Membranes were washed extensively at 65°C with Church wash-buffer (40 mM Na2HPO4 pH 7.2, 1% SDS). Hybridization signals were quantified with a Phoshor Imager (Typhoon Amersham). The following primers were used to generate probes for H19-ICR methylation analysis:
MboI-meF: ATGCAAATGAACCACTAGGAGTTTAG
MboI-meB: TCCACGAGGTACCAGCCTAGA
SacI-meF: CTCAAGAGCTATCTCAGGTATCTGAC
SacI-meB: TTAAGATGACAGTCACCAGCGC

Chromatin immunoprecipitation (ChIP) and analysis of gene expression

cDNA synthesis was performed using Superscript II RNase H-reverse transcriptase (Invitrogen) according to the manufacturers instructions using oligo (dT)12-18 primers. ChIP was performed as described in the Upstate protocol (http://www.upstate.com), except that cells were cross-linked using 2% formaldehyde for 5 minutes at room temperature. Real-time quantification of transcripts or precipitated DNA sequences was performed on Opticon II PCR machines (MJ Research) using Platinum Taq DNA polymerase (Invitrogen) and SYBR Green (Sigma-Aldrich) using the following cycling conditions: 2 min 94°C, 45 cycles of 30 sec 94°C, 1min 55°C, 15sec 72°C and 15sec 75°C (during which measurements were taken). Enrichment was calculated relative to a control gene and all values were normalized to input measurements. The following antibodies were used: anti-CTCF (N3) rabbit polyclonal (Heath submitted). Data was normalized to input and enrichment was measured over control gene (Amylase). The following primers were used for ChIP :
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
Amylase 99 CTCCTTGTACGGGTTGGT
Amylase 100 AATGATGTGCACAGCTGAA
Expression levels were normalized to levels measured for a control gene (Hprt). The following primers were used for mRNA expression analysis:
CTCF FRT-PCR: GTGCTGGCCAGATGGCGTAGAG
CTCF BRT-PCR: TTGTCATCGAGATCCGGCTCAGC
HPRT-B: GTTAAAGTTGAGAGATCATCTCCACC
HPRT_F: AGTGATAGATCCATTCCTATGACTGTAG
H19 BRT-PCR: TCTTCTTGATTCAGAACGAGACGG
H19 FRT-PCR: ACTTCTCTGCTGCTCTCTGGATC
IGF2 FRT-PCR: GTTCGGACCGCGGCTTCTACTTC
IGF2 BRT-PCR: GCAGCACTCTTCCACGATGCCAC
Snrpn FRT-PCR: CTGCTACTGCTAGCATTGCAG
Snrpn BRT-PCR: CCAGGTGGAGGAGCCATAATG
Gtl2 FRT-PCR: ACGCACAACACGTTGCAAC
Gtl2 BRT-PCR: CAGGTGTCTGTGTCCGTGTGTCC

Dnmt1 FRT-PCR: CAGATAGCTACCGGTTCTTCGG
Dnmt1 BRT-PCR: AGAGCTTAATCTCCAGGCCAATG
Igf2r FRT-PCR: GCTGGCCTTACTGCTGCATAAG
Igf2r BRT-PCR: TTGAGTACTTGTAGGACACGCCC
Grb10 FRT-PCR: TCGTACTGACACTGTGCCATC
Grb10 BRT-PCR: CATCCAGAGTGAAGAAGGTCTG
Lsh1 FRT-PCR: GGCCGAACAAACGGAGCCTGCGGTG
Lsh1 BRT-PCR: CAATTGCTGTTGTTCCATTTTAGTC

Lentivirus production and infection

Cre-lentivirus production and transduction of cells was done as described (Splinter et al, 2006). With the exception that confluent primary CTCF conditional knockout MEFs were transduced and that they were splitted 1:2 24h after transduction. Virus activity/functionality was tested by serial dilutions on CTCF conditional knockout (lox/lox) primary mouse embryonic fibroblasts (MEFs) and were scored for recombination after 4 days of infection by quantitative RT-PCR. The CAG-Cre construct was made by digesting Cre out of the Psp72Cre plasmid using BglII/XhoI and cloned into pNDCAG with BamHI/XhoI. Cre-ires-GFP virus was made as described (Sleutels unpublished).

Generation of conditional CTCF knockout mice with a M.m. castaneus distal chromosome 7

M.m. castaneus mice were backcrossed to *M.m. domesticus* (C57/Bl6) containing the CTCF conditional knockout allele. At each generation the presence of an *M.m. castaneus* allele at distal 7 was assayed using at least two SSLP and RFLP markers flanking the area. Oligo's used for SSLP and RFLP PCR analysis, if followed by digestion restriction enzymes were as indicated between brackets:
D7Nds4_F: GTGACAATACATTCCTGCTGT
D7Nds4_B: CTCAGATCTTATCTCTAGCAC
Drd4_F: TCAACCTGTGCGCCATGAGCGTG
Drd4_B: TGGTTGTAGCGCAGTGGCACGGT
Ins2 HindIII_F (HindIII): CATGTGGTGCTGTTCTGGGAGAT
Ins2 HindIII_B (HindIII): ATGGTGGCTGAAGTTGGTCTGTC
Igf2 NcoI_F (NcoI): TGGCTCTGCTGCTTAGGGAG
Igf2 NcoI_B (NcoI): AGTGCTGGACACTGGACCTGGT
D7mit167_F: CGTGTGAAGGCACACCTG
D7mit167_B: GAGCATCTGTGTGTGTGCCT
Slc22a18_F (KspI): ATCAACAGGACTTTTGCCCC
Slc22a18_B (KspI): CATGAAGAGACACGTTAGC

6

Discussion

Discussion

Goal of the thesis

CTCF is required for the proper regulation of many different processes, for example, the regulation of embryonic development, cell cycle progression, imprinted gene expression and X-chromosome inactivation. Correct expression of the protein throughout development is of fundamental importance for the survival of an organism. CTCF regulates these processes not only by activating and repressing genes, as a classical transcription factor would do, it is able to change chromatin structure, thereby regulating spatial proximity of genes and their regulatory elements. CTCF can also block an enhancer from activating a promoter and is the only protein identified so far in vertebrates that can perform such a function.

The fact that CTCF has such diverse roles makes it difficult to reveal a specific function in a distinct process. The variety of processes involving CTCF and their complexity require CTCF to interact with other proteins. The goal of the research described in this thesis was to get a better insight into the different processes regulated by CTCF. We primarily focused on the identification of interacting protein partners.

The identification of novel CTCF-interacting partners, provides evidence for the involvement of CTCF in processes that were not reported before. Our studies provide some clues how these processes are regulated. I discuss the implication of these results in the context of what is known about the properties and complex function of CTCF in regulating gene expression and chromatin organization.

Properties of CTCF in the cell – relationship to tagging approaches

Most transcription factors and chromatin proteins undergo rapid cycles of binding and unbinding on chromatin, with dwell times in the order of only a few seconds (Phair et al., 2004). By tagging CTCF with GFP its mobility in living cells was analyzed using FRAP. This analysis reveals that 70% of the CTCF protein present in the nucleus is immobile (Heath manuscript in preparation), indicating that CTCF is most of the time tightly bound to DNA, or is interacting with a nuclear matrix structure and/or nuclear compartments (Wiesmeijer et al., 2008). Indeed a significant portion of CTCF is present in an insoluble nuclear fraction, which represents a combination of nuclear matrix material and genomic DNA. DNA, packaged into chromatin is locally constrained but its spatial organization in the nucleus of living cells is considered to be dynamic. CTCF might therefore form the basis of chromatin loops and tether these to 'fixed' structures such as the nuclear matrix or nucleolar rim. In *Drosophila* and yeast, insulators are tethered to the nuclear lamina or nuclear pore complex (Ishii et al., 2002; Stewart et al., 2007). Our mass spectrometry analysis identifies components of these structures, however these factors were also identified in the BirA control experiments. Thus, it is difficult to judge whether CTCF associates with the nuclear matrix or nuclear substructures.

Several known CTCF interacting proteins are reported to be a component of the nuclear matrix. These include nucleophosmin, PARP, YY1 and the largest subunit of RNA polymerase II (Feuerstein and Mond, 1987; Mika and Rost, 2005; Vidakovic et al., 2004). For nucleophosmin as well as for the DNA/RNA binding factor YB1 a direct interaction between purified proteins

was not reported, leaving the possibility that the interaction is not direct but mediated via other proteins, or DNA (Chernukhin et al., 2000; Yusufzai et al., 2004).

Tight control of CTCF protein levels is required for proper development and cellular function. This is shown by the fact that both cells overexpressing and lacking CTCF have severe problems with proliferation. Modifying CTCF by a tagging approach might therefore be risky. Indeed, tagging the N-terminus of CTCF with GFP resulted in low expression levels of the tagged protein and a severe reduction in the viability of homozygous knock-in mice (Heath manuscript in preparation). Tagging CTCF at the C-terminus with the biotin-tag did not affect expression levels of the protein. This indicates that the size or the position of the tag is able to affect the stability and/or efficiency of the fusion protein production.

Mice expressing biotinylated CTCF cannot only be used to identify interacting proteins, but also to identify CTCF binding sites in the genome, in a temporal and tissue-specific manner. Identification of binding sites can be done using ChIP with streptavidin-beads. These are likely to give higher enrichment compared to antibodies, due to the high affinity of the biotin-streptavidin interaction. Another advantage of this approach is that it can be scaled up, without worrying about the quantities of antibody.

Fractionation profiles of nuclear extracts isolated from murine erythroid leukemia (MEL) and embryonic stem (ES) cells show CTCF to be present in complexes ranging from about 350 kD to 5 MD or larger. Other transcription factors like GATA1 and Ldb1 were shown to be present in protein complexes of similar sizes (Rodriguez et al., 2005). In MEL cells CTCF is predominately present in complexes larger than 670 kD while in ES cells CTCF is mostly present in complexes of 2,5 MD or more. This suggests that CTCF forms larger complexes in ES cells. Since we find multiple different CTCF-interacting proteins most of which have not been reported before, we propose that CTCF interacts with multiple components in a transient way. In the absence of DNA interactions are harder to detect, indicating that many interactions are, directly or indirectly, mediated via DNA. Several known interacting proteins show binding next to CTCF on specific sites. For example, YY1 binds in close proximity to CTCF on the X chromosome, UBF binds to the CTCF bound rDNA spacer promoter, Kaiso and nucleophosmin bind to the β-globin insulator and the large subunit of RNA polymerase II binds to some CTCF binding sites (Chernukhin et al., 2007; Defossez et al., 2005). In this way factors can modulate and fine tune CTCFs action on specific sites or mediate its insulator function, and vice versa.

The fact that only a few known CTCF interacting proteins where found back in our analysis could be due to the specificity of these interactions with CTCF to a particular cell type. For most published CTCF interacting proteins the amount of protein interacting with CTCF cannot be judged from immunoprecipitations that were performed since the authors do not indicate the enrichment of the pull down compared to the input. Furthermore all immunoprecipitations done so far to confirm or identify protein partners were performed in human cervical carcinoma (HeLa) cells. The observed interactions could be specific for that cell type. The exception is Yy1 that shows an interaction with CTCF at the endogenous level in ES cells (Donohoe et al., 2007). The interaction is weak and/or only a minority of Yy1 interacts with CTCF, however we were able to confirm this interaction.

We found chromatin remodeling and transcription (co) factors to bind to CTCF, as well as components involved in transcriptional elongation, cell cycle regulation, replication, repair, protein modification and rRNA transcription. These data point to a general role for CTCF in the nucleus. CTCF might form a platform for multiple proteins allowing multiple processes to take place.

CTCF in ES cells

CTCF is expected to gain binding to the ICR during differentiation of ES cells, since we report a (minor) loss of methylation at this site and a gain of ICR mediated interchromosomal interaction was reported during the process of ES cell maturation (Zhao et al., 2006). In agreement with this, chromatin in undifferentiated ES cells is reported to be globally decondensed and to contain a high fraction of only loosely bound architectural chromatin proteins, including linker histone H1 and HP1 (Meshorer and Misteli, 2006). It would be interesting to study the role of CTCF in ES cell differentiation in more detail and test whether a gain of binding upon differentiation is generally applicable to CTCF bound insulator sites. Moreover identification of CTCF interacting proteins in differentiated ES cells might teach us more about how CTCF can contribute to changes in chromatin structures. In pull downs from ES cells no transcription factors were found to interact with CTCF. On the other hand, the RNA polymerase I complex and factors involved in cell cycle regulation are exclusively identified in ES cells and not in the tissues that were used for pull downs. CTCF might interact mainly with transcription factors in a tissue specific manner. To maintain the pluripotent state of ES cells, tissue specific transcription factors might be more repressed in these cells and interactions could be less favorable.

Since CTCF is reported to contribute to overall chromatin structure it would be expected that ES cells lacking CTCF would not be able to differentiate. ES cells deleted for CTCF show a similar phenotype as observed in the MEFs; they stop dividing and can be maintained for multiple days until the cultures are taken over by cells that do express the protein. ES cells gradually deleting CTCF can be rescued upon expression of the mouse and chicken protein, however CTCFL is not able to rescue the deficiency (Sleutels unpublished). It would be interesting to determine whether CTCF protein with specific zinc-fingers deletions can also rescue these ES cells. A read-out for these rescue experiments can be simply a gain of cell division. The differentiation capabilities of these cells or inter- or intra-chromosomal interaction mediated by CTCF could also be analyzed.

CTCF and its role in regulating transcription by RNA polymerase I

Identification of an interaction between CTCF and UBF and a shared function in regulating RNA polymerase I was not expected, because only very small amounts of CTCF protein can be detected in the nucleolus. CTCF is more enriched in nucleoli of K562 cells, a line derived from patients with chronic myeloid leukemia. It was reported that upon differentiation of these cells endogenous CTCF is more enriched in the nucleolus. However we could not recapitulate this result (data not shown). We also do not observe an enrichment of UBF binding to CTCF in differentiated K562 cells (data not shown). The study done by Torrano et al. is performed using constructs overexpressing CTCF and adding compounds to stimulate differentiation (Torrano et al., 2006). This raises the question to what extend this system mimics the endogenous situation. However we do report multiple CTCF binding sites in the human rDNA repeat indicating that CTCF plays, also in humans, an important role in regulating rRNA transcription. It would be interesting to fuse endogenous CTCF to a nucleolar localization signal and analyze the impact of this with respect to rRNA transcription (Mekhail et al., 2007).

CTCF was shown to interact with another nucleolar protein, nucleophosmin. The interaction was suggested to take place at the borders of the nucleolus and to be involved in tethering insulator sequences to this site (Yusufzai et al., 2004). Although CTCF is largely excluded from the nucleolus we do show it binds to a nucleolar protein and to rDNA sequences. Whether

CTCF binding to the rDNA spacer promoter takes place in the nucleolus or at the borders of the nucleolus remains to be determined.

The nucleolus is a dynamic nuclear structure that assembles and disassembles during each mitotic cell division. Nucleolar proteins were demonstrated to rapidly associate with and dissociate from nucleolar components in a continuous exchange with the nucleoplasm (Chen and Huang, 2001). The functional significance of this is unclear but it might involve reactivation or modification of specific functional components. So in contradiction to CTCF, UBF is a very mobile protein not only exchanging rapidly between nucleolus and nucleoplasm but also rapidly associating and dissociating from it's binding to the rDNA. All RNA polymerase I associated components undergo rapid exchange at the promoter and stably associate with chromatin only when they are incorporated into an elongation complex (Dundr et al., 2002). An approximate stoichiometry of 100 UBF dimers for each rRNA gene was calculated (Kermekchiev et al., 1997). With about the same amount of UBF and CTCF molecules in a cell, UBF is in clear excess in the nucleolus compared to CTCF. The *in vitro* affinity of the UBF-CTCF interaction is quite high since the purified proteins remain bound to each other after high stringency washes.

It would be interesting to know whether regions exist where CTCF can bind to the rDNA locus indirectly by binding to UBF. Dimethyl 3,3'-dithiobispropionimidate (DTBP) is a protein-protein crosslinking agent that facilitates ChIP of proteins that do not directly contact the DNA but are recruited by protein-protein interactions (Kurdistani and Grunstein, 2003). This method can be used to analyze CTCF binding along the whole rDNA repeat. CTCF might be involved in the formation of single rDNA expression units by the formation of distinct loops. However the overall nucleolar structure does not change in the absence of CTCF as judged by a similar staining pattern of UBF and RNA polymerase I in the absence of CTCF. This indicates that CTCF does not play a major role in this in non-dividing or resting cells. However in activated T-cells the nucleolar organization is affected in the absence of CTCF as shown by reduced protein levels of fibrillarin and nucleophosmin (Heath submitted). Indicating that stimulatory signals or cell division are needed in order to detect the consequence of absence of CTCF on rDNA transcription and nucleolar organisation.

The binding strength of CTCF to the spacer promoter is comparable to other known CTCF sites although the repetitive nature of the rDNA should be taken into account. Mammalian cells contain about four hundred copies of tandemly repeated ribosomal RNA genes of which about half is actively transcribed (Grummt, 2003). If the repetitive nature is taken into account it can be concluded that only a few molecules of CTCF bind to the spacer promoter. Bandshifts show a higher affinity of CTCF for the spacer promoter fragment compared to the chicken lysozyme binding site. This indicates that the *in vivo* affinity is high but the amount of CTCF molecules present to bind to the spacer promoter is limiting, which could explain the low signal of CTCF in the nucleolus as detected by immunofluoresence. Whether CTCF binds to the spacer promoter on active and not on inactive rDNA repeats or on both is an interesting question that remains to be solved.

Many questions remain unanswered regarding the function of a CTCF-UBF interaction and the role of CTCF in regulating RNA polymerase I transcription. It would be interesting to know whether CTCF can bind the RNA polymerase I complex directly or only indirectly via an interaction with UBF (Figure 6.1). Furthermore the sequence of events regulating transcription from the rDNA spacer promoter remains to be determined. Since UBF and RNA polymerase I binding are down regulated specifically at the spacer promoter and enhancer region in the absence of CTCF, it seems likely that CTCF is involved in recruiting the proteins to these sites. The binding of UBF, RNA polymerase I and CTCF results in transcription from the spacer promoter.

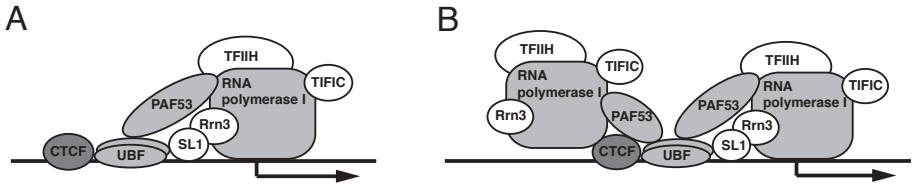

Figure 6.1 Mouse RNA polymerase I transcription initiation complex at the spacer promoter
A. CTCF binds to RNA polymerase I indirectly via its interaction with UBF. **B.** CTCF binds directly to UBF and RNA polymerase I. Grey oval represent factors that were identified from mass spectrometry analysis on CTCF pull downs. The polymerase-associated factor that is responsible for an interacting of UBF with RNA polymerase I is PAF53 and was found back in our mass spectrometry lists.

How exactly CTCF and UBF can enhance transcription from the spacer promoter is not clear. It seems likely that they are involved in recruiting and accumulating RNA polymerase I on this site. Enhancement of transcription of the 45S pre-rRNA was reported to depend on the enhancers and spacer promoter. However spacer promoter transcription is also involved in repression of 45S pre-rRNA transcription. Short intergenic RNAs that arise from the unstable spacer promoter transcript are involved in this (Mayer et al., 2006) (Figure 6.2). Whether these RNAs originate from spacer promoters that are adjacent to active or silent rDNA repeats is not known. The fact that CTCF activates transcription from the spacer promoter by maintaining RNA polymerase I and UBF binding and the observation that under silent conditions, in differentiated 3T3L1 cells, CTCF binding is reduced suggest that CTCF binds only to active rDNA repeats. This would than also indicate that spacer promoter transcripts are transcribed only from active rDNA repeats and suggests the transcripts to work *in trans* and target specific rDNA repeats for silencing upon binding to NoRC.

The similarities between the position of CTCF binding sites in the mouse and *Xenopus* rDNA repeat, together with the enhancer-blocking properties of this element in *Xenopus* and the histone marks present in mouse suggest that CTCF functions as a boundary at this site (Robinett et al., 1997). It could potentially block transcripts that are accidentally not terminated and pass through the IGS. Alternatively CTCF may function to bring the promoters and enhancers in close proximity and in this way provide specificity and directionality of the enhancers to activate the appropriate promoter. Such a function for a CTCF bound enhancer-blocking element has been described recently (Yoon et al., 2007). Additional evidence for CTCF to function as an enhancer-blocking boundary at the rDNA spacer promoter are the presence of Cohesin binding on this site and the ability of the site to block an enhancer from activating a promoter in an *in vitro* assay. It would also be interesting to know whether CTCF binds this site in a methylation sensitive manner. CTCF binding is reported to be methylation sensitive indicating that CTCF is likely to only bind the unmethylated (active) rDNA repeats.

