CTCF: Comprehending the complex functions of an 11-zinc-finger transcription factor

Printed in Optima Grafische Communicatie, Rotterdam.

ISBN: 978-90-8559-338-6

The work presented in this thesis was performed at the Department of Cell Biology at the Erasmus University Medical Center Rotterdam, The Netherlands. This department is part of the Medisch Genetisch Centrum Zuid-West Nederland (MGC).

Research performed in this thesis was supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO and NWO-ALW) and the European Science Foundation EUROCORES Programme EuroDYNA (ERAS-CT-2003-980409).

Cover image: Blue sky over St. Ives
Thesis layout: Helen

CTCF: Comprehending the complex functions of an 11-zinc-finger transcription factor

CTCF: Analyse van de complexe functies van een transcriptie factor met 11 zinc vingers

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
Donderdag 13 December 2007 om 9.00 uur

door

Helen Elizabeth Heath

geboren te Sutton Coldfield (England)

Promotiecommissie

Promoter:	Prof.dr. F. Grosveld
Overige leden:	Prof.dr. E.A. Dzierzak Dr. J.N.J. Philipsen Dr. R.W. Hendriks
Copromoter:	Dr.ir. N.J. Galjart

The pictures I contemplate painting would constitute a halfway state and an attempt to point out the direction of the future- Without arriving there completely.

Jackson Pollock.

For Dad.

Contents **Page**

List of abbreviations

3C	Chromosome Conformation Capture
4C	Chromosome Conformation Capture on Chip.
ACH	Active Chromatin Hub
AGM	Aorta-Gonad-Mesenephros
APC	Antigen presenting cell
ATP	Adenosine triphosphate
BORIS	Brother of the regulator of imprinted sites
bp	Base pairs
CD	Cluster differentiation
cDNA	Complimentary deoxyribonucleic acid
ChIP	Chromatin immunoprecipitation
CpG	Cytosine-guanine
CTCF	CCCTC-binding factor
DN	Double negative
DNA	Deoxyribonucleic acid
DNase I	Deoxyribonuclease I
DP	Double positive
E	Embryonic day
ES	Embryonic stem (cells)
FACS	Fluorescence-activated cell sorting
FCS	Fluorescence correlation spectroscopy
FISH	Fluorescent *in situ* hybridisation
FRAP	Fluorescence recovery after photobleaching
GFP	Green fluorescent protein
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HMT	Histone methyltransferase
HS	Hypersensitive site
HSC	Haematopoietic stem cell
ICR	Imprinting control region
IF	Immunofluorescence
Ifn-γ	Interferon-γ
IL	Interleukin
ISP	Immature single positive
kb	Kilo base pairs
kD	Kilo Dalton
KI	Knock in
KO	Knock out
LacZ	Bacterial β-galactosidase
LCR	Locus control region
MEF	Mouse embryonic fibroblast
mRNA	Messenger ribonucleic acid
PCR	Polymerase chain reaction
RNAi	Ribonucleic acid interference
S/MAR	Scaffold/matrix associated region
TCR	T-cell receptor
Th-1/2	T helper type-1/2

The scope of the thesis

In a multi-cellular organism, every somatic cell nucleus broadly contains the same sequence of DNA, yet clearly most cells are very different to each other. Specific sets of genes encoding proteins become activated whereas others are repressed. Within the genome, independently regulated genes are often found in close proximity to other genes that have different patterns of expression. How specific gene loci are organised in nuclear space is only recently emerging. CTCF is a protein that has been strongly implicated in mediating many distinct processes of gene regulation, including transcription, chromatin structure, and the structural organisation of gene loci. The aim of this thesis was to investigate the function of the CTCF protein *in vivo*, in particular the role of CTCF in regulating cellular proliferation, differentiation and the organisation of gene loci within the nucleus.

The introduction aims to give an overview of the information required to understand the foundations of studies presented and discussed in this thesis. The transcription or activation of genes occurs in the cell nucleus and requires specific modifications of chromatin. Chapter 1 describes the formation of chromatin and key factors that modify this structure. How transcription is initiated, and influenced by cis-regulatory elements is also discussed. Since the initial characterisation as a transcription factor, many structural and regulatory functions have been attributed to CTCF, as detailed in chapter 2, which imply CTCF is a key regulator of development and cell viability. The haematopoietic system is used in this thesis as a model for investigating the function of CTCF in two distinct lineages. In chapter 3 the development of erythrocytes and T-lymphocytes is introduced.

Chapters 4 and 5 describe the experiments used to address questions regarding CTCF function at the β-globin locus and during T-cell differentiation respectively. Published data strongly demonstrate the clustering of cis-regulatory elements into an active chromatin hub (ACH) facilitates transcriptional activation of the β-globin locus. Studies presented in this thesis reveal CTCF is required for normal ACH formation, but suggest additional elements may be required to regulate β-globin gene transcription. By deleting CTCF in T-cells we show that CTCF is necessary for the expansion of T-cells within the thymus. Moreover, CTCF is required for cytokine expression in T-helper type-2 cells and it is shown that CTCF may mediate T-cell receptor signaling.

Chapter 6 describes the generation of GFP-CTCF knock-in mice. The data show two distinct dynamic populations of CTCF exist within the nucleus, with the majority of the protein rendered largely immobile. GFP-CTCF is demonstrated to be a functional protein, and can substitute for endogenous CTCF in embryonic stem cells.

In the concluding chapter 7, the data presented in this thesis and the future implications of it will be discussed.

Chapter 1

The organisation of chromatin and transcriptional regulation

Chapter 1: The organisation of genomic DNA.

General introduction

The DNA of a eukaryotic cell is contained within the nucleus. Essentially, DNA can be represented as a linear array of four different nucleotides arranged as a sequence. Each nucleotide is known as a base, and the helical double-stranded DNA structure in cells is arranged due to coupling of adenine-thymine and cytosine-guanine bases on opposite strands of DNA. Functional units of nucleotides that encode proteins are called genes. On average, the genome of a mouse contains almost 22,000 genes. It is noteworthy that genes only occupy ~5% of the total amount of DNA. One of the most remarkable features of DNA in a multi-cellular organism is the scale. If observed as linear, the DNA in every cell would measure around two metres in length (Turner 2002). These facts immediately suggest that organisation of DNA within a nucleus is very important in order to fit into the 3-10μM diameter compartment (Razin et al. 2007). It is well documented that DNA adopts a higher-order complex structure together with specialised proteins to form chromatin, the sequential compaction of which allows DNA to be accommodated within nuclear space. Chromatin is dynamic, the compaction and de-condensation of which correlates broadly with areas of inactive and active gene expression, respectively, although this relationship is not absolute.

Modifications to chromatin, either the DNA-associated proteins or the DNA itself, that do not alter the DNA sequence, are known as epigenetic modifications and these can have profound effects on gene expression. Epigenetics is defined as heritable changes in phenotype that are not attributable to alterations in DNA sequence. It remains debatable how epigenetic modifications are remembered through subsequent cell divisions and indeed whether or not epigenetic marks are the cause or consequence of memory (Kouzarides 2007). In a multi-cellular organism, every somatic cell nucleus broadly contains the same sequence of DNA, yet in each cell specific sets of genes encoding proteins required for that particular cell fate and function at that particular stage of development and/or differentiation are expressed. Spatial and temporal expression of genes is mediated co-ordinately by direct modifications of chromatin, modifications of the DNA and interaction of the DNA with numerous general and cell type specific proteins, some of which function to remodel the chromatin structure.

Chromatin structure and histones

Chromatin is the DNA-protein complex that exists in the nucleus. The fundamental unit of chromatin is the nucleosome, which is organised in repeating arrays. A nucleosome is described as a short stretch of 146 ± 2bp of DNA (Noll and Kornberg 1977) wrapped 1.65 left-handed superhelical turns around a core subunit of specialised proteins called histones (Luger et al. 1997). The histone core forms a globular structure with an octameric configuration comprising two molecules of each histone protein H2a, H2b, H3 and H4. H2A-H2B dimers bind DNA on entry and exit from the nucleosome whereas DNA wrapping around the core interacts with the tetrameric H3-H4 structure.

The central part of each histone protein has a histone fold domain comprising 3 α-helicies and 2 loop structures (Arents et al. 1991) however the amino terminal ends consists of 15-30 residues that are apparently unstructured. These ends are called histone tails (Kornberg and Lorch 1999). Histone H2a and H2b, in contrast to the other histone proteins, also have a carboxy-terminal tail (Turner 2002). Histone tails are key sites of modifications that can alter the structure of chromatin (see below). The complex of DNA together with the histone core is known as the nucleosome core particle (NCP) and migrates with a relative molecular mass of 210 kD (Luger 2006). Histone proteins and the configuration of the NCP are among the most conserved structures across species. In addition to the canonical core histones, variant histone forms encoded by distinct genes are now being recognised to have important structural and functional implications (Kamakaka and Biggins 2005, Li, 2007).

Histones require very little sequence specificity to interact with DNA (Oudet et al. 1975, Khorasanizadeh, 2004) and are able to package virtually any DNA. One notable exception is G-quadruplex or so-called G-DNA or G4-DNA, where sequences of guanine bases induce complex folding of the DNA which is not free to wrap around the histone octamer (Oganesian and Bryan 2007). The significance of G4-DNA is an emerging aspect of gene regulation. Although the association of DNA and histones appears to be sequence independent, there is some degree of favouritism regarding formation and positioning of nucleosomes on DNA according to rotational preferences induced by double stranded nucleotides (Pina et al. 1990). On average there are 14 points of contact between DNA and the histone core (Luger and Richmond 1998) generating over 120 direct atomic interactions (Luger 2003). The interactions between histones and DNA are restricted to the phosphodiester backbone of the DNA. The basic histone proteins contain many positively charged arginine and lysine residues that are tightly attracted to the negatively charged phosphate groups of DNA.

Packaging of DNA into a nucleosome induces a 5-10 fold compaction of the DNA (Weintraub and Groudine 1976). The positioning of histones along the DNA occurs with remarkable regularity, spaced approximately at 200 ± 40bp intervals (McGhee and Felsenfeld 1980). The regions of DNA separating these structures and not enclosed within a nucleosome is known as linker DNA. The length of linker DNA is variable (10-80 bp) and the variation is thought to be important for regulating gene expression (Spadafora et al. 1976, Felsenfeld, 2003). Linker DNA is associated with a fifth histone called H1 which helps stabilise subsequent folding of the polynucleosome string through a helical configuration into a 30nm fibre resulting in a further 50-fold compaction of the DNA. The significance of H1 for inducing compaction to 30nm fibres is unresolved, however, linker histones do stabilise pre-existing compact fibres (Schwarz and Hansen 1994). High mobility group proteins (HMGs) are abundant chromatin-associated proteins that compete with linker histones for binding, thereby affecting the structure of chromatin (Catez et al. 2004, Bianchi, 2005).

Each region of linker DNA is occupied by a single histone H1 molecule. Histone H1 is not required to form the core nucleosome structure but does attract modifications that coordinate chromatin structure and regulate gene expression (Fan et al. 2005b). Further condensation of the chromatin through 60-130nm fibres known as chromonemas (Belmont et al. 1989) generate chromosomes (fig 1.1). In terms of weight, chromosomes are comprised of 50% DNA, with histones and other chromatin associated proteins accounting for the residual half (Hayes 2002). It is not fully resolved how compaction of chromatin is configured nor the exact disposition of chromatin higher order structure. Interestingly, the existence of 30nm fibres remains elusive *in vivo* (Tremethick 2007). Analysis of chromatin structure *in vitro* suggests the chromatin configuration becomes less condensed representing the 10nm structure described as ‘beads on a string’ in the presence of low-ionic strength buffers. In conditions close to physiological concentrations however, compaction of the chromatin into 30nm fibres are visualised (Hansen 2002). The greatest degree of chromatin compaction is observed in heterochromatin (see below) and mitotic chromosomes. The compaction of interphase chromosomes during mitosis results in a 10,000-fold reduction in chromosome length (Li et al. 1998). Despite the opinion regarding chromatin compaction as generally inhibitory to transcription and other processes that require access to DNA, mitotic chromatin appears to be quite dynamic in structure and remains accessible to transcription factors (Chen et al. 2005).

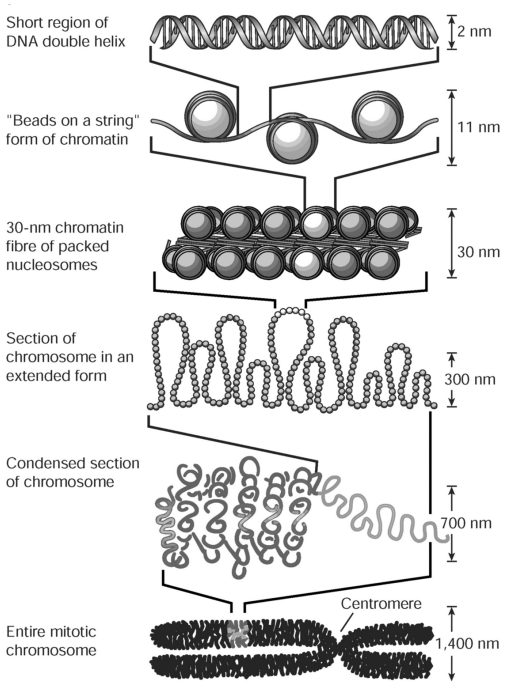


Figure 1.1. The organisation of DNA within the chromatin structure
Chromatin is a complex of the double-stranded DNA helix with proteins called histones. The sequential packaging of DNA and proteins to more compact configurations generates the distinguishable structure of mitotic chromosomes. Figure taken from (Felsenfeld and Groudine 2003).

Histone variants

The four core histone proteins as described above are encoded by multiple copies of histone gene cassettes that are expressed during S phase of the cell cycle. Newly synthesised $(H3\text{-}H4)_2$ tetramers carrying specific covalent modifications followed by two H2A-H2B dimers are incorporated during replication of DNA behind the replication fork (Akey and Luger 2003) (Loyola and Almouzni 2004, Jin, 2005). In addition to the major histone proteins, histone variants provide another degree of alteration to chromatin that can influence gene expression.

Variant histone forms are encoded by distinct genes and present both structural and functional consequences of incorporation to the nucleosome (Suto et al. 2000, Chakravarthy, 2005). Genes encoding histone variants are normally single-copy, intron-

containing genes, which are expressed constitutively, unlike core histones, and are integrated into nucleosomes in a DNA replication-independent manner (Kamakaka and Biggins 2005). Interestingly not all canonical histone proteins seem to have functional variants (Akhmanova et al. 1996) whereas others have many variant forms (Li et al. 2007). Histone variants are usually characteristic features of higher vertebrate genomes, however the H2A.Z and H2A.X variants of H2A are widely conserved (Raisner and Madhani 2006, Redon, 2002) and essential across eukaryotic species (van Daal and Elgin 1992, Liu, 1996, Faast, 2001). There are five predominant H2A-type histones, including the core histone H2A and three major histone H3-type histones. Another H2A variant (H2AvD) is known in Drosophila (Redon et al. 2002) and mammals possess a testis-specific H3 variant H3.4 (Kamakaka and Biggins 2005).

Incorporation of H2A and H3 variants in the histone octamer affect nucleosomal interactions that influence transcriptional competence of specific genes, regulate genome stability through chromosome segregation or DNA replication and repair, and characterise specific chromosomal regions such as macroH2A1 and H2A-Bbd on the inactive and active X chromosomes respectively (Costanzi and Pehrson 1998, Chadwick, 2001) and the H3 variant CENP-A at centromeres (Blower et al. 2002). H2A.X is a canonical histone protein in yeast (Li et al. 2007) however in mammalian cells, this variant is linked to transcriptional repression. Phosphorylation of H2A.X (γ-H2A.X) is a recognition site of DNA damage (Fillingham et al. 2006). H2A.Z has reported roles in transcriptional activation and repression and is a feature of insulator regions. Recently H2A.Z was identified as a marker of chromatin domain boundaries in human cells (see below, (Fan et al. 2002, Barski, 2007). H3 and variant H3.3 diverge only in four amino acids, which are not thought to alter the overall structure of the protein or its incorporation into the nucleosome. Nevertheless, localisation of H3.3 at transcriptionally active regions and enrichment for covalent modifications associated with open chromatin is strikingly opposite to that of H3 (McKittrick et al. 2004).

Histones are not confined to chromatin. The details of extra-chromatic histone activity are outside the scope of both my research and this thesis, however histones seem to be multi-functional in dimensions that exist outside the structure of chromatin, even outside the nucleus (Parseghian and Luhrs 2006).

Modifications and remodeling of chromatin

The histone proteins within the nucleosome complex are considered to be general repressors of gene activity (Grunstein 1997, Boeger, 2005). Nucleosomes and the higher order packaging of chromatin renders the DNA broadly inaccessible to regulatory protein complexes required for processes such as DNA repair, recombination, replication and transcription. Nevertheless, all these processes must occur on the

chromatin template. Nucleosome structure and their interaction with DNA are therefore stable but importantly not static.

Covalent modifications of core histones and DNA generate molecular markers distinguishing active, potentially active and inactive chromatin. Eukaryotes have four distinct mechanisms for allowing plasticity of nucleosomes within the chromatin template; covalent modifications of histone proteins, reorganisation of nucleosome position by ATP-dependent chromatin remodeling complexes, incorporation of variant histone molecules both into the nucleosome core and linker histone variants and covalent modification of DNA itself. Histone modifications and remodeling are widely conserved, however the incorporation of variant histones and DNA methylation seem restricted to more complicated genomes.

Covalent histone modifications

Covalent modifications of histone proteins occur on both histone tails and the globular histone core (Cosgrove et al. 2004, Mersfelder, 2006, Kouzarides, 2007). There are currently eleven different classes of identified histone modifications (table 1.1 (Kouzarides 2007, Hassan, 2006) and many enzymes are known to catalyse post-translational histone modifications (de la Cruz et al. 2005). Mutations in these enzyme complexes are often associated with malignancies in humans (Santos-Rosa and Caldas 2005), and aberrant patterns of histone modifications have been identified as hallmarks of human cancers (Fraga et al. 2005).

Covalent histone modifications enable an otherwise largely regular histone octamer to diversify dramatically resulting in profound effects on chromatin both structurally and functionally. All known covalent histone modifications with the exception of methylation and ubiquitylation influence the interaction of DNA with histones by changing the electrostatic charge of nucleosomes (Li et al. 2007). Interestingly, histone tails are not thought to be required for the structure of the nucleosome itself as removal of histone tails by digestion with trypsin has no effect on the stability of the nucleosome (Whitlock and Simpson 1977, Ausio, 1989). However it is increasingly clear that histone tail-mediated internucleosome interactions and higher order chromatin structure are affected by modifications (Shogren-Knaak et al. 2006, Polach et al. 2000, Tse et al. 1998). The significance of core histone modifications in mediating DNA interactions with the nucleosome remains unclear; however with the increasing number of core histone modifications identified it is likely that we will see the importance of these modifications in the near future (Xu et al. 2005).

Chromatin Modification	Residue Modified	Functions regulated
Acetylation	H2A (K5) H2B (K12, K15) H3 (K9, K14, K18, K23, K56y) H4 (K5, K8, K12, K16)	Transcription, Repair, Replication, Condensation
Methylation (lysine) mono/di/tri-me	H3 (K4, K9, K27, K36, K79) H4 (K20)	Transcription, Repair
Methylation (arginine) Mono/di(active)di(repress)	H3 (R2, R8, R17, R26) H4 (R3)	Transcription
Phosphorylation (serine)	H2B (S14) H3 (S10, S28) H4 (S1)	Transcription, Repair, Replication, Condensation
Phosphorylation (thrionine)	H3 (T3)	
Ubiquitylation	H2A (K119) H2B (K120 h/123 y)	Transcription, Repair
ADP ribosylation (arginine)	H2B (E2 mono-ar)	Transcription
Sumoylation	H2A H2B H4	Transcription
Deimination (conversion of argine-citrulline) mono-me-R not di/tri-me-R	H3 H4	Transcription
Proline isomerization (distortion of proline conformation)	H3 (P30y, P38y)	Transcription
Biotinylation	H2A (K9, K13, K125, K127,K 129) H3 (K4, K9, K18) H4 (K8, K12)	Gene silencing, Mitotic condensation, Gene silencing

Table 1.1. An overview of modifications identified on histones and their functions (see text for details). Only residue modifications with known modifying enzymes in human and yeast (*Saccharomyces cerevisiae*) are listed, with the exception of biotinylation that is dependent on exogenous biotin, based on (Kouzarides 2007, Hassan, 2006).

The function of covalent histone modifications can be broadly categorised as active or repressive modifications in the context of transcriptional activity. Acetylation is one of the most commonly studied post-translational modifications of histones and was first identified over forty years ago to correlate with active transcription (Allfrey et al. 1964, Grunstein, 1997, Roh et al. 2005). Lysine residues are invariably the substrate for histone acetyl transferase (HAT) complexes of which there are two main types; type A (nuclear) and type B (cytoplasmic) that function in transcriptional activation and

acetylation of newly synthesised histones respectively (Narlikar et al. 2002). There are three major acetyl transferase protein families, GCN5-related acetyl-transferase (GNAT), MOZ, YBF2/SAS3, SAS2, TIP60 protein family (MYST) and CREB-binding protein/p300 (CBP/p300) (Sterner and Berger 2000, Ehrenhofer-Murray, 2004, Santos-Rosa, 2005). In addition there are also general transcription factor HATs and HATs which are nuclear receptor cofactors. HATs are unable to bind DNA directly and therefore operate as transcriptional co-activators. In addition to transcriptional activation, histone acetylation has been implicated in regulating transcriptional initiation, through the stabilization of basal transcription factors to promoters and with transcriptional elongation, DNA replication and DNA damage repair.

The extent of acetylated histones throughout the genome is balanced by the opposing modification of deacetylation. Specific protein complexes can functionally reverse most covalent modifications identified to date, with the exception of methylated arginines (see below). Histone deacetylases (HDACs) remove acetyl groups from lysines, and hypoactelylation of histones correlates with transcriptional inactivity and gene repression. HDACs form parts of multi-subunit protein complexes associated with chromatin remodeling (see below). In mammals, three identified groups of HDACs are characterised according to homologous corresponding proteins in yeast. Mammalian homologues of Rpd3 and Hda1 define HDAC class I and II respectively. Class I HDACs are constitutively nuclear and ubiquitously expressed whereas class II HDACs have restricted expression patterns and characteristically transfer between the nucleus and cytoplasm. The enzymatic component of both groups of HDACs depends on a zinc molecule. Class III HDACs, also referred to as SIR-HDACs, are distinguished by a functional requirement for nicotinamide adenine dinucleotide (NAD^+) for deacetylation activity that subsequently reduces NAD^+ into *O*-acetyl-ADP-ribose and nicotinamide (NAM) (Blander and Guarente 2004).

Acetylation of lysines and phosphorylation of serines on histones are specifically correlated with active chromatin. In contrast, ubiquitylation of lysines and methylation of both lysines and arginines can function in either transcriptional activation or repression depending on the amino acid modified (Kouzarides 2007). Methylation of histones introduces a further level of complexity, as lysines can be mono- di- or tri-methylated, although tri-methylation has not been identified on arginines. The regulation of arginine methylation is not fully understood; despite characterisation of protein arginine methyl-transferases (PRMTs), a demethylating complex has not yet been identified. The existence of an arginine demethylase is presumed following characterisation of lysine demethylases (Kouzarides 2007, Shi et al. 2004).

Histone modifications are never single entities, and not every available amino acid may be modified nor every histone octamer will be modified in the same way. Histone modifications have been shown to function cooperatively and influence each other in a manner suggesting a histone code (Strahl and Allis 2000, Jenuwein & Allis 2001, Iizuka & Smith 2003). The code hypothesis proposes one way of viewing the combination of different covalent modifications and how in conjunction with each other, recognition motifs are set allowing for interactions with different protein complexes that can remodel histone-DNA interactions and/or recruit transcription factors to promoter regions.

Chromatin remodeling complexes

Histone-DNA interactions can also be modified via ATP-hydrolysis. Large multi-protein complexes that mediate the structure of chromatin in an ATP-dependent manner either by physical alteration of nucleosome position and/or attracting different histone modification enzymes are known as chromatin remodeling complexes (Ehrenhofer-Murray 2004). Chromatin remodeling complexes also mediate the complete displacement of histones from transcriptionally active genes and exchange of histone variants (Kimura and Cook 2001, Clayton et al. 2006). These transient alterations in chromatin facilitate transcription factor accessibility and are required for other epigenetic modifications such as DNA methylation.

The ATPase subunit is common to all protein complexes that remodel chromatin. On the basis of sequence features and identity of the ATPase, three families of remodeling complexes are distinguished, SWI/SNF, ISWI and Mi-2. The SWI/SNF complex is the prototypical chromatin-remodeling complex, originally identified in *S.cerevisiae* (Cote et al. 1994) after identification of *SWI2/SNF2* gene mutations that affected multiple gene expression resulting in the mating type switching (swi) and sucrose non-fermenting (snf) phenotype. Two Swi.Snf-like complexes are known in humans, BAF and PBAF that possess specific ATPase subunits (BRG1 and brm) and associated factors (BAFs) (Wang et al. 1996). ISWI is a smaller complex in relation to SWI/SNF, however in contrast to SWI/SNF, which requires BRG1, BAF47 and BAF155/170 for minimal catalytic activity *in vitro*, the ATPase of ISWI alone is sufficient for chromatin remodeling (Phelan et al. 1999, Langst, 1999). SWI/SNF complexes are associated with transcriptional activity and histone acetylation, whereas the ISWI complex is considered to be involved in repressive chromatin modifications due to the ISWI-mediated induction of nucleosome sliding *in cis*, which generates inaccessible chromatin conformations over neighbouring DNA sequences. The apparent contrasting activities of SWI/SNF and ISWI complexes possibly reflect their different substrate

specificities; chromatin remodeling by ISWI is dependent on histone tails, whereas SWI/SNF is able to remodel tail-less histones (Langst and Becker 2001).

The ATPase subunit of Mi-2 belongs to the 2 MDa nucleosome remodeling and histone deacetylation (NuRD) complex (Xue et al. 1998), and like ISWI has intrinsic chromatin remodeling activity. The Mi-2 subunit is distinguished from the other ATPases by a pair of chromodomains, designating this group as the chromo-ATPase/helicases-DNA binding domain (CHD) type. Mi-2 has two isoforms, α (CHD3) and β (CHD4). The NuRD complex contains HDAC1 and 2 and is able to deacetylate nucleosomes but not free histone proteins (Wade et al. 1999).

ATPase complexes and histone modifying proteins are unable to bind to DNA, and instead depend upon associated transcription factors for targeting to the necessary sequences. The differentiation between activating and repressing complexes can not simply be viewed according to transcription factor interaction, as transcription factors such as GATA-1 can attract both HAT and HDAC-containing complexes (Rodriguez et al. 2005). The influence of chromatin remodelling complexes and their associated factors can only be acknowledged within a specific context or developmental stage; Mi-2 directly interacts with HDACs, implying repressive activity of NuRD, however in T-cells, Mi-2β is required for recruiting the p300 histone acetyltransferase to the *Cd4* enhancer, directly regulating CD4 expression (Williams et al. 2004).

DNA methylation

Methylated DNA is often considered as refractory to transcriptional activity, either by directly inhibiting transcription factor binding or by attracting methyl-binding proteins that in turn attract transcriptional repressor complexes such as HDACs and chromatin remodeling complexes (Nan et al. 1998, Feng, 2001). DNA methylation mainly occurs as a stable, non-random covalent modification of cytosine at position 5 (Geiman and Robertson 2002). Methylation of DNA in mammals occurs principally on CpG dinucleotides by DNA methyltransferases DNMT1, DNMT3a and DNMT3b. Five DNMTs have been identified (DNMT1, DNMT2, DNMT3a, DNMT3b, DNMT3L), however only three are known to catalyse the addition of a methyl group to DNA.

Maintenance of methylated sites following DNA replication is achieved by DNMT1, which restores DNA methylation on hemi-methylated CpG dinucleotides and is essential for retaining patterns of methylation in proliferating cells (Bestor 1992). Following fertilisation, the male pronucleus undergoes demethylation followed by widespread demethylation of both maternal and paternal genomes before implantation (Howlett and Reik 1991, {Mayer, 2000). Re-establishment of DNA methylation in the embryo is the function of *de novo* methyltransferases DNMT3a and 3b (Okano et al. 1999). Correct establishment and maintenance of DNA methylation is essential for

genomic stability, embryonic development and viability. Interestingly, chromatin remodeling complexes, DNA helicases and histone modifications can all influence DNA methylation (Li 2002).

An association between histone methylation and DNA methylation has been known for several years based on mutagenesis of *defective in methylation-5* (*dim-5*), a H3-K9 methyltransferase in the fungus *Neurospora crassa* that was shown to be critical for DNA methylation (Tamaru and Selker 2001). DNA methylation is known to be both necessary and sufficient for establishing chromatin structure and histone modifications during development (Hashimshony et al. 2003).