Transcripts from the spacer promoter were shown to be important for rDNA silencing by tethering NoRC resulting in heterochromatin formation (Mayer et al., 2006) (Figure 6.2). The question is why and when silencing needs to be controlled by spacer transcripts. The link of epigenetic control of rDNA transcription with cell growth and proliferation is not fully understood. The ratio of active and inactive rDNA repeats could be reset during each cell division or the number of active/inactive rDNA repeats could be slightly variable during the cell cycle. Limiting the number of active rRNA genes might decrease DNA damage and stabilize rDNA by repressing homologous recombination. Altogether we propose that CTCF, by regulating spacer promoter

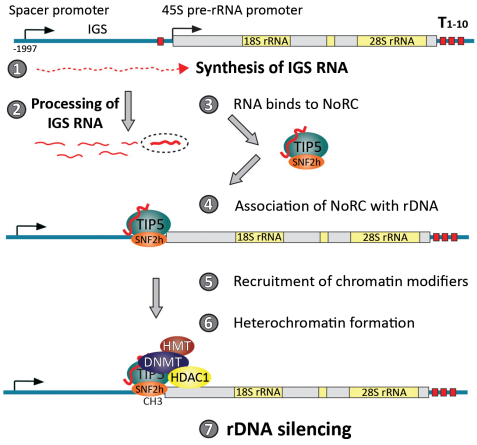

Figure 6.2 Model depicting the role of spacer promoter transcripts in rDNA inactivation
First, intergenic spacer (IGS) transcripts are synthesized by the spacer promoter (step 1) This IGS RNA is than processed into short intermediates (step 2). RNA that matches the rDNA promoter associates with the chromatin remodeling complex NoRC (step 3). Once bound to RNA, NoRC is recruited to the rDNA promoter (step 4). The large subunit of NoRC, TIP5, interacts with histone modifying enzymes, such as HDACs, HMTs, and DNMTs (step 5), leading to heterochromatic histone modifications and *de novo* DNA methylation (step 6). As a consequence, transcription complex formation is impaired and rRNA genes are silenced (step 7). (Adapted from (Grummt, 2007))

transcription is important for controlling silencing of rDNA repeats or plays a role in maintaining a balance between active and inactive repeats. In the absence of CTCF we could however not detect a difference in 45S rRNA compared to wild-type. This could be explained by the fact that cells without CTCF do not divide. Cell division is likely to be needed in order to clear the cell from RNA-NoRC complexes and to re-activate the heterochromatic silenced rDNA repeats. It would be interesting to determine transcription from the spacer promoter and gene promoter in CTCF absent T cells that are able to proliferate upon circumventing T-cell receptor signaling with Ionomycin and phorbol ester (PMA) (Heath unpublished observation). Interestingly *in vitro* activation of CTCF negative T cells does results in a defective nucleolar organization (Heath submitted).

CTCF and cell cycle regulation

CTCF plays important roles in cell growth control. Enhanced expression of CTCF induces growth retardation at multiple points during the cell cycle without apoptosis. The cell cycle profile of these cells is identical to untreated cells and viable cells could be maintained without dividing for several days (Rasko et al., 2001). Interestingly we show that MEFs deleted for CTCF show a similar phenotype as observed upon enhanced expression of CTCF. Also CTCF deficient T cells show a proliferative defect without any obvious block at a specific stage during the cell cycle. Although most cell types have an impaired proliferative capacity in the absence of CTCF some cell types like γδ T cells proliferate and develop normal without CTCF.

It is likely that multiple mechanisms are involved in CTCF dependent cell proliferation. CTCF was suggested to modulate the transcription of multiple genes involved in cell proliferation such as cMyc and Rb. Interestingly conditional expression of CTCF leads to a reduction in cMyc and up regulation of p27 and p19ARF in immature B cells (Qi et al., 2003). A misregulation of cMyc is however not detected in T cells deleted for CTCF. In those cells we do detect an up regulation of two major cell cycle inhibitors p21 and p27 and no change in levels of p53. Normally p21 and p27 are downregulated at the ISP stage of T cell development to allow transition from G1 to S-phase. We show binding of CTCF to the first intron of p21 in MEFs indicating that p21 is a direct target of CTCF. By binding to this site CTCF might regulate the transcription of the gene *in cis* for example

by interfering with the elongation of the transcript. This might explain the proliferative block observed in CTCF deficient cells.

The identified CTCF protein partners might solve another piece of the puzzle to clarify the role of CTCF in cellular proliferation. The identification of the T-cell specific factors TCF/Lef, does not explain the defects observed in T cell development in the absence of CTCF. Interestingly, multiple proteins that function during mitosis by associating with the spindle were identified in our analysis in ES cells. During mitosis CTCF is reported to remain bound to the chromosomes and to associate with the centrosome and mitotic spindle as shown by antibody stainings on fixed cells and imaging of GFP-CTCF expressing cells (Burke et al., 2005; Zhang et al., 2004). However in CTCF negative ES cells that are rescued with GFP-CTCF the protein could be detected only on mitotic chromosomes and not on centrosomes (Heath manuscript in preparation). Interestingly, not all CTCF binding sites are occupied by CTCF during mitosis (Burke et al., 2005; Komura et al., 2007). The specific localization of CTCF on mitotic chromosomes together with the identification of a subunit of the Condensin complex suggests that CTCF is perhaps required for regulating the condensation of chromatin at mitosis.

Centrosomes are important for the completion of cytokinesis and cell cycle progression through G1 into S phase (Hinchcliffe et al., 2001). They assemble and organize the mitotic bipolar spindle that ensures accurate chromosome segregation during mitosis. The association of CTCF with mitotic centrosomes is microtubule independent as shown by nocodazole treatment indicating that CTCF bind via different means (Zhang et al., 2004). The interacting proteins identified from our analysis could be involved in recruiting CTCF to the centrosome. It would be very interesting to examine the role of CTCF during mitosis in more detail.

The CP190 protein that binds to CTCF in *Drosophila* is important for proper insulator function and was originally identified as a protein binding to centrosomes. The protein seems to be unique in Drosophila since no homologs are detected in mouse or human. Despite its localization to the centrosomes the mutant shows no defect in centrosome or microtubule organization during cell division indicating that CP190 is not essential for these processes (Butcher et al., 2004).

The effect of CTCF on rRNA transcriptional regulation, its influence on cell cycle regulating target genes and its potential involvement in mitosis are likely to contribute to the proliferative defect observed in CTCF negative cells.

CTCF in organizing chromatin structure

The CTCF interacting proteins Nucleophosmin, Kaiso and CHD8 have been implicated in regulating its insulator function. However this regulatory function of Kaiso and Nucleophosmin have been demonstrated only at specific insulator sites. The recent mapping of all CTCF binding sites throughout the human genome can now be used to address the question whether regulation of CTCF insulator function by these proteins is generally applicable or specific to certain sites. The mapping of CTCF binding sites in the human also highlights the importance of CTCF for proper regulation of a large part of the genome. The abundance of its binding sites implies a role for CTCF not only in regulating the expression of specific genes, but also of whole chromatin domains. This is reflected by the fact that 45% of the CTCF binding sites map to intergenic regions where CTCF is likely to function as an insulator, while 20% of its binding sites maps to promoter regions (Barski et al., 2007).

Whether the intergenic CTCF binding sites all function as insulators or also include enhancers remains to be determined. Insulators and enhancers are both shown to coincide with

incorporation of the histone variant H2A.Z, high levels of H3K4 methylation and H3K9 mono-methylation (Barski et al., 2007). In the absence of CTCF a local loss of histone H3 acetylation and a gain of H3K27 dimethylation can be observed in the β-globin locus (Splinter et al., 2006). A local loss of H3 acetylation, H4 acetylation, H3K4 dimethylation and a loss of the histone variant H2A.Z is observed in the absence of CTCF on the rDNA spacer promoter. This indicates a role for CTCF in maintaining local histone marks, most likely by interacting with chromatin remodeling factors. In this way CTCF is likely to regulate gene expression by locally opening up chromatin thereby influencing the accessibility of DNA for RNA polymerase.

The CTCF interacting protein CHD8 is reported to be part of the MLL1 complex that includes the WD40-repeat protein WDR5 and is involved in methylation of lysine 4 of histone H3 (Dou et al., 2005; Ishihara et al., 2006). However direct evidence for a role of CHD8 in regulating CTCF dependent histone modification is lacking. Interestingly we identified the WDR5 component of this complex to co-purify with CTCF. WDR5 is required for binding of the methyltransferase complex to the methylated lysine 4 of histone H3 (Wysocka et al., 2005). It would therefore be very interesting to test whether this protein binds to CTCF and whether it is involved in methylating H3K4 at its binding sites. A role for WDR5 in transcription factor dependent regulation of gene expression was shown recently. An interaction between WDR5 and Pax7 resulted in local H3K4 trimethylation at its binding site followed by transcriptional activation (McKinnell et al., 2008). Beside CTCF, the β-globin cHS4 boundary protein USF was shown to interact with CHD8. The potential CTCF interacting protein Snf2h and WDR5 were also identified as USF interacting proteins among other chromatin remodeling factors. Downregulation of USF results in a local loss of H3R4 methylation and decreased acetylation of H3 and H4 (Huang et al., 2007). Proteins binding to USF like PRMT1, and the histone acetyltransferases (HATs) PCAF and SRC-1 were suggested to be important in this.

Comparing the identified CTCF interacting proteins with proteins that are involved in regulating the function of other boundary- or architectural factors can learn us more about how insulators function. Multiple factors that form boundaries between silenced and active chromatin domains in *Drosophila*, yeast and mouse like SATB1, LSD1 and Dot1 were shown to function by distinct mechanism but all involve recruitment of histone modifiying complexes (Altaf et al., 2007; Lan et al., 2007).

Boundary function and enhancer-blocking abilities were found to be separate functions, but how these two functions are performed in still unclear (Kurshakova et al., 2007; Recillas-Targa et al., 2002). An enhancer-blocking insulator was suggested to participate in pairing the correct enhancer and promoter and conferring directionality to the enhancer, rather than solely as an element that blocks inappropriate enhancer-activated transcription (Krebs and Dunaway, 1998). Recent experiments show that enhancer-blocking insulators can function by directly interacting with the regulated promoter and enhancer elements (Yoon et al., 2007). CTCF shares its enhancer-blocking abilities with the *Drosophila* proteins Su(Hw), Zw5, BEAF32 and GAGA factor. Not much is known about the functioning or interacting proteins of Zw5 and BEAF32. The two proteins were reported to bind to each other and BEAF32 was shown to bind to the heterochromatin protein D1 (Blanton et al., 2003; Cuvier et al., 2002).

Su(Hw) can organize chromatin structure and stabilize its insulator function by recruiting multiple proteins and a RNA component (Lei and Corces, 2006). The fact that Su(Hw) and dCTCF are present together in insulator bodies and share a common interacting protein, CP190, is interesting (Gerasimova et al., 2007; Mohan et al., 2007). This indicates that insulators can function together and share interacting proteins.

GAGA is on the other hand known to interact with various other proteins. These proteins

function in regulating polycomb group complexes and transcriptional abilities by GAGA and include tramtrack, pipsqueak, lolal/batman, SAP18 (Sin3a/ HDAC co-repressor complex), PRC1 (polycomb group), NURF and Ssrp1/FACT (Adkins et al., 2006; Lehmann, 2004). Recent experiments revealed a role for GAGA and FACT in directing the replacement of histone H3 with its variant H3.3 (Nakayama et al., 2007). FACT is involved in nucleosome rearrangements and disassembly. It was proposed that this replacement is achieved through the disassembly of a nucleosome by the GAGA-FACT complex followed by HIRA mediated deposition of a H3.3-containing nucleosome. This is suggested to be important to counteract the spreading of silent chromatin and marks boundaries of *cis*-regulatory domains (Mito et al., 2007). It would be very interesting to find out whether these mechanisms are evolutionarily conserved in vertebrates. Interestingly we detect the Ssrp1 and Cdc68 components of the FACT complex to be pulled down by CTCF. A potential function of a CTCF-FACT interaction could be similar of that of the GAGA-FACT complex in incorporating histone variants at specific sites. Alternatively the interaction could be important during transcription initiation or elongation when FACT associates with RNA polymerase II. Factors involved in elongation were also shown to be pulled down by CTCF and include the PAF-complex, MRG15 and Tceb.

The recently reported finding that CTCF binding is often accompanied by the presence of the histone variant H2A.Z is confirmed by our experiments. H2A.Z binding to CTCF sites seems to be not specific to intergenic sites i.e. potential insulator sites. The CTCF bound Igf2/H19 ICR does not show significant H2A.Z levels, but consisted with this variant being also a mark of active enhancers it was found to be enriched at the CTCF bound H19 enhancers (Verona et al., 2008; Yoon et al., 2007). Furthermore we can detect an enrichment of H2A.Z at the CTCF binding site in the cMyc promoter. Recent studies suggest that histone turnover helps to maintain continuous access to sequence-specific DNA-binding proteins that regulate epigenetic inheritance, providing a dynamic alternative to histone-marking models for the propagation of active chromatin. Interestingly it was reported recently that transcription factors like p53 and cMyc can provide positioning cues that direct the location of H2A.Z-containing nucleosomes. By doing so these factors can regulate transcription in part depending on the position of H2A.Z incorporation (Gevry et al., 2007), raising the question whether CTCF can also regulate transcription in part by prefentially positioning H2A.Z within chromatin. H2A.Z deposition is not only reduced in the absence of CTCF but also in the absence of SRCAP and some histone acetyltransferases (Wong et al., 2007; Zhang et al., 2005) Global depletion of H2A.Z from mammalian cells causes a highly unstable genome caused by defects in the chromosome segregation process (Rangasamy et al., 2004). SRCAP is the complex involved in catalyzing the incorporation of H2A.Z in chromatin (Ruhl et al., 2006). The potential CTCF interacting SWI/SNF chromatin remodeling protein BAF57 was shown to be part of the SRCAP complex, suggesting that CTCF could be involved in targeting SRCAP to specific promoter or insulator regions and in this way regulates incorporation of H2A.Z.

Chromatin structure is not only influenced by post-translational modifications of histones but also by the RNAi machinery (yeast, plants and *Drosophila*) and binding of structural non-histone proteins. The HMG-box family of proteins comprises such a group of structural proteins.

Interestingly we report an interaction of two HMG-box (HMGB) containing proteins with CTCF, UBF and the not yet characterized binding to Lef1. Interestingly two other proteins copurifing with CTCF, Ssrp1 and BAF57, both share a 50-60% homology with the first HMG-box of UBF. This suggest a common theme between CTCF and its binding to non-sequence specific architectural HMGB proteins. The HMGB proteins are abundant and highly mobile proteins in the cell nucleus that influence chromatin structure and enhance the accessibility of binding sites to regulatory factors. Due to their remarkable DNA bending activity, HMGB proteins can

increase the structural flexibility of DNA, which could help CTCF in the formation of chromatin loops. Members of the HMGB family act as versatile modulators of chromatin function they can facilitate nucleosome remodeling (Grasser et al., 2007). One mechanism by which HMGB proteins could prime the nucleosome for migration is to loosen the wrapped DNA and so enhance accessibility to chromatin-remodeling complexes and possibly also to transcription factors like CTCF (Travers, 2003). The interaction of CTCF with another chromatin architectural protein HP1 is also interesting.

The recent finding that half of the CTCF binding sites are enriched for Cohesin and that Cohesin contributes to CTCF insulator function suggests that Cohesin is the factor involved in structural and functional organization of chromatin (Parelho et al., 2008; Stedman et al., 2008; Wendt et al., 2008). However an interaction between the two proteins could not be detected indicating that not only direct but also indirect interactions via other proteins are important for mediating CTCF function.

The data presented in this thesis have enhanced our understanding of the complex functioning of CTCF. The protein partners we identified provide insight in CTCF function and provide an exciting challenge to unravel the precise function of specific interactions.

References

Adkins, N.L., Hagerman, T.A. and Georgel, P. (2006) GAGA protein: a multi-faceted transcription factor. *Biochem Cell Biol*, **84**, 559-567.

Agalioti, T., Lomvardas, S., Parekh, B., Yie, J., Maniatis, T. and Thanos, D. (2000) Ordered recruitment of chromatin modifying and general transcription factors to the IFN-beta promoter. *Cell*, **103**, 667-678.

Agresti, A. and Bianchi, M.E. (2003) HMGB proteins and gene expression. *Curr Opin Genet Dev*, **13**, 170-178.

Ainscough, J.F., John, R.M., Barton, S.C. and Surani, M.A. (2000) A skeletal muscle-specific mouse Igf2 repressor lies 40 kb downstream of the gene. *Development*, **127**, 3923-3930.

Akhmanova, A., Mausset-Bonnefont, A.L., van Cappellen, W., Keijzer, N., Hoogenraad, C.C., Stepanova, T., Drabek, K., van der Wees, J., Mommaas, M., Onderwater, J., van der Meulen, H., Tanenbaum, M.E., Medema, R.H., Hoogerbrugge, J., Vreeburg, J., Uringa, E.J., Grootegoed, J.A., Grosveld, F. and Galjart, N. (2005) The microtubule plus-end-tracking protein CLIP-170 associates with the spermatid manchette and is essential for spermatogenesis. *Genes Dev*, **19**, 2501-2515.

Akhmanova, A., Verkerk, T., Langeveld, A., Grosveld, F. and Galjart, N. (2000) Characterisation of transcriptionally active and inactive chromatin domains in neurons. *J Cell Sci*, **113 Pt 24**, 4463-4474.

Altaf, M., Utley, R.T., Lacoste, N., Tan, S., Briggs, S.D. and Cote, J. (2007) Interplay of chromatin modifiers on a short basic patch of histone H4 tail defines the boundary of telomeric heterochromatin. *Mol Cell*, **28**, 1002-1014.

Alvarez, J.D., Yasui, D.H., Niida, H., Joh, T., Loh, D.Y. and Kohwi-Shigematsu, T. (2000) The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. *Genes Dev*, **14**, 521-535.

Andrews, N.C. and Faller, D.V. (1991) A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. *Nucleic Acids Res*, **19**, 2499.

Andrews, P.D., Ovechkina, Y., Morrice, N., Wagenbach, M., Duncan, K., Wordeman, L. and Swedlow, J.R. (2004) Aurora B regulates MCAK at the mitotic centromere. *Dev Cell*, **6**, 253-268.

Arabi, A., Wu, S., Ridderstrale, K., Bierhoff, H., Shiue, C., Fatyol, K., Fahlen, S., Hydbring, P., Soderberg, O., Grummt, I., Larsson, L.G. and Wright, A.P. (2005) c-Myc associates with ribosomal DNA and activates RNA polymerase I transcription. *Nat Cell Biol*, **7**, 303-310.

Aravind, L. (2001) The WWE domain: a common interaction module in protein ubiquitination and ADP ribosylation. *Trends Biochem Sci*, **26**, 273-275.

Aravind, L. and Landsman, D. (1998) AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Res*, **26**, 4413-4421.

Armstrong, J.A., Bieker, J.J. and Emerson, B.M. (1998) A SWI/SNF-related chromatin remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLF in vitro. *Cell*, **95**, 93-104.

Arnold, R., Burcin, M., Kaiser, B., Muller, M. and Renkawitz, R. (1996) DNA bending by the silencer protein NeP1 is modulated by TR and RXR. *Nucleic Acids Res*, **24**, 2640-2647.

Arnosti, D.N. and Kulkarni, M.M. (2005) Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J Cell Biochem*, **94**, 890-898.

Azzam, H.S., Grinberg, A., Lui, K., Shen, H., Shores, E.W. and Love, P.E. (1998) CD5 expression is developmentally regulated by T cell receptor (TCR) signals and TCR avidity. *J Exp Med*, **188**, 2301-2311.

Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M. and Teichmann, S.A. (2004) Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol*, **14**, 283-291.

Banditt, M., Koller, T. and Sogo, J.M. (1999) Transcriptional activity and chromatin structure of enhancer-deleted rRNA genes in Saccharomyces cerevisiae. *Mol Cell Biol*, **19**, 4953-4960.

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.

Barnden, M.J., Allison, J., Heath, W.R. and Carbone, F.R. (1998) Defective TCR expression in transgenic mice constructed using cDNA-based alpha- and beta-chain genes under the control of heterologous regulatory elements. *Immunol Cell Biol*, **76**, 34-40.

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.

Bazett-Jones, D.P., Leblanc, B., Herfort, M. and Moss, T. (1994) Short-range DNA looping by the Xenopus HMG-box transcription factor, xUBF. *Science*, **264**, 1134-1137.

Bednar, J., Horowitz, R.A., Grigoryev, S.A., Carruthers, L.M., Hansen, J.C., Koster, A.J. and Woodcock, C.L. (1998) Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc Natl Acad Sci U S A*, **95**, 14173-14178.

Bell, A.C., West, A.G. and Felsenfeld, G. (1999) The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*, **98**, 387-396.

Bell, A.C., West, A.G. and Felsenfeld, G. (2001) Insulators and boundaries: versatile regulatory elements in the eukaryotic. *Science*, **291**, 447-450.

Bell, S.P., Learned, R.M., Jantzen, H.M. and Tjian, R. (1988) Functional cooperativity between transcription factors UBF1 and SL1 mediates human ribosomal RNA synthesis. *Science*, **241**, 1192-1197.

Belotserkovskaya, R., Oh, S., Bondarenko, V.A., Orphanides, G., Studitsky, V.M. and Reinberg, D. (2003) FACT facilitates transcription-dependent nucleosome alteration. *Science*, **301**, 1090-1093.

Bendelac, A., Matzinger, P., Seder, R.A., Paul, W.E. and Schwartz, R.H. (1992) Activation events during thymic selection. *J Exp Med*, **175**, 731-742.