Monoallelic gene expression

The majority of genes in diploid organisms are expressed from both alleles, however in mammals, some of the most important developmental fate decisions are made and regulated through the specific preferential expression of genes from one allele. Classical examples of such processes include X-chromosome inactivation, imprinted gene expression and allelic exclusion, which occurs in B- and T-lymphocytes and murine olfactory neurons (Yang and Kuroda 2007). Allelic exclusion is a mechanism to generate diversity and specificity of B- and T-lymphocyte antigen receptors (B/TCR) (see chapter 3). Strict regulation and timing of allelic exclusion is essential for the development of both B- and T-lymphocytes. Receptor diversity is also required in olfactory neurons, which have the potential to express any one of more than 1300 odorant receptor genes (Chess et al. 1994). In contrast, genes that are typically 'imprinted' do not encode receptors, but do regulate growth with many imprinted genes being essential for late-embryonic and post-natal development (Hurst and McVean 1998). Genomic imprinting describes the parental-specific expression of autosomal genes. The transcriptional capacity of a given imprinted gene is determined according to maternal or paternal inheritance. There are currently over 70 identified imprinted genes (Sleutels and Barlow 2001). One of the most outstanding examples of regional gene expression occurs in female cells where an entire chromosome is inactivated. X-inactivation was initially described in mice over 45 years ago as a hypothesis for transcriptional dosage compensation between XY males and XX females (Lyon 1961).

The mechanisms and regulation of monoallelic expression are perhaps surprisingly similar for each example given both the specificity of genes/genetic regions influenced, and the apparent randomness of allelic inactivation choice particularly for X-inactivation and allelic exclusion. An entire field of research is devoted to elucidating X-inactivation choice, and the matter of whether or not this is a random process remains highly controversial and beyond the scope of this thesis. The signals or characteristic

features that determine which *Tcr* allele should be silenced are also unknown. Epigenetic modifications that determine genomic imprinting are the only example of no choice, developmental fate is decided in gametes and this is maintained in somatic cells throughout life.

Asynchronous DNA replication, differential DNA methylation and chromatin modifications together with non-coding RNAs and even nuclear localisation are common themes to all examples of monoallelic expression (reviewed in (Goldmit and Bergman 2004). X-inactivation is achieved through coordinated and sequential activities of *X-inactive specific transcript* (*Xist*) RNA coating of the chromosome, and RNA polymerase II exclusion. Changes in histone modifications including loss of H3K9 acetylation and H3K4 hypomethylation occur, which is augmented by polycomb group (PcG) protein expression, H3K27 tri-methylation and H3K9 methylation. DNA methylation of CpG residues compounds the repressed state (Peters et al. 2003). The contribution of different histone methyltransferases to the induction and maintenance of X-inactivation indicates the complexity of regulation required in this phenomenon (Ohhata et al. 2004). Significantly, not all genes on the designated inactive X are silent, and the mechanisms of so-called 'escape gene' expression are only recently beginning to be addressed (see chapter 2 for details). Differential patterns of DNA methylation are a central feature of imprinted gene expression as not only is DNA methylation a stable, heritable modification via the action of DNA methyltransferases, but DNA methylation obstructs the binding of key transcription factors and other regulatory proteins, reciprocal effects of this are observed for example at the *Igf2-H19* locus (Hark et al. 2000, Bell et al. 2000). Monoallelic expression of olfactory receptor genes depends strongly on interactions between genes and the enhancer element *H*. It remains unclear if promoter-enhancer interactions are the defining event to maintain gene expression from that allele and mark the other allele for repression (Serizawa et al. 2003, Lomvardas et al. 2006).

Euchromatin and heterochromatin

A differential distribution of DNA within the nucleus has long been acknowledged. Early cytogenetic studies revealed two distinct arrangements of chromatin, euchromatin and heterochromatin. Typically, areas of heterochromatin contain repetitive sequences of DNA that replicate late in S-phase, and have a low gene density. Heterochromatic structures are more compact during interphase and are found in discrete areas of the nucleus; the nuclear periphery, at the centromeres and at telomeric ends of individual chromosomes. By contrast, euchromatin is gene rich and undergoes early S-phase replication. Euchromatin is characterised by an open, transcriptionally permissive chromatin structure that is distributed throughout the interphase nucleus (table 1.2). It is now recognised that euchromatin and

heterochromatin comprise structurally and transcriptionally very different examples of chromatin, however there is evidence of transcription and regions of decondensed chromatin occurring in otherwise typically heterochromatic areas and also silent or repressed genes are found within regions of open chromatin (Gilbert et al. 2004). The typical characteristics of heterochromatin have in the past led to the view of heterochromatin having little functional significance consisting of 'junk' DNA. This idea is no longer valid, particularly in light of the fact many vital genes have been mapped to heterochromatic regions in Drosophila (Devlin et al. 1990, Dimitri et al. 2003).

Areas of heterochromatin and euchromatin are not irreversibly stuck with respect to transcriptional status, chromatin modifications or nuclear position. The formation of heterochromatin outside of constitutively heterochromatic domains such as centromeres is important for regulating temporal and spatial expression of certain genes during cellular differentiation and in particular is associated with X-inactivation and imprinted gene expression in mammals and silencing of homeotic gene loci by the PcG proteins (Quina et al. 2006). Recent analysis of histone modifications at centromeric domains have determined that chromatin conformation is distinct from that observed in both euchromatin and heterochromatin, which may specifically reflect the role of centromeres mediating mitosis in the nucleus (Sullivan and Karpen 2004).

Feature	**Euchromatin**	**Heterochromatin**
Staining after trypsin-giemsa / packaging in interphase	Isopycnotic / dispersed, decondensed	Heteropycnotic/ condensed
DNase sensitivity	Sensitive	Less sensitive
Replication timing	Early/throughout S-phase	Late S-phase
Presence of genes	High/variable density, house-keeping genes	Low density, lack house-keeping genes
DNA sequence	Predominantly unique	Largely repetitive (satellites, transposons)
Transcriptional activity	Poised or active	Poised or silent
Meiotic (reciprocal) recombination	Normal frequency	Infrequent
Characteristic modifications	Hyperacetylated H3 and H4 Methylated H3K4 CpG hypomethylation	Hypoacetylated H3 and H4 Methylated H3K9, HP-1 CpG hypermethylation

Table 1.2. Adapted from (Richards and Elgin 2002 and Holmquist & Ashley, 2006).

Interesting correlations exist between the position of active and inactive gene loci within the interphase nucleus and the proximity to heterochromatic regions. The location of a gene locus close to or within a heterochromatic domain is associated with transcriptional repression. This has been observed using both transgenes and analysis of endogenous loci. In T-lymphocytes, the T-helper subsets (Th-1 and Th-2) of naïve CD4 cells are distinguished by IGFγ^+/IL-4$^-$ and IL-4$^+$/IFNγ^- expression respectively. Progressive polarisation into Th1 or Th2 lineage occurs over several cell divisions. This prevents expression of the opposing cytokines (for further discussion, see chapter 3). Repositioning of the silenced IL-4 (Th1) and IFNγ (Th2) to centromeric heterochromatin is seen only in polarised cells, not in naïve CD4 cells. Increasing evidence shows that this is a dynamic process possibly regulated by transcriptional activity, which can relocate chromatin domains away from or towards heterochromatin. Transcriptional activation may 'move' loci to a more transcriptionally permissive nuclear environment whereas silencing of an active locus can be achieved by positioning close to heterochromatic regions. However repositioning of a locus might also be a consequence, not simply a cause of silencing. For example, transcription of the recombination-activating genes (*Rag*)-1 and deoxynucleotidyl terminal transferase (*Dntt* or *TdT*) genes in immature thymocytes is irreversibly repressed upon T-cell receptor engagement in CD4 and CD8 T-cells via the acquisition of repressive histone modifications; deacetylation of H3K9, hypomethylation of H3K4 and gain of methylation at H3K9. H3K9 methylation attracts binding of HP-1, a component of silent chromatin. Binding of HP-1 is thought to support heterochromatin formation at this locus (Bannister et al. 2001). The timing of transcriptional silencing and relocation of these loci to heterochromatin indicated a delay of around twelve hours between silencing and repositioning, suggesting the loci are already silent when they become incorporated into heterochromatin.

Movement of active gene loci to a heterochromatic region can be considered as one way to negatively regulate gene expression, if binding of specific proteins to the DNA regulatory sequences of a given gene, or interacting protein partners can function to associate their target genes with heterochromatin. Ikaros is a DNA-binding protein specifically expressed in haematopoietic and lymphoid cells and is required for correct progression of both B- and T-cell lineages (Wang et al. 1996). Ikaros expression has been tightly linked with transcriptional repression, and is a prime candidate for identifying factors that can target specific genes to centromeric and pericentromeric heterochromatin. Studies by Brown et al revealed Ikaros co-localises with γ-satellite DNA in mice, which is a marker for repetitive heterochromatic elements (Brown et al. 2001). The ability of Ikaros to bind to both centromeric repetitive DNA as well as cis-regulatory elements of lymphoid restricted genes suggests Ikaros may attract gene loci to

heterochromatic regions. Interestingly, Ikaros-mediated relocation of silent genes to pericentromeric chromatin occurs in mitogen-activated cells, and not resting splenocytes (Brown et al. 1997).

The influence of heterochromatin on the expression of genes is seen through the phenomena of position effect variegation (PEV). PEV describes the abnormal silencing of a gene locus that has been repositioned, for example by translocation, to an area of heterochromatin in a population of cells that would usually express the gene. Initial studies in Drosophila showed that the proximity of a gene to heterochromatin strongly influences the expression of that gene (Ananiev and Gvozdev 1974). Similar analyses in mice demonstrated that PEV occurs also in mammals. PEV is thought to occur by spreading of repressive histone modifications and binding of transcriptional repressors to the active gene, eventually silencing transcription. Interestingly, the β-globin locus control region (LCR) was the first cis-acting element identified as capable of conferring position independent expression on a linked transgene when inserted into transcriptionally silent chromatin (Grosveld et al. 1987). This is now a required feature of sequences to be called an LCR (see below).

Insulator sequences (see below) can also block the spread of heterochromatin into active gene loci and as such these sequences are thought to function as important regulatory boundaries between active and inactive chromatin domains. The properties of both LCRs and insulators imply that endogenous genes and linked transgenes may be protected from PEV due to looping of chromatin domains away from the surrounding heterochromatin. Relocation of a locus away from heterochromatin does not always require transcriptional activation. Histone acetylation modifications associated with open chromatin are observed on the β-globin locus before expression of the globin genes, suggesting the modifications to make transcription possible are enough and that transcription itself is not required to move chromatin domains away from a heterochromatic state (Ragoczy et al. 2003). Although movement of a gene locus away from heterochromatin appears to be a prerequisite for transcription, it is not sufficient to initiate gene activation (see below).

DNA loops, nuclear matrix or scaffold, and associated regions

In the interphase nucleus, topological arrangement of the genome into chromatin compacts the chromatin array, but also regulates spatial and temporal expression of certain genes. Chromatin is partitioned at 5-200 kb intervals into loops (Bode et al. 2003) by elements that may demarcate regions of active and inactive gene expression. Factors that organise chromatin loops play a very important role not only in

gene regulation but perhaps in the maintenance of the nuclear structure as a whole (Earnshaw 1988).

Several proteins involved in gene regulation are known to bind the nuclear matrix via specialised DNA sequences, matrix or scaffold associated regions (MARs/SARs) (table 1.3). The terms MARs and SARs to describe the attachment points of DNA to the nuclear matrix or scaffold are interchangeable and as such are collectively addressed as S/MARs. The distinction between SARs and MARs refers to the metaphase or interphase cell extraction of these regions respectively. What constitutes the nuclear matrix from the nuclear scaffold appears to depend on the method of isolation. The nuclear matrix was initially characterised as an insoluble, filamentous protein complex, resistant to DNase treatment, detergent and high salt extraction. The precise nature and composition of the nuclear matrix is still largely unresolved, and this together with technical difficulty in isolating the structure leads to continued debate over the existence of a nuclear skeleton. Besides chromatin and related proteins, the nucleus is a highly territorial structure comprised of functionally distinct regions such as nucleoli (see below). Interestingly, in the absence of any chromatin, the other functional elements of the nucleus remain intact, suggesting that a skeleton network or matrix of some description remains (Nickerson 2001).

The nuclear matrix has been defined as a complex of RNA and non-histone proteins that forms a network-like base to which chromatin can attach. S/MARs are sequences of DNA (200-300bp) that remain attached to the nuclear matrix following extraction. There is no definitive consensus sequence defining S/MARs, however they have a predominance of A+T base pairs (70%) and frequently contain topoisomerase II cleavage sites (Bode et al. 1992). Base unpairing regions (BURs) are key features of S/MARs and as such S/MARs have a high unwinding potential which may relieve stress in the DNA fibre generated during transcription. S/MARs emerge regularly in the genome; on average there is one MAR in every 30-100Kb of DNA (Bode et al. 2003), however within the 90Kb that includes the globin locus, eight S/MAR sites have been identified and at the 3'end of the *Cd8β* gene, two MARs are located just 4.5Kb apart (Cunningham et al. 1994).

By physically generating chromatin loops, S/MARs have been seen as insulator sequences, regions that distinguish chromatin domains into active and inactive gene loci. A number of proteins have been identified to associate with S/MARs, including lamins, which may be important for the structure of the network (Luderus et al. 1992), Nucleolin, a major nucleolar protein (Dickinson and Kohwi-Shigematsu 1995) and Bright, a factor primarily expressed in B-cells (Herrscher et al. 1995). S/MARs have been implicated in both transcriptional activation and repression (Maya-Mendoza et al. 2003).

Scaffold/Matrix Attachment Region (S/MAR) Binding Proteins	Expression Pattern	Reference
B-cell regulator of IgH transcription (Bright)	B-cells	Herrscher, 1995
CCAAT displacement protein (CDP; Cut/Cux)	Ubiquitous except terminally differentiated cells	Kaul-Ghanekar, 2004
CCCTC-binding factor (CTCF)	Ubiquitous**	Dunn, 2003
Histone H1	Ubiquitous	Izaurralde, 1989
High Mobility Group 1 and 2 (HMG1/Y)	Inreased in tumors and cells with increased metastatic potential	Liu, 1999
Ku autoantigen	MAR activity identified in breast carcinoma cell lines	Galande, 1999
Lamins type A and B1	Ubiquitous, MAR studies in rat liver	Luderus, 1992
Methyl CpG binding protein 2 (MeCP2)	Ubiquitous	Stratling, 1999
Nucleolin	Major nucleolar protien, ubiquitous	Dickinson, 1995
Poly (ADP-ribose) polymerase (PARP)	MAR activity identified in breast carcinoma cell lines	Galande, 1999
p53*	Ubiquitous	Gohler, 2005, Singh, 2007
Scaffold Attachment Factor A and B (SAF-A/B; p120, hnRNP)	Ubiquitous	Fackelmayer, 1994 Renz, 1996
Special AT-rich Binding Protein 1 and 2 (SATB1/2)	Predominantly T-cells	Dickinson, 1992
Scaffold/matrix-associated region-1-binding protein (Smar-1)	Predominantly T-cells	Kaul-Ghanekar, 2004
Topoisomerase II	Ubiquitous	Adachi, 1989
Nuclear Matrix Protein-1/ Ying-Yang1 (NMP1,YY1)	Ubiquitous	Tan, 1998 Ogino, 2002

Table 1.3. S/MAR binding proteins and their expression pattern.

The majority of S/MAR binding proteins currently identified are ubiquitously expressed with the exception of tissue-specific proteins found in B- and T-lymphocytes. *Mutated p53 binds S/MAR sequences opposed to wt p53 (Gohler et al, 2005; Singh et al, 2007). **CTCF is not expressed during certain stages of spermatogenesis (Loukinov et al 2002).

Significantly, not all S/MARs function as tethering points for chromatin. Selective association of S/MARs with the nuclear matrix can therefore alter the structural arrangement of chromatin and potentially regulate gene expression. Cell-type specific matrix associated proteins may help this regulation. SATB1 (special AT-rich binding protein 1) is a S/MAR binding protein predominantly expressed in thymocytes (Dickinson et al. 1992). In cells that express SATB1, it has been shown to occupy S/MARs positioned at the base of chromatin loops in contrast to non-expressing cells that reposition the SATB1 binding S/MARs to the looping chromatin domain. SATB1 appears to be sensitive to the base unwinding potential of a S/MAR, a feature that may be mediated by its homeodomain. SATB1 has been identified to recruit chromatin-remodeling complexes, regulating histone modifications and gene expression over long distances (Yasui et al. 2002). A characteristic nuclear localisation in T-cells indicates a role for SATB1 in maintaining nuclear architecture (Cai et al. 2003). Matrix associated regions and their binding proteins are therefore important components of the chromatin structure and its plasticity, but also are tightly involved in regulating gene expression.

Transcription and regulation of eukaryotic gene expression

An open chromatin conformation mediated by covalent modifications of nucleosomes together with chromatin-remodeling complexes is necessary but not sufficient for genes to become active. In this instance, gene activity refers to transcription. Transcriptional regulation is a critical component of cellular function. The majority of genes studied encode proteins but many non-coding genes are present in the genome and these too have important structural and catalytic roles in the cell.

Basal transcription machinery and transcriptional initiation

Three RNA polymerases (RNA pol I, II, III) are responsible for transcribing nuclear genes. Transcription is the enzymatic process of copying DNA into RNA, which in the case of RNA pol II generates messenger RNA (mRNA) that is translated into polypeptide templates for functional proteins. Eukaryotic RNA polymerase I synthesises ribosomal RNA used in the production of ribosomes. RNA pol I activity is confined to specialised areas within the nucleus known as nucleoli (see below). RNA pol III transcribes small stable RNAs involved in RNA processing such as transfer RNA (tRNA), 7SL RNA, U6 small nuclear RNA (snRNA), as well as 5S rRNA (Paule and White 2000). All three RNA polymerases are intrinsically unable to identify their respective target promoters, and require direction from general transcription factors (GTFs) to guide them.

Each eukaryotic RNA polymerase is a complex of around twelve subunits whose sequences are largely conserved between species, but the composition of subunits

varies according to the type of polymerase; five subunits are shared between all three polymerases (Martinez 2002). Transcription can be regulated at any of three individual stages, initiation, elongation and termination. In the case of RNA pol II, five GTFs (TFIIB, -D, -E, -F and -H) are the minimal requirement for promoter recognition and accessibility (Conaway and Conaway 1997) together with a large multi-subunit complex called Mediator (Conaway et al. 2005). TFIIA is necessary for 5S gene transcription. Specific initiation requires recognition of core promoter DNA motifs including the TATA-box, initiator (Inr), TFIIB response element (BRE) and upstream promoter elements CCAAT and CACC boxes. Experiments *in vitro* identified the core promoter as the minimal sequence required to direct transcription and as such this is referred to as basal transcription. TATA-binding protein (TBP) is involved in transcription by all three polymerases (Fan et al. 2005a).

Sequential assembly of GTFs and pol II at core promoter sequences generate a pre-initiation complex (PIC). Each of the GTFs with the exception of TFIIB, are multi-subunit complexes. Binding of TFIID to the core promoter is the first step in transcription. TBP is a component of TFIID, which induces a physical change in promoter DNA forcing it to bend. TFIIA and –B then assemble and stabilise TFIID interactions. Addition of TFIIF functions as a checkpoint ensuring correct promoter complex assembly (Svejstrup 2004). A minimal complex of pol II-TFIIB-TFIIF is sufficient to initiate transcription, however, TFIIE and TFIIH are important for promoter melting; opening the promoter template (Reese 2003). TFIIH in particular is an interesting protein complex involved in at least five distinct pathways including transactivation of hormone-responsive genes, DNA repair, RNA pol I and II transcription (Rochette-Egly et al. 1997, Schaeffer, 1993, Iben et al. 2002). The capacity of TFIIH to open promoters is an ATP-dependent process achieved via two DNA helicases subunits XPB and XPD. CDK7 is a kinase catalytic subunit of TFIIH that phosphorylates tandem heptapeptide Tyr-Ser-Pro-Thr-Ser-Pro-Ser repeats of the carboxy-terminal domain (CTD) of the largest subunit of RNA pol II. Phosphorylation of CTD is essential for triggering the transition from transcriptional initiation to elongation *in vivo* (Dahmus 1995).

The transcriptional machinery accompanying each of the RNA polymerases must access DNA, which is packaged into chromatin as described above. It is now accepted that nucleosomes are removed from activated promoters, allowing transcription initiation complexes to assemble on effectively naked DNA *in vivo* (Boeger et al. 2003; Boeger et al. 2005). Genes transcribed by both pol II and III are associated with nucleosomes, whereas active rDNA repeats are nucleosome deficient when transcribed and repackaged following replication (Paule and White 2000). Large proportions of both active and inactive rRNA genes exist which can hamper analysis of chromatin conformation of these repeat arrays.

Cis-regulatory elements

Promoters as described above are small 100-200bp sequences located directly upstream of the transcriptional start site of genes that attract specific transcription factors and transcription machinery to position, assemble and initiate correct transcription. Mammalian cells are frequently lacking in defining elements of core promoters (TATA-boxes, Inrs) where the PIC compiles (Smale 2001), instead mammalian core promoters are often associated with stretches of G/C rich sequence (CpG islands) that escape methylation. DnaseI hypersensitive sites (HSs) are 200-300bp sequences that have an increased susceptibility to DnaseI digestion in comparison to the surrounding chromatin. These sites frequently encompass binding sites for specific transcription factors, and designate areas of chromatin that are potentially active. Sensitivity occurs via multi-protein interactions on the template that disturb the position of nucleosome. Clusters of HSs are indicative of locus control regions (LCRs) (see below).

Enhancer elements are sequences of assorted transcription factor binding sites that provide additional specificity and increase the rate of transcriptional activation. Structurally, enhancers are similar to promoters, however enhancers are typically longer in sequence, can be positioned many kilobases from transcription start sites and their activity is orientation independent; enhancers can be found either upstream or downstream of their specific genes or even within genes or overlapping promoters (Blackwood and Kadonaga 1998). For many years the precise mechanism of how enhancer-binding proteins made contact with promoters remained elusive and several different models were hypothesised, including tracking of protein complexes along the chromatin fibre from enhancers to promoters, and progressive polymerisation of proteins that would stretch from enhancers to promoters (Li et al. 2002). It is now widely accepted that long-range chromatin interactions between proteins bound to regulatory elements and promoters occur via looping out the intervening chromatin (Blanton et al. 2003, Tolhuis et al. 2002). Silencer elements configure in much the same way as enhancers, however they attract repressive protein complexes and function to inhibit transcription from cis-linked promoters.

An additional level of regulating gene expression occurs in the form of locus control regions (LCRs). LCRs are specialised regions of chromatin that are similar to enhancers in many ways. The structure of LCRs, composed of varying numbers of DNase I hypersensitive sites often incorporating enhancers or enhancer blocking/insulator elements and its functional definition of a sequence able to drive expression of a linked transgene at a level equivalent to its endogenous counterpart from any position in the genome (West and Fraser 2005) determines what separates LCRs

from other regulatory elements. LCRs were first functionally characterized in transgenic mice at the β-globin locus (Grosveld et al. 1987). Many LCRs are now known for different genes (Li et al. 2002). Despite the requirement for a linked LCR to confer position-independent expression of a transgene, protecting against position effect variegation, LCRs do not necessarily control the chromatin environment of their target genes; this instead most likely is conferred by recruitment of specific transcription factors such as GATA1 and erythroid krüppel–like transcription factor EKLF, which attract chromatin remodeling complexes to the β-globin locus (Drissen et al. 2004, Vakoc et al. 2005).

Genes are activated and repressed according to developmental cues, which will be specific for each distinct cell type. Different genes are required to express proteins constitutively while others have developmental restriction. Within the genome, constitutively expressed genes are juxtaposed with genes encoding specialised proteins, and frequently the regulatory elements of these genes will overlap. Insulator sequence elements were initially defined as boundaries positioned between domains of open and condensed chromatin, and were found to have two defining characteristics; the ability to inhibit the spread of heterochromatin and the ability to prevent enhancer-promoter interactions when placed in between the two (enhancer blocking) (West et al. 2002). These defining features of insulators are now revised, and it is accepted that enhancer-blocking and heterochromatin barrier functions are separable at least in vertebrates (Recillas-Targa et al. 2002). The activity of insulator elements is mediated by the interacting protein(s). CTCF is the archetypal factor shown to be required for enhancer blocking activity in vertebrates (Bell et al. 1999). Increasingly, it is clear that CTCF does not directly regulate all known vertebrate enhancer blocking sequences (see chapter 2). Interestingly, the transcription factor YY-1 is known to bind to insulator sequences within the imprinted *Peg3* gene in a methylation-sensitive manner (Kim et al. 2003). Differentially regulated insulator types exist; imprinted insulators are regulated in part by DNA methylation affecting the binding of proteins as described for CTCF and YY-1. At some insulator sequences, CTCF-mediated enhancer blocking activity is influenced by CTCF interaction with thyroid hormone receptor (TR) molecules. CTCF/TR complexes retain the capacity to prevent enhancer-promoter interactions, but this is sensitive to the presence of thyroid hormone (Lutz et al. 2003).

A territorial yet dynamic nucleus

Immunofluorescence analysis of an interphase eukaryotic nucleus reveals distinct nuclear processes such as transcription, ribosomal DNA processing and DNA replication each occurs within sub-nuclear structures (fig 1.2) (Stein et al. 2003).

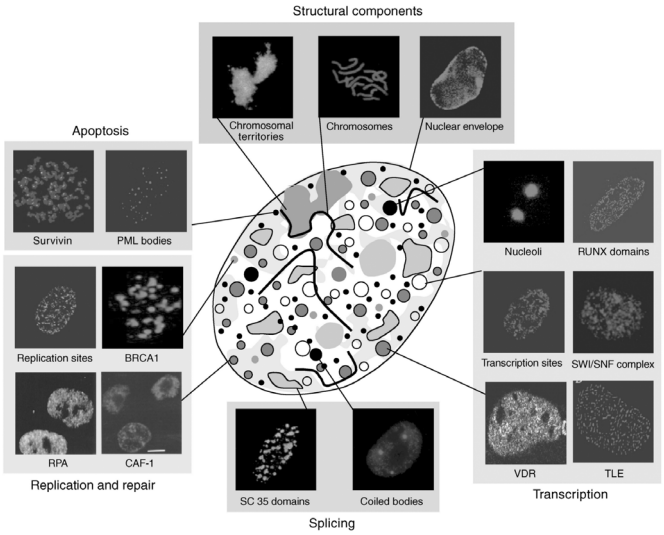


Figure 1.2. Taken from (Stein et al. 2003).
Different domains of the nucleoplasm can be defined as accumulations of specific proteins as revealed by indirect immunofluorescence. Abbreviations: chromatin assembly factor-1 (CAF-1), promyelocytic leukaemia (PML) bodies, transducin-like enhancer (TLE), vitamin D_3 receptor (VDR). For details see (Stein et al. 2003).

All of these structures must continually re-assemble after mitosis. Highly condensed mitotic chromosomes begin to expand once again during telophase, and nucleoli emerge. Whole-chromosome FISH analysis has revealed individual chromosomes occupy a defined space within the interphase nucleus, the chromosome territory (CT) (Cremer and Cremer 2001; Cremer et al. 2006). CTs are considered as sub-nuclear structures; chromosomes consistently localise in similar positions relative to the center or periphery of the nucleus. Interestingly, chromosomes seem to have preferential positions with reference to other chromosomes (Parada and Misteli 2002). The relative position of gene loci according to their CT as observed by FISH suggests there is a correlation between gene activity and nuclear position, where active genes are preferentially found

at the exterior of chromatin domains although transcriptionally active genes are also located within the CT (Dietzel et al. 1999, Verschure, 1999). The model for CT organisation proposes interchromatin domains (ICD) exist and physically separate distinct territories. Recently, an alternative FISH protocol (cryo-FISH) was developed, allowing fixed sucrose embedded cells to be cut into 150nm thick slices, which are then used for standard FISH (Branco and Pombo 2006). The authors reveal CTs in human lymphocytes are not as separated from each other as previously proposed, arguing against the existence of ICDs.

Live cell analysis using GFP/lac-repressor constructs has revealed the interphase chromatin is not static, but is in a state of continual movement (Chubb et al. 2002). The understanding of nuclear organisation and how nuclear processes actually occur has been greatly advanced by the use of (green) fluorescent protein live-cell markers. GFP has in the past been the protein marker of choice (Tsien 1998), although utilising markers with different excitation wavelengths permits more detailed analysis particularly regarding protein-protein interactions. Time-lapse imaging is used to understand the behaviour of a protein, for example at different stages of the cell cycle, under different growth/stimulation conditions in an attempt to understand the importance of localisation and protein dynamics both within the nucleus, and exchanges between nucleoplasm and cytoplasm. The dynamics of a (G)FP tagged protein can be analysed in time as a measure of the recovery of fluorescent molecules into a previously bleached area (FRAP) (Houtsmuller and Vermeulen 2001), indicating the relative mobility of a given protein. Additionally, fluorescent correlation spectroscopy (FCS) measures the number of actual fluorescent molecules passing through a specified volume. In this way, data can be extrapolated to determine the amount of specific protein that is present in nuclear/cellular space (Bacia et al. 2006). Importantly, these techniques are only reliable if the tagged protein of interest is functional and behaves as the endogenous protein in terms of localisation and expression levels.

It is widely believed that most nuclear factors, certainly most transcription factors are highly dynamic and in a constant flux between chromatin and the nucleoplasm (Becker et al. 2002, Misteli, 2001). Interestingly, although core histone proteins are part of the same octameric complex, H3 and H4 were extremely immobile proteins whereas H2A and B had a small fraction of molecules that exchanged quickly. In the case of H2B, the dynamics of the mobile fraction was transcription-dependent (Kimura and Cook 2001). Proteins for the most part are apparently readily mobile throughout the different nuclear compartments. Ribosomal RNA (rRNA) synthesis occurs exclusively within the nucleolus, however GFP-tagging of essential rRNA factors revealed a rapid exchange of these proteins (upstream binding factor (UBF) 1, UBF2, and TAF_I48) occurs between the

nucleus and sites of rDNA transcription in the nucleolus (Iben et al. 2002). The dynamics of proteins both within the interphase nucleus and during mitosis is an emerging aspect of our understanding of gene regulation and nuclear architecture. The mechanisms of epigenetic regulation of gene expression during interphase are relatively well established. The importance of specific protein interactions with mitotic chromosomes and how this contributes to cellular memory are currently unknown.

Chapter 2

What is CTCF and what does it do?

Chapter 2: What is CTCF and what does it do?

Maintenance of genetic and epigenetic integrity of the genome is essential. Epigenetic control of gene expression as described above regulates how essential cellular processes including transcription, proliferation, differentiation, cellular senescence, aging and death occur in a multi-cellular system. CTCF has been identified at the forefront of many of these processes. CCCTC-binding factor (CTCF) was originally identified as a transcription factor that negatively regulates the chicken c-myc gene (Lobanenkov et al. 1990). At the same time it was observed that another protein Nep1 bound to and regulated transcriptional activity of the chicken lysozyme silencer (Baniahmad et al. 1990). Nep1, initially thought to be the human homologue of chicken CTCF, was later confirmed to be identical to chicken CTCF (Burcin et al. 1997). CTCF is now known to mediate the transcription of a number of different genes, both positively and negatively, in the classical fashion of a transcription factor, by recognising and directly interacting with promoter sequences (Kuzmin et al. 2005, De La Rosa-Velazquez et al. 2007, Renaud et al. 2005). Interestingly this includes recognition of the germ cell-specific TFIIAα/β-like factor (ALF) promoter (Kim et al. 2006). CTCF binds within introns, exons and intergenic sequences in humans (Renaud et al. 2005, Vetchinova et al. 2006) with the majority of identified CTCF binding sites apparently located far from transcriptional start sites (Kim et al. 2007).

In this chapter I will introduce CTCF and the current knowledge regarding CTCF structure and functions will be discussed.

General features of CTCF

The *Ctcf* gene has been conserved throughout over 500 million years of deviation between vertebrate and invertebrate species; orthologous *Ctcf* sequences have been identified in genomes ranging from mosquito to human (Gray and Coates 2005). In mice, the gene coding for *Ctcf* is located on chromosome 8 and consists of at least two non-coding exons and 10 coding exons that generate a 728 amino acid protein with a predicted relative molecular mass of 82kDa. However, it is well documented that CTCF migrates in SDS-PAGE at around 130kDa due to effects from the amino and carboxy-terminal domains (Klenova et al. 1997). The genomic structure of CTCF is highly conserved particularly in mammals, where the exon-intron junctions in mice and humans are identical (Ohlsson et al. 2001). Recent analysis of zebrafish CTCF revealed structural similarities to human CTCF genomic organisation, which puts it closer to mammalian CTCF than chicken (Pugacheva et al. 2006).

The CTCF protein in all vertebrates features three distinct regulatory domains, of which the central zinc-finger domain consisting of 10 C_2H_2 and one C_2HC-type zinc-fingers is strictly preserved with 100% amino acid identity between avian and mammalian CTCF proteins (Filippova et al. 1996) and 98% identity comparing to

zebrafish and *Xenopus laevis* (Burke et al. 2002a). The eleven zinc-finger domain is also present in invertebrates, however the conservation of sequence is less consistent. The zinc-finger domain mediates CpG methylation-sensitive binding of CTCF to DNA and interactions with itself and other proteins. Importantly, CTCF is reported to bind diverse DNA sequences, a characteristic feature, which allows interaction between CTCF and many different regulatory elements, contributing to the multifunctional nature of the protein. The apparent lack of consensus binding site for CTCF has been partially explained by deletion analysis of individual zinc-fingers, revealing particular combinations of zinc-fingers recognise some target sites while different zinc-finger contacts are associated with other target site interactions (Baniahmad et al. 1990, Filippova et al. 1996, Quitschke et al. 2000). As such CTCF is considered to be a multivalent protein. Ostensibly, the CTCF protein has quite a flexible topology and structure. The apparent selective use of zinc-fingers suggests that further control either within the 11-zinc-finger domain as a whole, or indeed the entire protein must be present to enable conformational plasticity of the CTCF-DNA complex (Pugacheva et al. 2005). How this is achieved remains unknown. A wide range of regulatory sequences, some of which bear little sequence homology to each other or to comparative regulatory elements in related species, have been identified to bind CTCF. Chromatin immunoprecipitation followed by detection with genome-tilling microarrays identified 13,804 CTCF binding sites in the human genome and a predicted consensus sequence was proposed (Kim et al. 2007). Comparisons of conserved sequences in the human genome also led to the establishment of a vertebrate consensus sequence (Xie et al. 2007). Recently, a proposed consensus binding sequence for CTCF was reported in *Drosophila*. In this study, the authors compared the two consensus sequences of human and drosophila to each other, revealing a conserved AGGNGGC consensus sequence (Holohan et al. 2007). Although no agreed consensus sequence can encompass all established CTCF binding sites, within the 50bp DNase footprint of CTCF, a smaller characteristic motif occurs (Filippova et al. 2001). The study by Holohan *et al* hypothesised binding of CTCF to specific DNA may be more consistent and conserved than previous analysis suggests. Importantly, sequences identified in *Drosophila* to bind CTCF function as target sites for both *Drosophila* and chicken CTCF proteins, but the Drosophila CTCF protein does not interact with all vertebrate target sites (Moon et al. 2005, Klenova et al. 2002) suggesting functional evolution of the CTCF protein.

Consistent with its classification as a transcription factor, CTCF encompasses two additional transcriptional regulatory domains, which flank the central zinc-finger region. Transcriptional activation and repression activity has been associated with both the amino and carboxy-terminal domains, the specificity of action seems to be target sequence and cell-type dependent (Drueppel et al. 2004, Vostrov & Quitschke. 1997, Lutz et al. 2000).

It is not clear how domains outside of the zinc-finger region influence other CTCF functions such as enhancer blocking or mediating chromatin loop formation. CTCF is currently known to be phosphorylated and subject to poly(ADP-ribosylation) *in vivo* (Yu et al. 2004, Klenova, 2001), although other as yet unclassified modifications of CTCF may also occur (Zhang et al. 2004). Poly(ADP-ribosylation) describes the covalent addition of ADP-ribose polymers onto various target proteins, which in the case of CTCF occurs preferentially on the N-terminal domain. Attachment of linear or branched poly(ADP-ribose) units is directed almost exclusively to the γ-carboxyl group of glutamic acid residues by poly(ADP-ribose) polymerases such as PARP1. There are at least seven PARP family members, which may have redundant functions, complicating analysis of PARPs and proteins that are poly(ADP-ribosyl)ated. The balance of poly(ADP-ribosylation) is maintained by poly(ADP-ribose) glycohydrolase (PARG), which releases free ADP-ribose units by hydrolysing glycosidic bonds of ribose-ribose polymers (Soldani and Scovassi 2002, Rouleau, 2004). Poly(ADP-ribosylation) is a crucial modification which occurs on many different proteins from transcription factors to histones, and as such is essential for regulating numerous cellular processes including DNA repair, apoptosis, transcription, replication, differentiation and has a role in chromatin structure (Kraus and Lis 2003, Tulin et al. 2003, Soldani et al. 2001).

Interestingly, accumulation of the CTCF protein to the nucleolus has been shown to occur in differentiating cells in a poly(ADP-ribosylation)-dependent manner, and is proposed to negatively regulate ribosomal DNA (rDNA) transcription causing an arrest in cellular proliferation. Inhibition of poly(ADP-ribosylation) with 3-aminobenzamide (3-ABA), a PAR polymerase inhibitor, alleviated repression of rDNA transcription and impedes nucleolar translocation of the protein (Torrano et al. 2006). Poly(ADP-ribosylation) of CTCF appears to correlate with CTCF interaction with target sites and mediates insulator activity of CTCF target sites as shown by *H19* ICR insulator sensitivity to 3-ABA, (Yu et al. 2004). Detection of the PAR-modified CTCF protein with anti-PAR-polymer antibody revealed this increased the molecular mass of CTCF to 180kDa. It is significant to note that the authors of this study concede only proteins with more than 10 poly(ADP-ribose) units can be detected with this antibody. It is also possible that polymer chains less than 10 units in length attach to CTCF and mediates its function and/or structure. The 180kDa form of CTCF is not detectable in all cell types and it remains to be seen if other CTCF-dependent insulators also recruit poly(ADP-ribosyl)ated CTCF.

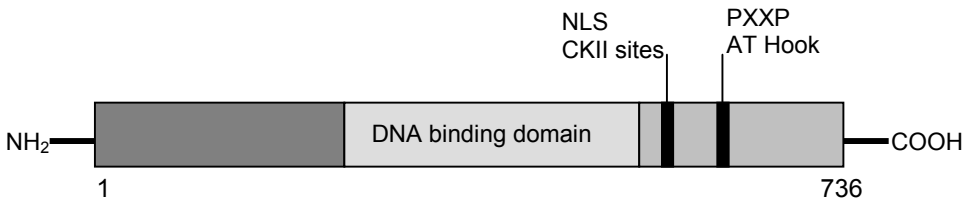


Figure 2.1: Schematic representation of the murine CTCF protein. The DNA binding domain (mid grey) is composed of 11 zinc fingers. Known casein kinase II (CKII) phosphorylation sites, PXXP and AT hook motifs are indicated together with the location of the nuclear localisation signal (NLS). Numbers indicate amino acids. See text for details.

Within the C-terminal domain, four prominent motifs are present that influence CTCF protein function. As a constitutively nuclear protein, CTCF requires import to the nucleus from the cytoplasm. Translocation into the nucleus is regulated by importins that usually bind to the NLS (Macara 2001). The CTCF NLS is located in the 8th coding exon. Mis-localisation of CTCF in the cytoplasm has been described in cases of invasive breast carcinoma, however a correlation between this and carcinogenesis remains undefined (Rakha et al. 2004) (see below). It remains unknown which importins function to locate the CTCF protein within the nucleus.

As mentioned above, CTCF is a phosphoprotein *in vitro* and *in vivo* (Klenova et al. 2001). A SKKEDSSDSE motif is unerringly conserved in vertebrates although in *X.laevis*, one aspartic acid (D) residue is replaced by glycine (G) and one serine (S) is changed to a threonine (T) (Burke et al. 2002). It is possible the D-G exchange may alter the conformation of the protein changing from polar to non-polar side chains that favour the outside and inside of proteins respectively. The serine residues within this motif have been identified as substrates for casein kinase II (Klenova et al. 2001). This study revealed by substitution of serines that phosphorylation of CTCF by casein kinase II is not required for binding DNA *in vitro* or the localisation of CTCF within the nucleus, but does effect transcriptional repression and cell cycle progression. It has been proposed that phosphorylation of CTCF may specify interactions between protein partners or regulate the dimerisation potential of CTCF that in turn effects its functional capacity.

Two further short motifs are located within the C-terminal of CTCF; a potential DNA-binding and chromatin-remodelling motif of KRRGRP, known as an AT-hook sequence, and a doublet of the PXXP motif distinctive of Src homology 3 (SH3)-domain binding proteins. Functionality of the PXXP sequence has been suggested to mediate CTCF interaction with BIN-1, a MYC-associated protein, which may regulate repressing capabilities of the CTCF C-terminal domain *in vitro* via essential prolines in this motif (Ohlsson et al. 2001).

More than a transcription factor

It is evident that the regulation of gene expression by CTCF extends far beyond that of a typical transcription factor. CTCF establishes hormone-responsive transcriptional regulatory elements and formation of enhancer-blocking boundary elements (Lutz et al. 2003, Burke et al. 2002, Recillas-Targa et al. 2006). Binding sites for CTCF are found in several examples to be in close proximity to thyroid hormone response elements (TREs) and the location of these composite regulatory elements does not necessarily correspond to transcription initiation sites nor is the spacing and orientation between CTCF binding sites and TREs consistent (Burke et al. 2002b). In at least two cases, the chicken *lysozyme* gene and human *c-myc*, CTCF binding sites and associated TREs are positioned in between an enhancer and its respective gene promoter. Enhancer-blocking activity of CTCF at these sites is mediated by the presence of thyroid hormone triiodothyronine (T3) where by the enhancer-blocking capacity of CTCF is disrupted in the presence of T3. Further analysis of the chicken lysozyme gene revealed CTCF binding *per se* is not sufficient for enhancer-blocking activity, as CTCF remains bound to chromatin even in addition of thyroid hormone and attracts acetylation modifications to histone H4 (Lutz et al. 2003). Interestingly, not all TREs function in the same manner; TRE 144 coordinates CTCF dependent transcriptional repression in the presence of T3 and is therefore considered as a negative TRE (Awad et al. 1999). Importantly, it seems that not all nuclear hormone receptor (NHR) binding sites function in synergy with CTCF and not all CTCF binding sites require T3 or nuclear hormone receptors for CTCF activity (Szabo et al. 2006) (see below).

CTCF, insulators and enhancer blocking

Chromatin insulators are largely characterised by two separable functions; enhancer blocking and protection against chromatin position effects. The ability of insulators to shield a transgene against long-term silencing is another important function known as silencer-blocking or anti-silencing effects (Pikaart et al. 1998, Yao et al. 2003, Hino et al. 2006). CTCF is increasingly linked to the regulation of epigenetics and long-range chromatin interactions mediating gene expression. The interactions between CTCF and diverse insulator sequences have been used to propose a regulatory role for CTCF in higher order chromatin organisation and nuclear dynamics.

CTCF was the first and currently only identified vertebrate protein able to interact with and regulate enhancer blocking activity of a variety of different insulator sequences (Bell et al. 1999). As such it was hypothesised that CTCF may function to delineate the genome into regulated functional domains. In complex genomes, the spatial and temporal integrity of gene regulation is critical. It is widely believed that

shielding neighbouring gene loci from inappropriate regulatory elements is achieved in part by insulator sequences and their associated binding proteins However, many genes have overlapping regulatory elements or anti-sense transcripts belonging to differentially regulated genes, so considering the genome as demarcated into domains is essentially misleading. The concept of functional expression modules (de Laat and Grosveld 2003) clarifies the genome into a manageable model, which retains the significance of spatial and temporal gene regulation.

Integral to assigning the role of establishing functional expression modules to CTCF is the identification of CTCF binding sites and their positions within the genome. Analysis of the chicken, mouse and human *β-globin* locus revealed flanking CTCF binding sites, strongly suggesting an archetypal insulator role for CTCF at this locus protecting temporal and spatial regulation of β-like globin genes from inappropriate activation by distal enhancers (Farrell et al. 2002). CTCF binding sites from the *β-globin* locus have been classified as insulators *in vitro* although the activity of these insulators varies significantly (Bell et al. 1999, Farrell et al. 2002, Bulger et al. 1999, Wai et al. 2003). Currently, three CTCF binding sites are known to be located 5' of the *β-globin* locus and one is positioned at the 3'HS1, all sites correspond to DNase I hypersensitive sites shown to cluster into a formation known as the Active Chromatin Hub (ACH) (Tolhuis et al. 2002) (see chapter 3 for details). No sites are identified within the β-globin domain (Yusufzai et al. 2004). Importantly, formation of the normal ACH is CTCF dependent as when one CTCF binding site is mutated, configuration of the ACH is disrupted without transcriptional effect on the β-globin genes (chapter 4, Splinter et al. 2006). Histone H3 acetylation is lost from a mutated 3'HS1 site which can no longer bind CTCF and is replaced by repressive histone modifications. This change in chromatin conformation occurs at all β-globin CTCF binding sites in the absence of CTCF, but the change is specifically local to CTCF binding sites, and does not induce the entire locus into a closed state. Questions therefore remain over the function of these CTCF binding sites at the *β-globin* locus and whether CTCF binding sites are indicative of functional expression modules (Kim et al. 2007).

Recent data indicates three previously unknown examples of gene loci that are surrounded by CTCF binding sites, α-polypeptide H^+/K^+ exchanging ATPase (*ATP4a*), myelin-associated glycoprotein (*MAG*) and *NIFIE14* encoding a transmembrane protein (Vetchinova et al. 2006), however it is clear that not all loci are regulated by flanking CTCF binding sites. The importance of considering functional expression modules rather than domains is demonstrated at the α-globin gene locus, which resides in a region of open chromatin regardless of transcriptional activation. CTCF dependent enhancer-blocking elements have been identified in the α-globin locus and CTCF binds to these sites in both erythroid and non-erythroid cells (Valadez-Graham et al. 2004).

Chromatin structure around the *c-myc* gene is part of a 160Kb domain flanked by MAR sequences that apparently insulate *c-myc* from the adjacent *pvt1* gene. The active *c-myc* gene is associated with acetylated histones, flanked by repressive chromatin modifications that comprise no other expressed genes. CTCF interaction with the c-myc MARs was not detected. A sequence 5' to the *c-myc* gene was identified to have insulator activity *in vitro* and was found to bind CTCF, but no CTCF site 3' to *c-myc* was identified. The c-myc insulator element (MINE) comprises separable enhancer blocking and barrier actions, which are CTCF-dependent and independent respectively and does not associate with the nuclear matrix. Importantly, the mechanism of the MINE activity seems to be distinct from other insulator elements such as chicken β-globin HS4 as it does not attract an open chromatin configuration and is located outside of the c-myc hyperacetylated domain. Several potential CTCF binding sites have been identified *in silico* and *in vitro*, but only two are significantly conserved between humans and mice and are associated with DNase I hypersensitive sites. Both CTCF binding sites are 5' to *c-myc,* and CTCF binds constitutively *in vivo* within the c-myc loop regardless of transcriptional activity, suggesting CTCF *per se* may not directly repress *c-myc* transcription (Lobanenkov et al. 1990, Filippova et al. 1996) and other factors are required for the regulation of c-myc (Gombert et al. 2003).

There are a number of complex loci, which have been identified as having CTCF-dependent enhancer blocking elements, however the significance of CTCF binding and possible mechanisms of action remain largely elusive. CTCF binding sites have been identified within the mouse rDNA locus in between the repeating units of ribosomal genes (SvdN unpublished data). In *Xenopus*, a sequence known as the repeat organiser (RO) was shown to have enhancer-blocking activity *in vitro* using transfected K562 cells (Bell et al. 1999, Robinett et al. 1997).This sequence was also identified to bind CTCF. Organisation of *rDNA* transcription is a highly complicated arrangement, as not all ribosomal genes will be transcribing at any one time and in a population of cells, not all cells will have identical rDNA locus chromatin configurations. Unravelling how CTCF functions at this locus will be of great interest.

Very little is known regarding repetitive sequences as regulated insulator elements. It is known that Alu-like repeating sequences flanking the human keratin-18 (K18) gene function as insulators in transgenic mice (Willoughby et al. 2000), and this activity requires RNA polymerase III. Whether there is a CTCF involvement remains open to speculation. Alu-like repeating sequences have been identified within CTCF binding sites (Vetchinova et al. 2006).

CTCF-mediated enhancer blocking activity seems to be largely conserved from *Drosophila* to human (Moon et al. 2005). The enhancer blocking capacity of insulators has been determined as separable to the action of preventing the aberrant acquisition of

repressive chromatin modifications, as CTCF binding is necessary for one function but not the other (Recillas-Targa et al. 2002). CTCF does not bind all insulators with enhancer blocking activity (Magdinier et al. 2004, Holohan et al. 2007, Gomos-Klein et al. 2007), indicating at least in some loci, enhancer-blocking activity is a CTCF-independent process and some insulators function in a CTCF-independent manner. *In vitro* silencer blocking activity by the chicken β-globin 5'HS4 insulator is CTCF independent (Yao et al. 2003), however *in vivo* silencer-blocking activity of CTCF has been identified to regulate the imprinted expression of alternative Wilms' tumour suppressor gene-1 (AWT1) and Wilms' tumour suppressor gene anti-sense transcripts (WT1-AS) (Hancock et al. 2007). Interesting connections have been made between the enhancer blocking capacity of insulator sequences and their interaction with the nuclear matrix. Several insulator sequences including the chicken β-globin HS4 and the *Drosophila* insulator gypsy are known to link chromatin to the nuclear matrix (Yusufzai and Felsenfeld 2004, Byrd et al. 2003). CTCF itself has been identified as a MAR-binding protein (Dunn et al. 2003) and association of CTCF with the nuclear matrix correlates with enhancer-blocking activity of insulators although this relationship is not absolute (Gombert et al. 2003). It has been proposed that insulators cannot contact the nuclear matrix in the absence of CTCF binding, which in turn leads to loss of enhancer-blocking activity, but again this can not be considered as a general rule. Significantly, the anti-silencing capacity of insulators may not require either CTCF or linking to the nuclear matrix (Hino et al. 2006).

With the exceptions of the mouse *β-globin* locus and imprinted gene loci, insulator activity is most commonly studied using *in vitro* cell culture systems. Importantly, this does not always reflect the *in vivo* function as CTCF-binding insulator sequences as defined *in vitro* can activate transcription *in vivo* (Vostrov and Quitschke 1997). The identification of new vertebrate proteins that interact with and regulate insulator activity will help to enhance our understanding of both how long-range gene regulation occurs, and may uncover mechanisms of CTCF-mediated insulator activity.

CTCF, enhancer-blocking and imprinted gene expression

The ability of CTCF to mediate the formation of chromatin loops *in cis* and interact with DNA in a CpG methylation-sensitive manner has been used as a model for the regulation of imprinted gene expression. The majority of genes in somatic cells are expressed from both maternally and paternally inherited alleles; genomic imprinting describes the inheritance of monoallelic gene expression in male and female cells where expression of a gene occurs exclusively from one allele depending on the parental origin, therefore imprinted genes have haploid expression despite the diploid nature of the

locus. In female cells, loci on one X chromosome are completely silenced (X-inactivation, see below). About seventy mammalian genes are known to be imprinted (Yoon et al. 2005). DNA methylation is an integral feature of imprinted gene regulation (Li et al. 1993) and until recently it was unclear how differential methylation patterns on these loci were maintained.

In mice, the most well characterised study of CTCF-mediated imprinted gene expression has come from analysis of the *Igf2/H19* imprinted locus. Insulin-like growth factor type 2 (*Igf2*) is an embryonic mitogen and *H19* generates a non-coding RNA of undefined function. A considerable proportion of imprinted genes produce non-coding RNAs, which may have important roles in the regulation of gene expression *in cis* (Sleutels and Barlow 2002). *Igf2* is positioned approximately 80Kb upstream of the *H19* on mouse chromosome 7. These factors are expressed solely from the paternal and maternal allele respectively (Murrell et al. 2004). Transcription at this locus is known to be dependent on two distinct regulatory elements; enhancers located downstream of *H19* are used to activate both *H19* and *Igf2* and a region of intervening sequence located 2.4Kb upstream of the *H19* transcriptional start site is critical for regulating imprinted expression of these genes. This element was termed the imprinting control region (ICR) or differentially methylated domain (DMD) and was later shown to bind CTCF in a methylation-sensitive manner (Leighton et al. 1995, Thorvaldsen et al. 1998, Hark et al. 2000, Bell et al. 2000). Importantly, binding of CTCF to the DMD occurs only on the maternal allele where this region is unmethylated and *H19* is transcribed. Co-ordinately, the *Igf2* gene is silenced on the maternal allele. CTCF interaction with the maternal DMD prevents the downstream enhancers activating *Igf2,* as such this element was proposed to be an insulator and was shown to have enhancer-blocking activity *in vitro* and *in vivo* (Hark et al. 2000, Bell et al. 2000). The *H19* DMD is an example of an enhancer-blocking element that does not associate with the nuclear matrix but matrix attachment of this locus does occur in a cell-type and parent-of-origin specific manner (Weber et al. 2003). Recent analysis of this locus has revealed CTCF interaction with the maternal ICR mediates clustering of the ICR with matrix attachment region (MAR) 3 (Kurukuti et al. 2006). On the paternal allele the DMD is hypermethylated and CpG methylation is incompatible to CTCF binding. Methylation of the paternal DMD coincides with silencing of *H19* by spreading of methylation repressive chromatin modifications onto the *H19* promoter. Activation of *Igf2* by the downstream enhancers apparently occurs by default.

Using 3C technology, it has been shown that the *Igf2/H19* locus forms chromatin loops that are in part mediated by CTCF (Murrell et al. 2004). Two distinct loops occur on the paternal and maternal alleles where differentially methylated regions (DMR) 1 and 2 interact with the unmethylated maternal and methylated paternal ICR respectively. The DMR1-ICR interaction positions *Igf2* into a silent loop where as on the paternal allele,

DMR2-ICR interactions bring the enhancers in close proximity to the *Igf2* promoter allowing transcriptional activation. It is unclear if CTCF binds to DMR2 but binding of CTCF to DMR1 has been shown by ChIP (Kurukuti et al. 2006). The importance of CTCF in mediating the imprinted expression patterns of these genes is demonstrated when the ICR binding CTCF is deleted. Deletion of the maternal ICR induces biallelic expression of *Igf2,* and this is not observed if the paternal ICR is absent. Loss of *Igf2* imprinting is also observed when levels of CTCF are reduced using RNA interference (RNAi) (Ling et al. 2006).

The loss of CTCF binding to the ICR either by mutation, deletion or knockdown correlates with a gain of methylation, which raised the question of whether DNA methylation is dependent on the absence of CTCF binding and if CTCF protects sequences from acquiring methylation. Two independent and apparently contradictory studies have addressed this issue. Reduction of CTCF in oocytes using RNAi revealed a requirement for CTCF in both the establishment and maintenance of differential methylation patterns of the *Igf2/H19* locus (Fedoriw et al. 2004). However, mutation of the four CTCF binding sites within the ICR in female and male germ lines did not result in a gain of methylation or indeed any change in methylation status of either paternal (hypermethylated) or maternal (hypomethylated) mutated alleles (Szabo et al. 2004). This study indicated that establishment of imprinting could be CTCF-independent but importantly, after fertilisation, the absence of CTCF binding led to acquisition of methylation on the maternal allele, linking CTCF to a role in methylation inhibition. Significantly, by mutating the CTCF binding sites, Szabó et al suggested that other proteins, which might bind CTCF sites such as BORIS, also could not be responsible for methylation patterns pre-implantation. This study does not rule out the possibility of other CTCF-dependent regulatory elements that may effect the establishment of imprinted gene expression of this locus (Cerrato et al. 2003). Likewise, a global loss of CTCF as done with RNAi will also affect other CTCF binding sites in the region and may be indicative of a more widespread effect of CTCF loss in nuclear organisation.

It is important to consider that not all imprinted gene loci are regulated in the same way. Indeed, even for a single imprinted locus, the pattern of expression can change depending on the developmental stage or particular tissue analysed. To conclude that CTCF regulates imprinted gene expression is an exaggerated generalisation of the story. CTCF has at least been implicated in the imprinted expression of several loci (Hikichi et al. 2003, Yoon et al. 2005, Hancock et al. 2007, Fitzpatrick et al. 2007) but the mechanisms of how CTCF functions and its precise involvement are far from being understood. ICRs are common to a number of imprinted genes and together with non-coding RNAs are thought to encompass how imprinted gene expression is regulated in general (Pauler and Barlow 2006).

CTCF and escape genes; X chromosome inactivation

Female mammalian somatic cells require dosage compensation to balance the expression of X-linked genes with that seen in males, and this is achieved by silencing one X-chromosome (for details see ch.1) (Lyon 1961). The first evidence for CTCF involvement in X-inactivation came from studies identifying CTCF binding sites in a region known as the choice/imprinting center located 5' of *Tsix* (Chao et al. 2002). Using chromatin immunoprecipitation, CTCF was found to bind to *Tsix* and it was proposed that binding of CTCF indicates the future active X-chromosome (Xa). Interestingly, CTCF binding sites also flank the promoter region of *Xist*. Silencing of the *Xist* promoter is mediated by full length *Tsix* as truncations impede hetrochromatinisation of the *Xist* promoter (Navarro et al. 2006). This implies a dual role for CTCF, as an activator of *Tsix*, but also required for *Xist* transcription that generates the *Xist* nuclear RNA, which initiates X-inactivation. Recent data reveal that binding sites for YY1 are in close proximity to CTCF binding sites within the X inactivation center and that YY1 is a necessary co-factor for CTCF activity in X-inactivation, particularly the transactivation of *Tsix* (Donohoe et al. 2007). It is not yet known if paired CTCF-YY1 sequences are a general element of imprinted or autosomal gene regulation.

The process of X-inactivation requires choice of which X becomes inactivated in female cells, and counting of the number of X chromosomes to ensure that male cells which only have one X keep their single X chromosome active. A 1.2Kb region of tandem repeats located 750bp downstream of the major *Tsix* promoter has been identified as critical for regulating counting (Vigneau et al. 2006). It was previously revealed that CTCF has several binding sites within this *DXPas34* element (Chao et al. 2002), however the mechanisms and function of *DXPas34* and CTCF binding remain unknown.

Not all X-linked genes on the inactive X are silenced. The proportion of active genes on the inactive X chromosome differs significantly between just seven genes in mice and 10-20% of genes humans (Cohen and Lee 2002, Carrel et al, 1999). A CTCF-dependent model for escape genes has recently been proposed as binding sites for CTCF are positioned 5' to three genes, human *EIF2S3* and mouse *Eif2s2x* and *Jarid1C* genes that remain active despite nearby inactive genes (Filippova et al. 2005). However, targeted insertion of cHS4, a known CTCF binding site and enhancer-blocker element, either side of the mouse *Hprt* locus in ES cells failed to protect the transgene against either random or imprinted X inactivation (Ciavatta et al. 2006). Although this data appears contradictory to that by Filippova et al, it is important to remember CTCF activity may be target site specific or more likely that additional sequences within the regions of the identified escape genes are required to establish transcriptional activity in the context of the inactive X chromosome.