Bender, M.A., Bulger, M., Close, J. and Groudine, M. (2000) Beta-globin gene switching and DNase I sensitivity of the endogenous beta-globin locus in mice do not require the locus control region. *Mol Cell*, **5**, 387-393.

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.

Bender, M.A., Reik, A., Close, J., Telling, A., Epner, E., Fiering, S., Hardison, R. and Groudine, M. (1998) Description and targeted deletion of 5' hypersensitive site 5 and 6 of the mouse beta-globin locus control region. *Blood*, **92**, 4394-4403.

Berger, S.L. (2007) The complex language of chromatin regulation during transcription. *Nature*, **447**, 407-412.

Bernardi, R. and Pandolfi, P.P. (2003) Role of PML and the PML-nuclear body in the control of programmed cell death. *Oncogene*, **22**, 9048-9057.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S.L. and Lander, E.S. (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, **125**, 315-326.

Betz, J.L., Chang, M., Washburn, T.M., Porter, S.E., Mueller, C.L. and Jaehning, J.A. (2002) Phenotypic analysis of Paf1/RNA polymerase II complex mutations reveals connections to cell cycle regulation, protein synthesis, and lipid and nucleic acid metabolism. *Mol Genet Genomics*, **268**, 272-285.

Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev*, **16**, 6-21.

Bird, A.P., Taggart, M.H. and Gehring, C.A. (1981) Methylated and unmethylated ribosomal RNA genes in the mouse. *J. Mol. Biol.*, **152**, 1-17.

Bird, A.P. and Wolffe, A.P. (1999) Methylation-induced repression--belts, braces, and chromatin. *Cell*, **99**, 451-454.

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

Boisvert, F.M., van Koningsbruggen, S., Navascues, J. and Lamond, A.I. (2007) The multifunctional nucleolus. *Nat Rev Mol Cell Biol*, **8**, 574-585.

Bouhouche, N., Syvanen, M. and Kado, C.I. (2000) The origin of prokaryotic C2H2 zinc finger regulators. *Trends Microbiol*, **8**, 77-81.

Boumil, R.M., Ogawa, Y., Sun, B.K., Huynh, K.D. and Lee, J.T. (2006) Differential methylation of Xite and CTCF sites in Tsix mirrors the pattern of X-inactivation choice in mice. *Mol Cell Biol*, **26**, 2109-2117.

Boyer, L.A., Latek, R.R. and Peterson, C.L. (2004) The SANT domain: a unique histone-tail-binding module? *Nat Rev Mol Cell Biol*, **5**, 158-163.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S. and Crawford, G.E. (2008) High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell*, **132**, 311-322.

Bradsher, J., Auriol, J., Proietti de Santis, L., Iben, S., Vonesch, J.L., Grummt, I. and Egly, J.M. (2002) CSB is a component of RNA pol I transcription. *Mol Cell*, **10**, 819-829.

Brown, J.M., Leach, J., Reittie, J.E., Atzberger, A., Lee-Prudhoe, J., Wood, W.G., Higgs, D.R., Iborra, F.J. and Buckle, V.J. (2006) Coregulated human globin genes are frequently in spatial proximity when active. *J Cell Biol*, **172**, 177-187.

Bruce, K., Myers, F.A., Mantouvalou, E., Lefevre, P., Greaves, I., Bonifer, C., Tremethick, D.J., Thorne, A.W. and Crane-Robinson, C. (2005) The replacement histone H2A.Z in a hyperacetylated form is a feature of active genes in the chicken. *Nucleic Acids Res*, **33**, 5633-5639.

Bruno, M., Flaus, A., Stockdale, C., Rencurel, C., Ferreira, H. and Owen-Hughes, T. (2003) Histone H2A/H2B dimer exchange by ATP-dependent chromatin remodeling activities. *Mol Cell*, **12**, 1599-1606.

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.

Buratowski, S. (2003) The CTD code. *Nat Struct Biol*, **10**, 679-680.

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.

Burgers, W.A., Fuks, F. and Kouzarides, T. (2002) DNA methyltransferases get connected to chromatin. *Trends Genet*, **18**, 275-277.

Burke, L.J., Zhang, R., Bartkuhn, M., Tiwari, V.K., Tavoosidana, G., Kurukuti, S., Weth, C., Leers, J., Galjart, N., Ohlsson, R. and Renkawitz, R. (2005) CTCF binding and higher order chromatin structure of the H19 locus are maintained in mitotic chromatin. *Embo J*, **24**, 3291-3300.

Burns, L.G. and Peterson, C.L. (1997) The yeast SWI-SNF complex facilitates binding of a transcriptional activator to nucleosomal sites in vivo. *Mol Cell Biol*, **17**, 4811-4819.

Butcher, R.D., Chodagam, S., Basto, R., Wakefield, J.G., Henderson, D.S., Raff, J.W. and Whitfield, W.G. (2004) The Drosophila centrosome-associated protein CP190 is essential for viability but not for cell division. *J Cell Sci*, **117**, 1191-1199.

Cai, H.N. and Shen, P. (2001) Effects of cis arrangement of chromatin insulators on enhancer-blocking activity. *Science*, **291**, 493-495.

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

Calikowski, T.T. and Meier, I. (2006) Isolation of nuclear proteins. *Methods Mol Biol*, **323**, 393-402.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S. and Zhang, Y. (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*, **298**, 1039-1043.

Capco, D.G., Wan, K.M. and Penman, S. (1982) The nuclear matrix: three-dimensional architecture and protein composition. *Cell*, **29**, 847-858.

Capelson, M. and Corces, V.G. (2005) The ubiquitin ligase dTopors directs the nuclear organization of a chromatin insulator. *Mol Cell*, **20**, 105-116.

Caspary, T., Cleary, M.A., Baker, C.C., Guan, X.J. and Tilghman, S.M. (1998) Multiple mechanisms regulate imprinting of the mouse distal chromosome 7 gene cluster. *Mol Cell Biol*, **18**, 3466-3474.

Cavanaugh, A.H., Hempel, W.M., Taylor, L.J., Rogalsky, V., Todorov, G. and Rothblum, L.I. (1995) Activity of RNA polymerase I transcription factor UBF blocked by Rb gene product. *Nature*, **374**, 177-180.

Chambeyron, S. and Bickmore, W.A. (2004) Does looping and clustering in the nucleus regulate gene expression? *Curr Opin Cell Biol*, **16**,

256-262.
Changolkar, L.N., Costanzi, C., Leu, N.A., Chen, D., McLaughlin, K.J. and Pehrson, J.R. (2007) Developmental changes in histone macroH2A1-mediated gene regulation. *Mol Cell Biol*, **27**, 2758-2764.
Changolkar, L.N. and Pehrson, J.R. (2006) macroH2A1 histone variants are depleted on active genes but concentrated on the inactive X chromosome. *Mol Cell Biol*, **26**, 4410-4420.
Chao, W., Huynh, K.D., Spencer, R.J., Davidow, L.S. and Lee, J.T. (2002) CTCF, a candidate trans-acting factor for X-inactivation choice. *Science*, **295**, 345-347.
Charalambous, M., Menheniott, T.R., Bennett, W.R., Kelly, S.M., Dell, G., Dandolo, L. and Ward, A. (2004) An enhancer element at the Igf2/H19 locus drives gene expression in both imprinted and non-imprinted tissues. *Dev Biol*, **271**, 488-497.
Chen, D., Belmont, A.S. and Huang, S. (2004) Upstream binding factor association induces large-scale chromatin decondensation. *Proc Natl Acad Sci U S A*, **101**, 15106-15111.
Chen, D. and Huang, S. (2001) Nucleolar components involved in ribosome biogenesis cycle between the nucleolus and nucleoplasm in interphase cells. *J Cell Biol*, **153**, 169-176.
Chen, M. and Shen, X. (2007) Nuclear actin and actin-related proteins in chromatin dynamics. *Curr Opin Cell Biol*, **19**, 326-330.
Chernukhin, I., Shamsuddin, S., Kang, S.Y., Bergstrom, R., Kwon, Y.W., Yu, W., Whitehead, J., Mukhopadhyay, R., Docquier, F., Farrar, D., Morrison, I., Vigneron, M., Wu, S.Y., Chiang, C.M., Loukinov, D., Lobanenkov, V., Ohlsson, R. and Klenova, E. (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.
Cheutin, T., McNairn, A.J., Jenuwein, T., Gilbert, D.M., Singh, P.B. and Misteli, T. (2003) Maintenance of stable heterochromatin domains by dynamic HP1 binding. *Science*, **299**, 721-725.
Cheutin, T., O'Donohue, M.F., Beorchia, A., Vandelaer, M., Kaplan, H., Defever, B., Ploton, D. and Thiry, M. (2002) Three-dimensional organization of active rRNA genes within the nucleolus. *J Cell Sci*, **115**, 3297-3307.
Chi, T.H., Wan, M., Zhao, K., Taniuchi, I., Chen, L., Littman, D.R. and Crabtree, G.R. (2002) Reciprocal regulation of CD4/CD8 expression by SWI/SNF-like BAF complexes. *Nature*, **418**, 195-199.
Ching, R.W., Dellaire, G., Eskiw, C.H. and Bazett-Jones, D.P. (2005) PML bodies: a meeting place for genomic loci? *J Cell Sci*, **118**, 847-854.
Chung, J.H., Bell, A.C. and Felsenfeld, G. (1997) Characterization of the chicken beta-globin insulator. *Proc Natl Acad Sci U S A*, **94**, 575-580.
Chung, J.H., Whiteley, M. and Felsenfeld, G. (1993) A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. *Cell*, **74**, 505-514.
Cobb, B.S., Hertweck, A., Smith, J., O'Connor, E., Graf, D., Cook, T., Smale, S.T., Sakaguchi, S., Livesey, F.J., Fisher, A.G. and Merkenschlager, M. (2006) A role for Dicer in immune regulation. *J Exp Med*, **203**, 2519-2527.
Comai, L., Tanese, N. and Tjian, R. (1992) The TATA-binding protein and associated factors are integral components of the RNA polymerase I transcription factor, SL1. *Cell*, **68**, 965-976.
Conconi, A., Widmer, R.M., Koller, T. and Sogo, J.M. (1989) Two different chromatin structures coexist in ribosomal RNA genes throughout the cell cycle. *Cell*, **57**, 753-761.
Constancia, M., Dean, W., Lopes, S., Moore, T., Kelsey, G. and Reik, W. (2000) Deletion of a silencer element in Igf2 results in loss of imprinting independent of H19. *Nat Genet*, **26**, 203-206.
Copenhaver, G.P., Putnam, C.D., Denton, M.L. and Pikaard, C.S. (1994) The RNA polymerase I transcription factor UBF is a sequence-tolerant HMG-box protein that can recognize structured nucleic acids. *Nucleic Acids Res*, **22**, 2651-2657.
Cosma, M.P., Tanaka, T. and Nasmyth, K. (1999) Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter. *Cell*, **97**, 299-311.
Cote, J., Quinn, J., Workman, J.L. and Peterson, C.L. (1994) Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science*, **265**, 53-60.
Cremer, T. and Cremer, C. (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet*, **2**, 292-301.
Cuvier, O., Hart, C.M., Kas, E. and Laemmli, U.K. (2002) Identification of a multicopy chromatin boundary element at the borders of silenced chromosomal domains. *Chromosoma*, **110**, 519-531.
Davies, K., Bowden, L., Smith, P., Dean, W., Hill, D., Furuumi, H., Sasaki, H., Cattanach, B. and Reik, W. (2002) Disruption of mesodermal enhancers for Igf2 in the minute mutant. *Development*, **129**, 1657-1668.
Davis, T.L., Yang, G.J., McCarrey, J.R. and Bartolomei, M.S. (2000) The H19 methylation imprint is erased and re-established differentially on the parental alleles during male germ cell development. *Hum Mol Genet*, **9**, 2885-2894.
de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., Grosveld, F. and Strouboulis, J. (2003) Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc Natl Acad Sci U S A*, **100**, 7480-7485.
De La Rosa-Velazquez, I.A., Rincon-Arano, H., Benitez-Bribiesca, L. and Recillas-Targa, F. (2007) Epigenetic regulation of the human retinoblastoma tumor suppressor gene promoter by CTCF. *Cancer Res*, **67**, 2577-2585.
De Winter, R.F. and Moss, T. (1986) Spacer promoters are essential for efficient enhancement of X. laevis ribosomal transcription. *Cell*, **44**, 313-318.
Dean, A. (2006) On a chromosome far, far away: LCRs and gene expression. *Trends Genet*, **22**, 38-45.
Defossez, P.A., Kelly, K.F., Filion, G.J., Perez-Torrado, R., Magdinier, F., Menoni, H., Nordgaard, C.L., Daniel, J.M. and Gilson, E. (2005) The human enhancer blocker CTC-binding factor interacts with the transcription factor Kaiso. *J Biol Chem*, **280**, 43017-43023.
Dekker, J., Rippe, K., Dekker, M. and Kleckner, N. (2002) Capturing chromosome conformation. *Science*, **295**, 1306-1311.
Delenda, C. (2004) Lentiviral vectors: optimization of packaging, transduction and gene expression. *J Gene Med*, **6 Suppl 1**, S125-138.
Dellaire, G. and Bazett-Jones, D.P. (2004) PML nuclear bodies: dynamic sensors of DNA damage and cellular stress. *Bioessays*, **26**, 963-977.

Dellaire, G. and Bazett-Jones, D.P. (2007) Beyond repair foci: subnuclear domains and the cellular response to DNA damage. *Cell Cycle*, **6**, 1864-1872.
Dickinson, L.A., Joh, T., Kohwi, Y. and Kohwi-Shigematsu, T. (1992) A tissue-specific MAR/SAR DNA-binding protein with unusual binding site recognition. *Cell*, **70**, 631-645.
Dillon, N. and Festenstein, R. (2002) Unravelling heterochromatin: competition between positive and negative factors regulates accessibility. *Trends Genet*, **18**, 252-258.
Dintilhac, A. and Bernues, J. (2002) HMGB1 interacts with many apparently unrelated proteins by recognizing short amino acid sequences. *J Biol Chem*, **277**, 7021-7028.
Docquier, F., Farrar, D., D'Arcy, V., Chernukhin, I., Robinson, A.F., Loukinov, D., Vatolin, S., Pack, S., Mackay, A., Harris, R.A., Dorricott, H., O'Hare, M.J., Lobanenkov, V. and Klenova, E. (2005) Heightened expression of CTCF in breast cancer cells is associated with resistance to apoptosis. *Cancer Res*, **65**, 5112-5122.
Dolznig, H., Boulme, F., Stangl, K., Deiner, E.M., Mikulits, W., Beug, H. and Mullner, E.W. (2001) Establishment of normal, terminally differentiating mouse erythroid progenitors: molecular characterization by cDNA arrays. *Faseb J*, **15**, 1442-1444.
Donohoe, M.E., Zhang, L.F., Xu, N., Shi, Y. and Lee, J.T. (2007) Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch. *Mol Cell*, **25**, 43-56.
Dou, Y., Milne, T.A., Tackett, A.J., Smith, E.R., Fukuda, A., Wysocka, J., Allis, C.D., Chait, B.T., Hess, J.L. and Roeder, R.G. (2005) Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell*, **121**, 873-885.
Dougherty, W.G. and Parks, T.D. (1989) Molecular genetic and biochemical evidence for the involvement of the heptapeptide cleavage sequence in determining the reaction profile at two tobacco etch virus cleavage sites in cell-free assays. *Virology*, **172**, 145-155.
Dousset, T., Wang, C., Verheggen, C., Chen, D., Hernandez-Verdun, D. and Huang, S. (2000) Initiation of nucleolar assembly is independent of RNA polymerase I transcription. *Mol Biol Cell*, **11**, 2705-2717.
Doyen, C.M., Montel, F., Gautier, T., Menoni, H., Claudet, C., Delacour-Larose, M., Angelov, D., Hamiche, A., Bednar, J., Faivre-Moskalenko, C., Bouvet, P. and Dimitrov, S. (2006) Dissection of the unusual structural and functional properties of the variant H2A.Bbd nucleosome. *Embo J*, **25**, 4234-4244.
Doyon, Y., Selleck, W., Lane, W.S., Tan, S. and Cote, J. (2004) Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans. *Mol Cell Biol*, **24**, 1884-1896.
Drewell, R.A., Arney, K.L., Arima, T., Barton, S.C., Brenton, J.D. and Surani, M.A. (2002) Novel conserved elements upstream of the H19 gene are transcribed and act as mesodermal enhancers. *Development*, **129**, 1205-1213.
Drewell, R.A., Brenton, J.D., Ainscough, J.F., Barton, S.C., Hilton, K.J., Arney, K.L., Dandolo, L. and Surani, M.A. (2000) Deletion of a silencer element disrupts H19 imprinting independently of a DNA methylation epigenetic switch. *Development*, **127**, 3419-3428.
Driegen, S., Ferreira, R., van Zon, A., Strouboulis, J., Jaegle, M., Grosveld, F., Philipsen, S. and Meijer, D. (2005) A generic tool for biotinylation of tagged proteins in transgenic mice. *Transgenic Res*, **14**, 477-482.
Dumortier, H., Klein Gunnewiek, J., Roussel, J.P., van Aarssen, Y., Briand, J.P., van Venrooij, W.J. and Muller, S. (1998) At least three linear regions but not the zinc-finger domain of U1C protein are exposed at the surface of the protein in solution and on the human spliceosomal U1 snRNP particle. *Nucleic Acids Res*, **26**, 5486-5491.
Dundr, M., Hoffmann-Rohrer, U., Hu, Q., Grummt, I., Rothblum, L.I., Phair, R.D. and Misteli, T. (2002) A kinetic framework for a mammalian RNA polymerase in vivo. *Science*, **298**, 1623-1626.
Dunn, K.L., Zhao, H. and Davie, J.R. (2003) The insulator binding protein CTCF associates with the nuclear matrix. *Exp Cell Res*, **288**, 218-223.
Eberharter, A. and Becker, P.B. (2004) ATP-dependent nucleosome remodelling: factors and functions. *J Cell Sci*, **117**, 3707-3711.
Edwards, C.A. and Ferguson-Smith, A.C. (2007) Mechanisms regulating imprinted genes in clusters. *Curr Opin Cell Biol*, **19**, 281-289.
Eisen, J.A., Sweder, K.S. and Hanawalt, P.C. (1995) Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions. *Nucleic Acids Res*, **23**, 2715-2723.
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. and Groudine, M. (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.
Espada, J., Ballestar, E., Santoro, R., Fraga, M.F., Villar-Garea, A., Nemeth, A., Lopez-Serra, L., Ropero, S., Aranda, A., Orozco, H., Moreno, V., Juarranz, A., Stockert, J.C., Langst, G., Grummt, I., Bickmore, W. and Esteller, M. (2007) Epigenetic disruption of ribosomal RNA genes and nucleolar architecture in DNA methyltransferase 1 (Dnmt1) deficient cells. *Nucleic Acids Res*, **35**, 2191-2198.
Falvo, J.V., Thanos, D. and Maniatis, T. (1995) Reversal of intrinsic DNA bends in the IFN beta gene enhancer by transcription factors and the architectural protein HMG I(Y). *Cell*, **83**, 1101-1111.
Fan, J.Y., Rangasamy, D., Luger, K. and Tremethick, D.J. (2004) H2A.Z alters the nucleosome surface to promote HP1alpha-mediated chromatin fiber folding. *Mol Cell*, **16**, 655-661.
Fan, T., Hagan, J.P., Kozlov, S.V., Stewart, C.L. and Muegge, K. (2005a) Lsh controls silencing of the imprinted Cdkn1c gene. *Development*, **132**, 635-644.
Fan, Y., Nikitina, T., Zhao, J., Fleury, T.J., Bhattacharyya, R., Bouhassira, E.E., Stein, A., Woodcock, C.L. and Skoultchi, A.I. (2005b) Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell*, **123**, 1199-1212.
Faro-Trindade, I. and Cook, P.R. (2006) Transcription factories: structures conserved during differentiation and evolution. *Biochem Soc Trans*, **34**, 1133-1137.
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.
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.

Feil, R., Walter, J., Allen, N.D. and Reik, W. (1994) Developmental control of allelic methylation in the imprinted mouse Igf2 and H19 genes. *Development*, **120**, 2933-2943.

Ferguson-Smith, A.C., Sasaki, H., Cattanach, B.M. and Surani, M.A. (1993) Parental-origin-specific epigenetic modification of the mouse H19 gene. *Nature*, **362**, 751-755.

Feuerstein, N. and Mond, J.J. (1987) "Numatrin," a nuclear matrix protein associated with induction of proliferation in B lymphocytes. *J Biol Chem*, **262**, 11389-11397.

Feuerstein, N., Spiegel, S. and Mond, J.J. (1988) The nuclear matrix protein, numatrin (B23), is associated with growth factor-induced mitogenesis in Swiss 3T3 fibroblasts and with T lymphocyte proliferation stimulated by lectins and anti-T cell antigen receptor antibody. *J Cell Biol*, **107**, 1629-1642.

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

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.

Filippova, G.N., Lindblom, A., Meincke, L.J., Klenova, E.M., Neiman, P.E., Collins, S.J., Doggett, N.A. and Lobanenkov, V.V. (1998) A widely expressed transcription factor with multiple DNA sequence specificity, CTCF, is localized at chromosome segment 16q22.1 within one of the smallest regions of overlap for common deletions in breast and prostate cancers. *Genes Chromosomes Cancer*, **22**, 26-36.

Filippova, G.N., Qi, C.F., Ulmer, J.E., Moore, J.M., Ward, M.D., Hu, Y.J., Loukinov, D.I., Pugacheva, E.M., Klenova, E.M., Grundy, P.E., Feinberg, A.P., Cleton-Jansen, A.M., Moerland, E.W., Cornelisse, C.J., Suzuki, H., Komiya, A., Lindblom, A., Dorion-Bonnet, F., Neiman, P.E., Morse, H.C., 3rd, Collins, S.J. and Lobanenkov, V.V. (2002) Tumor-associated zinc finger mutations in the CTCF transcription factor selectively alter tts DNA-binding specificity. *Cancer Res*, **62**, 48-52.

Filippova, G.N., Thienes, C.P., Penn, B.H., Cho, D.H., Hu, Y.J., Moore, J.M., Klesert, T.R., Lobanenkov, V.V. and Tapscott, S.J. (2001) CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. *Nat Genet*, **28**, 335-343.

Fitzpatrick, G.V., Pugacheva, E.M., Shin, J.Y., Abdullaev, Z., Yang, Y., Khatod, K., Lobanenkov, V.V. and Higgins, M.J. (2007) Allele-specific binding of CTCF to the multipartite imprinting control region KvDMR1. *Mol Cell Biol*, **27**, 2636-2647.

Flanagan, J.F., Mi, L.Z., Chruszcz, M., Cymborowski, M., Clines, K.L., Kim, Y., Minor, W., Rastinejad, F. and Khorasanizadeh, S. (2005) Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature*, **438**, 1181-1185.

Forrester, W.C., Epner, E., Driscoll, M.C., Enver, T., Brice, M., Papayannopoulou, T. and Groudine, M. (1990) A deletion of the human beta-globin locus activation region causes a major alteration in chromatin structure and replication across the entire beta-globin locus. *Genes Dev*, **4**, 1637-1649.

Fraser, P. and Bickmore, W. (2007) Nuclear organization of the genome and the potential for gene regulation. *Nature*, **447**, 413-417.

Fujimoto, S., Matsunaga, S., Yonemura, M., Uchiyama, S., Azuma, T. and Fukui, K. (2004) Identification of a novel plant MAR DNA binding protein localized on chromosomal surfaces. *Plant Mol Biol*, **56**, 225-239.

Fuks, F. (2005) DNA methylation and histone modifications: teaming up to silence genes. *Curr Opin Genet Dev*, **15**, 490-495.

Fuss, S.H., Omura, M. and Mombaerts, P. (2007) Local and cis effects of the H element on expression of odorant receptor genes in mouse. *Cell*, **130**, 373-384.

Galande, S., Purbey, P.K., Notani, D. and Kumar, P.P. (2007) The third dimension of gene regulation: organization of dynamic chromatin loopscape by SATB1. *Curr Opin Genet Dev*, **17**, 408-414.

Gao, J., Li, T. and Lu, L. (2007) Functional role of CCCTC binding factor in insulin-stimulated cell proliferation. *Cell Prolif*, **40**, 795-808.

Garcia, S.N., Kirtane, B.M., Podlutsky, A.J., Pereira-Smith, O.M. and Tominaga, K. (2007) Mrg15 null and heterozygous mouse embryonic fibroblasts exhibit DNA-repair defects post exposure to gamma ionizing radiation. *FEBS Lett*, **581**, 5275-5281.

Garrett, F.E., Emelyanov, A.V., Sepulveda, M.A., Flanagan, P., Volpi, S., Li, F., Loukinov, D., Eckhardt, L.A., Lobanenkov, V.V. and Birshtein, B.K. (2005) Chromatin architecture near a potential 3' end of the igh locus involves modular regulation of histone modifications during B-Cell development and in vivo occupancy at CTCF sites. *Mol Cell Biol*, **25**, 1511-1525.

Gerasimova, T.I., Byrd, K. and Corces, V.G. (2000) A chromatin insulator determines the nuclear localization of DNA. *Mol Cell*, **6**, 1025-1035.

Gerasimova, T.I. and Corces, V.G. (1998) Polycomb and trithorax group proteins mediate the function of a chromatin insulator. *Cell*, **92**, 511-521.

Gerasimova, T.I., Lei, E.P., Bushey, A.M. and Corces, V.G. (2007) Coordinated control of dCTCF and gypsy chromatin insulators in Drosophila. *Mol Cell*, **28**, 761-772.

Gevry, N., Chan, H.M., Laflamme, L., Livingston, D.M. and Gaudreau, L. (2007) p21 transcription is regulated by differential localization of histone H2A.Z. *Genes Dev*, **21**, 1869-1881.

Giese, K., Cox, J. and Grosschedl, R. (1992) The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. *Cell*, **69**, 185-195.

Gilbert, N., Thomson, I., Boyle, S., Allan, J., Ramsahoye, B. and Bickmore, W.A. (2007) DNA methylation affects nuclear organization, histone modifications, and linker histone binding but not chromatin compaction. *J Cell Biol*, **177**, 401-411.

Glibetic, M., Taylor, L., Larson, D., Hannan, R., Sells, B. and Rothblum, L. (1995) The RNA polymerase I transcription factor UBF is the product of a primary response gene. *J Biol Chem*, **270**, 4209-4212.

Gombert, W.M., Farris, S.D., Rubio, E.D., Morey-Rosler, K.M., Schubach, W.H. and Krumm, A. (2003) The c-myc insulator element and matrix attachment regions define the c-myc chromosomal domain. *Mol Cell Biol*, **23**, 9338-9348.

Gomos-Klein, J., Harrow, F., Alarcon, J. and Ortiz, B.D. (2007) CTCF-independent, but not CTCF-dependent, elements significantly contribute to TCR-alpha locus control region activity. *J Immunol*, **179**, 1088-1095.

Grasser, K.D., Launholt, D. and Grasser, M. (2007) High mobility group proteins of the plant HMGB family: dynamic chromatin modulators. *Biochim Biophys Acta*, **1769**, 346-357.

Gribnau, J., Luikenhuis, S., Hochedlinger, K., Monkhorst, K. and Jaenisch, R. (2005) X chromosome choice occurs independently of

asynchronous replication timing. *J Cell Biol*, **168**, 365-373.
Grimaldi, G., Fiorentini, P. and Di Nocera, P.P. (1990) Spacer promoters are orientation-dependent activators of pre-rRNA transcription in Drosophila melanogaster. *Mol Cell Biol*, **10**, 4667-4677.
Grisendi, S., Mecucci, C., Falini, B. and Pandolfi, P.P. (2006) Nucleophosmin and cancer. *Nat Rev Cancer*, **6**, 493-505.
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.
Grueneberg, D.A., Pablo, L., Hu, K.Q., August, P., Weng, Z. and Papkoff, J. (2003) A functional screen in human cells identifies UBF2 as an RNA polymerase II transcription factor that enhances the beta-catenin signaling pathway. *Mol Cell Biol*, **23**, 3936-3950.
Grummt, I. (2003) Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. *Genes Dev*, **17**, 1691-1702.
Grummt, I. (2007) Different epigenetic layers engage in complex crosstalk to define the epigenetic state of mammalian rRNA genes. *Hum Mol Genet*, **16 Spec No 1**, R21-27.
Haaf, T., Hayman, D.L. and Schmid, M. (1991) Quantitative determination of rDNA transcription units in vertebrate cells. *Exp Cell Res*, **193**, 78-86.
Haines, T.R., Rodenhiser, D.I. and Ainsworth, P.J. (2001) Allele-specific non-CpG methylation of the Nf1 gene during early mouse development. *Dev Biol*, **240**, 585-598.
Hamiche, A., Sandaltzopoulos, R., Gdula, D.A. and Wu, C. (1999) ATP-dependent histone octamer sliding mediated by the chromatin remodeling complex NURF. *Cell*, **97**, 833-842.
Hampsey, M. and Reinberg, D. (2003) Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation. *Cell*, **113**, 429-432.
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.
Hanada, K., Song, C.Z., Yamamoto, K., Yano, K., Maeda, Y., Yamaguchi, K. and Muramatsu, M. (1996) RNA polymerase I associated factor 53 binds to the nucleolar transcription factor UBF and functions in specific rDNA transcription. *Embo J*, **15**, 2217-2226.
Handwerger, K.E. and Gall, J.G. (2006) Subnuclear organelles: new insights into form and function. *Trends Cell Biol*, **16**, 19-26.
Hannan, K.M., Kennedy, B.K., Cavanaugh, A.H., Hannan, R.D., Hirschler-Laszkiewicz, I., Jefferson, L.S. and Rothblum, L.I. (2000) RNA polymerase I transcription in confluent cells: Rb downregulates rDNA transcription during confluence-induced cell cycle arrest. *Oncogene*, **19**, 3487-3497.
Hannan, R.D., Stefanovsky, V., Taylor, L., Moss, T. and Rothblum, L.I. (1996) Overexpression of the transcription factor UBF1 is sufficient to increase ribosomal DNA transcription in neonatal cardiomyocytes: implications for cardiac hypertrophy. *Proc Natl Acad Sci U S A*, **93**, 8750-8755.
Hans, F. and Dimitrov, S. (2001) Histone H3 phosphorylation and cell division. *Oncogene*, **20**, 3021-3027.
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.
Harrington, C.A. and Chikaraishi, D.M. (1987) Transcription of spacer sequences flanking the rat 45S ribosomal DNA gene. *Mol Cell Biol*, **7**, 314-325.
Harrison, S.C. (1991) A structural taxonomy of DNA-binding domains. *Nature*, **353**, 715-719.
Hassan, A.H., Prochasson, P., Neely, K.E., Galasinski, S.C., Chandy, M., Carrozza, M.J. and Workman, J.L. (2002) Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell*, **111**, 369-379.
Hata, K., Okano, M., Lei, H. and Li, E. (2002) Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development*, **129**, 1983-1993.
Hebert, M.D., Shpargel, K.B., Ospina, J.K., Tucker, K.E. and Matera, A.G. (2002) Coilin methylation regulates nuclear body formation. *Dev Cell*, **3**, 329-337.
Hendriks, R.W., de Bruijn, M.F., Maas, A., Dingjan, G.M., Karis, A. and Grosveld, F. (1996) Inactivation of Btk by insertion of lacZ reveals defects in B cell development only past the pre-B cell stage. *Embo J*, **15**, 4862-4872.
Heng, H.H., Goetze, S., Ye, C.J., Liu, G., Stevens, J.B., Bremer, S.W., Wykes, S.M., Bode, J. and Krawetz, S.A. (2004) Chromatin loops are selectively anchored using scaffold/matrix-attachment regions. *J Cell Sci*, **117**, 999-1008.
Henikoff, S. and Ahmad, K. (2005) Assembly of variant histones into chromatin. *Annu Rev Cell Dev Biol*, **21**, 133-153.
Henikoff, S., Furuyama, T. and Ahmad, K. (2004) Histone variants, nucleosome assembly and epigenetic inheritance. *Trends Genet*, **20**, 320-326.
Hikichi, T., Kohda, T., Kaneko-Ishino, T. and Ishino, F. (2003) Imprinting regulation of the murine Meg1/Grb10 and human GRB10 genes; roles of brain-specific promoters and mouse-specific CTCF-binding sites. *Nucleic Acids Res*, **31**, 1398-1406.
Hinchcliffe, E.H., Miller, F.J., Cham, M., Khodjakov, A. and Sluder, G. (2001) Requirement of a centrosomal activity for cell cycle progression through G1 into S phase. *Science*, **291**, 1547-1550.
Hirano, T. (1999) SMC-mediated chromosome mechanics: a conserved scheme from bacteria to vertebrates? *Genes Dev*, **13**, 11-19.
Hirano, T. (2002) The ABCs of SMC proteins: two-armed ATPases for chromosome condensation, cohesion, and repair. *Genes Dev*, **16**, 399-414.
Ho, Y., Elefant, F., Liebhaber, S.A. and Cooke, N.E. (2006) Locus control region transcription plays an active role in long-range gene activation. *Mol Cell*, **23**, 365-375.
Hoffmann, A. and Roeder, R.G. (1996) Cloning and characterization of human TAF20/15. Multiple interactions suggest a central role in TFIID complex formation. *J Biol Chem*, **271**, 18194-18202.
Hoffmann, M.J., Muller, M., Engers, R. and Schulz, W.A. (2006) Epigenetic control of CTCFL/BORIS and OCT4 expression in urogenital malignancies. *Biochem Pharmacol*, **72**, 1577-1588.
Hogan, C. and Varga-Weisz, P. (2007) The regulation of ATP-dependent nucleosome remodelling factors. *Mutat Res*, **618**, 41-51.
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.
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., Custer, M., Chen, G.A., Zhao, M., Nguyen, D.M., Barrett, J.C., Lobanenkov, V.V. and Schrump, D.S. (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.
Hoogenraad, C.C., Koekkoek, B., Akhmanova, A., Krugers, H., Dortland, B., Miedema, M., van Alphen, A., Kistler, W.M., Jaegle, M., Koutsourakis, M., Van Camp, N., Verhoye, M., van der Linden, A., Kaverina, I., Grosveld, F., De Zeeuw, C.I. and Galjart, N. (2002) Targeted mutation of Cyln2 in the Williams syndrome critical region links CLIP-115 haploinsufficiency to neurodevelopmental abnormalities in mice. *Nat Genet*, **32**, 116-127.
Hoogstraten, D., Nigg, A.L., Heath, H., Mullenders, L.H., van Driel, R., Hoeijmakers, J.H., Vermeulen, W. and Houtsmuller, A.B. (2002) Rapid switching of TFIIH between RNA polymerase I and II transcription and DNA repair in vivo. *Mol Cell*, **10**, 1163-1174.
Howell, C.Y., Bestor, T.H., Ding, F., Latham, K.E., Mertineit, C., Trasler, J.M. and Chaillet, J.R. (2001) Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. *Cell*, **104**, 829-838.
Howman, E.V., Fowler, K.J., Newson, A.J., Redward, S., MacDonald, A.C., Kalitsis, P. and Choo, K.H. (2000) Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. *Proc Natl Acad Sci U S A*, **97**, 1148-1153.
Hozak, P., Cook, P.R., Schofer, C., Mosgoller, W. and Wachtler, F. (1994) Site of transcription of ribosomal RNA and intranucleolar structure in HeLa cells. *J Cell Sci*, **107 (Pt 2)**, 639-648.
Huang, S., Li, X., Yusufzai, T.M., Qiu, Y. and Felsenfeld, G. (2007) USF1 recruits histone modification complexes and is critical for maintenance of a chromatin barrier. *Mol Cell Biol*, **27**, 7991-8002.
Huang, Y., Fang, J., Bedford, M.T., Zhang, Y. and Xu, R.M. (2006) Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science*, **312**, 748-751.
Iarovaia, O.V., Akopov, S.B., Nikolaev, L.G., Sverdlov, E.D. and Razin, S.V. (2005) Induction of transcription within chromosomal DNA loops flanked by MAR elements causes an association of loop DNA with the nuclear matrix. *Nucleic Acids Res*, **33**, 4157-4163.
Ishihara, K., Oshimura, M. and Nakao, M. (2006) CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Mol Cell*, **23**, 733-742.
Ishii, K., Arib, G., Lin, C., Van Houwe, G. and Laemmli, U.K. (2002) Chromatin boundaries in budding yeast: the nuclear pore connection. *Cell*, **109**, 551-562.
Iuchi, S. (2001) Three classes of C2H2 zinc finger proteins. *Cell Mol Life Sci*, **58**, 625-635.
Izaurralde, E., Kas, E. and Laemmli, U.K. (1989) Highly preferential nucleation of histone H1 assembly on scaffold-associated regions. *J Mol Biol*, **210**, 573-585.
Jansa, P. and Grummt, I. (1999) Mechanism of transcription termination: PTRF interacts with the largest subunit of RNA polymerase I and dissociates paused transcription complexes from yeast and mouse. *Mol Gen Genet*, **262**, 508-514.
Jantzen, H.M., Admon, A., Bell, S.P. and Tjian, R. (1990) Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins. *Nature*, **344**, 830-836.
Jantzen, H.M., Chow, A.M., King, D.S. and Tjian, R. (1992) Multiple domains of the RNA polymerase I activator hUBF interact with the TATA-binding protein complex hSL1 to mediate transcription. *Genes Dev*, **6**, 1950-1963.
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, J., Cai, Y., Li, B., Conaway, R.C., Workman, J.L., Conaway, J.W. and Kusch, T. (2005) In and out: histone variant exchange in chromatin. *Trends Biochem Sci*, **30**, 680-687.
Jones, B.K., Levorse, J. and Tilghman, S.M. (2001) Deletion of a nuclease-sensitive region between the Igf2 and H19 genes leads to Igf2 misregulation and increased adiposity. *Hum Mol Genet*, **10**, 807-814.
Jouvenot, Y., Poirier, F., Jami, J. and Paldi, A. (1999) Biallelic transcription of Igf2 and H19 in individual cells suggests a post-transcriptional contribution to genomic imprinting. *Curr Biol*, **9**, 1199-1202.
Junera, H.R., Masson, C., Geraud, G., Suja, J. and Hernandez-Verdun, D. (1997) Involvement of in situ conformation of ribosomal genes and selective distribution of upstream binding factor in rRNA transcription. *Mol Biol Cell*, **8**, 145-156.
Kahana, J.A. and Cleveland, D.W. (2001) Cell cycle. Some importin news about spindle assembly. *Science*, **291**, 1718-1719.
Kalab, P., Weis, K. and Heald, R. (2002) Visualization of a Ran-GTP gradient in interphase and mitotic Xenopus egg extracts. *Science*, **295**, 2452-2456.
Kamakaka, R.T. and Biggins, S. (2005) Histone variants: deviants? *Genes Dev*, **19**, 295-310.
Kanduri, C., Pant, V., Loukinov, D., Pugacheva, E., Qi, C.F., Wolffe, A., Ohlsson, R. and Lobanenkov, V.V. (2000) Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr Biol*, **10**, 853-856.
Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E. and Sasaki, H. (2004) Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature*, **429**, 900-903.
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.
Kellum, R. and Schedl, P. (1991) A position-effect assay for boundaries of higher order chromosomal domains. *Cell*, **64**, 941-950.
Kellum, R. and Schedl, P. (1992) A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Mol Cell Biol*, **12**, 2424-2431.
Kermekchiev, M., Workman, J.L. and Pikaard, C.S. (1997) Nucleosome binding by the polymerase I transactivator upstream binding factor displaces linker histone H1. *Mol Cell Biol*, **17**, 5833-5842.
Kim, J., Kollhoff, A., Bergmann, A. and Stubbs, L. (2003) Methylation-sensitive binding of transcription factor YY1 to an insulator sequence within the paternally expressed imprinted gene, Peg3. *Hum Mol Genet*, **12**, 233-245.
Kim, J.D., Hinz, A.K., Bergmann, A., Huang, J.M., Ovcharenko, I., Stubbs, L. and Kim, J. (2006) Identification of clustered YY1 binding sites in imprinting control regions. *Genome Res*, **16**, 901-911.
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.
Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D. and Ren, B. (2005) A high-resolution map of active promoters in the human genome. *Nature*, **436**, 876-880.

Kireeva, M.L., Hancock, B., Cremona, G.H., Walter, W., Studitsky, V.M. and Kashlev, M. (2005) Nature of the nucleosomal barrier to RNA polymerase II. *Mol Cell*, **18**, 97-108.

Kisielow, P., Bluthmann, H., Staerz, U.D., Steinmetz, M. and von Boehmer, H. (1988) Tolerance in T-cell-receptor transgenic mice involves deletion of nonmature CD4+8+ thymocytes. *Nature*, **333**, 742-746.

Klein, J. and Grummt, I. (1999) Cell cycle-dependent regulation of RNA polymerase I transcription: the nucleolar transcription factor UBF is inactive in mitosis and early G1. *Proc Natl Acad Sci U S A*, **96**, 6096-6101.

Klenova, E., Chernukhin, I., Inoue, T., Shamsuddin, S. and Norton, J. (2002) Immunoprecipitation techniques for the analysis of transcription factor complexes. *Methods*, **26**, 254-259.

Klenova, E. and Ohlsson, R. (2005) Poly(ADP-ribosyl)ation and epigenetics. Is CTCF PARt of the plot? *Cell Cycle*, **4**, 96-101.

Klenova, E., Scott, A.C., Roberts, J., Shamsuddin, S., Lovejoy, E.A., Bergmann, S., Bubb, V.J., Royer, H.D. and Quinn, J.P. (2004) YB-1 and CTCF differentially regulate the 5-HTT polymorphic intron 2 enhancer which predisposes to a variety of neurological disorders. *J Neurosci*, **24**, 5966-5973.

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., Morse, H.C., 3rd, Neiman, P.E. and Lobanenkov, V.V. (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., Fagerlie, S., Filippova, G.N., Kretzner, L., Goodwin, G.H., Loring, G., Neiman, P.E. and Lobanenkov, V.V. (1998) Characterization of the chicken CTCF genomic locus, and initial study of the cell cycle-regulated promoter of the gene. *J Biol Chem*, **273**, 26571-26579.

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.

Klenova, E.M., Nicolas, R.H., U, S., Carne, A.F., Lee, R.E., Lobanenkov, V.V. and Goodwin, G.H. (1997) Molecular weight abnormalities of the CTCF transcription factor: CTCF migrates aberrantly in SDS-PAGE and the size of the expressed protein is affected by the UTRs and sequences within the coding region of the CTCF gene. *Nucleic Acids Res*, **25**, 466-474.

Koide, T., Ainscough, J., Wijgerde, M. and Surani, M.A. (1994) Comparative analysis of Igf-2/H19 imprinted domain: identification of a highly conserved intergenic DNase I hypersensitive region. *Genomics*, **24**, 1-8.