Parallels between X-inactivation and autosomal imprinting are becoming increasingly apparent. As CTCF is known to mediate the imprinted gene expression of

some loci in a methylation-sensitive manner and differentially methylated regions within X chromosomes are now being identified, comparisons are intriguing (Boumil et al. 2006). Monoallelic expression is a characteristic feature of both X-inactivation and autosomal imprinting, however other loci including immunoglobulin genes, T-cell and olfactory receptors undergo allelic exclusion (Goldmit and Bergman 2004). The function of CTCF in mediating allelic exclusion and locus contraction is currently being investigated. Asynchronous replication timing is a feature of all monallelically expressed loci. Recent data suggest CTCF is involved in regulating the replication timing of the *Igf2/H19* locus, however it is not clear if CTCF-mediated replication timing occurs genome wide (Bergstrom et al. 2007). Interesting differences between X-inactivation and autosomal imprinting also occur. CpG DNA methylation strongly inhibits CTCF interaction with the autosomal imprinted *Igf2/H19* locus (Hark et al. 2000, Bell et al, 2000), however binding of CTCF to *Tsix* is only partially disrupted by CpG methylation (Chao et al. 2002).

Both X-inactivation and imprinted gene expression are connected to non-coding RNA transcripts. The silencing of one X chromosome requires production of a non-coding RNA, while a number of imprinted genes generate both sense and anti-sense non-coding transcripts. Non-coding RNAs are important regulators of gene expression, attracting the formation of heterochromatin including DNA methylation. The presence of non-coding RNAs is thought to be a key aspect of normal gene regulation (Holohan et al. 2007). CTCF interaction with intergenic transcripts has been linked to the regulation of the *DM1* locus associated with myotonic dystrophy (DM). CTG repeat expansion of the 3' non-coding region of the protein kinase *DMPK* gene and the accompanying heterochromatin is constrained by CTCF binding sites (Filippova et al. 2001, Cho et al, 2005). CTG repeats are a normal feature of this locus, however in pathological conditions, this repeat unit expands from between 5-38 repeats up to thousands. Repeat expansion carries accompanying heterochromatin modifications over the neighbouring *Six5* gene suppressing transcription (Otten and Tapscott 1995) and associated DNA methylation is thought to prevent CTCF binding thereby disrupting the insulator function of CTCF at this locus. It has been proposed elsewhere that CTCF may function as a barrier to non-coding transcription in *Drosophila* by blocking RNA polymerase II progression, and that this may facilitate the differential regulation of transcriptional units found in close proximity (Holohan et al. 2007).

CTCF and intrachromosomal interactions

The importance of intrachromosomal interactions has been appreciated for a long time. The classical example of functional trans-interactions is the generation of nucleoli. Around 400 ribosomal genes are arranged in tandem repeats that cluster at

defined positions within five pairs of chromosomes; 13, 14, 15, 21 and 22 in human diploid cells. These clusters are called nucleolus organising regions (NORs) and NORs on different chromosomes come together during interphase and help initiate formation of nucleoli (Anastassova-Kristeva 1977). Despite this, the mechanisms and implications of interchromosomal interactions are poorly understood. Techniques such as 3-C that enable analysis of chromatin interactions *in* cis have been adapted into chromatin conformation capture-on-chip (4-C) (Simonis et al. 2006), which together with DNA immunofluorescence (DNA-FISH) can identify chromatin interactions occurring *in* trans. Intrachromosomal interactions have recently been used to describe the regulated monoallelic expression patterns of interferon-γ and interleukin genes in distinct subsets of T-helper cells (Spilianakis et al. 2005). CTCF has been shown to mediate interchromosomal interactions between the *Igf2/H19* locus on mouse chromosome 7 and an intergenic sequence between the *Wsb1* and *Nf1* genes on mouse chromosome 11 (Ling et al. 2006). It remains to be seen if other loci also interact *in* trans in a CTCF-dependent manner. Random X-inactivation involving counting and choice mechanisms also requires trans X-X interaction and possibly X-autosome interactions (Bacher et al. 2006, Xu et al. 2006). CTCF is proposed to mediate the physical association between X-chromosomes (Xu et al. 2007). A fascinating common theme of known examples of trans-interactions with exception of NOR clusters is monoallelic expression. It remains to be seen how extensive the relationship between regulating monoallelic expression and chromosomal interactions *in trans* becomes and whether or not CTCF is a common factor in both.

CTCF; a tumour-suppressor gene?

In human cells, CTCF maps to the long arm of chromosome 16, in a region frequently deleted in sporadic breast and prostate tumours, 16q22.1 (Filippova et al. 1998). As such CTCF was proposed to be a tumour suppressor gene. CTCF expression is closely linked to cell-cycle regulation and there is evidence suggesting CTCF controls many genes involved in cell-cycle progression. Identified mutations in CTCF effect the zinc-finger region and binding to growth promoting but not growth-neutral promoters (Klenova et al. 2002). The precise role for CTCF controlling cellular proliferation is unclear, as ectopic expression of CTCF results in cell growth inhibition without apoptosis (Rasko et al. 2001, Docquier et al. 2005) and overexpression in K562 cells induces erythroid differentiation together with restricting proliferation. In at least two loci, CTCF binding is important for protecting promoters of tumour suppressor genes from acquiring methylation and repression (De La Rosa-Velazquez et al. 2007, Butcher et al. 2004). Current data regarding CTCF as a potential tumour suppressor gene seem to link mis-regulation of CTCF rather than CTCF mutations to be involved in tumourigenesis. Loss

or mutations of CTCF transcripts occur as a very late marker in Wilms' tumour development (Hancock et al. 2007). Interestingly, mutations in CTCF are rare and not correlated to loss of imprinting of *Igf2* in Wilms' tumours (Cui et al. 2001). Analysis of CTCF mutations in invasive breast carcinomas and familial breast cancer revealed no specific correlation and concluded CTCF mutation is not a critical risk factor (Aulmann et al. 2003, Zhou et al. 2004). Cytoplasmic accumulation of CTCF was found in the majority of invasive breast carcinoma analysed, however this mis-localisation of CTCF did not show a relationship to tumour type (Rakha et al. 2004). Clearly, further detailed analysis of CTCF deficient cells and tumours are required to understand the complexities linking CTCF and tumourigenesis. To date, the effect of post-translational modifications on CTCF activity have not been associated with tumours. CTCF is expressed in tumour cell lines such as K562 and HeLa cells, suggesting CTCF modifications or mis-targeting opposed to loss of CTCF occurs in malignant cells. Expression of the closely related CTCF protein BORIS is found to be over-expressed in cancer and is expressed in human tumour cell lines (Hong et al. 2005). The significance of BORIS expression in cancer is unknown, but may deregulate expression of CTCF target genes by competing for CTCF target sites. Loss of CTCF is not enough to induce activation of BORIS in somatic cells (this thesis).

Protein partners of CTCF

Given the numerous roles for CTCF, especially mediating histone modifications and chromatin structure, you could expect this to be assisted by many different cell-type specific or gene-specific interaction partners. A surprisingly low number of protein partners have been identified and characterised as specifically interacting with CTCF. To date, known CTCF-interacting proteins are ubiquitously expressed, as is CTCF.

Y-box binding protein 1 (YB-1) is an RNA/DNA-binding factor that is proposed to mediate transcriptional repression in a co-operative manner with CTCF (Chernukhin et al. 2000). Functional association of CTCF-YB1 complexes are not fully understood, however correct expression of the serotonin transporter gene (*5-HHT*) requires YB1 binding to an intronic variable number tandem repeat (VNTR) motif, and this binding is disrupted by CTCF (Klenova et al. 2004).

Kaiso is a zinc-finger transcription factor of the POZ (pox virus and zinc-finger) family. Target sites for Kaiso have been identified close to CTCF binding sites. Significantly, Kaiso interactions with DNA are not CpG methylation sensitive. Within the human β-globin locus, one unmethylated Kaiso site is present near the CTCF-mediated insulator 5'HS5, and binding of Kaiso prevents enhancer-blocking activity by CTCF possibly by disrupting binding of CTCF to its target site (Defossez et al. 2005). Interestingly, evidence suggests that Kaiso is able to recognise aberrantly CpG methylated CTCF binding sites in the promoter of the human *retinoblastoma* (*Rb*) gene,

leading to epigenetic silencing of the promoter (De La Rosa-Velazquez et al. 2007). It is not clear if Kaiso interaction is a general repressor of transcriptional activation or enhancer-blocking activity by CTCF or if this is restricted to modify the functions of CTCF in a temporal or cell-type specific manner.

CTCF has been shown to co-purify with nucleophosmin (B23), which is a protein associated with the granular component of the nucleolus (Yusufzai et al. 2004). Localisation of CTCF in nucleoli is observed in cells stimulated to go into terminal differentiation or growth arrest and apoptosis (Torrano et al. 2006). Both CTCF and B23 are associated with insulator sites *in vivo*, and these sites can be tethered to the nucleolus via CTCF (Yusufzai et al. 2004).

Eukaryotic RNA polymerase II is a multi-subunit complex responsible for transcription of all protein coding genes and some non-coding RNAs. The largest subunit has been shown to interact with CTCF *in vivo* (Chernukhin et al. 2007). Differential phosphorylation, either hypophosphorylation (IIa) or hyperphosphorylation (IIo) of the pol II large subunit occurs depending on whether pol II is associated with initiation (IIa) or elongating (IIo) transcripts. Co-immunoprecipitations (CoIP) revealed CTCF interacts with both IIa and IIo types, although a slightly improved affinity was observed for IIa, but this may be cell-type or cell cycle dependent. Similar CoIP experiments in this study suggested CTCF is not precipitated from nuclear extracts of HeLa cells using antibodies to histone H2A and H3, p53, Sp1, TBP or Rb1. CTCF also does not form a complex with TFIIH or MYC (Chernukhin et al. 2000).

SNF2-like chromodomain helicase protein CHD8 is a member of the CHD chromatin remodeling family of proteins that include Mi-2β (CHD4). Pull-down analysis showed direct interactions between CTCF and CHD8 *in vitro* and *in vivo* and both factors are required for enhancer-blocking activity of *H19* DMD in transfection assays (Ishihara et al. 2006). Importantly, when CHD8 levels are reduced by knockdown, CTCF remains bound to the *H19* DMD indicating that imprinted expression of *Igf2/H19* is dependent on both CTCF and CHD8, revealing the first direct mechanistic association between imprinting and chromatin remodeling. Knockdown of CHD8 also induced changes in DNA methylation and hypoacetylation of histones around CTCF binding sites.

YY1 is the first protein shown to interact with CTCF and mediate the function of CTCF in X-inactivation, co-operating in the transcriptional activation of *Tsix* (Donohoe et al. 2007) (see above).

One particular class of potential CTCF-interacting proteins is conspicuous by its absence, histone acetyltransferases (HAT). The 3'HS1 CTCF binding site of the mouse *β-globin* locus is positioned within a region of condensed chromatin, but is itself a site of histone acetylation (Bulger et al. 2003). Local gains of repressive histone modifications occurred at all known CTCF binding sites within the mouse *β-globin* locus when CTCF

was depleted (Splinter et al. 2006). This suggests CTCF can attract HAT complexes, which maintain an open chromatin structure, however no direct evidence for CTCF and HAT interactions have been identified. The loss of CTCF may have disrupted other as yet unknown protein interactions either directly linked to CTCF or protein complexes that bind close to CTCF such as USF (West et al. 2004), causing the gain of histone methylation at CTCF binding sites. Considering loss of CTCF induces conformational changes to the *β-globin* locus, it is possible that CTCF-mediated looping brings elements with HAT activity into the region of CTCF binding sites. It has been reported previously that CTCF interacts with the transcriptional co-repressor protein complex Sin3A, contributing to the transcriptional repressive action of CTCF by associated HDAC activities (Lutz et al. 2000). However, independent analysis to identify CTCF-HDAC interactions by co-immunoprecipitation using breast cancer cells failed to confirm specific CTCF interactions with either HDAC1 or HDAC2, but did not exclude the possibility of CTCF interaction with other co-repressor complexes harbouring HDAC3 such as N-CoR/SMRT (Dunn et al. 2003). N-CoR has been identified as a mediator of DNA methylation-dependent transcriptional repression through its interaction with Kaiso (Yoon et al. 2003), which provides an interesting mechanistic link between CTCF, Kaiso and epigenetic silencing.

One would imagine cell-type specific proteins must also be potential binding partners. Using homologous recombination in ES cells, a biotin tag has been engineered into the endogenous CTCF locus in mice (SvdN unpublished), which will enable identification of specific CTCF-interacting partners to be identified and characterised from any cell type or developmental stage.

Nuclear localisation and dynamics of CTCF, interphase and mitosis

CTCF is expressed constitutively as a nuclear protein in the vast majority of cells at all stages of development. Localisation of CTCF during interphase appears in a distinct pattern, which is largely diffuse throughout the nucleus and excluded from the nucleolus, but accumulation patterns can be seen as if CTCF is concentrated in many small areas (Kantidze et al. 2007). As CTCF is known to mediate chromosomal interactions and transcriptional activity *in trans,* it is tempting to speculate that clusters of CTCF binding sites accumulate in the nucleus to regulate nuclear architecture as has been proposed for the *gypsy* insulator in *Drosophila* (Yusufzai et al. 2004). It is not yet known if CTCF functionally clusters in this way, indications from CTCF deficient cells suggest no obvious structural abnormality based on DAPI signals, but this is not to suggest a more detailed analysis would not find impairments.

The subnuclear localisation of CTCF is not effected by phosphorylation (Klenova et al, 2001) although localisation of CTCF to the nucleolus is observed in a poly(ADP)ribosylation dependent manner (Torrano et al. 2006). Nucleolar accumulation of CTCF was observed when cells are induced to differentiate *in vitro* and this is dependent on both poly(ADP)ribosylation, ongoing rDNA transcription and protein synthesis, implying a dynamic regulation of CTCF in and out of the nucleolus. The *in vivo* functional significance of nucleolar CTCF remain questionable. In UR61 rat neuron-like cells, transfection of GFP-CTCF revealed specific localisation of CTCF within the nucleolus. CTCF has been shown to co-localise and co-purify with nucleophosmin (Yusufzai et al. 2004), a component of the granular compartment. Significantly, CTCF can specifically interact with UBF *in vitro* and *in vivo* (SvdN unpublished), but no co-localisation with UBF was observed in UR61 cells, indicating this interaction may reflect the type of cells analysed. Chicken erythrocytes do not enucleate as mammalian red blood cells do, but genome wide silencing does occur, as such they are an interesting model for chromatin structure and nuclear organisation during differentiation. Kantidze et al (Kantidze et al. 2007) describe compartmentalisation of CTCF in both embryonic erythroblast and mature erythrocyte nuclei, but not nucleoli. Comparisons between these studies (Kantidze et al. 2007, Torrano et al, 2006), highlight the important difference of transformed cancer cell lines and normal cells in culture. The physiological and epigenetic distinctions between terminally differentiating cells that stop proliferation and differentiation and differentiating cells, which remain transcriptionally active, should also be considered. CTCF and other transcription factors become dissociated from DNA in terminally differentiated erythrocytes, whereas in proliferating erythroblasts, the majority of CTCF is bound to DNA (Kantidze et al. 2007).

Transcriptional activity is not a requirement for CTCF binding to all target sites. During mitosis, a proportion of CTCF remains bound to particular target sites despite transcriptional inactivity (Burke et al. 2005). The localisation and dynamics of CTCF during mitosis is of considerable interest given the growth arrest phenotype in both CTCF over-expressing and deficient cells. CTCF associates with mitotic chromosomes throughout mitosis, and by immunofluorescent analysis and biochemical fractionation shown to co-localise with centrosomes and the midbody (Zhang et al. 2004). The function of CTCF binding to mitotic chromosomes is unclear; CTCF may remain bound to enforce a memory pattern of epigenetic modifications. Alternatively, CTCF might have a structural role, mediating the tight condensation of chromosomes or their correct alignment required for mitosis to occur. Interestingly, the association of CTCF with mitotic centrosomes is microtubule independent, as treatment with nocodazol, a microtubule depolymerising factor, does not disturb CTCF-centrosomal interactions (Zhang et al. 2004). The precise role of CTCF during mitosis needs careful consideration

as under certain cell culturing conditions, CTCF deficient cells can divide (this thesis). Importantly a significant amount of GFP-CTCF is released from chromatin as chromosomes align at the start of metaphase, and this 'free' GFP-CTCF is divided between the two daughter cells and the newly formed nuclei (this thesis).

The number and location of potential CTCF binding sites is becoming a topic of great speculation and significance. Identification of CTCF binding sites in the genome has previously been perturbed by the lack of a primary structure consensus sequence for CTCF, however, through combinations of two-dimensional electrophoretic mobility shift assays (2D-EMSA), chromatin immunoprecipitation (ChIP) and microarrays (ChIP-on chip), there is increasing information which maps CTCF binding sites in humans (Mukhopadhyay et al. 2004, Vetchinova et al. 2006, Kim et al. 2007). Ascertaining a figure for the number of CTCF binding sites present in the genome is useful to assess the possible function of CTCF in organising higher order chromatin interactions. To date, estimations range between 4000 and 30,000 CTCF binding sites in humans (Mukhopadhyay et al. 2004, Vetchinova et al. 2006, Kim et al. 2007). According to an average chromatin domain size estimation of 80-300Kb, this would equate to 10,000-40,000 insulator boundary elements in the human genome (Heng et al. 2001, Vetchinova et al. 2006).

Functionally related proteins?

To date, only one CTCF-like protein has been identified in mammals. *Brother of the Regulator of Imprinted Sites* (BORIS) is homologous to CTCF only in the zinc-finger DNA-binding domain, the remaining sequences shares very little identity to CTCF (Loukinov et al. 2002). BORIS is able to bind to CTCF target sequences in tumour cells (Vatolin et al. 2005), and is expressed at a stage during spermatogenesis where CTCF expression is reduced. This stage coincides with the erasure and re-establishment of methylation patterns. As such, BORIS has been implicated in mediating reprogramming of epigenetic modifications, notably erasure of methylation in male germ cells (Loukinov et al. 2002). However, the role of BORIS or CTCF in regulating imprinted methylation patterns in germ cells is controversial (Szabo et al. 2004). BORIS is also implicated in demethylation of cancer-testis genes in somatic cells and the *de novo* methylation of imprinted genes in the male germ line (Vatolin et al. 2005, Jelinic et al. 2006). To date, BORIS is known to have a highly restricted pattern of expression, confined to 5-methylcytosine-deficient adult male germ cells, human tumour cell lines and CTCF negative cells (Vatolin et al. 2005). The reciprocal pattern of expression with CTCF in spermatocytes is interesting, no known BORIS-specific target sites have been identified implying some sort of regulation or feedback between the two proteins exist. Significantly, there is no direct correlation to CTCF absence and BORIS expression in

somatic cells; CTCF deficient T-cells both in the thymus and periphery do not aberrantly express BORIS (this thesis).

What actually regulates CTCF?

During the cell cycle, *Ctcf* mRNA levels increase at S-G_2. However, ectopic expression of *Ctcf* leads to arrest at apparently indiscriminate points of the cell cycle (Rasko et al. 2001, Klenova et al. 1998). Loss of CTCF effects the G_1-S transition, resulting in an extended G_1 phase in T-lymphocytes (this thesis). The factors and modifications of CTCF protein that determine its regulation are not fully understood. Determining specific regulatory factors that control the transcription of *Ctcf* is complicated by the fact CTCF is an essential protein required for cell viability; analysis such as chromatin immunoprecipitation can identify protein complexes that associate with the *Ctcf* promoter and other regulatory regions, but the follow up mutational analysis can be problematic. A number of potential binding sites have been identified in the *Ctcf* promoter, which are conserved in vertebrates, including potential binding sites for YY-1, GATA-1 and p53 (Pugacheva et al. 2006) although direct evidence for the functional interaction of any of these proteins with *Ctcf* regulatory elements is lacking. The YY-1 element is strictly conserved from zebrafish to humans and is associated with activation of *Ctcf*.

Few cells have been identified as or modified to be CTCF deficient *in vivo*. Absence of CTCF results in global impairment of cell regulatory systems leading to cell death, as such very little is known about factors that influence the expression of CTCF *in vivo*. Expression levels of CTCF require strict regulation; studies indicate severe impairment of cellular functions, which are not necessarily different, depending on overexpression or knockdown/knockout (Qi et al. 2003, Li. 2007). CTCF function and activity may differ depending on the cell type and stage in differentiation you analyse. Down-regulation of CTCF is required for differentiation of human monocyte-derived dendritic cells (DCs), and enforced expression in bone marrow-derived DCs in irradiated mice impairs proliferation and promotes apoptosis. However, enforced CTCF expression also resulted in an increase of plasmacytoid DCs, suggesting CTCF is specifically involved in regulating differentiation of dendritic cells. The mechanisms of this are currently unknown (Koesters et al. 2007). *Ctcf* knockdown inhibits differentiation of human myeloid leukaemia cells (Torrano et al. 2005). However, conditional deletion of CTCF as described in this thesis has revealed under specific conditions, T-lymphocytes can undergo TCR rearrangements, divide and differentiate in the absence of CTCF (this thesis). The indication that CTCF deficient cells can divide when given appropriate signals will be of significant benefit to the understanding of CTCF function and the regulation of this protein.

Chapter 3

Haematopoiesis; Erythrocytes and T-lymphocytes

Chapter 3: Haematopoiesis; Erythrocytes and T-lymphocytes.

Introduction to haematopoiesis

An extensive overview of the haematopoietic system, its origins and regulatory factors is beyond the scope of this thesis, however general information is useful for understanding the data presented. The haematopoietic system is used in this thesis as a model to study the function of CTCF in two distinct cell lineages, erythrocytes and T-lymphocytes.

Haematopoiesis is the process of generating all cellular populations and their derivatives that are found in blood. Blood is a multifunctional tissue, the cellular composition of which can be broadly divided into three categories of cell types, erythroid-megakaryocytic, lymphoid and myeloid. This categorisation as described by Katsura (Katsura, 2002) is controversial. Erythroid-megakaryocytic cells are derived from myeloid progenitors, therefore the lymphoid-myeloid distinction of cells is more generally considered. At least twelve different functional categories of mature cells and cell fragments (fig 3.1) are derived from a single pluripotent self-renewing cell called the haematopoietic stem cell (HSC) (Lemischka 1992, Katsura, 2002).

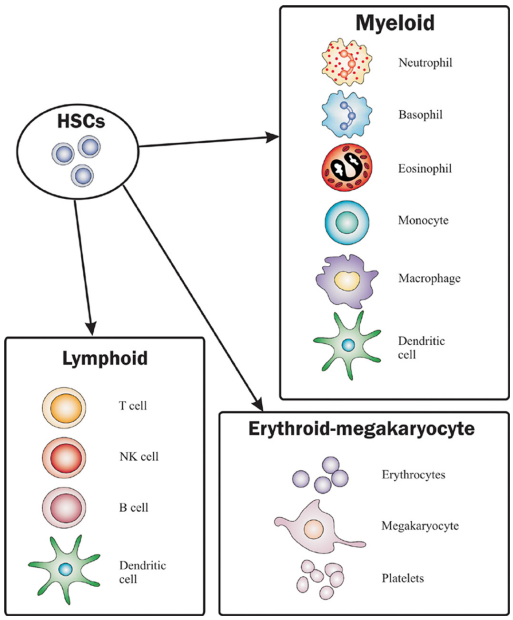


Figure 3.1. Classification of different haematopoietic cell types
Adapted from (Katsura 2002).

HSCs are very rare cells (1-10 cells per 100,000 cells) in the bone marrow (BM) (Abkowitz et al. 2000). The self-renewing HSC potential is maintained by one daughter cell following division of the original pluripotent stem cell, while the other daughter cell progresses along a pathway of differentiation.

The first step of differentiation gives rise to cells that retain the potential to become any haematopoietic cell type, but have crucially lost their capacity to self-renew. These cells are termed mulit-lineage precursors (MLPs). It is widely documented that following MLPs, differentiation proceeds into two separate lineage-defining precursors, the common lymphoid progenitor (CLP) or common myeloid progenitor (CMP) (Prohaska et al. 2002). CLPs are described as committed precursors with the potential to generate all cells of the lymphoid lineage including natural killer (NK) cells *in vivo* (Kondo et al. 1997). Further lineage restriction occurs as CLPs differentiate to B/T/NK cells. The precise nature of the B/T/NK lineage precursor remains controversial (Zuniga-Pflucker and Schmitt 2005). The model of CLPs is not without question and the actual identity of a lymphoid progenitor remains a matter of debate. Surprisingly CLPs defined in this context represent a very small percentage (0.02%) of BM cells. CMPs by comparison comprise 0.2% of total BM cells. They are defined as progenitors able to give rise to monocytes, granulocytes, erythrocytes and megakaryocytes based on in vitro colony forming unit (CFU) culture and *in vivo* transplantation assays (Akashi et al. 2000). As such they are also referred to as colony forming unit-granulocyte, erythrocyte, monocyte, megakaryocyte (CFU-GEMM) cells (Richmond et al. 2005) that differentiate to granulocyte/monocytes (GMPs) and megakaryocyte/erythroid (MEPs) cells (Akashi et al. 2000). The expression of specific transcription factors appear key to several cell fate decisions. Studies of a dominant negative form of the *Ikaros* gene revealed a lack of both B- and T-cells, suggesting the existence of a common precursor for both cell types (Georgopoulos et al. 1994). The function of *Ikaros* as a defining feature of CLPs is however inaccurate, as T-cells can exist *in vivo* in the absence of Ikaros expression (Wang et al. 1996). The precise differentiation potential of CLPs/CMPs remain under question particularly with respect to monocyte differentiation that has been observed both from CLPs and CMPs (Kee and Murre 1998). The identity and reality of CLPs as lymphoid restricted is challenged by the identification of bipotent myeloid/lymphoid precursors, so called common myeloid lymphoid precursors (CMLPs) (Kawamoto et al. 1997, Katsura, 2002) although these are only present at an early stage of development.

Differentiation into functional cell types proceeds along an ordered lineage pathway, which at a certain points becomes restricted and committed. The self-renewal, differentiation and survival potential of stem cells, and the development of subsequent progenitors and committed cell types is tightly regulated by cooperative signals from surrounding cells, cytokines and intrinsic expression of particular transcription factors.

Recent findings have proposed a hierarchical expression of transcription factors that is essential for coordinated HSC differentiation into progenitor cells and particular lineage fates. Transcription factors such as PU1, GATA-1, -2, and Pax5 have each been shown to essentially mediate the lineage direction of given precursor cells (Zhu and Emerson 2002, Akashi, 2005). Ectopic expression of GATA1 in CLPs is sufficient to redirect differentiation towards the megakaryocyte/erythroid lineages (Iwasaki et al. 2003). Interestingly, reprogramming of non-haematopoietic HeLa cells to express erythroid-specific genes is also induced by ectopic expression of GATA-1 (Layon et al. 2007) suggesting a significant role for GATA1 in mediating transcription of erythroid-specific genes, even in cells that seemingly have no priming potential for red cell expression profiles. These studies suggest that transcription factors may be the primary determinants of lineage plasticity.

Primitive and definitive haematopoiesis are distinct waves of haematopoietic cell production at different stages of development, which occur in several distinct tissues throughout life. Primitive haematopoiesis in mice originates from mesodermal precursors that form the blood islands in the visceral yolk sac. Blood vessels originating from endothelial cells (angioblasts) and haematopoietic cells are proposed to derive from a common precursor known as the hemangioblast (Huber et al. 2004). The yolk sac is the first site of haematopoiesis in mammalian embryos (Palis et al. 1999). Primitive haematopoiesis begins around mouse embryonic day (E) 7.25. The majority of haematopoietic cells generated are large, nucleated primitive erythroblasts. Primitive macrophages and some megakaryocytes are also generated (Lensch and Daley 2004). Primitive pro-erythroblasts circulate to and divide within the newly formed embryonic vascular system about a day later. Progressive differentiation is accompanied by gradual loss of proliferation, repression of gene expression and reduction in cell size, culminating in cell cycle arrest and enucleation (McGrath et al. 2003, {Kingsley, 2004). Primitive erythrocytes are known to remain in circulation for a short time after birth.

Independent to yolk sac-derived haematopoiesis, definitive haematopoiesis derives from distinct HSCs located within the embryonic aorta, gonads, mesonephros (AGM) around E10.5 (Medvinsky and Dzierzak 1996, Zhu, 2002), where they proliferate before migrating to the fetal liver and to a lesser extent the spleen. Interestingly, primitive haematopoietic cells are unable to support repopulation of irradiated adult recipient mice and thus can not be considered as true HSCs. In contrast, AGM derived HSCs can fully repopulate the complete haematopoietic system, demonstrating these stem cells are genuinely pluripotent. By E12.5, circulating primitive erythroblasts are accompanied by enucleated definitive red blood cells released from the fetal liver, which becomes the major site of haematopoiesis until birth, after which the spleen and bone marrow permanently function as the sites of definitive haematopoiesis in mice. Interestingly,

definitive haematopoietic progenitors are also found within the placenta, and this site was recently shown to be a new HSC niche (Alvarez-Silva et al. 2003, Ottersbach, 2005). T-lymphocytes are the only differentiated haematopoietic cell type not generated in the bone marrow. Progenitor cells derived from HSCs migrate to the developing thymus during embryogenesis and continue to populate the thymus into adult life. Commitment to the T-cell lineage is defined in the thymus (see below).

Haematopoiesis: erythropoiesis

Erythrocytes constitute the vast majority of cells in the blood and are responsible for the transport of oxygen and carbon dioxide around the body. As terminally differentiated cells, erythrocytes can not divide, but they also do not persist indefinitely. In mice, erythrocytes circulate on average for 60 days before being consumed by specialised cells in the spleen and liver. The haemeoglobin-associated iron that becomes recycled within the body. A continual production of red blood cells (erythropoiesis) is therefore required to maintain the supply of oxygen and exchange of carbon dioxide in tissues. The process of erythropoiesis is characterised by progressive differentiation of committed erythroid precursor cells towards the enucleated red blood cell (fig 3.2 Dessypris 1998).

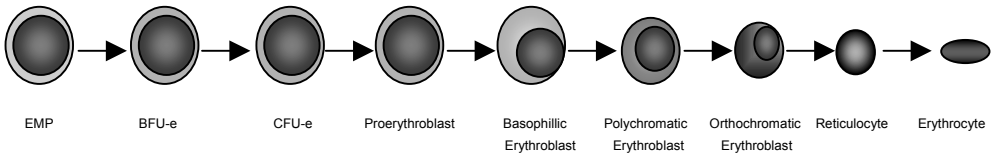


Figure 3.2. Adapted from ([215] Dessypris 1998).

Erythroid/megakaryocyte progenitors (EMP) contribute to both the erythroid and megakaryocyte lineages. Erythroid commitment is sealed at the BFU-e and CFU-e stage. Proerythroblasts indicate the initiation of terminal differentiation, which concludes with enucleated erythrocytes.

In vitro assays of bone marrow cells in semi-solid medium revealed distinct colonies emerging from different lineage restricted precursor cells according to type of growth factor present (Wong et al. 1986). Burst-forming units-erythroid (BFU-e) and colony forming units-erythroid (CFU-e) retain substantial proliferative capacity, which progressively declines as differentiation occurs. Patterns of gene expression define progression of the red cell lineage, in particular the characteristic production of α- and β-globin proteins that combine to form haemoglobin, which accumulates throughout

differentiation. The loss of cellular components such as the endoplasmic reticulum and mitochondria occurs and macrophages that accompany red blood cell precursors ingest the ejected nucleus, leaving a cell almost entirely composed of haemoglobin. Macrophages are also important as supporting cells, regulating correct transcription factor expression and lineage commitment during erythropoiesis at the site of erythroblast islands, where definitive erythropoiesis occurs within the fetal liver and bone marrow (Iavarone et al. 2004).

Haemoglobin, globin gene structure and expression

Haemoglobin is a hetero-tetrameric protein that consists of two α-like and two β-like globin chains, each of which is attached to one heme group containing an iron atom in the center. Different α- and β-globin genes are expressed at distinct stages of mammalian development to coincide with the requirement for haemoglobin with higher oxygen affinity in the embryo compared to the adult. Each of the globin gene loci, including the specialised muscle-specific globin protein myoglobin, is believed to derive from a common ancestral locus around 450 million years ago (Hardison 1996). The factors regulating correct temporal and spatial expression of globin genes are of great medical significance. Hereditary blood disorders, such as sickle cell anaemia and thalassemias are the most common genetically inherited diseases with over 250 million people world wide known to carry mutations or deletions within globin loci. Studies of such disorders have identified several important cis- and trans-acting factors that contribute to transcriptional regulation of these multi-gene loci, in particular the identification of the β-globin locus control region (LCR) (Grosveld et al. 1987). The functional significance of the β-globin LCR will be discussed later.

In mammals, globin gene structure is highly conserved, especially between humans and mice. Interestingly, epigenetic modifications and the chromatin environment around the α- and β-globin loci are significantly different, indicating distinct regulatory mechanisms are used to maintain the specific temporal yet coordinated expression patterns that occur (table 3.1) (Brown et al. 2001). Furthermore, the globin genes positioned 5' in both loci are expressed in the embryo; where the globin genes located 3' encode the prominent adult α- and β-globin proteins (fig 3.3). This led to a model of globin gene switching, which initially linked the temporal transition between embryonic and adult globin expression with the change from primitive to definitive erythropoiesis (Stamatoyannopoulos 1991, Ingram, 1972). However this model is no longer valid as primitive enucleated erythrocytes are found as early as E12.5 in mice (Kingsley et al. 2004).

Features of chromatin at the β-globin locus	Non-erythroid cells	Erythroid cells
Hyperacetylation H3 and H4	-	+
DNase hypersensitivity of LCR	-	+
Replication timing	Late S-phase	Early S-phase
Location in interphase nucleus	Heterochromatin	Euchromatin
Transcriptional activity of globin genes	Repressed	Active or poised

Table 3.1. Differences between chromatin states of the β-globin locus during differentiation. Adapted from (Schubeler et al. 2000, Francastel, 2001).

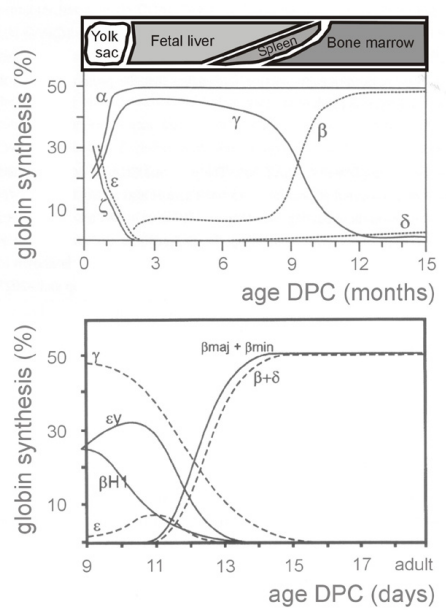


Figure 3.3. Developmental patterns of globin gene expression

The developmental regulation of human globin synthesis is shown (top panel). α- and ζ-genes are located in the α-globin locus (see text for details). The lower panel shows the expression pattern of individual mouse globin genes (solid lines) and human β-globin transgenes when expressed in mice (dashed line).

Both α- and β-globin gene loci are readily accessible model systems of spatial and temporal transcriptional regulation and cellular differentiation *in vivo*.

α-globin locus

The mouse α-globin gene locus is positioned in the telomeric region of chromosome 11 and consists of three functional developmentally regulated genes: one embryonic ζ globin gene, primarily expressed in embryonic blood cells and two fetal/adult α (α1, α2) genes, which are also expressed in embryonic blood and maintain expression in fetal liver and adult bone marrow. In mouse and human, α-like globin genes are expressed in the order of their chromosomal arrangement (5'ζ-α1-α2-3') (Zhou et al. 2006, Leder, 1981) and undergo one switching event, the change from ζ to α globin. Interestingly, the α-globin gene cluster is in an open chromatin environment in both erythroid and non-erythroid cells, and has overlapping genes that are both erythroid-specific and ubiquitously expressed (Sjakste et al. 2000, Litt, 2001).

A pre-existing sub-conformation consisting of tissue-restricted chromatin loops has been proposed, where by the ubiquitously expressed genes within the α-globin locus are constitutively clustered together in non-erythroid cells. Differentiation towards the erythroid lineage changes the configuration of the locus, allowing α-globin genes to cluster with the positive regulatory elements of the active housekeeping genes. As such, the mouse α-globin locus is proposed to form a transcription factory where highly expressed yet distinctly regulated genes co-localise with common regulatory elements (Caron et al. 2001). Erythroid-specific transcription factors such as GATA1 and NF-E2 have been shown to interact with the major regulatory element HS26 and α1/α2 promoters, which could suggest a role for GATA1 in mediating long-range chromatin interactions.

β-globin locus

In contrast to the α-globin locus, mouse β-globin genes are not expressed in the order of their chromosomal arrangement (Kingsley et al. 2006). A similar but not identical conservation of the β-globin locus is observed between human and mouse loci. In humans, five β-like globin genes are developmentally regulated and expressed in their chromatin order 5' ε-^{G}γ-^{A}γ-δ-β 3'. Conversely in mice, the β-globin locus consists of four functional genes ordered 5' εγ-βH1-β1-β2 3' (Trimborn et al 1999), however βH1 is the first transcribed gene, followed by a switch to εγ expression before generation of β1 and

β2 transcripts (Kingsley et al. 2006). β1 and β2 are also referred to as β-major and β-minor respectively. The surrounding chromatin environment also differs between α- and β-globin gene loci. In both mouse and human, the β-globin gene cluster is embedded within genes encoding olfactory receptor proteins on chromosome 7. As such, the β-globin locus provides a mechanism of understanding how chromatin modifications effect differential temporal and spatial expression of a multi-gene locus within the context of distinctly regulated neighbouring genes in mammals.

Central to the control of β-globin expression in all vertebrates is an array of erythroid specific DNase I hypersensitive sites located 5' to the locus which form the Locus Control Region (LCR). Additionally, in mice, two more regions of DNase I hypersensitivity are positioned outside of the locus within olfactory receptor genes 85/84kb and 62/60kb 5' of the εγ cap site (Farrell et al. 2000, Bulger, 2003). Both human and mouse β-globin loci are flanked by 3' hypersensitive sites. The single 3'HS1 is located nearly 70kb downstream of εγ (Tuan et al. 1985). Specific *cis* regulatory regions, such as promoters, enhancers and silencers individually influence transcriptional activity of each of the β-globin genes. Individual gene promoters and enhancers are characterised by DNase I HSs that are both erythroid and developmentally restricted. Binding sites for erythroid specific transcription factors such as GATA1, erythroid Krüppel-like factor (EKLF) and NF-E2 are found in the promoter regions, and regulate local chromatin structure by attracting histone acetyltransferases such as CREB-binding protein (CBP)/p300 (Zhang and Bieker 1998, Hung, 2001, Letting, 2003, Blobel, 1998). Interestingly, binding sites for ubiquitously expressed factors such as Yin Yang 1 (YY1) and SP1 are also found in the ε- and γ-globin promoters and the LCR. YY1 is a truly multifunctional protein that interacts with both transcriptional activator and repressor complexes (Affar el et al. 2006). The significance of YY1 in regulating the β-globin locus and transcription remains unclear. In contrast, regulation of β-globin expression has been linked to the phosphorylation status of SP1 (Feng and Kan 2005). Phosphorylation is acquired upon differentiation and the accompanying switch from γ- to β-globin expression, furthermore phosphorylated SP1 no longer binds to target HS2, HS3, HS4 and β-globin promoter sites suggesting that SP1 is required to negatively regulate expression of β-like globin genes.

In humans, intergenic sequences between γ- and δ-globin genes are shown to harbour a regulatory element that regulates γ^A to β-globin switching (Chakalova et al. 2005). This element is associated with a large protein complex consisting of Ikaros, and chromatin remodelling proteins, which may function in a context dependent manner (O'Neill et al. 1999, O'Neill, 2000). This data suggests LCR-gene contacts can be regulated by conformational changes 5' to the δ-globin gene, an area currently not known to contain CTCF target sites. Analysis of deletions that remove the Ikaros binding

element and other flanking regions indicate β-globin gene expression is repressed and γ-globin expression persists in adults (Chakalova et al. 2005). The timing of globin gene switching is disturbed in the absence of Ikaros in both humans and mice (Lopez et al. 2002).

Several matrix attachment regions (MARs) have been identified within the β-globin locus and have been proposed to mediate globin gene switching (Yan and Qian 1998, Case, 1999). Special A-T-rich binding protein 1 (SATB1) is predominantly expressed in T-cells but is known to interact with a regulatory element 3' to the γ^A-globin gene (Cunningham et al. 1994). Recently, SATB1 has been shown to bind MARs at LCR HS2 and within the ε-globin promoter in human erythroleukemia K562 cells (Wen et al. 2005). In this context, SATB1 forms a complex with the transcriptional co-activator CBP and is a positive regulator of ε-globin expression, attracting H3 and H4 acetylation and reducing H3-K9 methylation. Concurrently, γ-globin gene activity is repressed via increasing levels of histone methylation with no apparent change in acetylation. Reducing SATB1 expression resulted in reciprocal data, suggesting SATB1 and associated proteins are key regulators of differential ε- and γ-globin expression. CTCF is known to interact with MAR sequences and currently four identified CTCF binding sites (HS-85, HS-62, LCR HS5 and 3'HS1) are present in and around the mouse β-globin locus (Farrell et al. 2002, Bulger, 2003). It is of interest to note that sites occupied by SATB1 do not associate with CTCF and to date CTCF has no known direct role in mediating globin gene switching.

Understanding the mechanisms of coordinated and developmentally regulated expression of both α- and β-like globin gene expression has significant clinical relevance, and several studies have identified key *cis* and *trans*-regulatory factors. The majority of haemoglobinopathies occur due to deletions or mutations in either the α- or β-like globin genes or their regulatory elements (Collins and Weissman 1984). Optimally functional haemoglobin has a tetrameric configuration invariably of two α-globin chains and either two γ- or two β-globin chains depending on the developmental stage, which are linked to heme groups. Aberrant expression of γ-globin proteins in humans induces a condition called hereditary persistence of fetal haemoglobin (HPFH), which describes a failure to down regulate γ-globin proteins in the adult but manifests no negative clinical consequences for patients. On the contrary, HFPH patients frequently also present with other haemoglobinopathies such as thalassemia or sickle-cell anemia that on their own would cause severe medical problems, but the effects are compensated by the HPFH phenotype. Mutations that effect GATA1 and NF-E2 interaction with target sites in the γ-globin promoter stimulates continued γ-globin expression (Berry et al. 1992, Cunningham, 1994, Collins, 1985) indicating sequences 5' and 3' of γ-globin are critical to both γ- and β-globin expression.

Sickle cell anemia is characterised by degeneration of erythrocyte structure. The distinctive sickle shape occurs as the haemoglobin, which constitutes around 90% of an erythrocyte, polymerises in its deoxygenated form. Functional haemoglobin proteins are soluble when carrying oxygen or carbon dioxide, however, a single A-T base pair substitution in the 6th coding exon of the β-globin gene creates a new glutamine amino acid in place of the normal valine, which reduces the solubility of deoxygenated but not oxygenated haemoglobin (Ingram 1956, Kaul, 1996). Sickle cells attach to and block endothelial blood vessels, inducing varying degrees of anemia and impairing the flow of oxygen to the tissues. In turn, the body responds by releasing more oxygen from erythrocytes, which paradoxically generates more deoxygenated sickle cells that eventually cause necrotic lesions in the peripheral tissues.

α-/β-Thalassemia describes an anemic disease where an imbalance of either α- or β-globin chains occurs because of inadequate protein levels. In the case of β-thalassemia, large deletions of the β-globin locus including regulatory regions or point mutations in the β-globin gene are responsible. Interestingly, β-thalassemic patients are not able to compensate for a lack of β-globin with continued γ-globin synthesis despite expressing γ-globin during fetal life as normal. The resulting disproportionate expression of α-globin still accumulates within red cells and associates with heme, but functional haemoglobin is not generated. Subsequent anemia develops in two ways, firstly accumulations of α-globin protein triggers cell death disrupting erythropoiesis in the bone marrow and secondly, excess α-globin predisposes peripheral erythrocytes to lysis via increasing levels of reactive oxygen species (Bank 2005). Analysis of a Dutch patient with γδβ-thalassemia, which manifests no expression of any β-like globin genes revealed no mutations or deletions of either the β-globin promoter sequence or the β-globin gene. However a large 100kb deletion encompassing ε- and γ-globin genes and regulatory regions was found (Kioussis et al. 1983, Taramelli, 1986) indicating distal *cis* regulatory DNA elements were necessary for β-globin expression. Identification of clustered erythroid-specific DNase I HSs positioned upstream of the β-globin locus led to the characterisation of the archetypal locus control region (LCR), which when linked to a β-globin gene conferred copy number-dependent, position-independent, tissue-specific expression in transgenic mice (Grosveld et al. 1987).

β-globin LCR

The conservation of sequences between human and mouse allowed the homologous mouse β-globin LCR to be identified (Moon and Ley 1990). Interestingly, several differences in LCR structure and apparent function occur between species. Potential links between changes in chromatin structure and modifications with

transcriptional activation have been extensively studied in both chicken and mammalian species to determine the influence of the LCR on these processes, which are both necessary for globin gene expression. Important distinctions between the chicken and mammalian β-globin loci are apparent; not least the proportion of acetylated histones present. In chicken, the entire locus is marked by histone hyperacetylation and DNase hypersensitivity, which is constrained at either end by insulator sequences (Litt et al. 2001). In contrast, the murine β-globin locus has regions of both hyper- and hypoacetylation within the region defined by flanking insulator sequences (Bulger et al. 2003). Deletion of the murine LCR significantly reduced globin gene transcription. However, DNase sensitivity and chromatin modifications were not affected, suggesting the LCR is redundant for mediating chromatin structure in mice (Schubeler et al. 2001). Chromatin organisation of the β-globin locus is in part mediated by CTCF but independently of active LCR-β-globin gene contacts (see below) (Palstra et al. 2003).

LCR deletion effects vary between human and mouse loci (Schubeler et al. 2001). Murine erythroleukeamia (MEL) cells harbouring chromosome 11 from humans where β-globin genes are situated, can express β-globin genes normally, but unlike the endogenous mouse locus, hyperacetylation and promoter remodelling in hybrid cells is strictly dependent on the presence of the LCR. This is an important consideration as studies using the mouse globin locus and their results should not be extrapolated as representative of a general model or applied to humans without careful consideration.

Regulation of globin gene expression in nuclear space

Gene expression in eukaryotes is regulated by sequences located tens or hundreds of kilobases away from the specific target genes via protein interaction with these sequences mediating long-range modifications of the chromatin structure (Ptashne 1986, Bulger, 2002). The murine β-globin locus spans over 200kb, with the LCR positioned over 50kb away from the β-globin gene itself. The locus is surrounded by insulator sequences in humans and mice. The flanking insulator sequences are DNase HSs that interact with CTCF. These CTCF binding sites were identified to cluster specifically in progenitor erythroid cells independently of β-globin gene expression. They participate in the regulation of LCR interactions with active β-globin genes, leading to the idea of chromatin hub (CH) and active chromatin hub (ACH) formation respectively (Tolhuis et al. 2002, Palstra, 2003), whereby intervening sequences are looped out (fig 3.4).

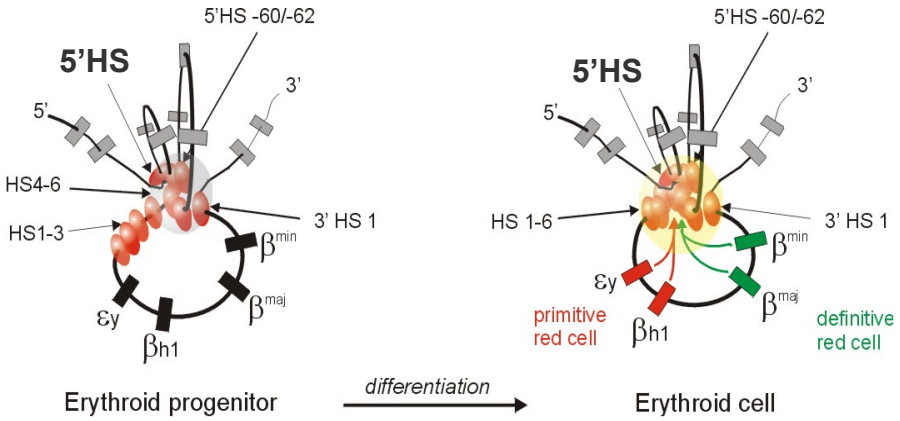


Figure 3.4. A 2-D representation of the 3-D mouse β-globin locus configuration
From (Palstra et al. 2003).

The conformation of the chromatin environment around the mouse β-globin locus is modified in a dynamic and developmentally regulated manner (see text for details). The β-globin genes are flanked by constitutive HSs. In erythroid progenitor cells, these sites cluster together and the intervening sequences including the developmentally regulated HSs of the LCR and β-globin genes are looped out, forming a chromatin hub. Upon differentiation, the proteins bound to the LCR and globin genes interact dynamically with each other and the constitutive HSs.

The significance of CTCF-mediated (A)CH formation and its role in regulating β-globin gene transcription is currently unclear (Splinter et al. 2006), but it appears transcription is independent to chromatin loop formation, currently identified to be CTCF-dependent. This is in accordance with previous data suggesting that the 3'HS1, which is essential for ACH formation, is not required for LCR-gene contacts, and globin gene transcription appears normal in its absence (Strouboulis et al. 1992). It remains to be seen if other CTCF-dependent or independent loops are formed at this locus that influence transcriptional activity.

Importantly, the erythroid-specific transcription factor erythroid krüppel-like factor (EKLF) is necessary for β-globin gene transcription but not the establishment of the CH (Drissen et al. 2004). Therefore either the transition of the locus into or the stabilisation of an active CH is EKLF-dependent. Interestingly, in erythroid progenitor cells when the CH is initially formed, EKLF is expressed but does not participate in the structure, suggesting EKLF is modified either to become involved or to prevent it activating globin expression too early. It would be interesting to analyse the binding of EKLF to the β-

globin locus in the absence of CTCF, to ascertain the stabilising role of EKLF and given that transcription of β-globin requires EKLF but apparently not CTCF. Other *cis*-regulatory elements apart from the LCR and β-globin promoter are proposed to affect ACH formation in the human β-globin locus (Patrinos et al. 2004).

Leuko/Lymphopoiesis: Generating immunity.

Leukocytes are classified into three groups (granulocytes, monocytes and lymphocytes) by their appearance under the microscope. Leukocytes such as basophils, eosinophils and neutrophils characteristically contain a granular cytoplasm and are therefore referred to collectively as granulocytes or polymorphonuclear granulocytes (PMNs). Mast cells are also granulocytes, but are not present in the circulation despite their haematopoietic origin. Monocytes are without granules and found in the circulation. Upon leaving the bloodstream, monocytes differentiate into macrophages, one of the main phagocytic cell types important for innate immunity (see below). There are two major classes of lymphocyte, B- and T-lymphocytes, which emerge from the bone marrow and thymus respectively. T-cells undergo further differentiation in peripheral lymphoid tissues such as the spleen and lymph nodes upon leaving the thymus, generating distinct cell types involved in inflammatory, allergic and cytotoxic responses.

All leukocytes are specialised blood cells involved in the recognition and prevention of and response to infections and as a defence protecting the body against environmental antigens. A functional protective immune system is critical for survival. The immune system can be broadly separated into two categories, innate and acquired immunity. These two types of immune response are very different and are performed by different sets of cells, but interact very closely generating a robust immune response (Fearon and Locksley 1996). The innate arm of the immune system can be considered as the first line of defence, involved in initiating inflammation and is dependent upon direct recognition of pathogens by phagocytic cells (macrophages, monocytes and PMNs). Eosinophils although considered as phagocytic cells play a major role in the defence against larger pathogens that cannot be internalised, such as parasitic worms or helminths. The action of eosinophils is also mediated by cytokines secreted by T-helper cells (see below). The innate immune response is important during the early stages of infection but is a direct, non-specific response and does not change upon repeated exposure to pathogens.

B- and T-lymphocytes broadly represent the acquired or adaptive arm of immunity, mediating immediate and long-term antigen-specific responses against antigens identified by the innate immune response as potentially harmful. However subsets of T-cells such as $\gamma\delta$ T cells and Th-17 cells also participate in innate immunity (see below). Activation and regulation of the acquired immune response is a major role of innate immunity. Adaptive immunity requires specific recognition of pathogenic antigens, both extracellular and intracellular using antigen receptors, the B-cell receptor (BCR) and the T-cell receptor (TCR). Both B- and T-cells have highly specialised and

cooperative immunological functions to recognise and protect the body from infection and generate memory lymphocytes, which respond rapidly upon re-infection. B-lymphocytes are antibody-producing cells found circulating in the plasma that eradicate pathogens via macrophage-mediated phagocytosis. In contrast, direct interaction between cytotoxic T-lymphocytes and pathogens or indirect responses triggered by T-helper cells function to remove infections. The development of both B- and T-lymphocytes proceeds through very strict checkpoints. This chapter will focus on the development of T-lymphocytes, which requires tight regulation to prevent autoimmune diseases such as multiple sclerosis, autoimmune diabetes, asthma and inflammatory bowel disease (IBD) such as Crohn's disease (Norman and Hickey 2005). T-lymphocyte differentiation and factors influencing the progression of the T-cell lineage into distinct subsets of mature T-cells will be discussed.

T-lymphocyte development:

T-lineage commitment and differentiation within the thymus

As with all haematopoietic cells, lymphocytes are ultimately derived from HSCs found in the adult bone marrow. T-lymphocytes however are unique to the haematopoietic lineage in that they are the only cells not produced in the bone marrow. The vast majority of functional T-lymphocytes found in the circulation and peripheral lymphoid tissues are derived from cells generated within the thymus, however a minimal proportion of the peripheral T-lymphocyte population has been described to originate from an extrathymic source (Rocha et al. 1995, Leclercq, 1996).

Differentiation from HSC to committed T-lymphocyte potentially involves several intermediate stages. However, the identity of the T-lymphocyte precursor cell remains controversial (Zuniga-Pflucker and Schmitt 2005, Bhandoola, 2007). During early embryogenesis, the fetal liver is the preliminary site of T-cell development prior to development of the thymus that occurs between embryonic days (E) 10-13.5 (Manley and Blackburn 2003). Specific cell types required for initiating thymus formation are shown to be present at E10 with organogenesis beginning at E11. The thyroid gland, parathyroids and thymic lobes develop in close association until eventual bilateral development of the thymus occurs and leads to separation from the parathyroids at E13.5 (Miosge and Goodnow 2005). The earliest thymus colonising cells appear at E11-12. It has been shown that these cells are not capable of generating B-cells, but do have the capacity for T/NK/DC lineage differentiation, suggesting B-lineage potential is restricted extrathymically (Harman et al. 2005, {Masuda, 2005). This restriction seems not to continue in adults; early lymphocytes in the thymus known as thymic lymphoid precursor (TLP) or early thymic progenitor (ETP) cells retain the potential to differentiate

along all lines of lymphoid development as B-lymphocytes, T-lymphocytes, lymphoid dendritic cells (DC) and natural killer (NK) cells before T-cell specification is set (Zuniga-Pflucker and Lenardo 1996, Rothenberg, 2005, Bhandoola, 2007). A number of hematopoietic progenitor cells located in the blood, bone marrow and thymus have the capacity to differentiate towards the T-cell lineage (Bhandoola et al. 2007). However committed T-cell development is not assured until a relatively late stage of differentiation when cells have resided in the thymus for several days. The thymic microenvironment strongly supports the potential of multipotent progenitors developing along the T-lineage, as direct transfer of HSCs into the thymus allows efficient generation of T-cells. In contrast, HSCs injected into the circulation is much less efficient, requiring differentiation and homing of differentiated cells to the thymus. How these circulating progenitor cells become targeted to and located within the thymus is regulated in part by expression of phagocyte glycoprotein-1 (PGP1; CD44), P-selectin glycoprotein ligand-1 (PSGL1) and integrins. Significantly, HSCs do not express the CC chemokine receptor CCR9, shown to be important for directing cells to settle in the thymus. The earliest CCR9 expressing progenitor cell is the early lymphoid progenitor (ELP). The contributions of ELPs to defining specific T-cell progenitors remain unknown (Bhandoola et al. 2007). Colonisation of the thymus by progenitor cells occurs at the cortico-medullary junction and Notch/Delta-like signals provided by thymic stromal cells within the cortex are essential for subsequent T-cell differentiation *in vivo* (Sambandam et al. 2005, Tan, 2005).

Lymphocytes within the thymus are widely and broadly classified according to expression of the major histocompatibility (MHC) (Germain 1994) receptor accessory molecules CD4 and CD8[1]. Double negative (DN) cells express neither CD4 nor CD8, double positive (DP) cells express both, and single positive (SP) cells express either CD4 or CD8 specifically. The earliest identified intrathymic cells to have T-cell potential are designated double negative 1 (DN1), early thymic progenitors (ETPs) or DN1a-e cells (Rothenberg 2007, Porritt, 2004). ETPs can strictly be distinguished from true DN cells as some cells that fall under the ETP classification also express low levels of CD4. However, as both $CD4^{low}$ and $CD4^{-}$ intrathymic progenitor cells have the same functional capacity, these terms may be interchangeable (Bhandoola et al. 2007). A more accurate classification of thymic cells and their subsequent differentiated states considers the expression of the interleukin 2 receptor alpha chain (IL-2Rα; CD25), CD44, CD3, CD8 and c-Kit (fig 3.5) (Sebzda et al. 1999).