Komura, J., Ikehata, H. and Ono, T. (2007) Chromatin fine structure of the c-MYC insulator element/DNase I-hypersensitive site I is not preserved during mitosis. *Proc Natl Acad Sci U S A*, **104**, 15741-15746.

Kooren, J., Palstra, R.J., Klous, P., Splinter, E., von Lindern, M., Grosveld, F. and de Laat, W. (2007) Beta-globin active chromatin Hub formation in differentiating erythroid cells and in p45 NF-E2 knock-out mice. *J Biol Chem*, **282**, 16544-16552.

Kornberg, R.D. (1974) Chromatin structure: a repeating unit of histones and DNA. *Science*, **184**, 868-871.

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

Krebs, J.E. and Dunaway, M. (1998) The scs and scs' insulator elements impart a cis requirement on enhancer-promoter interactions. *Mol Cell*, **1**, 301-308.

Krogan, N.J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Ryan, O.W., Golshani, A., Johnston, M., Greenblatt, J.F. and Shilatifard, A. (2003) The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol Cell*, **11**, 721-729.

Kruger, W., Peterson, C.L., Sil, A., Coburn, C., Arents, G., Moudrianakis, E.N. and Herskowitz, I. (1995) Amino acid substitutions in the structured domains of histones H3 and H4 partially relieve the requirement of the yeast SWI/SNF complex for transcription. *Genes Dev*, **9**, 2770-2779.

Kuhn, A. and Grummt, I. (1987) A novel promoter in the mouse rDNA spacer is active in vivo and in vitro. *Embo J*, **6**, 3487-3492.

Kumar, P.P., Bischof, O., Purbey, P.K., Notani, D., Urlaub, H., Dejean, A. and Galande, S. (2007) Functional interaction between PML and SATB1 regulates chromatin-loop architecture and transcription of the MHC class I locus. *Nat Cell Biol*, **9**, 45-56.

Kurasawa, Y., Earnshaw, W.C., Mochizuki, Y., Dohmae, N. and Todokoro, K. (2004) Essential roles of KIF4 and its binding partner PRC1 in organized central spindle midzone formation. *Embo J*, **23**, 3237-3248.

Kurdistani, S.K. and Grunstein, M. (2003) In vivo protein-protein and protein-DNA crosslinking for genomewide binding microarray. *Methods*, **31**, 90-95.

Kurshakova, M., Maksimenko, O., Golovnin, A., Pulina, M., Georgieva, S., Georgiev, P. and Krasnov, A. (2007) Evolutionarily conserved E(y)2/Sus1 protein is essential for the barrier activity of Su(Hw)-dependent insulators in Drosophila. *Mol Cell*, **27**, 332-338.

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.

Lamond, A.I. and Spector, D.L. (2003) Nuclear speckles: a model for nuclear organelles. *Nat Rev Mol Cell Biol*, **4**, 605-612.

Lan, F., Zaratiegui, M., Villen, J., Vaughn, M.W., Verdel, A., Huarte, M., Shi, Y., Gygi, S.P., Moazed, D., Martienssen, R.A. and Shi, Y. (2007) S. pombe LSD1 homologs regulate heterochromatin propagation and euchromatic gene transcription. *Mol Cell*, **26**, 89-101.

Langst, G., Becker, P.B. and Grummt, I. (1998) TTF-I determines the chromatin architecture of the active rDNA promoter. *Embo J*, **17**, 3135-3145.

Langst, G., Bonte, E.J., Corona, D.F. and Becker, P.B. (1999) Nucleosome movement by CHRAC and ISWI without disruption or trans-displacement of the histone octamer. *Cell*, **97**, 843-852.

Langst, G., Schatz, T., Langowski, J. and Grummt, I. (1997) Structural analysis of mouse rDNA: coincidence between nuclease hypersensitive sites, DNA curvature and regulatory elements in the intergenic spacer. *Nucleic Acids Res*, **25**, 511-517.

Lansbergen, G., Grigoriev, I., Mimori-Kiyosue, Y., Ohtsuka, T., Higa, S., Kitajima, I., Demmers, J., Galjart, N., Houtsmuller, A.B., Grosveld, F. and Akhmanova, A. (2006) CLASPs attach microtubule plus ends to the cell cortex through a complex with LL5beta. *Dev Cell*, **11**, 21-32.

Lee, G.R., Fields, P.E., Griffin, T.J. and Flavell, R.A. (2003) Regulation of the Th2 cytokine locus by a locus control region. *Immunity*, **19**, 145-153.

Lee, J., Inoue, K., Ono, R., Ogonuki, N., Kohda, T., Kaneko-Ishino, T., Ogura, A. and Ishino, F. (2002) Erasing genomic imprinting memory in

mouse clone embryos produced from day 11.5 primordial germ cells. *Development*, **129**, 1807-1817.
Lee, J.H., Cook, J.R., Yang, Z.H., Mirochnitchenko, O., Gunderson, S.I., Felix, A.M., Herth, N., Hoffmann, R. and Pestka, S. (2005) PRMT7, a new protein arginine methyltransferase that synthesizes symmetric dimethylarginine. *J Biol Chem*, **280**, 3656-3664.
Lee, J.S., Shukla, A., Schneider, J., Swanson, S.K., Washburn, M.P., Florens, L., Bhaumik, S.R. and Shilatifard, A. (2007) Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. *Cell*, **131**, 1084-1096.
Lee, J.T. (2005) Regulation of X-chromosome counting by Tsix and Xite sequences. *Science*, **309**, 768-771.
Lee, J.T., Davidow, L.S. and Warshawsky, D. (1999) Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat Genet*, **21**, 400-404.
Lee, J.T. and Lu, N. (1999) Targeted mutagenesis of Tsix leads to nonrandom X inactivation. *Cell*, **99**, 47-57.
Lee, P.P., Fitzpatrick, D.R., Beard, C., Jessup, H.K., Lehar, S., Makar, K.W., Perez-Melgosa, M., Sweetser, M.T., Schlissel, M.S., Nguyen, S., Cherry, S.R., Tsai, J.H., Tucker, S.M., Weaver, W.M., Kelso, A., Jaenisch, R. and Wilson, C.B. (2001) A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival. *Immunity*, **15**, 763-774.
Lehmann, M. (2004) Anything else but GAGA: a nonhistone protein complex reshapes chromatin structure. *Trends Genet*, **20**, 15-22.
Lei, E.P. and Corces, V.G. (2006) RNA interference machinery influences the nuclear organization of a chromatin insulator. *Nat Genet*, **38**, 936-941.
Leighton, P.A., Ingram, R.S., Eggenschwiler, J., Efstratiadis, A. and Tilghman, S.M. (1995a) Disruption of imprinting caused by deletion of the H19 gene region in mice. *Nature*, **375**, 34-39.
Leighton, P.A., Saam, J.R., Ingram, R.S., Stewart, C.L. and Tilghman, S.M. (1995b) An enhancer deletion affects both H19 and Igf2 expression. *Genes Dev*, **9**, 2079-2089.
Leung, A.K., Gerlich, D., Miller, G., Lyon, C., Lam, Y.W., Lleres, D., Daigle, N., Zomerdijk, J., Ellenberg, J. and Lamond, A.I. (2004) Quantitative kinetic analysis of nucleolar breakdown and reassembly during mitosis in live human cells. *J Cell Biol*, **166**, 787-800.
Lewis, A. and Murrell, A. (2004) Genomic imprinting: CTCF protects the boundaries. *Curr Biol*, **14**, R284-286.
Li, B., Carey, M. and Workman, J.L. (2007) The role of chromatin during transcription. *Cell*, **128**, 707-719.
Li, J., Langst, G. and Grummt, I. (2006) NoRC-dependent nucleosome positioning silences rRNA genes. *Embo J*, **25**, 5735-5741.
Li, J., Santoro, R., Koberna, K. and Grummt, I. (2005) The chromatin remodeling complex NoRC controls replication timing of rRNA genes. *Embo J*, **24**, 120-127.
Liakopoulos, D., Kusch, J., Grava, S., Vogel, J. and Barral, Y. (2003) Asymmetric loading of Kar9 onto spindle poles and microtubules ensures proper spindle alignment. *Cell*, **112**, 561-574.
Ling, J.Q., Li, T., Hu, J.F., Vu, T.H., Chen, H.L., Qiu, X.W., Cherry, A.M. and Hoffman, A.R. (2006) CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science*, **312**, 269-272.
Lo, S.J., Lee, C.C. and Lai, H.J. (2006) The nucleolus: reviewing oldies to have new understandings. *Cell Res*, **16**, 530-538.
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.
Lomvardas, S., Barnea, G., Pisapia, D.J., Mendelsohn, M., Kirkland, J. and Axel, R. (2006) Interchromosomal interactions and olfactory receptor choice. *Cell*, **126**, 403-413.
Long, B.H., Huang, C.Y. and Pogo, A.O. (1979) Isolation and characterization of the nuclear matrix in Friend erythroleukemia cells: chromatin and hnRNA interactions with the nuclear matrix. *Cell*, **18**, 1079-1090.
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., Pera, R.R., Schneider, D.T., Summersgill, B., Shipley, J., McIntyre, A., van der Spek, P., Schoenmakers, E. and Oosterhuis, J.W. (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.
Lopes, S., Lewis, A., Hajkova, P., Dean, W., Oswald, J., Forne, T., Murrell, A., Constancia, M., Bartolomei, M., Walter, J. and Reik, W. (2003) Epigenetic modifications in an imprinting cluster are controlled by a hierarchy of DMRs suggesting long-range chromatin interactions. *Hum Mol Genet*, **12**, 295-305.
Lorch, Y., Zhang, M. and Kornberg, R.D. (2001) RSC unravels the nucleosome. *Mol Cell*, **7**, 89-95.
Loukinov, D.I., Pugacheva, E., Vatolin, S., Pack, S.D., Moon, H., Chernukhin, I., Mannan, P., Larsson, E., Kanduri, C., Vostrov, A.A., Cui, H., Niemitz, E.L., Rasko, J.E., Docquier, F.M., Kistler, M., Breen, J.J., Zhuang, Z., Quitschke, W.W., Renkawitz, R., Klenova, E.M., Feinberg, A.P., Ohlsson, R., Morse, H.C., 3rd and Lobanenkov, V.V. (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.
Love, J.J., Li, X., Case, D.A., Giese, K., Grosschedl, R. and Wright, P.E. (1995) Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature*, **376**, 791-795.
Lupo, R., Breiling, A., Bianchi, M.E. and Orlando, V. (2001) Drosophila chromosome condensation proteins Topoisomerase II and Barren colocalize with Polycomb and maintain Fab-7 PRE silencing. *Mol Cell*, **7**, 127-136.
Lutz, M., Baniahmad, A. and Renkawitz, R. (2000a) Modulation of thyroid hormone receptor silencing function by co-repressors and a synergizing transcription factor. *Biochem Soc Trans*, **28**, 386-389.
Lutz, M., Burke, L.J., Barreto, G., Goeman, F., Greb, H., Arnold, R., Schultheiss, H., Brehm, A., Kouzarides, T., Lobanenkov, V. and Renkawitz, R. (2000b) Transcriptional repression by the insulator protein CTCF involves histone deacetylases. *Nucleic Acids Res*, **28**, 1707-1713.
Lutz, M., Burke, L.J., LeFevre, P., Myers, F.A., Thorne, A.W., Crane-Robinson, C., Bonifer, C., Filippova, G.N., Lobanenkov, V. and Renkawitz, R. (2003) Thyroid hormone-regulated enhancer blocking: cooperation of CTCF and thyroid hormone receptor. *Embo J*, **22**, 1579-1587.
Mackay, J.P., Sunde, M., Lowry, J.A., Crossley, M. and Matthews, J.M. (2007) Protein interactions: is seeing believing? *Trends Biochem Sci*, **32**, 530-531.
Magdinier, F., Yusufzai, T.M. and Felsenfeld, G. (2004) Both CTCF-dependent and -independent insulators are found between the mouse T cell receptor alpha and Dad1 genes. *J Biol Chem*, **279**, 25381-25389.
Mahy, N.L., Perry, P.E. and Bickmore, W.A. (2002) Gene density and transcription influence the localization of chromatin outside of

chromosome territories detectable by FISH. *J Cell Biol*, **159**, 753-763.
Mais, C., Wright, J.E., Prieto, J.L., Raggett, S.L. and McStay, B. (2005) UBF-binding site arrays form pseudo-NORs and sequester the RNA polymerase I transcription machinery. *Genes Dev*, **19**, 50-64.
Mann, B., Gelos, M., Siedow, A., Hanski, M.L., Gratchev, A., Ilyas, M., Bodmer, W.F., Moyer, M.P., Riecken, E.O., Buhr, H.J. and Hanski, C. (1999) Target genes of beta-catenin-T cell-factor/lymphoid-enhancer-factor signaling in human colorectal carcinomas. *Proc Natl Acad Sci U S A*, **96**, 1603-1608.
Martin, C.H., Mayeda, C.A., Davis, C.A., Ericsson, C.L., Knafels, J.D., Mathog, D.R., Celniker, S.E., Lewis, E.B. and Palazzolo, M.J. (1995) Complete sequence of the bithorax complex of Drosophila. *Proc Natl Acad Sci U S A*, **92**, 8398-8402.
Martin, D.I. (2001) Transcriptional enhancers--on/off gene regulation as an adaptation to silencing in higher eukaryotic nuclei. *Trends Genet*, **17**, 444-448.
Mayer, C., Schmitz, K.M., Li, J., Grummt, I. and Santoro, R. (2006) Intergenic transcripts regulate the epigenetic state of rRNA genes. *Mol Cell*, **22**, 351-361.
McKinnell, I.W., Ishibashi, J., Le Grand, F., Punch, V.G., Addicks, G.C., Greenblatt, J.F., Dilworth, F.J. and Rudnicki, M.A. (2008) Pax7 activates myogenic genes by recruitment of a histone methyltransferase complex. *Nat Cell Biol*, **10**, 77-84.
Meder, V.S., Boeglin, M., de Murcia, G. and Schreiber, V. (2005) PARP-1 and PARP-2 interact with nucleophosmin/B23 and accumulate in transcriptionally active nucleoli. *J Cell Sci*, **118**, 211-222.
Meier, N., Krpic, S., Rodriguez, P., Strouboulis, J., Monti, M., Krijgsveld, J., Gering, M., Patient, R., Hostert, A. and Grosveld, F. (2006) Novel binding partners of Ldb1 are required for haematopoietic development. *Development*, **133**, 4913-4923.
Mekhail, K., Rivero-Lopez, L., Al-Masri, A., Brandon, C., Khacho, M. and Lee, S. (2007) Identification of a common subnuclear localization signal. *Mol Biol Cell*, **18**, 3966-3977.
Meneghini, M.D., Wu, M. and Madhani, H.D. (2003) Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. *Cell*, **112**, 725-736.
Mermoud, J.E., Cohen, P.T. and Lamond, A.I. (1994) Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. *Embo J*, **13**, 5679-5688.
Meshorer, E. and Misteli, T. (2006) Chromatin in pluripotent embryonic stem cells and differentiation. *Nat Rev Mol Cell Biol*, **7**, 540-546.
Mika, S. and Rost, B. (2005) NMPdb: Database of Nuclear Matrix Proteins. *Nucleic Acids Res*, **33**, D160-163.
Miller, G., Panov, K.I., Friedrich, J.K., Trinkle-Mulcahy, L., Lamond, A.I. and Zomerdijk, J.C. (2001) hRRN3 is essential in the SL1-mediated recruitment of RNA Polymerase I to rRNA gene promoters. *Embo J*, **20**, 1373-1382.
Mito, Y., Henikoff, J.G. and Henikoff, S. (2007) Histone replacement marks the boundaries of cis-regulatory domains. *Science*, **315**, 1408-1411.
Miyoshi, H., Blomer, U., Takahashi, M., Gage, F.H. and Verma, I.M. (1998) Development of a self-inactivating lentivirus vector. *J Virol*, **72**, 8150-8157.
Mohan, M., Bartkuhn, M., Herold, M., Philippen, A., Heinl, N., Bardenhagen, I., Leers, J., White, R.A., Renkawitz-Pohl, R., Saumweber, H. and Renkawitz, R. (2007) The Drosophila insulator proteins CTCF and CP190 link enhancer blocking to body patterning. *Embo J*, **26**, 4203-4214.
Molina, T.J., Perrot, J.Y., Penninger, J., Ramos, A., Audouin, J., Briand, P., Mak, T.W. and Diebold, J. (1998) Differential requirement for p56lck in fetal and adult thymopoiesis. *J Immunol*, **160**, 3828-3834.
Moon, H., Filippova, G., Loukinov, D., Pugacheva, E., Chen, Q., Smith, S.T., Munhall, A., Grewe, B., Bartkuhn, M., Arnold, R., Burke, L.J., Renkawitz-Pohl, R., Ohlsson, R., Zhou, J., Renkawitz, R. and Lobanenkov, V. (2005) CTCF is conserved from Drosophila to humans and confers enhancer blocking of the Fab-8 insulator. *EMBO Rep*, **6**, 165-170.
Moore, W., Zhang, C. and Clarke, P.R. (2002) Targeting of RCC1 to chromosomes is required for proper mitotic spindle assembly in human cells. *Curr Biol*, **12**, 1442-1447.
Mueller, C.L. and Jaehning, J.A. (2002) Ctr9, Rtf1, and Leo1 are components of the Paf1/RNA polymerase II complex. *Mol Cell Biol*, **22**, 1971-1980.
Mueller, C.L., Porter, S.E., Hoffman, M.G. and Jaehning, J.A. (2004) The Paf1 complex has functions independent of actively transcribing RNA polymerase II. *Mol Cell*, **14**, 447-456.
Mukhopadhyay, R., Yu, W., Whitehead, J., Xu, J., Lezcano, M., Pack, S., Kanduri, C., Kanduri, M., Ginjala, V., Vostrov, A., Quitschke, W., Chernukhin, I., Klenova, E., Lobanenkov, V. and Ohlsson, R. (2004) The binding sites for the chromatin insulator protein CTCF map to DNA methylation-free domains genome-wide. *Genome Res*, **14**, 1594-1602.
Muravyova, E., Golovnin, A., Gracheva, E., Parshikov, A., Belenkaya, T., Pirrotta, V. and Georgiev, P. (2001) Loss of insulator activity by paired Su(Hw) chromatin insulators. *Science*, **291**, 495-498.
Murrell, A., Heeson, S., Bowden, L., Constancia, M., Dean, W., Kelsey, G. and Reik, W. (2001) An intragenic methylated region in the imprinted Igf2 gene augments transcription. *EMBO Rep*, **2**, 1101-1106.
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.
Myant, K. and Stancheva, I. (2008) LSH cooperates with DNA methyltransferases to repress transcription. *Mol Cell Biol*, **28**, 215-226.
Nabirochkin, S., Ossokina, M. and Heidmann, T. (1998) A nuclear matrix/scaffold attachment region co-localizes with the gypsy retrotransposon insulator sequence. *J Biol Chem*, **273**, 2473-2479.
Nakayama, T., Nishioka, K., Dong, Y.X., Shimojima, T. and Hirose, S. (2007) Drosophila GAGA factor directs histone H3.3 replacement that prevents the heterochromatin spreading. *Genes Dev*, **21**, 552-561.
Naldini, L., Blomer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F.H., Verma, I.M. and Trono, D. (1996) In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science*, **272**, 263-267.
Narlikar, G.J., Fan, H.Y. and Kingston, R.E. (2002) Cooperation between complexes that regulate chromatin structure and transcription. *Cell*, **108**, 475-487.
Navarro, P., Page, D.R., Avner, P. and Rougeulle, C. (2006) Tsix-mediated epigenetic switch of a CTCF-flanked region of the Xist promoter determines the Xist transcription program. *Genes Dev*, **20**, 2787-2792.
Newton, K., Petfalski, E., Tollervey, D. and Caceres, J.F. (2003) Fibrillarin is essential for early development and required for accumulation

of an intron-encoded small nucleolar RNA in the mouse. *Mol Cell Biol*, **23**, 8519-8527.
Ng, R.K. and Gurdon, J.B. (2008) Epigenetic memory of an active gene state depends on histone H3.3 incorporation into chromatin in the absence of transcription. *Nat Cell Biol*, **10**, 102-109.
Nielsen, A.L., Oulad-Abdelghani, M., Ortiz, J.A., Remboutsika, E., Chambon, P. and Losson, R. (2001) Heterochromatin formation in mammalian cells: interaction between histones and HP1 proteins. *Mol Cell*, **7**, 729-739.
Noma, K., Allis, C.D. and Grewal, S.I. (2001) Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science*, **293**, 1150-1155.
O'Brien, L.L. and Wiese, C. (2006) TPX2 is required for postmitotic nuclear assembly in cell-free Xenopus laevis egg extracts. *J Cell Biol*, **173**, 685-694.
O'Mahony, D.J., Smith, S.D., Xie, W. and Rothblum, L.I. (1992) Analysis of the phosphorylation, DNA-binding and dimerization properties of the RNA polymerase I transcription factors UBF1 and UBF2. *Nucleic Acids Res*, **20**, 1301-1308.
O'Sullivan, A.C., Sullivan, G.J. and McStay, B. (2002) UBF binding in vivo is not restricted to regulatory sequences within the vertebrate ribosomal DNA repeat. *Mol Cell Biol*, **22**, 657-668.
Oh, R., Ho, R., Mar, L., Gertsenstein, M., Paderova, J., Hsien, J., Squire, J.A., Higgins, M.J., Nagy, A. and Lefebvre, L. (2008) Epigenetic and phenotypic consequences of a truncation disrupting the imprinted domain on distal mouse chromosome 7. *Mol Cell Biol*, **28**, 1092-1103.
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.
Okamura, R.M., Sigvardsson, M., Galceran, J., Verbeek, S., Clevers, H. and Grosschedl, R. (1998) Redundant regulation of T cell differentiation and TCRalpha gene expression by the transcription factors LEF-1 and TCF-1. *Immunity*, **8**, 11-20.
Okano, M., Bell, D.W., Haber, D.A. and Li, E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, **99**, 247-257.
Oki, M., Valenzuela, L., Chiba, T., Ito, T. and Kamakaka, R.T. (2004) Barrier proteins remodel and modify chromatin to restrict silenced domains. *Mol Cell Biol*, **24**, 1956-1967.
Old, L.J. (2001) Cancer/testis (CT) antigens - a new link between gametogenesis and cancer. *Cancer Immun*, **1**, 1.
Orphanides, G., Wu, W.H., Lane, W.S., Hampsey, M. and Reinberg, D. (1999) The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. *Nature*, **400**, 284-288.
Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W. and Fraser, P. (2004) Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet*, **36**, 1065-1071.
Osoegawa, K., Tateno, M., Woon, P.Y., Frengen, E., Mammoser, A.G., Catanese, J.J., Hayashizaki, Y. and de Jong, P.J. (2000) Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome Res*, **10**, 116-128.
Paalman, M.H., Henderson, S.L. and Sollner-Webb, B. (1995) Stimulation of the mouse rRNA gene promoter by a distal spacer promoter. *Mol Cell Biol*, **15**, 4648-4656.
Pai, S.Y., Truitt, M.L., Ting, C.N., Leiden, J.M., Glimcher, L.H. and Ho, I.C. (2003) Critical roles for transcription factor GATA-3 in thymocyte development. *Immunity*, **19**, 863-875.
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.
Panov, K.I., Friedrich, J.K., Russell, J. and Zomerdijk, J.C. (2006) UBF activates RNA polymerase I transcription by stimulating promoter escape. *Embo J*, **25**, 3310-3322.
Panov, K.I., Friedrich, J.K. and Zomerdijk, J.C. (2001) A step subsequent to preinitiation complex assembly at the ribosomal RNA gene promoter is rate limiting for human RNA polymerase I-dependent transcription. *Mol Cell Biol*, **21**, 2641-2649.
Papoulas, O., Daubresse, G., Armstrong, J.A., Jin, J., Scott, M.P. and Tamkun, J.W. (2001) The HMG-domain protein BAP111 is important for the function of the BRM chromatin-remodeling complex in vivo. *Proc Natl Acad Sci U S A*, **98**, 5728-5733.
Parada, L. and Misteli, T. (2002) Chromosome positioning in the interphase nucleus. *Trends Cell Biol*, **12**, 425-432.
Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., Cobb, B.S., Yokomori, K., Dillon, N., Aragon, L., Fisher, A.G. and Merkenschlager, M. (2008) Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms. *Cell*.
Park, K.Y., Sellars, E.A., Grinberg, A., Huang, S.P. and Pfeifer, K. (2004) The H19 differentially methylated region marks the parental origin of a heterologous locus without gametic DNA methylation. *Mol Cell Biol*, **24**, 3588-3595.
Pathak, R.U., Rangaraj, N., Kallappagoudar, S., Mishra, K. and Mishra, R.K. (2007) Boundary element-associated factor 32B connects chromatin domains to the nuclear matrix. *Mol Cell Biol*, **27**, 4796-4806.
Paulsen, M., Takada, S., Youngson, N.A., Benchaib, M., Charlier, C., Segers, K., Georges, M. and Ferguson-Smith, A.C. (2001) Comparative sequence analysis of the imprinted Dlk1-Gtl2 locus in three mammalian species reveals highly conserved genomic elements and refines comparison with the Igf2-H19 region. *Genome Res*, **11**, 2085-2094.
Pavri, R., Zhu, B., Li, G., Trojer, P., Mandal, S., Shilatifard, A. and Reinberg, D. (2006) Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. *Cell*, **125**, 703-717.
Pederson, T. (2000) Half a century of "the nuclear matrix". *Mol Biol Cell*, **11**, 799-805.
Pedone, P.V., Pikaart, M.J., Cerrato, F., Vernucci, M., Ungaro, P., Bruni, C.B. and Riccio, A. (1999) Role of histone acetylation and DNA methylation in the maintenance of the imprinted expression of the H19 and Igf2 genes. *FEBS Lett*, **458**, 45-50.
Pemov, A., Bavykin, S. and Hamlin, J.L. (1998) Attachment to the nuclear matrix mediates specific alterations in chromatin structure. *Proc Natl Acad Sci U S A*, **95**, 14757-14762.
Penheiter, K.L., Washburn, T.M., Porter, S.E., Hoffman, M.G. and Jaehning, J.A. (2005) A posttranscriptional role for the yeast Paf1-RNA polymerase II complex is revealed by identification of primary targets. *Mol Cell*, **20**, 213-223.
Percipalle, P., Fomproix, N., Cavellan, E., Voit, R., Reimer, G., Kruger, T., Thyberg, J., Scheer, U., Grummt, I. and Farrants, A.K. (2006) The chromatin remodelling complex WSTF-SNF2h interacts with nuclear myosin 1 and has a role in RNA polymerase I transcription. *EMBO Rep*, **7**, 525-530.
Peterson, C.L. and Herskowitz, I. (1992) Characterization of the yeast SWI1, SWI2, and SWI3 genes, which encode a global activator of

transcription. *Cell*, **68**, 573-583.
Peyroche, G., Milkereit, P., Bischler, N., Tschochner, H., Schultz, P., Sentenac, A., Carles, C. and Riva, M. (2000) The recruitment of RNA polymerase I on rDNA is mediated by the interaction of the A43 subunit with Rrn3. *Embo J*, **19**, 5473-5482.
Phair, R.D., Scaffidi, P., Elbi, C., Vecerova, J., Dey, A., Ozato, K., Brown, D.T., Hager, G., Bustin, M. and Misteli, T. (2004) Global nature of dynamic protein-chromatin interactions in vivo: three-dimensional genome scanning and dynamic interaction networks of chromatin proteins. *Mol Cell Biol*, **24**, 6393-6402.
Poortinga, G., Hannan, K.M., Snelling, H., Walkley, C.R., Jenkins, A., Sharkey, K., Wall, M., Brandenburger, Y., Palatsides, M., Pearson, R.B., McArthur, G.A. and Hannan, R.D. (2004) MAD1 and c-MYC regulate UBF and rDNA transcription during granulocyte differentiation. *Embo J*, **23**, 3325-3335.
Postic, C. and Magnuson, M.A. (2000) DNA excision in liver by an albumin-Cre transgene occurs progressively with age. *Genesis*, **26**, 149-150.
Prawitt, D., Enklaar, T., Gartner-Rupprecht, B., Spangenberg, C., Oswald, M., Lausch, E., Schmidtke, P., Reutzel, D., Fees, S., Lucito, R., Korzon, M., Brozek, I., Limon, J., Housman, D.E., Pelletier, J. and Zabel, B. (2005) Microdeletion of target sites for insulator protein CTCF in a chromosome 11p15 imprinting center in Beckwith-Wiedemann syndrome and Wilms' tumor. *Proc Natl Acad Sci U S A*, **102**, 4085-4090.
Prieto, J.L. and McStay, B. (2007) Recruitment of factors linking transcription and processing of pre-rRNA to NOR chromatin is UBF-dependent and occurs independent of transcription in human cells. *Genes Dev*, **21**, 2041-2054.
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., Dawid, I.B. and Lobanenkov, V.V. (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.
Putnam, C.D., Copenhaver, G.P., Denton, M.L. and Pikaard, C.S. (1994) The RNA polymerase I transactivator upstream binding factor requires its dimerization domain and high-mobility-group (HMG) box 1 to bend, wrap, and positively supercoil enhancer DNA. *Mol Cell Biol*, **14**, 6476-6488.
Putnam, C.D. and Pikaard, C.S. (1992) Cooperative binding of the Xenopus RNA polymerase I transcription factor xUBF to repetitive ribosomal gene enhancers. *Mol Cell Biol*, **12**, 4970-4980.
Qi, C.F., Martensson, A., Mattioli, M., Dalla-Favera, R., Lobanenkov, V.V. and Morse, H.C., 3rd. (2003) CTCF functions as a critical regulator of cell-cycle arrest and death after ligation of the B cell receptor on immature B cells. *Proc Natl Acad Sci U S A*, **100**, 633-638.
Quinn, J., Fyrberg, A.M., Ganster, R.W., Schmidt, M.C. and Peterson, C.L. (1996) DNA-binding properties of the yeast SWI/SNF complex. *Nature*, **379**, 844-847.
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.
Ragoczy, T., Telling, A., Sawado, T., Groudine, M. and Kosak, S.T. (2003) A genetic analysis of chromosome territory looping: diverse roles for distal regulatory elements. *Chromosome Res*, **11**, 513-525.
Raisner, R.M. and Madhani, H.D. (2006) Patterning chromatin: form and function for H2A.Z variant nucleosomes. *Curr Opin Genet Dev*, **16**, 119-124.
Rand, E., Ben-Porath, I., Keshet, I. and Cedar, H. (2004) CTCF elements direct allele-specific undermethylation at the imprinted H19 locus. *Curr Biol*, **14**, 1007-1012.
Rangasamy, D., Greaves, I. and Tremethick, D.J. (2004) RNA interference demonstrates a novel role for H2A.Z in chromosome segregation. Nat Struct Mol Biol, 11, 650-655.
Rasko, J.E., Klenova, E.M., Leon, J., Filippova, G.N., Loukinov, D.I., Vatolin, S., Robinson, A.F., Hu, Y.J., Ulmer, J., Ward, M.D., Pugacheva, E.M., Neiman, P.E., Morse, H.C., 3rd, Collins, S.J. and Lobanenkov, V.V. (2001) Cell growth inhibition by the multifunctional multivalent zinc-finger factor CTCF. *Cancer Res*, **61**, 6002-6007.
Recht, J. and Osley, M.A. (1999) Mutations in both the structured domain and N-terminus of histone H2B bypass the requirement for Swi-Snf in yeast. *Embo J*, **18**, 229-240.
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.
Redon, C., Pilch, D., Rogakou, E., Sedelnikova, O., Newrock, K. and Bonner, W. (2002) Histone H2A variants H2AX and H2AZ. *Curr Opin Genet Dev*, **12**, 162-169.
Reese, J.C. (2003) Basal transcription factors. *Curr Opin Genet Dev*, **13**, 114-118.
Reeves, R. and Nissen, M.S. (1990) The A.T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. *J Biol Chem*, **265**, 8573-8582.
Reik, W. and Lewis, A. (2005) Co-evolution of X-chromosome inactivation and imprinting in mammals. *Nat Rev Genet*, **6**, 403-410.
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.
Risinger, J.I., Chandramouli, G.V., Maxwell, G.L., Custer, M., Pack, S., Loukinov, D., Aprelikova, O., Litzi, T., Schrump, D.S., Murphy, S.K., Berchuck, A., Lobanenkov, V. and Barrett, J.C. (2007) Global expression analysis of cancer/testis genes in uterine cancers reveals a high incidence of BORIS expression. *Clin Cancer Res*, **13**, 1713-1719.
Risinger, J.I., Maxwell, G.L., Chandramouli, G.V., Aprelikova, O., Litzi, T., Umar, A., Berchuck, A. and Barrett, J.C. (2005) Gene expression profiling of microsatellite unstable and microsatellite stable endometrial cancers indicates distinct pathways of aberrant signaling. *Cancer Res*, **65**, 5031-5037.
Robinett, C.C., O'Connor, A. and Dunaway, M. (1997) The repeat organizer, a specialized insulator element within the intergenic spacer of

the Xenopus rRNA genes. *Mol Cell Biol*, **17**, 2866-2875.
Rodriguez, P., Bonte, E., Krijgsveld, J., Kolodziej, K.E., Guyot, B., Heck, A.J., Vyas, P., de Boer, E., Grosveld, F. and Strouboulis, J. (2005) GATA-1 forms distinct activating and repressive complexes in erythroid cells. *Embo J*, **24**, 2354-2366.
Rosa, A.L., Wu, Y.Q., Kwabi-Addo, B., Coveler, K.J., Reid Sutton, V. and Shaffer, L.G. (2005) Allele-specific methylation of a functional CTCF binding site upstream of MEG3 in the human imprinted domain of 14q32. *Chromosome Res*, **13**, 809-818.
Rothenberg, E.V. and Taghon, T. (2005) Molecular genetics of T cell development. *Annu Rev Immunol*, **23**, 601-649.
Roussel, P., Andre, C., Comai, L. and Hernandez-Verdun, D. (1996) The rDNA transcription machinery is assembled during mitosis in active NORs and absent in inactive NORs. *J Cell Biol*, **133**, 235-246.
Rozenblatt-Rosen, O., Hughes, C.M., Nannepaga, S.J., Shanmugam, K.S., Copeland, T.D., Guszczynski, T., Resau, J.H. and Meyerson, M. (2005) The parafibromin tumor suppressor protein is part of a human Paf1 complex. *Mol Cell Biol*, **25**, 612-620.
Ruhl, D.D., Jin, J., Cai, Y., Swanson, S., Florens, L., Washburn, M.P., Conaway, R.C., Conaway, J.W. and Chrivia, J.C. (2006) Purification of a human SRCAP complex that remodels chromatin by incorporating the histone variant H2A.Z into nucleosomes. *Biochemistry*, **45**, 5671-5677.
Russell, J. and Zomerdijk, J.C. (2005) RNA-polymerase-I-directed rDNA transcription, life and works. *Trends Biochem Sci*, **30**, 87-96.
Sabbattini, P., Canzonetta, C., Sjoberg, M., Nikic, S., Georgiou, A., Kemball-Cook, G., Auner, H.W. and Dillon, N. (2007) A novel role for the Aurora B kinase in epigenetic marking of silent chromatin in differentiated postmitotic cells. *Embo J*, **26**, 4657-4669.
Saha, A., Wittmeyer, J. and Cairns, B.R. (2006) Chromatin remodelling: the industrial revolution of DNA around histones. *Nat Rev Mol Cell Biol*, **7**, 437-447.
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.
Sakai, K. and Miyazaki, J. (1997) A transgenic mouse line that retains Cre recombinase activity in mature oocytes irrespective of the cre transgene transmission. *Biochem Biophys Res Commun*, **237**, 318-324.
Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y. and Hume, D.A. (2007) Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat Rev Genet*, **8**, 424-436.
Sandmeier, J.J., French, S., Osheim, Y., Cheung, W.L., Gallo, C.M., Beyer, A.L. and Smith, J.S. (2002) RPD3 is required for the inactivation of yeast ribosomal DNA genes in stationary phase. *Embo J*, **21**, 4959-4968.
Santoro, R. and Grummt, I. (2001) Molecular mechanisms mediating methylation-dependent silencing of ribosomal gene transcription. *Mol Cell*, **8**, 719-725.
Santoro, R. and Grummt, I. (2005) Epigenetic mechanism of rRNA gene silencing: temporal order of NoRC-mediated histone modification, chromatin remodeling, and DNA methylation. *Mol Cell Biol*, **25**, 2539-2546.
Santoro, R., Li, J. and Grummt, I. (2002) The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription. *Nat Genet*, **32**, 393-396.
Sawado, T., Halow, J., Bender, M.A. and Groudine, M. (2003) The beta -globin locus control region (LCR) functions primarily by enhancing the transition from transcription initiation to elongation. *Genes Dev*, **17**, 1009-1018.
Schnapp, G., Schnapp, A., Rosenbauer, H. and Grummt, I. (1994) TIF-IC, a factor involved in both transcription initiation and elongation of RNA polymerase I. *Embo J*, **13**, 4028-4035.
Schneider, D.A., Michel, A., Sikes, M.L., Vu, L., Dodd, J.A., Salgia, S., Osheim, Y.N., Beyer, A.L. and Nomura, M. (2007) Transcription elongation by RNA polymerase I is linked to efficient rRNA processing and ribosome assembly. *Mol Cell*, **26**, 217-229.
Schnell, S., Demolliere, C., van den Berk, P. and Jacobs, H. (2006) Gimap4 accelerates T-cell death. *Blood*, **108**, 591-599.
Schoenfelder, S., Smits, G., Fraser, P., Reik, W. and Paro, R. (2007) Non-coding transcripts in the H19 imprinting control region mediate gene silencing in transgenic Drosophila. *EMBO Rep*, **8**, 1068-1073.
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.
Schubeler, D., Groudine, M. and Bender, M.A. (2001) The murine beta-globin locus control region regulates the rate of transcription but not the hyperacetylation of histones at the active genes. *Proc Natl Acad Sci U S A*, **98**, 11432-11437.
Schuetz, A., Allali-Hassani, A., Martin, F., Loppnau, P., Vedadi, M., Bochkarev, A., Plotnikov, A.N., Arrowsmith, C.H. and Min, J. (2006) Structural basis for molecular recognition and presentation of histone H3 by WDR5. *Embo J*, **25**, 4245-4252.
Schweinsberg, S., Hagstrom, K., Gohl, D., Schedl, P., Kumar, R.P., Mishra, R. and Karch, F. (2004) The enhancer-blocking activity of the Fab-7 boundary from the Drosophila bithorax complex requires GAGA-factor-binding sites. *Genetics*, **168**, 1371-1384.
Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., Tajima, S., Mitsuya, K., Okano, M. and Koseki, H. (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature*, **450**, 908-912.
Shaw, P. and Doonan, J. (2005) The nucleolus. Playing by different rules? *Cell Cycle*, **4**, 102-105.
Sheldon, K.E., Mauger, D.M. and Arndt, K.M. (2005) A Requirement for the Saccharomyces cerevisiae Paf1 complex in snoRNA 3' end formation. *Mol Cell*, **20**, 225-236.
Shen, X., Xiao, H., Ranallo, R., Wu, W.H. and Wu, C. (2003) Modulation of ATP-dependent chromatin-remodeling complexes by inositol polyphosphates. *Science*, **299**, 112-114.
Shin, J.Y., Fitzpatrick, G.V. and Higgins, M.J. (2008) Two distinct mechanisms of silencing by the KvDMR1 imprinting control region. *Embo J*, **27**, 168-178.
Shpargel, K.B., Ospina, J.K., Tucker, K.E., Matera, A.G. and Hebert, M.D. (2003) Control of Cajal body number is mediated by the coilin C-terminus. *J Cell Sci*, **116**, 303-312.
Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B. and de Laat, W. (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet*, **38**, 1348-1354.
Simonis, M., Kooren, J. and de Laat, W. (2007) An evaluation of 3C-based methods to capture DNA interactions. *Nat Methods*, **4**, 895-901.
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.