[1] Lymphocytes, like all WBCs, express many different molecules on their cell surface. The CD (cluster designation) nomenclature categorises these different molecules based upon clusters of monoclonal antibodies that recognise specific markers. This system is now used to indicate the molecule recognised by each specific 'CD' of monoclonal antibodies (Roitt, I., Brostoff, J., Male, D. Immunology. Fourth ed. Mosby. p2.2)

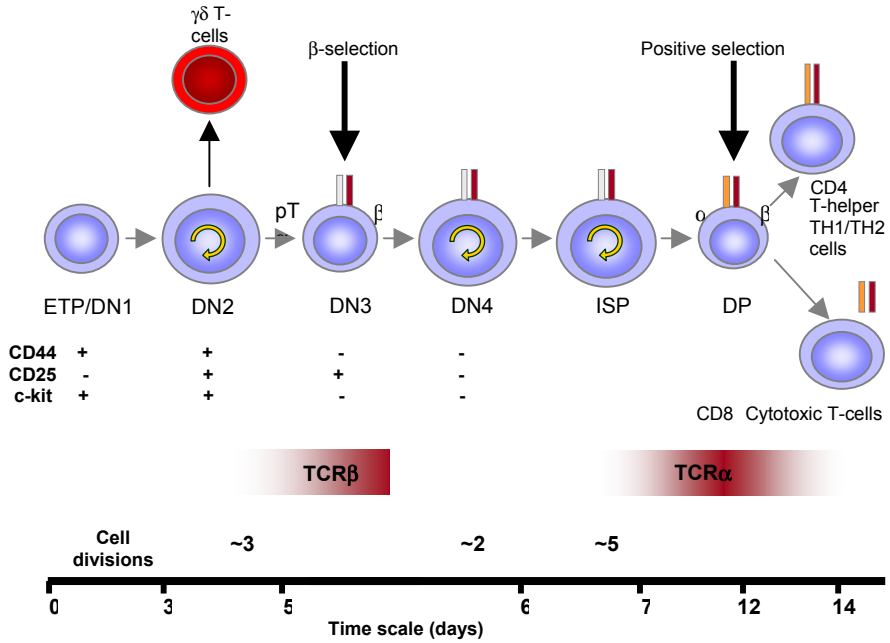


Figure 3.5. A schematic overview of T-cell differentiation pathways in the thymus Adapted from (Sebzda et al. 1999).

TCRγ rearrangements can be detected at DN2, indicating commitment to the γδ lineage is initiated at this point. However TCRδ gene expression, which is absolutely required for the generation of γδ T-cells is not detected until DN3. CD3 is a co-receptor molecule expressed from DN3 onwards as part of the (pre)-T-cell receptor. In mice, ISP cells express CD8, whereas the equivalent stage in human cells express CD4. See text for details.

Commitment to the pro-T-cell DN2 stage from ETPs is dependent on interleukin (IL) 7 expression by thymic stromal cells and is characterised by up-regulation of CD25 and initiation of T-cell receptor gene rearrangements, the earliest detection of which involves the *Tcrγ* gene locus at the DN2 stage (Capone et al. 1998). Fully committed T-cells are not generally recognised however until DN3 (Rothenberg 2007) when rearrangements of *Tcrγ*, *δ*, and *β* gene loci occur and CD44 expression is lost. A small percentage of DN3 cells express a TCR composed of γ and δ chains, which differentiate into the γδ T-cell lineage (see below). The vast majority of cells rearrange *Tcrβ* genes that are expressed on the cell surface forming a complex with the surrogate pre-Tα chain and the CD3 complex generating the pre-TCR. Signalling from the pre-TCR through Zap-70, calcium mobilisation and subsequent activation of NFAT and NFκB inhibits further recombination of the TCRβ allele, induces proliferation and down-regulation of CD25 to give DN4

($CD44^-$ $CD25^-$) cells and actuates CD4/CD8 expression (Wiest et al. 1999). Aspects of pre-TCR expression are significantly different to both αβ and γδ TCRs; functional expression of the pre-TCR results in apparently spontaneous inclusion in lipid rafts, microdomains enriched in glycolipids (Saint-Ruf et al. 2000). Pre-TCR signalling in contrast to αβ and γδ TCRs occurs constitutively and does not require ligand interactions mediated by thymic stromal cells (Irving et al. 1998). Failure to develop a functional pre-TCR leads to apoptosis (von Boehmer et al. 1999). This process is known as β-selection.

T-cell development in the thymus is dependent on *Tcr* gene loci rearrangements that strictly occur in a developmental stage-specific manner. Rearrangement of *Tcrβ* genes is a critical check-point for T-cell differentiation as only cells that have undergone functional *β*-gene rearrangement will down-regulate CD25, proliferate and continue differentiation towards the mature T-cell stage, initiating expression of the *Cd4* and *Cd8* co-receptor genes. Antigen receptor genes in both B- and T-cells are composed of separate variable (V), diversity (D), and joining (J) gene segments which rearrange by somatic recombination via the actions of recombination-activating genes (*Rag*). Recombination signal sequences (RSS) flank the V, D and J genes, and recombination is restricted between RSS of differing lengths (Tonegawa 1983). Two RAG proteins (RAG-1 and RAG-2) function co-ordinately, inducing site-specific double-strand breaks at the coding sequence/RSS border in both *Tcr* and *Bcr* loci (van Gent et al. 1996), which are repaired by non-homologous-end-joining DNA repair machinery (Weaver 1995). *Tcrγ* gene recombination can be detected as early as DN2, indicating functional *Rag* genes are expressed at this stage, however *Tcrβ* gene rearrangement does not normally occur until DN3. *Tcrβ* genes rearrange in a highly ordered process, D_β-to-J_β rearrangements precede V_β-to-DJ_β, which are mediated in part by the *Tcr*β enhancer (E_β) and $D_\beta 1$ promoter (Mathieu et al. 2000). Activation of the $D_\beta 1$ promoter ($PD_\beta 1$) by E_β is necessary for initial D_β-to-J_β rearrangements whereby E_β is proposed to mediate the formation of open chromatin modifications across the D_β, J_β and C_β domain. In contrast, rearrangements of the TCRβ variable (V_β) genes are thought to occur in an E_β-independent manner (Ryu et al. 2004). Interestingly, E_β and $PD_\beta 1$ have been shown by 3-C to functionally interact with each other at $D_\beta 1$ but not at J_β by forming chromatin loops (Oestreich et al. 2006).

The regulation of chromatin structure and changes in histone modifications are proposed to facilitate accessibility of RAG proteins to specific *Tcr* gene sites. Particular attention has been given to the role of histone H3 acetylation and the requirement of ATP-dependent chromatin remodelling complexes in mediating *Tcr* gene rearrangements. Hyperacetylation of H3 in both *Tcrβ* and *Tcrα* genes is strongly correlated with the stage at which rearrangements of these loci occur (McMurry and Krangel 2000, Winandy, 2005). Stringent control of both the initiation and termination of *Tcr* gene rearrangements is required as this determines cellular developmental potential.

The mechanisms by which the timing of *Tcr* gene rearrangements are mediated and coordinated with developmental transition are not yet fully understood. The cessation of VDJβ recombination, which coincides with a loss of V_β transcriptional activity and repressive chromatin formation, is required for progression from the DN to DP stage. It is of interest to note that forced transcriptional activity and modification of chromatin structure was not sufficient to re-induce V_β rearrangements, suggesting regulatory (transcriptional) factors or other epigenetic modifications were not present to fully recapitulate the developmentally appropriate gene regulatory network or cellular environment required (Jackson and Krangel 2006).

Progression though the DN stages also requires the regulated expression of specific transcription factors, notably GATA3 (Ting et al. 1996, Pai et al. 2003). In mice, T-cells differentiate from the DN stage through an intermediate $CD8^+$ expressing cell type known as immature single positive (ISP), a transient stage that is largely overlooked in most articles. Intense cellular proliferation occurs at this stage in order to generate the large numbers of DP cells required in the thymus. The transition between ISP and DP cells requires the down-regulation of cell-cycle regulators such as p21 and p27 (Tsukiyama et al. 2001, Wolfraim, 2004). Sequential up-regulation of *Cd4* gene expression defines the DP stage, at which point *Tcr*α Vα–to-Jα gene rearrangements begin. Histone acetylation and H3K4 methylation is found at promoter sequences important for regulating *Tcr*α recombination events initiated by the *Tcr*α enhancer (Eα), which affects long-range chromatin modifications and structure of the locus (Hawwari and Krangel 2005). It remains to be seen if CTCF participates in the long-range chromatin modifications of the *Tcr*α locus, however there is currently no evidence suggesting CTCF directly affects Tcrα expression (this thesis). Unlike *Tcr*β gene recombination, biallelic and multiple rearrangements of *Tcr*α genes occur, although there seems to be a matter of proximity between Vα and Jα genes determining recombination timing (Krangel 2007).

Functional rearrangement of *Tcr*α genes and the subsequent formation of diverse TCR$\alpha\beta$ complexes initiate selective pressure on the developing T-cells based upon TCR$\alpha\beta$ signalling and the capacity of specific TCRs to recognise antigens. Mature T-cells are distinguished according to their ability to recognise infected cells displaying peptides from the infecting pathogen on their cell surface. These peptides are distinct for each pathogen and accordingly are recognised specifically by the diverse TCR repertoire generated by TCR recombination. Cell surface antigen presentation is achieved by interactions between pathogen peptides and specific membrane glycoproteins, the major histocompatability complex (MHC) (Germain 1994). Antigen/MHC complexes are recognised by the TCR$\alpha\beta$, and this triggers the termination of RAG expression and *Tcr*α

gene recombination. Subsequently, expression of TCRαβ increases and cells initiate expression of CD69 on their surface, which is a marker for activated T- and B-lymphocytes. DP cells undergo both positive and negative selection at this stage based upon the reactivity of the Ag/MHC complex and TCRαβ signalling. DP cells are selected against if they do not receive any functional TCRαβ-mediated signal within 3-4 days, and as a result they will go into apoptosis by a process termed death by neglect (Sebzda et al. 1999). Self-reactive T-cells also undergo negative selection to maintain a population of T-lymphocytes designed specifically to recognise MHC complexes presenting non-self antigens, which also provides a mechanism of maintaining tolerance (Kappler et al. 1988). DP $CD69^+$ cells that express a fully functional and non-self reactive TCRαβ are positively selected and go forward to further differentiate into mature single positive (SP) T-cells.

Two classes of MHC molecules exist; MHC class I molecules are expressed in all nucleated cells in the body (Ellmeier et al. 1999). In contrast, expression of MHC class II molecules is relatively cell-type restricted. MHC class II molecules recognise extracellular pathogenic peptides that have been internalised either specifically (by B-lymphocytes) or non-specifically by dendritic cells (DC) or macrophages, which collectively are referred to as antigen presenting cells (APCs). MHC molecules class I and II also differ in their respective co-receptor molecules that in turn classifies the eventual cell lineage fate. Differentiation into $CD4^+$ SP and $CD8^+$ SP cells requires the expression of one co-receptor molecule to be repressed. T-cells that express an MHC class I restricted TCRαβ maintain CD8 expression and differentiate into cytotoxic $CD8^+$ SP cells. Conversely, the interaction of TCRαβ with MHC class II molecules corresponds with CD4 co-receptor expression and the differentiation of $CD4^+$ SP cells. The mechanism by which the TCR has specificity for MHC molecules and how this determines subsequent cell lineage direction remains unknown and controversial (Taniuchi et al. 2002). Studies by Hernández-Hoyos et al using transgenic mice revealed the expression of particular MHC molecules does not seal the fate of T-cell differentiation. Changes in expression levels of the src-family protein kinase p56 (LCK) during positive selection have been shown to redirect CD4/CD8 lineage determination. T-cells with an MHC class II restricted TCR can develop into functional $CD8^+$ cells when levels of LCK are reduced. On the other hand, when activity of LCK is increased, cells carrying an MHC class I restricted TCR are capable of CD4 differentiation (Hernandez-Hoyos et al. 2000). The zinc-finger transcription factor cKrox is known to be required and sufficient for directing thymocytes undergoing positive selection into CD4 lineage differentiation (Sun et al. 2005). The *Zfp67* gene encoding cKrox is up-regulated specifically during $CD4^+$ T-cell differentiation. Constitutive expression of cKrox transgenes effectively induces the same result as increased LCK activity, directing DP

cells with MHC class I TCRs to the CD4 cell lineage, whereas in the absence of cKrox, MHC class II-restricted DP cells develop into $CD8^+$ cells. The cKrox protein is also known as T helper-inducing POZ/Kruppel factor (Th-POK). Helper deficient (HD) mice are completely devoid of CD4 cells. This phenotype occurs from a spontaneous single nucleotide mutation in the DNA-binding domain of *Zfp67*, which functionally inactivates the cKrox/Th-POK protein (Dave et al. 1998, He, 2005, Sun, 2005).

The expression of both *Cd4* and *Cd8* genes is a tightly regulated process, requiring coordinated silencing at the DN stage, activity at the DP stage and subsequent independent transcriptional control in single positive cells. The *Cd8* locus consists of two separable genes (*Cd8*α and *Cd8*β), whereas the CD4 protein is the product of a single gene (Taniuchi et al. 2004). The epigenetic mechanisms behind *Cd4* and *Cd8* expression patterns are only recently emerging; interestingly DNA methylation is apparently not required for the silencing of *Cd4* expression (Lee et al. 2001, Wilson, 2005). *Cd4* expression is mediated in part by a promoter, a proximal enhancer and other enhancer-like elements (Krangel 2007), however the expression pattern of *Cd4* is underlined by an intronic *Cd4* silencer element that represses expression in both DN and CD8 SP cells (Sawada et al. 1994). Interestingly, the Runx family of transcription factors is essential to the function of this silencer element, and studies have revealed specific requirements for Runx1 and Runx3 interaction at the DN and SP stage respectively (Taniuchi et al. 2002), although some functional redundancy may occur in $CD8^+$ cells (Woolf et al. 2003). Changes in chromatin conformation at the *Cd4* locus occur via Runx interaction recruiting repressive chromatin remodelling complexes including the DNA-binding component of the SWI/SNF-like BAF complex (BAF57), histone methyltransferase Suv39h1 and HDAC1 (Chi et al. 2002, Telfer, 2004, Reed-Inderbitzin, 2006). A corresponding silencer for *Cd8* expression is unknown, however over-expression of the transcription factor GATA3 is known to repress development of $CD8^+$ SP cells (Nawijn et al. 2001) and this may also involve the recruitment of specific chromatin modifications. In conjunction with a role in *Cd4* repression, Runx 1 is also an activator of *Cd8* expression in DP cells via interaction with *Cd8* enhancers (Sato et al. 2005). Activation of *Cd4* requires hyperacetylation of histone H3 at the *Cd4* enhancer. Surprisingly, this is dependent upon Mi-2β (CHD4), a component of the nucleosome remodeling deacetylase (NuRD) complex (Williams et al. 2004), which in the context of T-cell development is required for attracting both the HeLa E-box binding protein (HEB) transcription factor and the p300 HAT to the enhancer.

Transcription factors required for T-lineage development

The development of T-cells is tightly regulated by the activity of multiple transcription factors at every stage from specifying precursor cells with T-lineage potential, through to mediating functions of definitive effector cells. A few of the key factors will be highlighted here.

Before T-lineage fate is committed, lymphocyte precursor cells strictly require the expression of *Ikaros*, a krüppel-like zinc-finger transcription factor, and PU.1 for the development of all lymphoid cells (Georgopoulos et al. 1994, Spain et al. 1999). PU.1 expression levels are critical for correct thymocyte development and importantly lineage direction. High overexpression of PU.1 can redirect precursor cells towards myeloid differentiation, whereas moderately increased levels arrest T-cell development at the pro-T stage (Anderson et al. 2002). The differentiation from lymphoid progenitors to early DN thymocytes also involves GATA3, the expression of which together with *c-myb* characterises the earliest emergence of distinguishable T-cell progenitors in the thymus (Ting et al. 1996, Hendriks et al. 1999, Allen et al. 1999). GATA3 is one of the most significant regulators of T-cell development; expressed throughout differentiation, but is down regulated at stages of TCR rearrangement when cells are in G_1 arrest. Significantly, the induction of *Gata3* in thymocytes is directly regulated by *c-myb* and is independent of TCR signaling (Maurice et al. 2007).

As described above, *Tcr* gene rearrangements are initiated at DN2-3, and β-selection occurs, whereby only T-cells completing functional *Tcr*β gene rearrangements can proceed towards the TCR$\alpha\beta$ lineage. The expression of a number of genes including *Rag1/2*, and components of the pre-TCR are essential. A regulatory cascade involving Notch interaction with recombination signal binding protein J (RBP-J) activates expression of both the preTα (pTα) chain and HES-1, a basic helix-loop-helix (bHLH) transcription factor that is necessary for the expansion of early T-cells during β-selection (Tomita et al. 1999). HLH proteins are key regulators of both B- and T-lymphocyte development, for review (Murre 2005). Class I bHLH proteins are known as E-proteins or E-box binding proteins in reference to their DNA-binding consensus sequence. E12 and E47 are E-proteins encoded by a single gene, however different proteins are produced by virtue of differential splicing. E47 is important in regulating the β-selection checkpoint by initiating cell cycle arrest at DN3 before *Tcr*β gene rearrangements begin, and in the absence of pre-TCR signaling, maintaining developmental arrest (Engel et al. 2001, Petersson et al. 2002).

Following β-selection, cell proliferation is induced, and continues until *Tcr*α genes begin rearrangement in DP cells. The progression from DN-DP occurs via an intermediate ISP stage. Defective *Tcr*α gene expression can result in accumulation of ISP cells, as observed in the absence of TCF-1, a T-cell specific high-mobility group (HMG) transcription factor (Roberts et al. 1997) required for *Tcr*α expression. HEB, another bHLH factor (Barndt et al. 1999, Barndt et al. 2000) also mediates the DN-DP transition, however redundancy between different E-proteins rescues the developmental block observed in the HEB knock out. Importantly, HEB is functional as a DNA-binding protein either as a homodimer or in a heterodimeric complex with the related factor E2A. Deficient E2A expression affects early T-lymphocyte development (Bain et al. 1997). Interestingly, GATA3 is an activator of E2A, which in turn negatively regulates expression of the CD5 glycoprotein in DP cells via binding of E2A to the CD5 promoter (Yang et al. 2004, Ling et al. 2007). As $\alpha\beta$TCR signaling is repressed by CD5, Ling et al suggest GATA3 mediates a positive feedback loop where $\alpha\beta$TCR activation induces GATA3 expression, leading to an increase of E2A and subsequent repression of CD5, which continues activation of the TCR.

Peripheral T-cell development

The differentiation into either CD4 or CD8 SP cells in the thymus precedes the loss of CD69 expression and up-regulation of CD62L, which is required for migration of mature SP cells from the thymus to the peripheral lymphoid tissues. Differentiated CD4 and CD8 T-cells that have undergone stringent selection within the thymus to ensure they specifically recognise non-self antigens presented by MHC molecules emigrate via the circulation to the spleen and lymph nodes without having previously encountered their specific antigen, and as such are referred to as naïve T-cells. Before naïve T-cells come across their specific cognate peptide-MHC, they circulate as quiescent cells arrested in G_0. Prior to activation, IL-2 expression is repressed in naïve T-cells in part by DNA methylation of the IL-2 promoter, and repressive chromatin modifications across the locus (Fitzpatrick and Wilson 2003).

Upon initial activation, involving TCR stimulation and recognition of the antigen together with co-stimulatory signals from APCs, naïve cells begin production of IL-2 (Wilson et al. 2002) following remodeling of the IL-2 locus that requires c-rel activity (Rao et al. 2003). The expression of IL-2 is known to mediate growth, differentiation and/or apoptosis in lymphocytes, dependent upon the composition of the IL-2 receptor (IL-2R) (Herblot et al. 1999). At least three subunits (α,β, and γ) make up the IL-2R, however the IL-2R α-chain, which is only expressed in peripheral T-cells in the presence of co-stimulatory signals from APCs, increases the affinity for IL-2 interaction and this initiates cell cycle release into G_1, progression and subsequent cellular proliferation. IL-2 is

known to be necessary for IL-2Rα expression in spleen and lymph nodes, but dispensable for expression of IL-2Rα in the thymus and bone marrow (Herblot et al. 1999).

Naïve T-cells differentiate into effector cells following 4-5 days of proliferation (fig 3.6) (Laurence and O'Shea 2007).

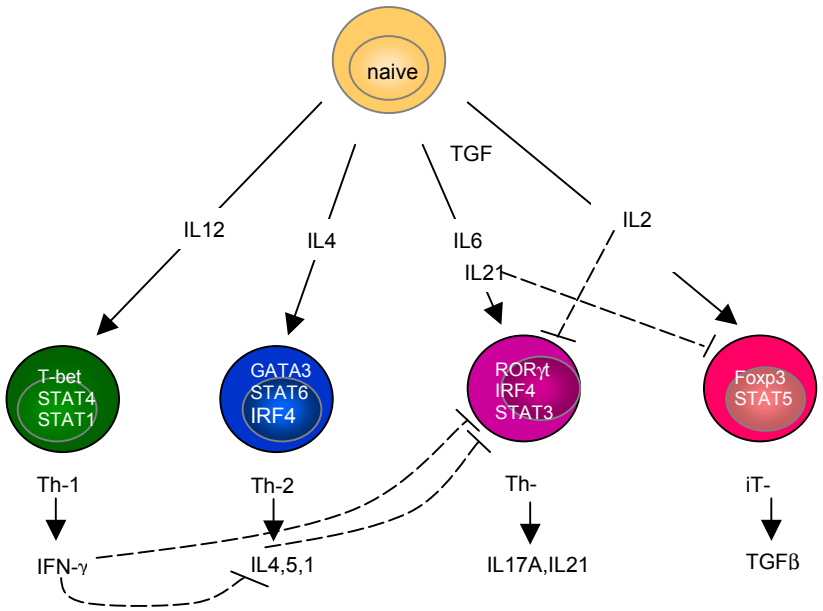


Figure 3.6. An overview of T-lineage differentiation in mice
Adapted from (Laurence and O'Shea 2007).

In the presence of APCs and cognate peptides, naïve T-cells differentiate in to distinct lineages characterised by their cytokine secretion. An important differential requirement for TGFβ in Th-17 development occurs between humans and mice. Inducible T-regulatory cells (iT-reg) are anti-inflammatory cells that develop under the influence of IL2. Th-17 cells are inflammatory cells directed by STAT3 signaling. For details see text and (Laurence and O'Shea 2007).

T-cells expressing the MHC class I molecules and CD8 co-receptor are committed to differentiate into cytotoxic T-cells. Importantly, cytotoxic lymphocytes (CTLs) include both $CD8^+$ cells together with natural killer (NK) cells and lymphokine-activated killer (LAK) cells. $CD8^+$ cells are a specialised cell type involved in the rapid removal of virus-infected cells after infection, secreting cytolytic effector molecules such as perforin and granzyme B that destroy the infected cell to which the $CD8^+$ cell has bound. Characteristically,

$CD8^+$ cells also express interferon-γ (Ifn-γ), and they are sometimes referred to as Tc1 cells (Asnagli and Murphy 2001), however $CD8^+$ cells can express IL-4, IL-5 and IL-13 when exogenous IL-4 is present, differentiating into so called Tc2 cells (Cerwenka et al. 1998). $CD4^+$ T-cells undergo further differentiation into T-helper (Th) cells of which there are three main classes, Th-1, Th-2 and Th-17. As mentioned above, activation of naïve T-cells involves a two-step signalling event; initial TCR-derived signals and a co-stimulation via accessory molecules expressed on APCs. Significantly, only activated DCs express sufficient levels of the CD80 (B7-1) and CD86 (B7-2) co-stimulators to induce effector cell differentiation. CD28 is the main co-stimulatory molecule on Th cells that interacts with CD80 and CD86 (Lenschow et al. 1996).

The characteristics and functions of Th-1 and Th-2 cells have been extensively reviewed (Abbas et al. 1996, Glimcher, 2000). Th-1 and Th-2 cells differ in the type of immune response elicited, the type of pathogens they respond to, and they are also associated with different kinds of immunopathology (table 3.2) (O'Garra and Arai 2000).

Features	T helper-1	T helper-2
Characteristic cytokines	IFN-γ, IL-2, IL-10, lymphotoxin	IL-4, Il-5, IL-10, IL-13
Characteristic gene expression	T-bet	GATA3
Pathogen eradication	Intracellular, bacteria, parasites, yeasts, fungi	Extracellular, parasitic worms
Immune response	Cell-mediated immunity, production of compliment fixing antibodies (IgG2a), NK activation, activation of cytotoxic $CD8^+$	Humoral immunity, activation of mast cells and eosinophils
Immunopathology	Autoimmune disease*	Allergy, atopy

Table 3.2. T-helper cell characteristics
Adapted from (O'Garra and Arai 2000).
* The induction of autoimmune disease is now characterised as a Th-17 cell immunopathology. It was previously thought that Th-17 cells were derived from Th-1 cells. Recent studies demonstrate the emergence of Th-17 cells directly from naïve $CD4^+$ cells (Harrington et al. 2005, Park et al. 2005). See text for details.

Importantly, T-helper cells are characterised by their cytokine expression profile, which under normal conditions is strictly restricted to either Th-1 or Th-2 cells.

Th-17 cells are a recently emerging aspect of $CD4^+$ T-cell biology. Previous studies of IL-17 expressing cells suggested they were derived from 'pre-Th-1' cells that

under certain conditions, possibly in the presence of IL-23 would diverge into Th-17 cells, whereas IL-12 stimulation allowed Th-1 differentiation. The expression of IL-17 is now known to occur directly from differentiated naïve $CD4^+$ cells, therefore Th-17 cells are a completely distinct effector cell type (Harrington et al. 2005, Park et al. 2005). Significantly, these studies identified IL-17-producing cells as pathogenic in experimental mouse models of inflammatory disease, and development of the disease could be impeded in mice deficient for IL-23, which initiates IL-17 expression. Interestingly, the developmental initiating factors, cytokine profile and functions of Th-17 cells are distinct between humans and mice. It also remains unclear if Th-17 cells are a terminally committed lineage of cells like Th-1 and Th-2 cells (Wilson et al. 2007).

Defining Th-1 and Th-2 cells

T-helper effector cell differentiation and their characteristic patterns of effector cytokine expression are mediated on several different levels, both intrinsic to the cell including chromatin modifications (Shoemaker et al. 2006), DNA methylation (Sanders 2006), chromatin structure (Agarwal and Rao 1998, Spilianakis et al. 2004, Lee et al. 2005), chromosomal interactions (Spilianakis et al. 2005), regulation of gene expression by transcription factors (Agnello et al. 2003, Yates et al. 2004, Szabo et al. 2000, Zheng & Flavell 1997) and nuclear repositioning of factors regulating cytokine expression (Hewitt et al. 2004), but also environmental factors such as hormones or nutrients (Spilianakis et al. 2005b). None of these processes occurs independently, and in some cases single proteins are implicated in many different modes of regulation.

Differentiation of naïve T-cells to effector Th cells is influenced by the concentration and type of antigen presented, the avidity of peptide-MHC interaction and the type of co-stimulation received. Importantly, an intricate regulatory system of cytokine signalling between APCs and Th cells and chromatin remodelling is central to effector cell differentiation and function. Naïve T-cells maintain distinguishing Th cytokines (Ifn-γ and IL-4) in a permissive yet inactive condition (Baguet and Bix 2004). Following antigen stimulation, specific cytokine signals direct differentiation into Th-1 or Th-2 cells, and both Ifn-γ and IL-4 loci are remodelled to enhance transcription from one loci but not the other ensuring rapid secretion of cytokines upon subsequent exposure to specific antigens.

IL-12 is necessary for the differentiation of naïve T-cells into the Th-1 lineage. APCs are the predominant source of IL-12 and are also the first line response to infection, providing a critical link between innate and adaptive immunity. The expression of IL-12 by APCs is initiated in response to the uptake of bacterial, viral or protozoal infection and also via interactions between CD40 cell surface molecules on APCs and

CD40L on activated T-cells (Agnello et al. 2003). Co-operative activity between IL-12 and an unrelated cytokine IL-18 induces Ifn-γ production from NK and T-cells. A feedback mechanism occurs where by IL-12 and Ifn-γ expression promotes Th1 differentiation, and subsequent Ifn-γ expression in Th-1 cells leads to increased IL-12 expression in APCs that continues the differentiation of Th-1 cells. Ifn-γ expression is imperative for Th-1 differentiation. Many characteristic functions and expression profiles are mutually exclusive between Th-1 and Th-2 cells (fig 3.4) (Rao and Avni 2000), however in the absence of Ifn-γ expression, the defining features become blurred as Th-1 cells isolated from Ifn-γ knock out mice express Th-2 cytokines when re-stimulated in the presence of IL-4 (Zhang et al. 2001). Ifn-γ expression is also important for expression of the IL-12Rβ2 chain (Szabo et al. 1997). The IL-12 receptor (IL-12R) is comprised of two subunits (β1 and β2) and is only expressed following T-cell activation, therefore IL-12R is repressed in naïve T-cells. IL-12Rβ2 is a specific marker of Th-1 cells and the component parts of the receptor are each necessary for normal Th-1 differentiation and responses *in vivo*. Further feedback loops occur consisting of Ifn-γ and the transcription factor T-box expressed in T-cells (T-bet) (Szabo et al. 2000) where T-bet directly regulates Ifn-γ via chromatin remodelling of the Ifn-γ locus together with an accessory factor HLX (Zheng et al. 2004) and indirectly promotes IL-12Rβ2 expression. IL-12 interaction with IL-12Rβ2 triggers tyrosine phosphorylation and Janus kinase-2 (Jak-2) activation, which in turn activates signal transducer and activator of transcription (STAT) 4. Importantly, STAT4 sustains T-bet expression allowing increased Ifn-γ expression (Szabo et al. 2003). Ifn-γ signaling also activates STAT1, and regulation of T-bet by IFN-γ requires IFN-γR signaling that occurs in a STAT1-dependent manner (Lighvani et al. 2001).