Sirri, V., Roussel, P. and Hernandez-Verdun, D. (1999) The mitotically phosphorylated form of the transcription termination factor TTF-1 is associated with the repressed rDNA transcription machinery. *J Cell Sci*, **112 (Pt 19)**, 3259-3268.
Sleutels, F. and Barlow, D.P. (2001) Investigation of elements sufficient to imprint the mouse Air promoter. *Mol Cell Biol*, **21**, 5008-5017.
Smith, M.M. (2002) Centromeres and variant histones: what, where, when and why? *Curr Opin Cell Biol*, **14**, 279-285.
Soullier, S., Jay, P., Poulat, F., Vanacker, J.M., Berta, P. and Laudet, V. (1999) Diversification pattern of the HMG and SOX family members during evolution. *J Mol Evol*, **48**, 517-527.
Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R. and Flavell, R.A. (2005) Interchromosomal associations between alternatively expressed loci. *Nature*, **435**, 637-645.
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.
Sportoletti, P., Grisendi, S., Majid, S.M., Cheng, K., Clohessy, J.G., Viale, A., Teruya-Feldstein, J. and Pandolfi, P.P. (2008) Npm1 is a haploinsufficient suppressor of myeloid and lymphoid malignancies in the mouse. *Blood*.
Stedman, W., Kang, H., Lin, S., Kissil, J.L., Bartolomei, M.S. and Lieberman, P.M. (2008) Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. *Embo J*.
Stefanovsky, V.Y., Pelletier, G., Bazett-Jones, D.P., Crane-Robinson, C. and Moss, T. (2001) DNA looping in the RNA polymerase I enhancesome is the result of non-cooperative in-phase bending by two UBF molecules. *Nucleic Acids Res*, **29**, 3241-3247.
Stewart, C.L., Roux, K.J. and Burke, B. (2007) Blurring the boundary: the nuclear envelope extends its reach. *Science*, **318**, 1408-1412.
Stock, J.K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A.G. and Pombo, A. (2007) Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol*, **9**, 1428-1435.
Strohner, R., Nemeth, A., Jansa, P., Hofmann-Rohrer, U., Santoro, R., Langst, G. and Grummt, I. (2001) NoRC--a novel member of mammalian ISWI-containing chromatin remodeling machines. *Embo J*, **20**, 4892-4900.
Sullivan, G.J., Bridger, J.M., Cuthbert, A.P., Newbold, R.F., Bickmore, W.A. and McStay, B. (2001) Human acrocentric chromosomes with transcriptionally silent nucleolar organizer regions associate with nucleoli. *Embo J*, **20**, 2867-2874.
Sullivan, G.J. and McStay, B. (1998) Dimerization and HMG box domains 1-3 present in Xenopus UBF are sufficient for its role in transcriptional enhancement. *Nucleic Acids Res*, **26**, 3555-3561.
Swigut, T. and Wysocka, J. (2007) H3K27 demethylases, at long last. *Cell*, **131**, 29-32.
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.
Szutorisz, H., Dillon, N. and Tora, L. (2005) The role of enhancers as centres for general transcription factor recruitment. *Trends Biochem Sci*, **30**, 593-599.
Takada, S., Paulsen, M., Tevendale, M., Tsai, C.E., Kelsey, G., Cattanach, B.M. and Ferguson-Smith, A.C. (2002) Epigenetic analysis of the Dlk1-Gtl2 imprinted domain on mouse chromosome 12: implications for imprinting control from comparison with Igf2-H19. *Hum Mol Genet*, **11**, 77-86.
Takagi, N. and Sasaki, M. (1975) Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. *Nature*, **256**, 640-642.
Takeuchi, T., Watanabe, Y., Takano-Shimizu, T. and Kondo, S. (2006) Roles of jumonji and jumonji family genes in chromatin regulation and development. *Dev Dyn*, **235**, 2449-2459.
Tetsu, O. and McCormick, F. (1999) Beta-catenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature*, **398**, 422-426.
Thorvaldsen, J.L., Fedoriw, A.M., Nguyen, S. and Bartolomei, M.S. (2006) Developmental profile of H19 differentially methylated domain (DMD) deletion alleles reveals multiple roles of the DMD in regulating allelic expression and DNA methylation at the imprinted H19/Igf2 locus. *Mol Cell Biol*, **26**, 1245-1258.
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.
Tong, J.K., Hassig, C.A., Schnitzler, G.R., Kingston, R.E. and Schreiber, S.L. (1998) Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature*, **395**, 917-921.
Torrano, V., Navascues, J., Docquier, F., Zhang, R., Burke, L.J., Chernukhin, I., Farrar, D., Leon, J., Berciano, M.T., Renkawitz, R., Klenova, E., Lafarga, M. and Delgado, M.D. (2006) Targeting of CTCF to the nucleolus inhibits nucleolar transcription through a poly(ADP-ribosyl)ation-dependent mechanism. *J Cell Sci*, **119**, 1746-1759.
Torres-Rosell, J., Machin, F., Farmer, S., Jarmuz, A., Eydmann, T., Dalgaard, J.Z. and Aragon, L. (2005) SMC5 and SMC6 genes are required for the segregation of repetitive chromosome regions. *Nat Cell Biol*, **7**, 412-419.
Tower, J., Henderson, S.L., Dougherty, K.M., Wejksnora, P.J. and Sollner-Webb, B. (1989) An RNA polymerase I promoter located in the CHO and mouse ribosomal DNA spacers: functional analysis and factor and sequence requirements. *Mol Cell Biol*, **9**, 1513-1525.
Travers, A.A. (2003) Priming the nucleosome: a role for HMGB proteins? *EMBO Rep*, **4**, 131-136.
Travis, A., Amsterdam, A., Belanger, C. and Grosschedl, R. (1991) LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor alpha enhancer function [corrected]. *Genes Dev*, **5**, 880-894.
Tsai, M.Y., Wang, S., Heidinger, J.M., Shumaker, D.K., Adam, S.A., Goldman, R.D. and Zheng, Y. (2006) A mitotic lamin B matrix induced by RanGTP required for spindle assembly. *Science*, **311**, 1887-1893.
Udvardy, A., Maine, E. and Schedl, P. (1985) The 87A7 chromomere. Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. *J Mol Biol*, **185**, 341-358.
Ueda, T., Abe, K., Miura, A., Yuzuriha, M., Zubair, M., Noguchi, M., Niwa, K., Kawase, Y., Kono, T., Matsuda, Y., Fujimoto, H., Shibata, H., Hayashizaki, Y. and Sasaki, H. (2000) The paternal methylation imprint of the mouse H19 locus is acquired in the gonocyte stage during foetal testis development. *Genes Cells*, **5**, 649-659.
van de Wetering, M., de Lau, W. and Clevers, H. (2002) WNT signaling and lymphocyte development. *Cell*, **109 Suppl**, S13-19.
van Genderen, C., Okamura, R.M., Farinas, I., Quo, R.G., Parslow, T.G., Bruhn, L. and Grosschedl, R. (1994) Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in LEF-1-deficient mice. *Genes Dev*, **8**, 2691-2703.
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., Risinger, J.I., Barrett, J.C. and Lobanenkov, V.V. (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.
Verbeek, S., Izon, D., Hofhuis, F., Robanus-Maandag, E., te Riele, H., van de Wetering, M., Oosterwegel, M., Wilson, A., MacDonald, H.R. and Clevers, H. (1995) An HMG-box-containing T-cell factor required for thymocyte differentiation. *Nature*, **374**, 70-74.
Verona, R.I., Thorvaldsen, J.L., Reese, K.J. and Bartolomei, M.S. (2008) The transcriptional status but not the imprinting control region determines allele-specific histone modifications at the imprinted H19 locus. *Mol Cell Biol*, **28**, 71-82.
Vidakovic, M., Grdovic, N., Quesada, P., Bode, J. and Poznanovic, G. (2004) Poly(ADP-ribose) polymerase-1: association with nuclear lamins in rodent liver cells. *J Cell Biochem*, **93**, 1155-1168.
Vignali, M., Hassan, A.H., Neely, K.E. and Workman, J.L. (2000) ATP-dependent chromatin-remodeling complexes. *Mol Cell Biol*, **20**, 1899-1910.
Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.M., Bollen, M., Esteller, M., Di Croce, L., de Launoit, Y. and Fuks, F. (2006) The Polycomb group protein EZH2 directly controls DNA methylation. *Nature*, **439**, 871-874.
Visintin, R. and Amon, A. (2000) The nucleolus: the magician's hat for cell cycle tricks. *Curr Opin Cell Biol*, **12**, 372-377.
Voit, R. and Grummt, I. (2001) Phosphorylation of UBF at serine 388 is required for interaction with RNA polymerase I and activation of rDNA transcription. *Proc Natl Acad Sci U S A*, **98**, 13631-13636.
Voit, R., Schafer, K. and Grummt, I. (1997) Mechanism of repression of RNA polymerase I transcription by the retinoblastoma protein. *Mol Cell Biol*, **17**, 4230-4237.
von Lindern, M., Deiner, E.M., Dolznig, H., Parren-Van Amelsvoort, M., Hayman, M.J., Mullner, E.W. and Beug, H. (2001) Leukemic transformation of normal murine erythroid progenitors: v- and c-ErbB act through signaling pathways activated by the EpoR and c-Kit in stress erythropoiesis. *Oncogene*, **20**, 3651-3664.
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.
Wai, A.W., Gillemans, N., Raguz-Bolognesi, S., Pruzina, S., Zafarana, G., Meijer, D., Philipsen, S. and Grosveld, F. (2003) HS5 of the human beta-globin locus control region: a developmental stage-specific border in erythroid cells. *Embo J*, **22**, 4489-4500.
Wallace, J.A. and Felsenfeld, G. (2007) We gather together: insulators and genome organization. *Curr Opin Genet Dev*, **17**, 400-407.
Wang, J., Zhang, J., Zheng, H., Li, J., Liu, D., Li, H., Samudrala, R., Yu, J. and Wong, G.K. (2004) Mouse transcriptome: neutral evolution of 'non-coding' complementary DNAs. *Nature*, **431**, 1 p following 757; discussion following 757.
Wang, W., Chi, T., Xue, Y., Zhou, S., Kuo, A. and Crabtree, G.R. (1998a) Architectural DNA binding by a high-mobility-group/kinesin-like subunit in mammalian SWI/SNF-related complexes. *Proc Natl Acad Sci U S A*, **95**, 492-498.
Wang, Z.G., Ruggero, D., Ronchetti, S., Zhong, S., Gaboli, M., Rivi, R. and Pandolfi, P.P. (1998b) PML is essential for multiple apoptotic pathways. *Nat Genet*, **20**, 266-272.
Ward, A., Fisher, R., Richardson, L., Pooler, J.A., Squire, S., Bates, P., Shaposhnikov, R., Hayward, N., Thurston, M. and Graham, C.F. (1997) Genomic regions regulating imprinting and insulin-like growth factor-II promoter 3 activity in transgenics: novel enhancer and silencer elements. *Genes Funct*, **1**, 25-36.
Warner, M.H., Roinick, K.L. and Arndt, K.M. (2007) Rtf1 is a multifunctional component of the Paf1 complex that regulates gene expression by directing cotranscriptional histone modification. *Mol Cell Biol*, **27**, 6103-6115.
Wattler, F., Wattler, S., Kelly, M., Skinner, H.B. and Nehls, M. (1999) Cloning, chromosomal location, and expression analysis of murine Smarce1-related, a new member of the high-mobility 365 group gene family. *Genomics*, **60**, 172-178.
Webber, A.L., Ingram, R.S., Levorse, J.M. and Tilghman, S.M. (1998) Location of enhancers is essential for the imprinting of H19 and Igf2 genes. *Nature*, **391**, 711-715.
Weber, M., Milligan, L., Delalbre, A., Antoine, E., Brunel, C., Cathala, G. and Forne, T. (2001) Extensive tissue-specific variation of allelic methylation in the Igf2 gene during mouse fetal development: relation to expression and imprinting. *Mech Dev*, **101**, 133-141.
Webster, K.E., O'Bryan, M.K., Fletcher, S., Crewther, P.E., Aapola, U., Craig, J., Harrison, D.K., Aung, H., Phutikanit, N., Lyle, R., Meachem, S.J., Antonarakis, S.E., de Kretser, D.M., Hedger, M.P., Peterson, P., Carroll, B.J. and Scott, H.S. (2005) Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. *Proc Natl Acad Sci U S A*, **102**, 4068-4073.
Wegner, M., Helftenbein, E., Muller, F., Meinecke, M., Muller, S. and Grummt, F. (1989) Identification of an amplification promoting DNA sequence from the hypotrichous ciliate Stylonychia lemnae. *Nucleic Acids Res*, **17**, 8783-8802.
Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., Yahata, K., Imamoto, F., Aburatani, H., Nakao, M., Imamoto, N., Maeshima, K., Shirahige, K. and Peters, J.M. (2008) Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*.
West, A.G., Gaszner, M. and Felsenfeld, G. (2002) Insulators: many functions, many mechanisms. *Genes Dev*, **16**, 271-288.
West, A.G., Huang, S., Gaszner, M., Litt, M.D. and Felsenfeld, G. (2004) Recruitment of histone modifications by USF proteins at a vertebrate barrier element. *Mol Cell*, **16**, 453-463.
Whetstine, J.R., Nottke, A., Lan, F., Huarte, M., Smolikov, S., Chen, Z., Spooner, E., Li, E., Zhang, G., Colaiacovo, M. and Shi, Y. (2006) Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell*, **125**, 467-481.
Wiesmeijer, K., Krouwels, I.M., Tanke, H.J. and Dirks, R.W. (2008) Chromatin movement visualized with photoactivable GFP-labeled histone H4. *Differentiation*, **76**, 83-90.
Williams, C.J., Naito, T., Arco, P.G., Seavitt, J.R., Cashman, S.M., De Souza, B., Qi, X., Keables, P., Von Andrian, U.H. and Georgopoulos, K. (2004) The chromatin remodeler Mi-2beta is required for CD4 expression and T cell development. *Immunity*, **20**, 719-733.
Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T. and Mann, M. (1996) Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature*, **379**, 466-469.
Wolfe, S.A., Nekludova, L. and Pabo, C.O. (2000) DNA recognition by Cys2His2 zinc finger proteins. *Annu Rev Biophys Biomol Struct*, **29**, 183-212.
Wolfer, A., Wilson, A., Nemir, M., MacDonald, H.R. and Radtke, F. (2002) Inactivation of Notch1 impairs VDJbeta rearrangement and allows pre-TCR-independent survival of early alpha beta Lineage Thymocytes. *Immunity*, **16**, 869-879.

Wong, M.M., Cox, L.K. and Chrivia, J.C. (2007) The chromatin remodeling protein, SRCAP, is critical for deposition of the histone variant H2A.Z at promoters. *J Biol Chem*, **282**, 26132-26139.
Workman, J.L. (2006) Nucleosome displacement in transcription. *Genes Dev*, **20**, 2009-2017.
Workman, J.L. and Kingston, R.E. (1998) Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu Rev Biochem*, **67**, 545-579.
Wurtele, H. and Chartrand, P. (2006) Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology. *Chromosome Res*, **14**, 477-495.
Wysocka, J., Swigut, T., Milne, T.A., Dou, Y., Zhang, X., Burlingame, A.L., Roeder, R.G., Brivanlou, A.H. and Allis, C.D. (2005) WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell*, **121**, 859-872.
Xi, H., Shulha, H.P., Lin, J.M., Vales, T.R., Fu, Y., Bodine, D.M., McKay, R.D., Chenoweth, J.G., Tesar, P.J., Furey, T.S., Ren, B., Weng, Z. and Crawford, G.E. (2007) Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genet*, **3**, e136.
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.
Xu, N., Donohoe, M.E., Silva, S.S. and Lee, J.T. (2007) Evidence that homologous X-chromosome pairing requires transcription and Ctcf protein. *Nat Genet*, **39**, 1390-1396.
Xu, N., Tsai, C.L. and Lee, J.T. (2006) Transient homologous chromosome pairing marks the onset of X inactivation. *Science*, **311**, 1149-1152.
Xue, Y., Wong, J., Moreno, G.T., Young, M.K., Cote, J. and Wang, W. (1998) NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol Cell*, **2**, 851-861.
Yang, C., Bolotin, E., Jiang, T., Sladek, F.M. and Martinez, E. (2007) Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene*, **389**, 52-65.
Yasui, D., Miyano, M., Cai, S., Varga-Weisz, P. and Kohwi-Shigematsu, T. (2002) SATB1 targets chromatin remodelling to regulate genes over long distances. *Nature*, **419**, 641-645.
Yoon, H.G., Choi, Y., Cole, P.A. and Wong, J. (2005) Reading and function of a histone code involved in targeting corepressor complexes for repression. *Mol Cell Biol*, **25**, 324-335.
Yoon, Y.S., Jeong, S., Rong, Q., Park, K.Y., Chung, J.H. and Pfeifer, K. (2007) Analysis of the H19ICR insulator. *Mol Cell Biol*, **27**, 3499-3510.
Youn, M.Y., Yoo, H.S., Kim, M.J., Hwang, S.Y., Choi, Y., Desiderio, S.V. and Yoo, J.Y. (2007) hCTR9, a component of Paf1 complex, participates in the transcription of interleukin 6-responsive genes through regulation of STAT3-DNA interactions. *J Biol Chem*, **282**, 34727-34734.
Yu, W., Ginjala, V., Pant, V., Chernukhin, I., Whitehead, J., Docquier, F., Farrar, D., Tavoosidana, G., Mukhopadhyay, R., Kanduri, C., Oshimura, M., Feinberg, A.P., Lobanenkov, V., Klenova, E. and Ohlsson, R. (2004) Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation. *Nat Genet*, **36**, 1105-1110.
Yuan, X., Feng, W., Imhof, A., Grummt, I. and Zhou, Y. (2007) Activation of RNA polymerase I transcription by cockayne syndrome group B protein and histone methyltransferase G9a. *Mol Cell*, **27**, 585-595.
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.
Zhang, H., Roberts, D.N. and Cairns, B.R. (2005) Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell*, **123**, 219-231.
Zhang, L.F., Huynh, K.D. and Lee, J.T. (2007) Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. *Cell*, **129**, 693-706.
Zhang, P., Du, J., Sun, B., Dong, X., Xu, G., Zhou, J., Huang, Q., Liu, Q., Hao, Q. and Ding, J. (2006) Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. *Nucleic Acids Res*, **34**, 6621-6628.
Zhang, R., Burke, L.J., Rasko, J.E., Lobanenkov, V. and Renkawitz, R. (2004) Dynamic association of the mammalian insulator protein CTCF with centrosomes and the midbody. *Exp Cell Res*, **294**, 86-93.
Zhao, H. and Dean, A. (2004) An insulator blocks spreading of histone acetylation and interferes with RNA polymerase II transfer between an enhancer and gene. *Nucleic Acids Res*, **32**, 4903-4919.
Zhao, K., Wang, W., Rando, O.J., Xue, Y., Swiderek, K., Kuo, A. and Crabtree, G.R. (1998) Rapid and phosphoinositol-dependent binding of the SWI/SNF-like BAF complex to chromatin after T lymphocyte receptor signaling. *Cell*, **95**, 625-636.
Zhao, Z., Tavoosidana, G., Sjolinder, M., Gondor, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K.S., Singh, U., Pant, V., Tiwari, V., Kurukuti, S. and Ohlsson, R. (2006) Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet*, **38**, 1341-1347.
Zhou, Y. and Grummt, I. (2005) The PHD finger/bromodomain of NoRC interacts with acetylated histone H4K16 and is sufficient for rDNA silencing. *Curr Biol*, **15**, 1434-1438.
Zhou, Y., Santoro, R. and Grummt, I. (2002) The chromatin remodeling complex NoRC targets HDAC1 to the ribosomal gene promoter and represses RNA polymerase I transcription. *Embo J*, **21**, 4632-4640.
Zimber, A., Nguyen, Q.D. and Gespach, C. (2004) Nuclear bodies and compartments: functional roles and cellular signalling in health and disease. *Cell Signal*, **16**, 1085-1104.
Zomerdijk, J.C., Beckmann, H., Comai, L. and Tjian, R. (1994) Assembly of transcriptionally active RNA polymerase I initiation factor SL1 from recombinant subunits. *Science*, **266**, 2015-2018.

7

Summary

Summary

All multi-cellular organisms start as a single fertilized egg. During development cells proliferate and differentiate, ultimately forming complete organs and an organism. Different types of cells exist within an organism, with highly specific functions and shapes. For example, muscle cells serve to exert force, blood cells serve to transport oxygen and carbon dioxide, and neurons serve to transmit signals. Despite enormous differences in appearance and function most cells share the same genetic material, which is present as a long strand of DNA that contains multiple genes. Not all genes are in use in all cells, some genes are turned 'on' others are turned 'off'. The genes that are active determine the function and identity of a cell. The human genome contains about 25.000 genes, which is only twice as much as the genome of a fruit fly. Genes are transcribed into mRNA, which is subsequently translated into a functional product, called protein. These are the functional units of the cell and represent the building blocks of the different compartments of a cell. Examples of proteins are enzymes, hormones and antibodies. By turning the transcription of genes 'on' or keeping it 'off' a cell decides over what mRNAs are synthesized and what proteins are produced. To a large extent this determines the unique protein composition of a cell and, hence, its phenotype.

The *Ctcf* gene codes for a protein that contains eleven zinc-fingers, with which it can bind to DNA or to other proteins. CTCF (CCCTC binding factor) belongs to the family of transcription factors, which means it is able to turn the transcription of specific genes 'on' or 'off'. Which specific genes are activated is tightly regulated and depends on the cell type and time point during development. A lack of correct transcriptional regulation can result in all kinds of diseases such as for example cancer. The different proteins it can bind to influence transcriptional regulation by CTCF. This thesis describes the work that others and I have done to understand how CTCF functions. Chapter 1 gives a broad introduction into the subject, which is needed to understand the experiments that are described in the remainder of the thesis. It gives an introduction to gene regulation and organisation of DNA in the cell nucleus, and what is known about CTCF.

To better understand the role of CTCF we generated so-called conditional *Ctcf* "knockout" mice, in which the *Ctcf* gene can be deleted in a spatio-temporally controlled manner. Studies in mice with this knockout allele are described in chapter 2. Deletion of the gene will stop the production of CTCF protein and cells then have to cope without CTCF. If we switch off the *Ctcf* gene very early in development, we do not obtain embryos without CTCF, indicating that CTCF is required for early stages of embryogenesis. This is consistent with published data. We then decided to delete the *Ctcf* gene in so-called T-cells (thymocytes) of the immune system of adult mice, since it is well known that mice can survive without an immune system, as long as they are not challenged by infections. We show that in the absence of CTCF proliferation and differentiation of T-cells in the thymus is blocked and no correct immune response can be build up in the absence of CTCF. Our data reveal a new role for CTCF in cell cycle control.

The variety and complexity of processes involving CTCF require that CTCF works together with multiple proteins. Thus, interaction of CTCF with other factors largely determines how and what CTCF does. Chapter 3 describes experiments that were done to identify as many CTCF interacting proteins as possible. Upon coupling of a small protein tag to CTCF, which can be labelled with biotin, we could easily affinity-purify biotinylated CTCF with a steptavidin-based affinity matrix (streptavidin binds tightly to biotin). By performing this purification under relatively mild conditions we were able to co-purify proteins that bind to CTCF from cellular extracts. Afterwards

CTCF protein partners were identified and analyzed by mass spectrometry. Embryonic stem cells expressing CTCF with the biotin label were injected into blastocysts to generate a *Ctcf* knock-in mouse that expressed biotinylated CTCF in all of its tissues. This mouse was used to identify CTCF protein partners in the different tissues.

We showed that CTCF binds to multiple different factors. Based on what is known about those proteins we discuss how they could function together with CTCF. This analysis is described in Chapter 3. In Chapter 4, I focus on a CTCF-interacting partner that came out of the biotin-pull down and that is called UBF (upstream binding factor). UBF is important for the activity of genes that code for ribosomal RNA (rRNA). This type of RNA is very important for formation of the ribosome, a megadalton complex that is absolutely essential for protein synthesis. We show that CTCF, as well as its homolog CTCFL, can bind with their zinc-fingers to the part of the UBF protein that is involved in DNA binding. Moreover we show that CTCF together with UBF and RNA polymerase I (the enzyme that ensures correct transcription of rRNA) bind to a piece of DNA that is important for correct regulation of rRNA transcription. In the absence of CTCF, binding of these proteins to DNA is reduced, indicating an important role for CTCF in rRNA transcription. Furthermore, we show that CTCF can organize chromatin in a very recognizable manner, by regulating the binding of other factors.

A unique function of CTCF is that it can block the inappropriate activation of a gene. CTCF does this by shielding a gene from activating sequences in the neighbourhood. I describe a cellular system that can be used to study this feature of CTCF in more detail in chapter 5. The system makes use of the fact that CTCF regulates two neighbouring genes, one coding for the growth factor Igf2 and another, H19, coding for an RNA that is not translated. These genes are imprinted, which means they are expressed from only one of the two parental copies. CTCF sits in between the two genes when they are inherited from the mother. This prevents activation of the Igf2 gene. When the two genes are inherited from the father CTCF is not bound in between the genes and Igf2 is active. Correct activity of the two genes is very important for foetal growth. By crossing conditional CTCF knockout mice with mice from another strain we can distinguish between the two parental copies. From these mice cells were isolated and tested. As these experiments are ongoing no firm conclusion can yet be drawn. However, our experiments do show that the activity of both genes is reduced in the absence of CTCF. This indicates that CTCF might have an additional role in regulating the activity of these genes. It appears that the cellular system can be used to gain a better insight into the role of CTCF in regulating imprinted genes.

In Chapter 6 the data described in this thesis are discussed. Based on our results together with what is known from experiments done by other people I speculate about the different regulatory mechanisms of CTCF. This results in suggestions for future research and experiments that can be performed.

Samenvatting

Alle multicellulaire organismen beginnen als een bevruchte eicel. Gedurende de ontwikkeling vindt vermeerdering en specialisatie van cellen plaats, die zich groeperen en organen vormen. Een organisme bestaat uit verschillende typen cellen met elk een specifieke functie en vorm. Spiercellen dienen bijvoorbeeld om kracht uit te oefenen, bloedcellen transporteren zuurstof en koolstofdioxide, en neuronen dienen om signalen in de hersenen door te geven. Ondanks een enorm verschil in uiterlijk en functie hebben deze cellen veel gemeen. Ze bevatten vrijwel allemaal het erfelijke materiaal in de vorm van een lange streng DNA waarop een groot aantal genen ligt. Niet al deze genen zijn in alle cellen in gebruik, sommige genen staan 'aan' en anderen staan 'uit'. De genen die aanstaan bepalen de functie en identiteit van een cel. Het humane genoom bevat ongeveer 25.000 genen, wat slechts twee keer zoveel als het genoom van een fruitvliegje. Genen worden afgeschreven tot mRNA, wat daarna vertaald wordt in een functioneel product, een eiwit. Deze eiwitten doen in de cel het werk en fungeren als bouwstenen van de verschillende onderdelen van de cel. Voorbeelden van eiwitten zijn enzymen, hormonen en antilichamen. Door het afschrijven (transcriptie) van genen 'aan' of 'uit' te zetten kan een cel beslissen welk mRNA gevormd wordt en welk type eiwit er geproduceerd wordt. Dit bepaald voor een groot gedeelte de unieke eiwit samenstelling en daarmee dus ook de functie en vorm van de cel.

Het *Ctcf* gen, codeert voor een eiwit met elf zinkvingers waarmee het DNA en andere eiwitten kan binden. CTCF (CCCTC binding factor) behoord tot de familie van transcriptiefactoren, wat betekend dat het bepaalde genen 'aan' en 'uit' kan zetten. Welke genen geactiveerd worden is strak gereguleerd qua celtype en tijdstip tijdens de ontwikkeling. Een gebrek aan correcte regulatie van gen transcriptie resulteert vaak in allerlei ziekten zoals b.v. kanker. Verschillende andere eiwitten kunnen aan CTCF binden en daardoor beïnvloeden hoe CTCF transcriptie van specifieke genen reguleert. Dit proefschrift beschrijft de experimenten die ik tezamen met andere mensen heb gedaan om het functioneren van CTCF beter te begrijpen. Hoofdstuk 1 geeft een brede introductie van het onderwerp, hierdoor kan de basis van de experimenten beschreven in dit proefschrift beter begrepen worden. Het hoofdstuk geeft een introductie over genregulatie en organisatie van het DNA in de celkern. Verder wordt uitgebreid ingegaan op wat tot nu toe bekend is over CTCF.

Om de rol van CTCF beter te begrijpen hebben we zogenaamde conditionele *Ctcf* "knock-out" muizen gemaakt waarin we het *Ctcf* gen kunnen uitschakelen. We kunnen zo zelf de tijd van, en het cel type waarin, het uitschakelen van het *Ctcf* gen plaats vind in de zich ontwikkelende muis bepalen. De studies die gedaan zijn met deze *Ctcf* knock-out muizen worden beschreven in hoofdstuk 2. Als we het *Ctcf* gen uitschakelen stopt de productie van CTCF eiwit en cellen moeten dan om zien te gaan met de afwezigheid van CTCF. Als we het *Ctcf* gen vanaf het begin van de ontwikkeling uitschakelen, worden er geen embryo's zonder CTCF geboren, dit geeft aan dat CTCF vereist is voor de vroege stadia in embryogenese. Dit komt overeen met al gepubliceerde data. Vervolgens hebben we het *Ctcf* gen uitgeschakeld in zogenaamde T-cellen van het immuunsysteem van een volwassen muis, aangezien het bekend is dat muizen kunnen overleven zonder een immuunsysteem, zo lang ze niet blootgesteld worden aan ziekteverwekkers. We laten zien dat in de afwezigheid van CTCF de differentiatie van T-cellen in de thymus geblokkeerd is en dat geen adequate afweer door het immuunsysteem opgebouwd wordt in de afwezigheid van CTCF. Onze data toont een nieuwe rol aan voor CTCF in regulatie van celdeling.

De variëteit en complexiteit van processen waarbij CTCF betrokken is vereisen dat CTCF

functioneert tezamen met meerdere eiwitten. Het contact van CTCF met andere eiwitten bepaald hoe en wat CTCF doet. In hoofdstuk 3 worden experimenten beschreven die gedaan zijn om eiwitten die aan CTCF binden te identificeren. Door CTCF aan een klein eiwit genaamd biotine te koppelen kunnen we CTCF zuiveren met een streptavidine affiniteit matrix (streptavidine bind sterk aan biotine). Door deze eiwitzuivering onder milde condities uit te voeren zijn we in staat geweest eiwitten die aan CTCF binden te isoleren uit cel extracten. Deze eiwitten zijn door middel van massaspectrometrie geanalyseerd en geïdentificeerd. Door embryonale stamcellen die CTCF met het biotine label bevatten in te brengen in blastocysten is een *Ctcf* "knock-in" muis gemaakt, die gebiotinyleerd CTCF aanmaakt in alle weefsels. Deze muis is gebruikt om CTCF eiwit partners te identificeren uit verschillende weefsels.

We laten zien dat CTCF aan veel verschillende factoren bind en geven aan de hand van wat al bekend is van deze eiwitten een discussie over hoe ze samen met CTCF kunnen werken. Deze analyse wordt beschreven in hoofdstuk 3. In hoofdstuk 4 ligt de focus op een CTCF interactie partner geïdentificeerd uit de biotine zuiveringen, UBF (upstream binding factor) genaamd. UBF is belangrijk voor de activiteit van genen die coderen voor ribosomaal RNA (rRNA). Ribosomaal RNA is belangrijk voor de vorming van het ribosoom, de plaats waar alle eiwitten in de cel aangemaakt worden. We laten zien dat zowel CTCF, als een homoloog CTCFL genaamd, met hun zinkvingers kunnen binden aan een gedeelte van het UBF eiwit wat van belang is voor DNA binding. Bovendien tonen we aan dat CTCF samen met UBF en RNA polymerase I (het enzym dat zorgt voor rRNA transcriptie) bindt aan een stukje DNA wat van belang is voor de correcte regulatie van rRNA transcriptie. In de afwezigheid van CTCF is binding van deze eiwitten aan DNA afgenomen wat aangeeft dat CTCF een belangrijke rol speelt in de regulatie van rRNA. Bovendien laten we zien dat CTCF chromatine kan organiseren op een erg herkenbare manier, door de binding van andere factoren te reguleren.

Een unieke functie van CTCF is dat het ongewenste activatie van een gen kan tegenhouden. CTCF doet dit door het gen af te schermen van activerende stukjes DNA die in de buurt aanwezig zijn. In hoofdstuk 5 beschrijf ik een celsysteem dat gebruikt kan worden om deze eigenschap van CTCF in meer detail te kunnen bestuderen. Het systeem maakt gebruik van het feit dat CTCF twee naast elkaar liggende genen reguleert, het ene gen codeert voor de groeifactor *Igf2* en het andere voor het niet-coderende *H19* gen. Deze genen zijn ingeprent, wat betekend dat slechts een van de twee ouderkopieën actief is. CTCF bevind zich tussen deze twee genen als ze overgeërfd worden van de moeder. Dit voorkomt de activatie van het *Igf2* gen. Indien de twee genen overgeërfd zijn van de vader is CTCF niet gebonden tussen deze genen en is het *Igf2* gen actief. Correcte activiteit van deze genen is erg belangrijk voor foetale groei. Door conditionele *Ctcf* knock-out muizen te kruisen met muizen van een ander ras kunnen we verschil maken tussen de twee ouderkopieën. Uit deze muizen zijn dan ook cellen geïsoleerd en getest. Aangezien deze experimenten nog niet afgerond zijn kunnen nog geen duidelijke conclusies getrokken worden. Onze experimenten tonen echter aan dat de activiteit van beide genen gereduceerd is in de afwezigheid van CTCF. Dit duidt op een mogelijk extra, nog onbekende, rol voor CTCF in de regulatie van de activiteit van deze genen. Het cellulaire systeem wat is opgezet kan gebruikt worden om meer duidelijkheid te krijgen over de rol van CTCF in de regulatie van ingeprente genen en met name die van *Igf2* en *H19*.

Tot slot geeft hoofdstuk 6 een discussie over de data die in dit proefschrift beschreven wordt. Op basis van onze resultaten en wat bekend is uit onderzoeken gedaan door andere mensen wordt gespeculeerd over de diverse regulatie mechanismen van CTCF. Aan de hand hiervan worden experimenten die in de toekomst gedaan kunnen worden uiteen gezet.

Curriculum vitae

Name: Suzanne van de Nobelen
Born: 4 January 1979, Gorinchem, The Netherlands

Education

1991-1996: HAVO, CSG Oude Hoven, Gorinchem

1996-2000: Bachelor
Biochemistry, HLO, Hogeschool van Utrecht, Utrecht
Graduation project: LDO, NV Organon, Oss
"Mammalian two hybrid system a cellular assay for the ZAP-70 interaction with the T-cell receptor zeta chain"
(Dr. A. Garritsen)

2000-2002: Master of Science
Biology, University of Utrecht, Utrecht
Graduation project: Molecular Neurosciences, University Medical Centre, Utrecht. "Identification of differentially expressed genes in the midbrain dopamine area of mice lacking Pitx3"
(Dr. M.P. Smidt)

Graduation paper: Physiological Chemistry, University Medical Centre, Utrecht. "The Igf2/H19 locus as a model system for the regulation of genomic imprinting"
(Dr. P.H. Steenbergh)

2003-2008: PhD-student
Department of Cell biology, Erasmus Medical Centre, Rotterdam
"Touched by CTCF: analysis of a multi-functional zinc-finger protein"
(Prof.dr. F.G. Grosveld and Dr.ir. N.J. Galjart)

Publications

Smits S, **van de Nobelen S**, Hornman K, von Oerthel L, Burbach J, Smidt M
'Signalling through phospholipase C beta 4 is not essential for midbrain dopaminergic neuron survival'
Neuroscience 2005 136(1):171-9

Dankwoord

In de vijf jaar van mijn promotieonderzoek hebben veel mensen bijgedragen aan het uiteindelijke resultaat, en deze mensen wil ik hierbij dan ook graag bedanken !

Ik wil beginnen met mijn promoter Frank Grosveld te bedanken voor de mogelijkheid om onderzoek te doen op zijn afdeling. Frank, ik heb veel geleerd van je adviezen en ideeën. Je heb me aangemoedigd kritisch te kijken naar mijn experimenten, resultaten en presentatie. Bedankt.

Niels, bedankt voor alles ! je optimisme, adviezen en het feit dat je altijd tijd vrij kon maken voor vragen en discussies. Ik heb veel geleerd van je kijk op sommige resultaten en de vrijheid die je gaf heeft ervoor gezorgd dat ik veel verschillende manieren heb geleerd om een vraagstelling experimenteel te beantwoorden. Mijn MacBook Pro is inmiddels alweer 1,5 jaar oud... zeker te weten toe aan vervanging? Sorry nog voor die ene keer dat ik wel heel hard moest lachen toen je op een snikhete zomerse dag het lab binnen kwam met een T-shirt daarover een overhemd met lange mouwen met daaronder een korte broek, twee melkflessen en van die flip-flop plastic teenslippers.

At first I want to thank the 'nuclear side' of the lab, our small CTCF-group. Happy we are not having to deal with depolymerizations and aggregates taking place on the other side of the lab/ nuclear membrane. Helen, I always thought as long as you are in the lab it will take a while before I have my defense. But in the end this happened to be not the case and time passed by very quickly. I am very happy that we could stay working together and shared the ups and downs of our projects. Sometimes we were more enthusiastic about results of the other than of our own. Thank you for your happiness, care and advise.

Frank, toen je hoorde dat ik een gedeelte van een tweeling was moest je wel even slikken. Bedankt voor alles wat je me geleerd hebt, het feit dat je altijd tijd voor me had, onze vele (wetenschappelijke) discussies en je enthousiasme en humor. Ik zal je bekende en elke keer weer grappige kreten als : " Ja en de jackpot gaat naar..", "sst daar heb je hem" , "Ja, suzie, jaa" zeker missen. Je gewoonte om alle eppen uitsluitend te nummeren van 0 (?) tot 10 of hoger heb ik toch maar niet overgenomen. En oh ja, nog bedankt voor de kaart vanaf je exotische vakantieadres :'ik ook aan jouw.' Je 'de realistische pessimist is beter af dan de optimist' en 'alle wetenschappers zijn manisch-depressief' hadden ook goede stellingen geweest.

Niet alleen mijn onderzoek maar zeker ook de sfeer maakte dat het elke dag weer leuk was om naar het lab te gaan. Het mag toch met recht gezegd worden dat lab 10.30 een van de gezelligste labs is. De gangmakers: Gideon, Jeffrey, Marco en Frank mede door jullie was het altijd gezellig en hebben we wat af gelachen met z'n allen. Niet alleen op het lab, maar ook bij alle andere activiteiten die we buiten werktijd met z'n allen deden maakte dat ik mij snel thuis voelde na mijn verhuizing naar Rotterdam. Een elk jaar weer terugkomend fenomeen waren de labvideo's voor the pantomime alhoewel het eerste jaar moeilijk te evenaren was.

Jeff: bedankt voor je interesse, optimisme en gezelligheid. Succes met de laatste loodjes ik denk toch dat de titel 'navigating towards a catastrophe' misschien iets te pessimistisch is. Gideon: mooie gelamineerde omslag hè? Soms raken er ook opeens dingen kwijt op het lab, raar. Marco: ik weet niet of het wel wat voor jouw is..., maar ik geef nog een feest. Naast Gideon en Marco nog andere alweer een tijdje ex-lab 10.30-ers. Marja: bedankt voor je gezelligheid en

je behulpzaamheid rondom mijn vragen over het proefschrift, Daniel: Jij had al snel Gideon's gewoonte overgenomen mij voor de gek houden en dat lukte je vaak aardig. Is het nou wel of niet waar van Boren? en dat je je proefschrift aan het schrijven bent? bedankt nog voor die (leuke) foto van mij samen met de exotische (bijna uitgestorven) Phthirus pubis. Is die al gesequenced? Michael: je had het niet altijd door, maar ik heb heel wat keren in een deuk gelegen om je grappige uitspraken. Veel succes met je overgang van de CLIPs en CLASPs naar CTCF. Ook bedank ik Katharina, Anna (Akhmanova), Anna (Szypowska), Soraya, Xenia en Filipe voor hun gezelligheid, interesse in mijn onderzoek en tips en adviezen. Natuurlijk ook niet te vergeten de 'lunchgroep' Nanda, Phebe, Rick en Bjorn. Hebben jullie het gezien ze hebben tegenwoordig ook oranje !

For all the chats, movies, drinks and dinners. Its great to have such a group of friends that share the ups and downs of a life as a PhD student. Thank you for all the fun, care and laughs. Helen. Athina and Cristina: now I can ask the nerve-wrecking question 'Who will be the next?' all the best with finishing your thesis and experiments. Ana: I am sure there will soon be a mouse born named Pereira, don't feed it too much 'stroopwafels'. Susana: for being interested and always willing to go out. Magda: for my defense (or is it the party?) being more important than starting your new job in sunny Marseille. Laura: have you spotted already a 'bouwvakkersdecolleté' in Madrid? Soon we might spot another 'cucaracha' (its such a happy song) on the CTCF side of the lab. I hope we will keep in touch and meet in the different European cities were we (will) live.

Besides people from our lab al lot of other members of the department contributed in different ways to this thesis and the nice atmosphere in the department to all of you: thank you!

Jeroen en Karel, de mannen van de mass spec. jullie hebben een grote bijdrage geleverd. Bedankt voor al jullie hulp! Natuurlijk konden de metingen nooit snel genoeg uitgevoerd worden. Jullie nieuwe strategie om te zeggen dat de data 3 maanden op zich laat wachten en dan na 1,5 maand al met een lijst aan te komen werkt goed, dan valt het altijd mee.

Lab 7.02/7.06: Erik, Robert-Jan, Wouter, Raymond en Sanja bedankt voor jullie interesse en hulp waar nodig. Lab 9.10: Bedankt voor jullie interesse en gezelligheid op het celkweek lab, Iris, Bas, Kim en Joost (vergeet je de volgende keer je tandenborstel en poster niet ?).

Tevens wil ik graag de leden van mijn kleine commissie bedanken voor hun suggesties en het kritisch lezen van dit proefschrift: Dies, Sjaak en Peter, dank je wel.

En niet te vergeten andere mensen die een niet onbelangrijke bijdrage hebben geleverd zoals de verzorgers van mijn muizen in het EDC, bedankt voor jullie goede zorgen. Ook de computermannen en Melle en Marike bedankt voor jullie ondersteuning.

My thesis contains data that is generated thanks to our collaboration with the lab of Prof. Rainer Renkawitz in Giessen, Germany. Thank you Rainer and Jorg, for help and discussions and of course I enjoyed very much the (ski)meetings in het Kleinwalsertal with labs from Rotterdam, Giessen and Marburg.

Naast mijn vrienden van het werk, die natuurlijk direct bij mijn onderzoek betrokken zijn, hebben de momenten met mijn andere vrienden voor de soms hard nodige afleiding gezorgd. Sarah, Henja en Judith ondanks alle drukte rondom jullie kleine mannetjes beleven jullie altijd geïnteresseerd in het wel en wee rondom mijn promotie. Nu heb ik weer meer tijd en hoop ik dat we wat meer gezellig met z'n allen uitgaan en bijkletsen. Henja stel het wel even uit tot na 7 mei dan kan je nog mee feesten. Bedankt dat jullie met interesse hebben meegeleefd met mijn promotie: Ria en Richard, Evelien en Rob, Marleen en Martin, Esther en Martin, Gemma en Gerrit, Jerome, Mees en Niels. Ook de mensen die in de beginjaren hebben meegeleefd Marco, Debby, Hans en Nel

bedankt.

Natuurlijk wil ik ook mijn familie bedanken. Waarvoor het vaak onduidelijk was wat nou eigenlijk het nut was van mijn onderzoek. Opa, ondanks dat mijn onderzoek niet direct zal leiden tot het genezen van ziekten weet ik dat u daardoor niet minder trots op me zult zijn. Ma en Frans ondanks dat jullie wilden onthouden en begrijpen wat ik nou precies aan het onderzoeken was viel dat meestal niet mee. Toch iets met 'moleculaire biologie' ? Bedankt dat jullie altijd achter mij staan, meeleven en me steunen bij mijn keuzes. Etienne en Henriette: dat rare werk van mij was vaak maar moeilijk te begrijpen. En ook de andere kant van de familie Pa, Cora en Oma van de Nobelen bedankt voor jullie belangstelling tijdens mijn promotietraject. Hopelijk geeft de Nederlandse samenvatting van mijn onderzoek een beetje verduidelijking waar ik nou eigenlijk al die jaren mee bezig was.

Alexander, je vertelt me regelmatig dat het ook best lekker is om af en toe gewoon 'niks' te doen. Bedankt voor je liefde, steun en geduld, je begrip als ik weer eens een weekend op het lab doorbracht en natuurlijk je hulp met de lay-out van mijn proefschrift. Ik kijk er naar uit om binnenkort weer wat meer tijd te hebben zodat we een verre reis kunnen gaan maken.

Suzanne