Extrinsic cytokine signals from DCs and other accessory cells also mediate the differentiation of Th-2 cells in a similar fashion to IL-12 for Th-1 cells. Antigen-stimulated T-cells are directed towards the Th-2 lineage by OX40, CD28 and IL-4, although IL-4 expression is the major factor polarising Th-2 differentiation (Yates et al. 2004). Transcription factors including nuclear factor of activated T-cells (NF-AT), AP-1 family members and cMAF regulate IL-4 expression in response to antigen stimulation via binding to the IL-4 proximal promoter (Rooney et al. 1995, Ho, 1996). Similar to IL-12 signaling, IL-4 binds to the IL-4R, which activates STAT6 via Jak-mediated tyrosine phosphorylation (Agnello et al. 2003). STAT6 in turn mediates the regulated expression of GATA3 by IL-4. Feedback regulation exists whereby IL-4 mediates GATA3 expression, which subsequently regulates IL-4 expression. IL-4 signal transduction by STAT6 is an integral part of Th-2 differentiation, however IL-2-mediated STAT5 activation is also important (Cote-Sierra et al. 2004, Zhu, 2003). Stat5 has two isoforms, a and b, that are both activated by IL-2. Interestingly, expression of a constitutively active

form of Stat5a can induce IL-4 expression in Th-1 cells in a GATA3 independent manner (Zhu et al. 2003). GATA3 is essential for Th-2 development, although it appears this is mediated in part by IL-4/STAT6 signaling, as STAT5 does not influence GATA3 expression. It is interesting to note that enforced expression of GATA3 is sufficient to induce Th-2 responses independently of IL-4/STAT6 signaling, implying GATA3 itself has a key role in Th-2 function (Ouyang et al. 1998). IL-4 production is dependent upon STAT5 signaling, and defects in IL-4 expression in the absence of STAT5 are not overcome by GATA3 expression (Cote-Sierra et al. 2004) suggesting a GATA3-independent regulation of IL-4 exists. Indeed, conditional deletion of *Gata3* from established Th-2 cells reduces but does not completely abolish the expression of IL-4 (Zhu et al. 2004). The IL-4/STAT6 and IL-2/STAT5 pathways are independent but optimal IL-4 production probably requires both pathways to work in conjunction (Zhu et al. 2006). Analysis of Stat5a deficient mice also revealed a key role for this signaling pathway in the repressive regulation of Th-1 differentiation (Takatori et al. 2005). STAT5a directly activates suppressor of cytokine signaling (SOCS) 3 by binding to the promoter, resulting in inhibition of STAT4, thereby preferentially directing Th-2 differentiation. In the absence of STAT5a, levels of SOCS3 are reduced, allowing IL-12-mediated activation of STAT4 and Th-1 differentiation.

The pathways regulating Th-1 and Th-2 differentiation are both mutually exclusive and reciprocally inhibitory. Ectopic retroviral expression of Gata3 in Th-1 cells suppresses Th1 development by down-regulation of Stat4, a process that normally occurs in Th-2 cells. Interestingly, retroviral expression of STAT4 together with GATA3 in Th-1 cells upholds the Th-1 phenotype, suggesting the control of STAT4 signaling is critical to determining Th-1/Th-2 fate. Conversely, retroviral T-bet expression in Th-2 cells induces Th-1 cytokine expression and represses the expression of Th-2 cytokines (Szabo et al. 2002). IL-12Rβ2 expression is however not sufficient to induce Ifn-γ expression in committed Th-2 cells, nor restrict the expression of IL-4 (Heath et al. 2000). Significantly, activation of Ifn-γ and repression of IL-4 by T-bet requires expression of Runx3, a transcription factor known to mediate the expression of CD8 in thymocytes (Taniuchi et al. 2002, Djuretic, 2007).

Intra- and interchromosomal interactions in T helper cells

The gene encoding IL-4 is located on mouse chromosome 11 within a multi-gene locus consisting of two other Th-2 cytokine genes, *Il-5* and *Il-13* that are expressed in a developmentally restricted manner (figure 3.7). The entire Th-2 cytokine locus spans over 120kb (Spilianakis and Flavell 2004). *Il-5* is transcribed in the opposing direction to both *Il13* and *Il-4* and separated from these genes by the constitutively expressed *Rad50* gene encoding a DNA-repair protein. Th-2 cells are characterised by the expression of

IL-4 together with IL-5 and IL-13, however IL-5 is also expressed in eosinophils. The coordinated activity of these genes is known to be regulated through chromatin modifications resulting in DNase I hypersensitive sites, gains of histone acetylation and DNA demethylation (Takemoto et al. 1998, Lee, 2002, Avni, 2002). Using 3-C analysis, Spilianakis & Flavell demonstrated that Th-2 cytokine genes are regulated by long-range chromatin interactions between the gene promoters and a locus control region (LCR) present in the 3' end of the *Rad50* gene and that these interactions are dependent on GATA3 and STAT6. Interestingly, GATA3-expressing retrovirus transfection of 3T3 fibroblasts together with ionomycin was used to analyse the contribution of GATA3 in the conformation of this locus, although the authors state that despite detectable chromatin interactions, fibroblasts are unable to support cytokine gene expression, suggesting within a given locus, chromatin interactions *per se* are not enough to mediate correct gene expression. Never the less, GATA3 was identified by chromatin IP (ChIP) and electrophoretic mobility-shift assays (EMSA) to bind *Rad50* hypersensitive site 7 (RHS7) that is essential for Th-2 cytokine expression and long-range chromatin interactions within this locus (Lee et al. 2005).

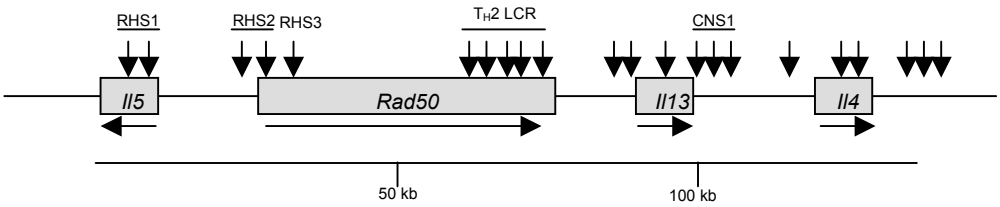


Figure 3.7: Schematic representation of the T_H2 cytokine gene locus.
Adapted from Lee et al 2005. The expression of cytokine genes is influenced by sequences in the ubiquitously expressed *Rad50* gene. DNase hypersensitive sites are indicated with downward arrows. The T_H2 LCR consists of RHS 4-7. Horizontal arrows represent the direction of transcription. Abbreviations, *Rad50* hypersensitive site (RHS), conserved non-coding sequence (CNS). See text for details (Lee et al. 2005).

The differentiation of Th-1 and Th-2 cells provides an intriguing system for analysing how differential gene expression patterns are coordinated. Naïve T-cells express low levels of both Ifn-γ and IL-4 from mouse chromosome 10 and 11 respectively, and differentiation into specific effector cells requires the silencing and activation maintenance of reciprocal loci. Spilianakis et al have proposed that this is regulated through intrachromosomal interactions between the Ifn-γ promoter and regulatory regions (RHS6 and RHS7) of the Th-2 cytokine locus in naïve T-cells based on 3-C analysis and fluorescence *in situ* hybridisation (FISH) (Spilianakis et al. 2005a).

GATA3 was previously identified to bind to RHS7 and mediate interchromosomal interactions at the Th-2 locus. Interestingly interchromosomal interactions are reduced in cells where RHS7 is deleted. The precise mechanism of these interactions, what factor(s) bind RHS6 and the potential role of GATA3 remains unknown.

Special AT-rich-binding protein 1 (SATB1) is a multi-functional, nuclear-matrix-associated protein predominantly expressed in T-cells. SATB1 regulates long-range transcriptional activation and repression of many different genes via interaction with chromatin remodeling complexes (Yasui et al. 2002). T-cell development is strongly impaired in the absence of SATB1, as thymocytes do not develop beyond the DP stage (Alvarez et al. 2000). An indication for the importance of SATB1 came from indirect immunofluorescence of SATB1 localisation in T-cells where a cage-like distribution throughout the nucleus was observed (Cai et al. 2003) and it was suggested SATB1 tethers chromatin loops to the nuclear matrix. Significantly, SATB1 is implicated in the regulation of interchromosomal interactions at the Th-2 locus and mediates Th-2 cytokine expression (Cai et al. 2006). RNAi depletion of SATB1 suggests both chromatin loop formation and cytokine activation are dependent upon SATB1, however in the absence of CTCF, activation of Th-2 cytokines is impaired despite SATB1 expression (this thesis). It is currently unclear if SATB1 remains functional in CTCF knock out cells, or indeed what chromosomal interactions remain or are mediated by CTCF at this locus.

γδ T-cells

The majority of knowledge regarding T-cell development comes from the analysis of TCRαβ cells, however these are not the only representatives in the T-cell army of immunological defence. All species known to possess αβ T-cells also have γδ T-cells (Kaufmann 1996). In both humans and mice, γδ T-cells constitute around 10% of all T-cells found in the circulation and peripheral lymphoid tissues, in mice however γδ T-cells importantly are the major population of T-cells present in skin and mucosal epithelium (Havran and Boismenu 1994). Mucosal epithelial tissues, such as the skin, tongue, lungs, intestine and genitourinary tract present an important porous barrier that regulates the prevention of pathogen invasion while mediating the exchange of nutrients. This is frequently the initial site of pathogen encounter, and as such the epithelia has a specialised immune cell niche environment, the intraepithelial lymphocyte compartment (IEL). Interestingly, despite the relative rarity of IELs in one given location, the mucosal epithelium constitutes large areas of the body; as such the total number of IELs is greatly in excess of the number of lymphocytes circulating in the blood (Kaufmann 1996).

γδ T-cells develop within the thymus, differentiating from DN1 cells. Developing T-cells have 4 TCR genes (α, β, δ, γ) that undergo separate and coordinated

rearrangements, the earliest detection of which involves the *Tcrγ* gene locus at the DN2 stage (Capone et al. 1998). Rearrangement of the *Tcrγ* locus is regulated by E2A proteins and commits T-cell differentiation to the γδ-cell lineage (Bain et al. 1999). Interestingly, the fate of pro-T cells ($CD25^+CD44^+$c-$kit^-CD3^-CD4^-CD8^-$) to differentiate as either αβ or γδ cells, is regulated by interleukin receptor-7 (IL-7R) expression levels (Kang et al. 2001). Understanding of γδ-T-cell developmental regulation is lacking, although subsequent rearrangement of *Tcrδ* genes is critical, as demonstrated by mice deficient for this locus (Girardi 2006). Very little is known regarding proteins expressed on the cell surface of γδ-cells and the γδ-TCR itself is currently the only distinguishing marker for these cells. Most γδ-cells are $CD4^-CD8^-$, however some express a CD8αα phenotype. Normal repression of *CD8α* and *β* genes in γδ-cells is mediated by DNA methylation (Lee et al. 2001). The γδ-*Tcr* loci undergo rearrangements by V(D)J recombination like αβ genes, and although lower numbers of variable gene segments are available, γδ-TCRs have higher potential for diversity due to the use of δ-chains with variable lengths and sequences at V/D/J joining sites that can accommodate multiple nucleotide insertions known as complimentary determining region 3 (CDR3) (Kaufmann 1996, Rock et al. 1994). In mice, six subsets of γδ-T-cell exist characterized by TCRs composed of differential variable chains (Girardi 2006). γδ $V\gamma5V\delta1^+$ cells initiate a wave of migration from the fetal thymus, settling in epidermal tissues. These cells have almost no further capacity for TCR diversity. Other TCR-restricted γδ-cell subsets migrate from the neonatal and adult thymus in a developmentally regulated manner. The expression of particular variable chains is thought to direct specific γδ-cells to particular mucosal areas, $V\gamma5^+$ cells are located in the skin, whereas $V\gamma6^+$ and $V\gamma7^+$ T-cells reside in the genitourinary tract and intestines respectively. $V\gamma1^+$ is the predominant re-circulating γδ-cell type found in the blood and lymphatic system.

Immunosurveillance by γδ cells is critical to controlling infections of bacterial and protozoal origin demonstrated by mouse mutants and treatment of mice with anti-TCRγδ monoclonal antibodies. The functional capacity γδ-T-cells classifies them as part of the innate immune system; rapid antigen recognition by restricted γδ-TCRs occurs for the most part independently of MHC molecules and does not require previous exposure to antigen, although differentiation into cells with a memory phenotype, characteristic of adaptive immunity also occurs with γδ cells (Holtmeier and Kabelitz 2005). Interestingly, the population of γδ cells increases significantly in αβ-T-cell deficient mice; suggesting γδ cell development is sensitive to the presence of αβ cells (Viney et al. 1994). Importantly, γδ cells are able to express 'characteristic' αβ $CD4^+$ Th-1 and Th-2 cytokines (Ifn-γ and IL-4) when activated by bacterial and helminth antigens accordingly (Ferrick et al. 1995).

An emerging aspect of γδ T-cells is their anti-tumour effect and the potential use of γδ cells in immunotherapy. The predominant site of γδ cells is the skin. Girardi et al

analysed the development of cutaneous malignancies in mice deficient for γδ cells by intradermal injections of carcinogenic mutagens and transformed cells (Girardi et al. 2001). This study revealed γδ cells impede the onset and progression of tumours and are important for cytolytic clearance of tumour cells. Furthermore, γδ cells are frequently found associated with a variety of tumours in humans. The possible clinical application of γδ cells in tumour immunosurveillance and protection against graft-versus-host disease is reviewed in (Lamb and Lopez 2005).

Chapter 4

CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus

Based on Genes and Development 20(17): 2349-2354.

Chapter 4

CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus.

Erik Splinter[1],[1] Helen Heath[1],[1] Jurgen Kooren, Robert-Jan Palstra, Petra Klous, Frank Grosveld, Niels Galjart, and Wouter de Laat

[1]Equal contribution

Department of Cell Biology and Genetics, Erasmus Medical Center, 3000 DR Rotterdam, The Netherlands.

Received March 14, 2006; revised version accepted June 12, 2006.

Summary

CTCF (CCCTC-binding factor) binds sites around the mouse *β-globin* locus that spatially cluster in the erythroid cell nucleus. We show that both conditional deletion of CTCF and targeted disruption of a DNA-binding site destabilize these long-range interactions and cause local loss of histone acetylation and gain of histone methylation, apparently without affecting transcription at the locus. Our data demonstrate that CTCF is directly involved in chromatin architecture and regulates local balance between active and repressive chromatin marks. We postulate that throughout the genome, relative position and stability of CTCF-mediated loops determine their effect on enhancer-promoter interactions, with gene insulation as one possible outcome.

Introduction

Chromatin insulators are DNA sequences that confer autonomous expression on genes by protecting them against inadvertent signals coming from neighbouring chromatin. CTCF (CCCTC-binding factor) is the prototype vertebrate protein exhibiting insulator activity (West et al. 2002) that can act as an enhancer-blocker or as a barrier against repressive forces from nearby heterochromatin (Defossez and Gilson 2002; Recillas-Targa et al. 2002). In vivo, CTCF binds to the imprinting control region of the *H19/insulin-like growth factor (Igf2)* locus, where it is thought to act as a methylation-sensitive enhancer blocker (Bell and Felsenfeld 2000; Hark et al. 2000). Moreover, CTCF binding sites have been found - and its insulator activity has been anticipated- at the imprinting center that determines choice of X inactivation (Chao et al. 2002), at boundaries of

domains that escape X inactivation (Filippova et al. 2005) and at sites flanking CTG/CAG repeats at the *DM1* locus (Filippova et al. 2001). CTCF was first defined as an insulator protein when it was found to be required for the enhancer-blocking activity of a hypersensitive site 5' of the chicken *β-globin* locus (5' HS4) (Bell et al. 1999). A similar CTCF-dependent insulator site was subsequently found at the 3' end of the locus and both sites coincide with erythroid-specific transitions in DNase I sensitivity of chromatin (Saitoh et al. 2000). Such observations suggested that CTCF partitions the genome in physically distinct domains of gene expression. The molecular mechanism underlying CTCF's insulating activity is still unknown.

CTCF-binding sites also flank the human and mouse *β-globin* locus (Fig. 1A), which contain a number of developmentally regulated, erythroid-specific *β-globin* genes and an upstream locus control region (LCR) required for high *β-globin* expression levels. In mice, three CTCF-binding sites have been identified upstream (HS-85, HS-62 and HS5) and one downstream (3' HS1) of the locus (Farrell et al. 2002; Bulger et al. 2003). Previously we have applied chromosome conformation capture (3C) technology (Dekker et al. 2002) to study long-range DNA interactions between these and other sites in the *β-globin* locus. In erythroid cells, the CTCF-binding sites (including HS-85, see below) participate in spatial interactions between the LCR and the active *β-globin* genes, and collectively form an Active Chromatin Hub (ACH) (Tolhuis et al. 2002). No such long-range DNA interactions were detected in non-erythroid cells. However, in established I/11 erythroid progenitor cells that do not yet show activated *β-globin* gene expression, contacts between the LCR and the genes are absent, but long-range DNA interactions already exist between the hypersensitive sites that contain CTCF-binding sites (Palstra et al. 2003). Here, we investigated the involvement of CTCF in the formation of these loops.

Results and Discussion

β-globin locus conformation in erythroid cells with reduced levels of CTCF protein

To investigate the role of CTCF in the formation of chromatin loops, we analyzed β-globin DNA contacts in cells lacking the CTCF protein. Analysis was focused on E12.5 erythroid progenitor cells because they can be expanded ex vivo (Dolznig et al. 2001) and lack stable LCR-gene contacts, therefore best reveal the interactions between outer hypersensitive sites. Chromatin immunoprecipitation (ChIP) experiments revealed that CTCF was bound in vivo to cognate sites in the ®-globin locus in these cells (Fig. 1F), while the protein was absent from HS5 and 3' HS1 in brain cells not showing these loops (Supplemental Figure 1). Since CTCF-null mice die early during embryogenesis, a conditional knock-out mouse model was generated by inserting two lox sites upstream and downstream of the first and last coding exon of CTCF, respectively. To delete CTCF,

fetal liver cells were isolated from lox/lox E12.5 embryos, cultured under conditions that select erythroid progenitors, and infected with a replication-deficient lentivirus expressing Cre recombinase (Fig. 1B). Heterozygous (lox/wt) cells from littermate embryos underwent the same treatment and served as controls. Cre recombination resulted in nearly 100% deletion of targeted CTCF alleles, with a reduction in mRNA and protein levels to 2-3% and 10-25%, respectively in lox/lox as compared to wild-type (Fig. 1C-E). CTCF binding to cognate sites in the *β-globin* locus was reduced but not completely abolished in lox/lox cells, as demonstrated by ChIP (Fig. 1F). To investigate *β-globin* locus conformation in these cells by 3C technology, we used a novel taqman probe-based quantitative polymerase chain reaction (Q-PCR) strategy to accurately quantify 3C ligation efficiencies (Supplementary Fig. 2).

We found that the structure of the *β-globin* locus in wild-type and lox/wt E12.5 progenitor cells was essentially the same as previously observed in l/11 progenitor cells (data not shown), with long-range interactions between the CTCF-binding sites HS-85, HS-62/60, HS4/5 and 3' HS1 (Fig. 1G). In lox/lox cells containing lower levels of CTCF protein, however, clearly reduced DNA-DNA interaction frequencies were observed specifically between the sites that normally bind CTCF (Fig. 1G-H). This is true for all combinations of binding sites, except for the interaction between 3 HS1 and HS-62 (but see below). The results demonstrated that CTCF is required for long-range DNA-DNA interactions between cognate binding sites in the *β-globin* locus. Gene expression analysis revealed the same low levels of expression for all *β-globin* genes in wild-type versus lox/lox progenitor cells (Supplementary Fig. 3). Moreover, we did not find activation of any of the mouse olfactory genes immediately surrounding the *β-globin* locus (MOR5B1-3 and MOR3B14) (data not shown). Hence, the reduction of CTCF protein to low levels had no appreciable effect on gene expression at, or around, the *β-globin* locus in erythroid cells representing a differentiation stage prior to LCR-mediated gene activation.

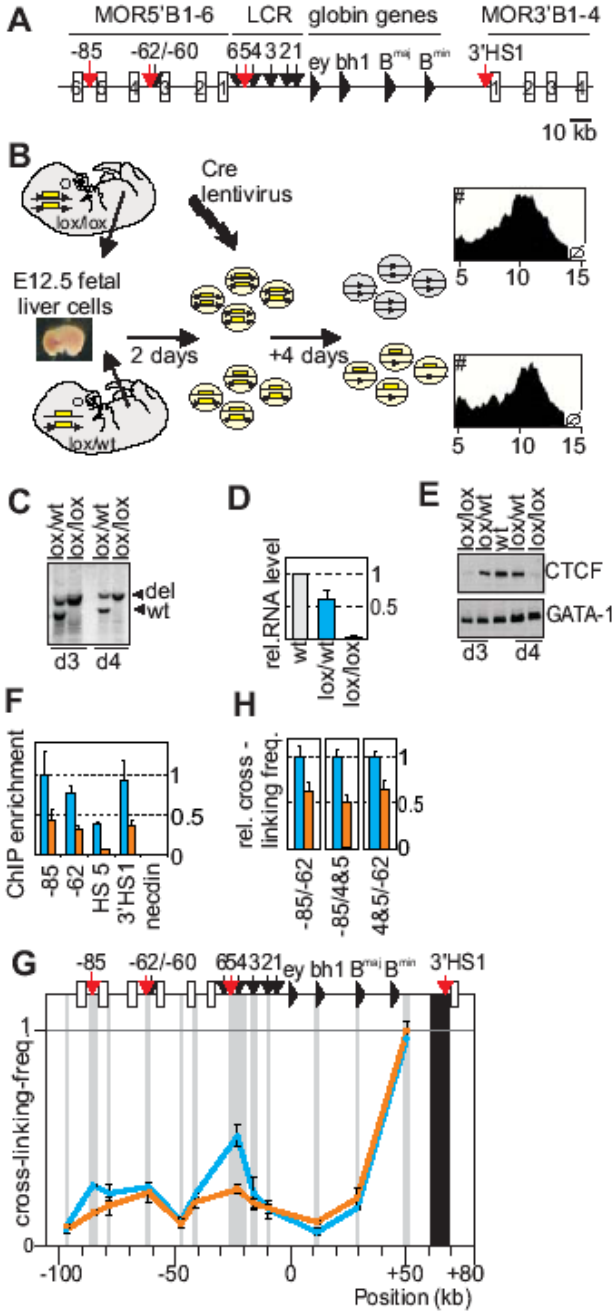

A
MOR5'B1-6
LCR
globin genes
MOR3'B1-4
-85
-62/-60
654321
ey bh1 Bmaj Bmin
3'HS1
10 kb
B
Cre
lentivirus
lox/lox
E12.5 fetal
liver cells
2 days
+4 days
lox/wt
5
10
15
C
lox/wt
lox/lox
lox/wt
lox/lox
del
wt
d3
d4
D
rel.RNA level
wt
lox/wt
lox/lox
1
0.5
E
lox/lox
lox/wt
wt
lox/wt
lox/lox
CTCF
GATA-1
d3
d4
F
ChIP enrichment
1
0.5
0
-85
-62
HS 5
3'HS1
necdin
H
rel. cross - linking freq.
1
0.5
0
-85/-62
-85/4&5
4&5/-62
G
-85
-62/-60
654321
ey bh1 Bmaj Bmin
3'HS1
cross-linking-freq
1
0
-100
-50
0
+50
+80
Position (kb)

Figure 1. Deleting CTCF in primary erythroid progenitors reduces the frequency of interactions between cognate binding sites in the β-globin locus

(A) Schematic presentation of the mouse *β-globin* locus. DNase I hypersensitive sites (arrows) and CTCF-binding sites (red) are indicated.
(B) Strategy to delete *Ctcf*. The conditional knock-out allele is composed of two lox sites (triangles) flanking the coding part of the *Ctcf* gene (yellow rectangle). Cells containing lox/lox or wt/lox *Ctcf* loci are infected with a Cre recombinase expressing lentivirus to delete the CTCF gene leaving a single lox site. The cellular size distribution is the same for homozygous and heterozygous conditional knock-out cells 4 days after infection.
(C) Southern blot analysis showing complete deletion of *Ctcf* conditional knock-out alleles in lox/wt and lox/lox cells, 3 and 4 days after infection.
(D) CTCF mRNA levels (as determined by quantitative RT-PCR) in untreated wild-type (level set to 1) and Cre-recombined lox/wt and lox/lox cells, 4 days after infection. Expression was normalized to HPRT. Standard deviation is indicated.
(E) Western blot analysis of CTCF protein and GATA-1 protein (control, stripped and re-hybridized blot), showing strongly reduced amounts of CTCF protein in lox/lox cells, 3 and 4 days after infection. (F) CTCF ChIP analysis, showing reduced CTCF-binding to cognate sites in the *β-globin* locus in lox/lox (orange) versus control cells (blue) (4 days after infection). Values were normalized to input signals and expressed as enrichment over necdin.
(G) 3C analysis, demonstrating reduced interaction frequencies between 3 HS1, HS4/5 and HS-85 in lox/lox cells (orange), as compared to control cells (blue) (4 days after virus infection). (H) 3C analysis, demonstrating reduced interaction frequencies in lox/lox cells between the other CTCF-binding sites in the *β-globin* locus.

Long-range interactions of 3' HS1 containing nucleotide changes that disrupt CTCF binding

The structural changes in the *β-globin* locus that we observed in cells with deleted CTCF may be a direct consequence of reduced protein-binding to the locus, or could be caused by secondary pathways that fail to act on the locus in the absence of sufficient CTCF. To investigate this, we disrupted CTCF-binding locally by changing four conserved nucleotides in the core CTCF-binding site of the endogenous 3'HS1 (Supplementary Fig. 4). Band-shift assays confirmed that these alterations completely abolished CTCF-binding *in vitro* (Supplementary Fig. 5). Targeting was performed in embryonic stem (ES) cells that were established from a cross between the two inbred strains 129 and C57BL/6 (B6); because of the genetic origin of the construct, targeting was directed to the B6 allele. Two additional, non-conserved, nucleotides were changed 70 base pairs downstream of the core CTCF-binding site to allow allele-specific analysis of CTCF-binding to 3'HS1 by chromatin immunoprecipitation (ChIP). Moreover, an extra *HindIII* restriction site was introduced ~850bp downstream of the CTCF-binding site, which enabled us to exclusively analyse DNA interactions of the targeted 3'HS1 by 3C. An independent control ES line was generated containing the extra *HindIII* site with the

normal 3'HS1. In each cell line the neomycin selection cassette was removed by transient expression of Cre recombinase, leaving behind a single lox site immediately downstream of the newly introduced *HindIII* site (Supplementary Fig. 4).

Definitive erythroid progenitors were generated from the ES cells in vitro to analyse the consequences of the targeted nucleotide changes in erythroid cells. We established two such ES-EP cell lines, ES-EP (Δ3'HS1) (or 'Δ') and the control line ES-EP(c) (or 'c') (Fig. 2A). After validation of the cells as a model system for erythroid differentiation (Supplimentary fig. 6; Carotta et al. 2004), we analysed CTCF-binding to mutated and wild-type 3' HS1 in vivo. In the control line ES-EP(c), CTCF bound strongly and equally well to 3' HS1 on both alleles. In ES-EP (Δ3' HS1) however, binding to 3' HS1 on the non-targeted 129 allele was the same as in ES-EP(c), but binding to the mutated 3' HS1 on the targeted B6 allele was completely abolished (Fig. 2B). Thus, the change of 4 nucleotides (nt) in the core CTCF-binding site prevented binding of CTCF to 3' HS1 also in vivo. Next we analysed whether disruption of CTCF binding at 3' HS1 affected its long-range DNA interactions in the *β-globin* locus. The extra *HindIII* restriction site introduced downstream of 3' HS1 was used to focus 3C analysis exclusively on the targeted B6 allele. In undifferentiated ES-EP(c), the wild-type *β-globin* B6 allele formed a chromatin hub typically observed in normal erythroid progenitor cells, with 3' HS1 interacting with HS4/5, HS-62 and HS-85 (Fig. 2C). In undifferentiated ES-EP (Δ3' HS1) however, the mutated 3' HS1 showed a dramatic drop in interaction frequencies with all these DNA elements, to levels similar to those observed in non-expressing fetal brain cells (Fig. 2C). Thus, disruption of CTCF binding to 3' HS1 severely destabilized the large chromatin loop containing the LCR and the globin genes in erythroid progenitor cells. The fact that the interaction with HS-62 was lost upon targeted disruption of CTCF-binding to 3' HS1, but not in our conditional CTCF knock-out experiments suggests that this interaction is more resistant than others to the reduction of levels of CTCF protein.

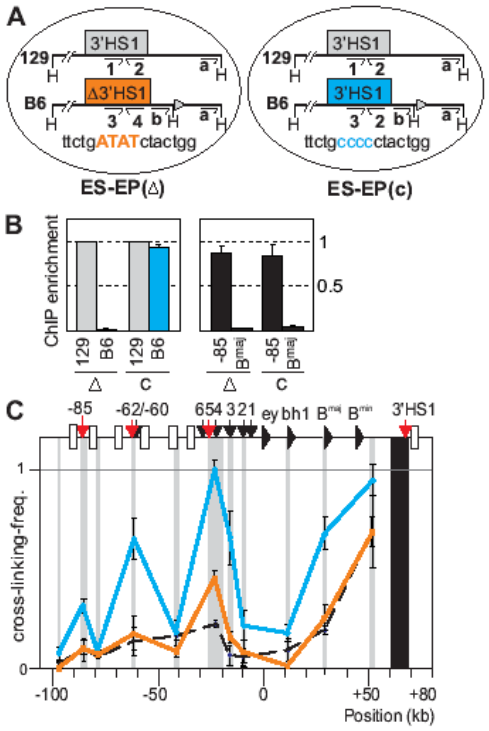


Figure 2. Targeted nucleotide changes in 3 HS1 disrupt CTCF-binding and reduce frequency of long-range 3 HS1 interactions

(A) Erythroid progenitor cell lines derived *in vitro* from ES cells. ES-EP("3 HS1) harbours 4 targeted nucleotide changes in the core CTCF-binding site of 3 HS1 on the B6 allele (orange). ES-EP(c) contains wild-type 3' HS1 on the B6 allele (blue). The non-targeted, intact, 129 allele is in grey. For ChIP, each 3 HS1 CTCF-binding site can be analysed with a unique primer pair (#1-4). An extra *HindIII* site targeted downstream of 3 HS1 allows exclusive 3C analysis of B6 allele (with 3C-primer b).
(B) Absence of CTCF-binding to mutated 3 HS1 in vivo. ChIP on undifferentiated ES-EP cell lines, with antibody against CTCF. Enrichment over necdin (normalized to input) was plotted; highest value set to 1. Left: 3 HS1 alleles in the two ES-EP lines (" and c); right: positive (HS-85) and negative (*β-major*) controls.
(C) 3C analysis with primer b to exclusively analyse 3 HS1-DNA interactions on the targeted B6 allele in undifferentiated ES-EP(c) (blue) and ES-EP ("3 HS1) cells (orange). Note that interaction frequencies with mutated 3 HS1 (orange) are reduced compared to wild type 3 HS1 (blue). Black hatched line: 3 HS1 interactions in fetal brain, analysed with primer a and plotted for comparison.

Expression of β-globin and surrounding olfactory receptor genes in the absence of CTCF-mediated chromatin loops

Since the large, CTCF-dependent, loops are formed only in human or mouse cells that are committed to, or highly express the *β-globin* genes (Palstra et al. 2003), we investigated the relationship between these loops and transcriptional regulation in detail. First, we analysed whether CTCF at 3' HS1 serves as an enhancer-blocker that prevents the inappropriate activation of downstream mouse olfactory receptor genes (*MOR*s) by the β-globin LCR in erythroid cells, as suggested previously (Farrell et al. 2002). For this, we compared mRNA levels of the *MOR3'B1-4* genes between differentiated ES-EP(c) and ES-EP (Δ3' HS1) cells when the LCR is fully active. We found no inappropriate activation of any of the downstream *MOR*s, nor of *MOR5'B3*, in the differentiated ES-EP (Δ3' HS1) cells (data not shown) and we concluded that insulator activity of CTCF at 3' HS1 is not required for blocking LCR-mediated activation of downstream MOR genes in ES-EP cells. Noteworthy, previously it was found that deletion of the complete HS5 from the endogenous locus also had no effect on expression of the surrounding *MOR* genes (Bulger et al. 2003). We envision that the transcription factor environment in erythroid cells does not support the activation of olfactory receptor genes. Next, we analysed whether the CTCF-dependent loops influence *β-globin* gene expression. Upon erythroid differentiation, the LCR forms stable contacts with the active *β-globin* genes and strongly up-regulates their transcription rate (Carter et al. 2002; Tolhuis et al. 2002). We reasoned that a shared presence on one chromatin loop anchored by CTCF in progenitor cells would decrease the spatial distance between LCR and genes, which may facilitate their productive interaction later during differentiation. If true, absence of such a pre-existing loop could possibly result in a delay of full *β-globin* gene activation. To test this, we compared the kinetics of LCR-mediated gene activation between the individual alleles of differentiating ES-EP(c) and ES-EP (Δ3' HS1) cells. Two sets of *β-major* intron primers were designed that allowed independent analysis of ongoing transcription from the B6 allele and 129 allele (Supplementary Fig. 7). ES-EP(c) and ES-EP (Δ3'HS1) cells were induced to undergo synchronous differentiation and RNA was collected at various time intervals. As expected, *β-major* transcription rates increased strongly upon differentiation. However, at each given stage of differentiation, we detected the same gene activity between the 129 and B6 allele, both in ES-EP(c) and ES-EP (Δ3'HS1) cells (Fig. 3A). Thus, the CTCF-dependent chromatin loop with 3' HS1 that topologically defines the *β-globin* locus in erythroid cells does not detectably influence the expression kinetics or levels of the *β-globin* and nearby *MOR* genes. This was also true for the embryonic *β-globin* genes *εγ* and *βH1*, which were off in both cell lines before and after differentiation (data not shown).

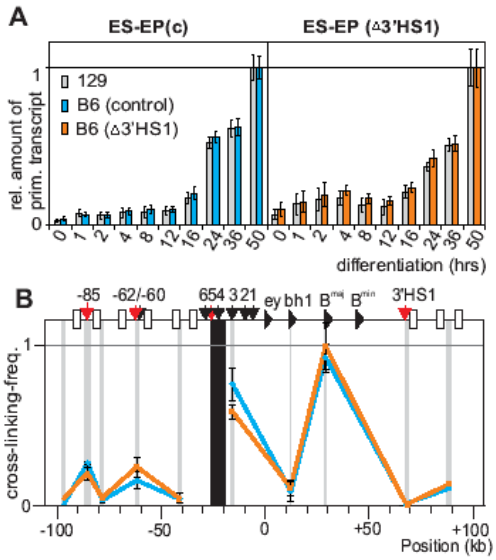


Figure 3. Targeted nucleotide changes in 3 HS1 do not affect ≤-globin gene expression

(A) Expression kinetics are the same between mutated B6 allele and intact 129 allele. Ongoing ≤major transcription was measured by quantitative RT-PCR, using 129 and B6 specific primers against intron 2 of the ≤major gene. Expression was normalized against HPRT expression levels. LCR-mediated up-regulation of *β-major* was not delayed in the absence of a pre-existing CTCF-dependent chromatin loop (compare B6 and 129 alleles in ES-EP ("3 HS1). X-axis: hours after induction of differentiation. Error bars represent standard-error of mean.
(B) Locus-wide, B6-specific, analysis of interaction frequencies with HS4/5 after differentiation. Graphs show that HS4/5 interactions with *β-major* and rest of the locus are not affected in ES-EP ("3 HS1) after differentiation. Note that primer a was used near 3 HS1, which on the B6 allele analyses a small (0.5 kb) fragment downstream of (i.e. not containing) 3 HS1, whereas on the 129 allele this primer would analyse an ~8kb fragment encompassing 3 HS1. The dramatic drop in interaction frequencies reassuringly shows that analysis is restricted to the B6 allele.

Establishment of LCR-gene contacts in the absence of a pre-existing loop with 3' HS1

The unaltered *β-globin* gene expression patterns from the targeted allele in ES-EP (Δ3' HS1) cells suggested that in the absence of a pre-existing chromatin loop, LCR-gene contacts can still be established normally upon erythroid differentiation. To test this, we searched for 129/B6 polymorphisms near restriction sites in the LCR that would allow allele-specific 3C analysis. This resulted in the design of a TaqmMan probe for a *HindIII*

fragment encompassing HS4 and HS5 that signals exclusively from the B6 allele (Supplementary Fig. 8). Although HS4 and HS5 are not prime candidates in the LCR to directly contact the genes, this relatively large *HindIII* fragment was previously shown to be representative of the complete LCR, since it displayed a prominent peak of interaction with the *β-major* gene upon erythroid differentiation (Tolhuis et al. 2002). In both differentiated ES-EP(c) and ES-EP (Δ3' HS1) cell lines, we found identical locus-wide interaction frequencies for HS4/5 between the B6 alleles containing either wild-type or mutated 3' HS1, and both showed a strong peak of interaction with the *β-major* gene (Fig. 3B). This demonstrated indeed that a pre-existing loop between upstream sites and 3' HS1 is dispensable for the establishment of stable LCR-gene contacts later during erythroid differentiation. Such a conclusion is in agreement with transgenic experiments showing full β-globin expression from constructs lacking 3' HS1 (Strouboulis et al. 1992).

Histone modifications in the absence of CTCF binding

3' HS1 was previously shown to be present in, and close to the 3' border of an open chromatin domain spanning ~145 kb around the *β-globin* locus in erythroid cells (Bulger et al. 2003). Within this domain, a large (~15kb) region of highly repetitive DNA is present approximately 3kb upstream of 3' HS1, that cannot be analysed for nuclease sensitivity, but likely is packed into compact chromatin. To further investigate this, we analysed histone modifications at and directly around 3' HS1 in ES-EP cells. Using an antibody that recognizes both di-methyl H3K9 and di-methyl H3K27, we found that these repressive marks were abundantly present on both sides of 3' HS1, but not inside the 3' HS1 (Fig. 4A). Conversely, acetylation of histone H3, a mark for open chromatin, was clearly enriched at 3' HS1 but not, or much less, at sites surrounding the hypersensitive site (Fig. 4C). These data argued against the existence of a large open chromatin domain extending across 3' HS1 and suggested that 3' HS1 forms an isolated entity of open chromatin. To address whether CTCF plays a role in the establishment of this pattern, we performed ChIP on ES-EP(c) and ES-EP (Δ3' HS1) cells and used allele-specific primer pairs to compare modifications at 3' HS1 on targeted, versus non-targeted alleles. In the control cell line, we found identical levels of di-meH3K9/K27 at 3' HS1 on the two alleles. In ES-EP (Δ3' HS1) however, loss of CTCF binding was accompanied by an increase of di-meH3K9/K27 and concomitant decrease of AcH3 at 3' HS1 (Fig. 4B-C). We found no indication for spreading of the methyl mark into the locus, either locally (Fig. 4, compare levels of enrichment between A and B), or at the *β-major* gene, which locates more inside the locus (Fig.4B, analysed by allele-specific primers). In fact, AcH3 levels at *β-major* were also similar for the targeted and non-targeted allele in ES-EP (Δ3' HS1) cells - two observations that were fully in agreement with our finding that *β-major* gene expression was not affected by disrupted CTCF binding to 3'HS1 (Fig. 3A).

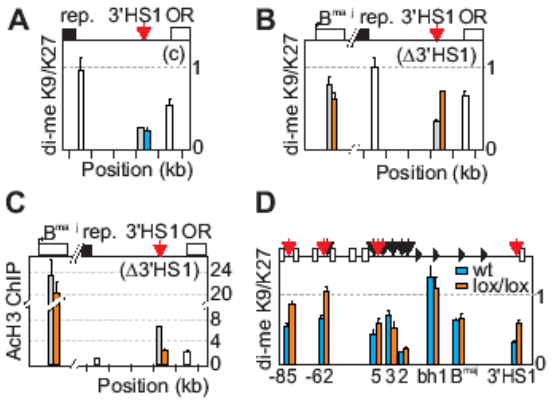


Figure 4. Histone modifications in the absence of CTCF binding

(A) ChIP enrichment for di-meH3K9/K27 2.8kb upstream, 1.2kb downstream and at 3 HS1 in undifferentiated control ES-EP (c) cells. White bars: non-allele-specific primers; grey bar: 129-allelespecific primer pair; blue bar: B6-allele specific primer pair. Values in 4A-C were normalized to input (therefore expressed per allele) and enrichment expressed relative to levels observed at inactive necdin (set to 1).
(B) ChIP enrichment for di-meH3K9/K27 at and around 3 HS1 and at ≤major in undifferentiated ES-EP ("3 HS1) cells. White bars: non-allele-specific primers; grey bars: 129-allele-specific primer pairs; orange bars: B6allele specific primer pairs. Note increased methylation only at mutated 3 HS1 (C) ChIP enrichment for acetylated histone H3 at and around 3 HS1 and at ≤major in undifferentiated ES-EP ("3 HS1) cells. Note decreased acetylation only at mutated 3 HS1
(D) ChIP enrichment for di-meH3K9/K27 at the *β-globin* locus in control (blue) and conditional CTCF knock-out (lox/lox) (orange) E12.5 erythroid progenitors cells. Note that reduced levels of CTCF cause a drop in di-meH3K9/K27 at the CTCF-binding sites HS-85, HS-62, HS5 and 3 HS1 (indicated by red arrows), but not elsewhere in the locus.

We considered the possibility that spreading of di-meH3K9/K27 into the locus requires disruption of CTCF binding to more sites than just 3' HS1. To investigate this, we compared di-meH3K9/K27 levels in E12.5 wild-type versus lox/lox conditional CTCF knock-out progenitor cells, the latter containing reduced amounts of CTCF (Fig. 1). We found that loss of CTCF binding to 3' HS1, HS5, HS-62 and HS-85 coincided with locally increased amounts of di-meH3K9/K27, while modification levels elsewhere in the locus appeared unaffected (Fig. 4D). Since CTCF-binding to *β-globin* sites was reduced but not absent in lox/lox cells, this leaves open the possibility that residual CTCF amounts prevent spreading of di-meH3K9/K27 into the locus. We concluded that CTCF regulates the balance between active and repressive chromatin modifications at its binding sites and we propose that CTCF-mediated acetylation of histones prevents their methylation.

Mechanistically, CTCF may directly attract histone acetyl-transferases (HATs), although current evidence for this interaction is lacking. Alternatively, CTCF mediated chromatin looping brings binding sites into spatial proximity with HATs bound elsewhere to the DNA (de Laat and Grosveld 2003). The observation that CTCF-binding was required for histone acetylation is interesting because previously these two events were previously claimed to be uncoupled (Recillas-Targa et al. 2002). Our data do not support the generality of boundaries demarcating expression domains, but fit better with the concept that genes maintain autonomous expression profiles mostly through their unique ability to productively interact with positive regulatory elements (Dillon and Sabbattini 2000; de Laat and Grosveld 2003).

CTCF organises higher-order chromatin structure

We have presented two independent lines of evidence that together firmly establish that CTCF functions in the formation of chromatin loops; removal of most CTCF protein, as well as targeted disruption of a CTCF-binding site, resulted in destabilization of long range contacts between cognate binding sites in the *β-globin* locus. CTCF is critical for the looped conformation present in erythroid progenitor cells, but is dispensable for LCR-gene contacts established later during differentiation. We, and others, previously have shown that the latter contacts depend on the transcription factors EKLF and GATA-1 (Drissen et al. 2004; Vakoc et al. 2005). Together, these studies begin to delineate the factors that act sequentially to form a functional *β-globin* ACH in differentiated erythroid cells where *β-globin* genes are fully expressed. Based on the observations that CTCF dependent chromatin loops are tissue-specific and evolutionary conserved between mouse and man, it seemed reasonable to expect that these loops would play a role in gene expression. Such function may exist but is beyond our current detection limits. An alternative view is that evolutionary selection against sites forming chromatin loops within a gene locus positions them outside the *β-globin* locus, without necessarily being selected to act, positively or negatively, on gene expression (Dillon and Sabbattini 2000).

We hypothesize that CTCF also organizes higher-order chromatin structure at other gene loci and we predict that such chromatin loops facilitate communication between genes and regulatory elements but can also lead to the exclusion of interactions between elements. In terms of transcriptional regulation, the final outcome of such chromatin loops will depend on the position of CTCF-binding sites relative to other regulatory elements and the genes, the concentration of the *trans*-acting factors involved, and the affinities of the (long-range) interactions. In Drosophila, a limited 3C analysis previously provided indications for a loop between the *scs* and *scs'* enhancer blocking elements (Blanton et al. 2003). Moreover, insulator proteins like suppressor of Hairy-wing (Su[Hw]) and Modifier of mdg4 2.2 (Mod[mdg4]2.2) have been found to coalesce into large foci, called insulator bodies. These bodies preferentially localize at

the nuclear periphery and are hypothesized to bring together distant insulator sites, with intervening chromatin fibres looped out to form isolated expression domains (Gerasimova et al. 2000). Our observations made on the CTCF protein provide high-resolution insight into the nature of such loops in mammals. It will be interesting to see if CTCF forms chromatin loops through multimerization of CTCF molecules bound to distinct DNA elements (Yusufzai et al. 2004), or whether this loop formation also involves other factors. Similarly, future experiments should provide insight whether CTCF-dependent chromatin looping occurs at a defined physical structure in the nucleus (Dunn et al. 2003; Yusufzai and Felsenfeld 2004; Yusufzai et al. 2004) or whether the base of such loops has a more fluid nature.

Materials and Methods

Generation of conditional CTCF knock-out mice and CTCF antibody

Targeting constructs and strategy for the generation of conditional *Ctcf* knock-out mice as well as the polyclonal antibody generated against CTCF will be described in detail as part of a study that addresses the role of CTCF in T cell development (H. Heath, manuscript in preparation).

Lentivirus production

Cre-lentivirus was produced by transient transfection of 293T cells according to standard protocols (Zufferey et al. 1997). Briefly, 293T cells were transfected with a 3:1:4 mixture of psPAX-2, pMD2G-VSVG (kind gifts of D. Trono, University of Geneva) and a transfer vector construct that is essentially as pRRLsin.sPPT.CMV.GFP.Wpre (Follenzi et al. 2002) but with CMV-Cre instead of CMV-GFP, using poly(ethylenimine) (PEI). After 24 hours medium was refreshed and virus-containing medium was harvested 48 and 72 hours after transfection. After filtration through a 0.45μm cellulose acetate filter, the virus stock was concentrated 1000 times by centrifugation at 19.4K rpm for 2 hours at 10°C in a SW28 rotor. Virus stocks were stored at -80°C. Virus activity/functionality was tested by serial dilutions on primary mouse embryonic fibroblasts (MEFs) containing loxP sites, which were scored for recombination after 4 days of infection by Southern blotting.

Culturing primary E12.5 fetal liver cells and lentiviral infection

Fetal livers were isolated from E12.5 embryos, resuspended in FCS with 10% DMSO by repeated pipetting and stored in liquid nitrogen until genotyping of embryos was completed. Per experiment, cells from three fetal livers of the same genotype were

pooled and cultured as described (Dolznig et al. 2001). After 2 days of culturing, cells were infected by adding Cre-lentivirus to medium and centrifugation of cell culture plates for 55 minutes at 2.5K rpm (37°C). Cre-mediated recombination efficiency was analysed 3 and 4 days after infection by standard Southern and western blotting techniques (antibody used to detect GATA-1: #N6, Santa-Cruz). CTCF RNA levels were analysed by quantitative RT-PCR (see below).

3C Analysis, Chromatin Immunoprecipitation and Analysis of gene expression

These PCR-based techniques were performed according to standard procedures (Splinter et al. 2004). For details, probe and primer sequences, see supplemental information.

In vitro differentiation of ES cells into ES-EPs and characterization of ES-EPs

Differentiation of ES cells into ES-EPs, expansion of ES-EPs and in vitro differentiation of ES-EPs into erythrocytes was performed as described (Carotta et al. 2004), except that embryoid bodies were formed in 4000-6000 hanging drop cultures (~200 ES cells/drop) that were pooled and disrupted after 6 days of culturing to generate ES-EPs. After 2 to 3 weeks of cultivation a homogeneous erythroid progenitor population was obtained, as determined by FACS analysis (Table S1), cell morphology and the capacity to respond massively to physiologically relevant stimuli to terminally differentiate into enucleated erythrocytes within 72 hours.

Acknowledgements

We thank A. de Wit, T. van Dijk and M. von Lindern for help with experiments. This work was supported by a grant from the division of Earth and Life Sciences (ALW) of the Netherlands Organization for Scientific Research (NWO), through the ESF EuroDYNA programme to N.G., EU grants LSHG-CT-2004-503433 (Epigenetic) and LSHM-CT-2003-504468 (Cells into Organs) and NWO grant 912.03.009 to F.G. and NWO grants 016-006-026 and 912-04-082 to W.d.L.

References

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

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

Blanton, J., M. Gaszner, and P. Schedl. 2003. Protein:protein interactions and the pairing of boundary elements in vivo. Genes Dev 17: 664-75.

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

Carotta, S., S. Pilat, A. Mairhofer, U. Schmidt, H. Dolznig, P. Steinlein, and H. Beug. 2004. Directed differentiation and mass cultivation of pure erythroid progenitors from mouse embryonic stem cells. Blood 104: 1873-80.

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

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

Cho, D.H., C.P. Thienes, S.E. Mahoney, E. Analau, G.N. Filippova, and S.J. Tapscott. 2005. Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. Mol Cell 20: 483-9.

de Laat, W. and F. Grosveld. 2003. Spatial organization of gene expression: the active chromatin hub. Chromosome Res 11: 447-59.

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

Dekker, J., K. Rippe, M. Dekker, and N. Kleckner. 2002. Capturing chromosome conformation. Science 295: 1306-11.

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

Dolznig, H., F. Boulme, K. Stangl, E.M. Deiner, W. Mikulits, H. Beug, and E.W. Mullner. 2001. Establishment of normal, terminally differentiating mouse erythroid progenitors: molecular characterization by cDNA arrays. Faseb J 15: 1442-4.

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

Dunn, K.L., H. Zhao, and J.R. Davie. 2003. The insulator binding protein CTCF associates with the nuclear matrix. Exp Cell Res 288: 218-23.

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

Fedoriw, A.M., P. Stein, P. Svoboda, R.M. Schultz, and M.S. Bartolomei. 2004. Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting. Science 303: 238-40.

Filippova, G.N., M.K. Cheng, J.M. Moore, J.P. Truong, Y.J. Hu, D.K. Nguyen, K.D. Tsuchiya, and C.M. Disteche. 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., C.P. Thienes, B.H. Penn, D.H. Cho, Y.J. Hu, J.M. Moore, T.R. Klesert, V.V. Lobanenkov, and S.J. Tapscott. 2001. CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. Nat Genet 28: 335-43.

Follenzi, A., G. Sabatino, A. Lombardo, C. Boccaccio, and L. Naldini. 2002. Efficient gene delivery and targeted expression to hepatocytes in vivo by improved lentiviral vectors. Hum Gene Ther 13: 243-60.

Gerasimova, T.I., K. Byrd, and V.G. Corces. 2000. A chromatin insulator determines the nuclear localization of DNA. Mol Cell 6: 1025-35. Hark, A.T., C.J. Schoenherr, D.J. Katz, R.S. Ingram, J.M. Levorse, and S.M. Tilghman. 2000. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature 405: 486-9.

Lachner, M. and T. Jenuwein. 2002. The many faces of histone lysine methylation. Curr Opin Cell Biol 14: 286-98.

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

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

Saitoh, N., A.C. Bell, F. Recillas-Targa, A.G. West, M. Simpson, M. Pikaart, and G. Felsenfeld. 2000. Structural and functional conservation at the boundaries of the chicken beta-globin domain. Embo J 19: 2315-22.

Splinter, E., F. Grosveld, and W. de Laat. 2004. 3C technology: analysing the spatial organization of genomic loci in vivo. Methods Enzymol 375: 493-507.

Strouboulis, J., N. Dillon, and F. Grosveld. 1992. Developmental regulation of a complete 70-kb human beta-globin locus in transgenic mice. Genes Dev 6: 185764.

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

Vakoc, C.R., D.L. Letting, N. Gheldof, T. Sawado, M.A. Bender, M. Groudine, M.J. Weiss, J. Dekker, and G.A. Blobel. 2005. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. Mol Cell 17: 45362.

West, A.G., M. Gaszner, and G. Felsenfeld. 2002. Insulators: many functions, many mechanisms. Genes Dev 16: 271-88. Yusufzai, T.M. and G. Felsenfeld. 2004. The 5'-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element. Proc Natl Acad Sci U S A 101: 8620-4.
Yusufzai, T.M., H. Tagami, Y. Nakatani, and G. Felsenfeld. 2004. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. Mol Cell 13: 291-8. Zufferey, R., D. Nagy, R.J. Mandel, L. Naldini, and D. Trono. 1997. Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. Nat Biotechnol 15: 8715.

Supplemental Material and Methods

3C Analysis

3C analysis was performed essentially as described (Splinter et al. 2004), using HindIII as the restriction enzyme. Quantitative real-time PCR (Opticon I, MJ Research) was performed with Platinum Taq DNA Polymerase (Invitrogen) and double-dye oligonucleotides (5'FAM + 3'TAMRA) as probes, using the following cycling conditions: 94^0C for 2 min and 44 cycles of 15 s at 94^0C and 90 s at 60^0C. Probe and primer sequences are listed in supplementary Table 2.

Chromatin Immunoprecipitation (ChIP)

ChIP was performed as described in the Upstate protocol (http://www.upstate.com), except that cells were cross-linked at 2% formaldehyde for 5 minutes at room temperature. Quantitative real-time PCR (Opticon I, MJ Research) was performed using SYBR Green (Sigma) and Platinum Taq DNA Polymerase (Invitrogen), under the following cycling conditions: 94^0C for 2 min, 44 cycles of 30 s at 94^0C, 60 s at 55^0C, 15 s at 72^0C and 15 s at 75^0C (during which measurements are taken). Enrichment was calculated relative to Necdin and values were normalized to input measurements. Primer sequences are listed in supplementary Table 3. Antibodies used: Anti-acetyl-Histone H3 (#06-599, Upstate); anti-di-methyl Histone H3 K9/K27 (ab7312, Abcam).

Lentivirus production and infection

Cre-lentivirus was produced by transient transfection of 293T cells according to standard protocols (Zufferey et al. 1997). 293T cells were transfected with a 3:1:4 mixture of psPAX-2, pMD2G-VSVG (kind gifts of D. Trono, University of Geneva) and a transfer vector construct that is essentially as pRRLsin.sPPT.CMV.GFP.Wpre (Follenzi et al. 2002) but with CMV-Cre instead of CMV-GFP, using poly(ethylenimine) (PEI). After 24 hours medium was refreshed and virus-containing medium was harvested 48 and 72 hours after transfection. After filtration through a 0.45μm cellulose acetate filter, the virus stock was concentrated 1000 times by centrifugation at 19.4K rpm for 2 hours at 10^0C in a SW28 rotor. Virus stocks were stored at -80^0C. Virus activity/functionality was tested by serial dilutions on primary mouse embryonic fibroblasts (MEFs) containing loxP sites, which were scored for recombination after 4 days of infection by Southern blotting. Fetal livers were isolated from E12.5 embryos, resuspended in FCS with 10% DMSO by repeated pipetting and stored in liquid nitrogen until genotyping of embryos was completed. Per experiment, cells from three fetal livers of the same genotype were pooled and cultured as described (Dolznig et al. 2001). After 2 days of culturing, cells were infected by adding Cre-lentivirus to medium and centrifugation of cell culture plates for 55 minutes at 2.5K rpm (37^0C). Cre-mediated recombination efficiency was analysed

by standard Southern en western blotting techniques (antibody used to detect GATA-1: #N6, Santa-Cruz). CTCF RNA levels were analysed by quantitative RT-PCR (see below).

Analysis of gene expression

Total RNA was isolated from cultured fetal liver cells or 0.5-1 x 10^6 of ES-EPs at the indicated time points using TRIzol reagent (Invitrogen) according to the manufacturers instructions. 1 µg of RNA was treated for 1 hour at room temperature with amplification grade DNAseI (Invitrogen) to remove genomic DNA contamination. An aliquot was used as a no RT control. cDNA synthesis was performed using Superscript II RNase H-Reverse transcriptase (Invitrogen) according to the manufacturers instructions using 200ng random hexamers as primers. Quantification of primary transcripts was performed on Opticon II real-time PCR machines (MJ research) using Platinum Taq (Invitrogen) and SYBR Green (Sigma), using the following PCR program: 2 min 94^0C, 45 cycles of 30 sec 94^0C, 1 min 62^0C, 15 sec 72^0C and 15 sec 75^0C (during which measurements are taken), followed by 10 minutes chain extension and a melting curve. Expression was normalized against HPRT expression levels. Primer sequences are listed in supplementary Table 3.

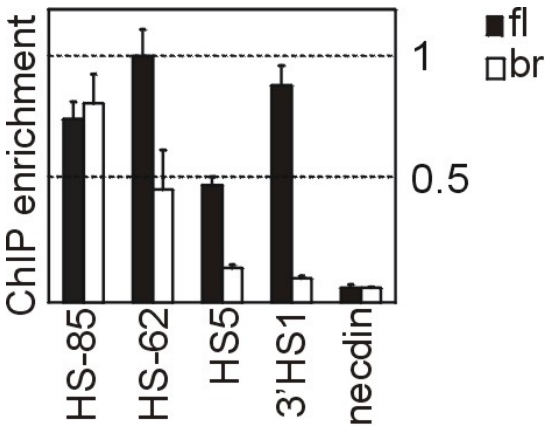
Targeting nucleotide changes to 3'HS1 in ES cells

The 3'HS1 targeting constructs were based on a 5.6 kb *BamHI-EcoRV* isolated from BAC RP23-370E12 (BACPAC Resources) (Supplementary Fig. 1). Site-directed mutagenesis was performed on an internal 683bp *NdeI-NdeI* fragment (coordinates: 67033-67716, see Supplementary Fig. 1), using QuikChange® Multi Site-Directed Mutagenesis Kit (Stratagene). Oligonucleotide used to change CTCF-binding site: CGGAAATCAGCGGAACACTTCTGATATCTACTGGTATGCAACAGG.

Oligonucleotide used to change 2 nucleotides 70 bp downstream of the core CTCF-binding site: CAGTTTATCCCAGTTACGTTTAGTTGACAACCTGAGAC. Before reintroduction into the *BamHI-EcoRV* targeting vector, the complete *NdeI* fragment was sequenced to confirm that only targeted nucleotides were changed. A TK-NEO cassette flanked by head-to-tail oriented loxP sites and containing a *HindIII* site immediately upstream one of the loxP sites was inserted as an *XbaI-SpeI* fragment into the *AvrII* site at position 68251 (Supplementary Fig. 1). For selection against random integration events, diphtheria toxin (DTA) (Yu et al. 2000) was cloned outside the region of homology. ES cells for targeting were isolated from 129xB6 F1 blastocysts and transfected with the *SalI* linearised targeting construct by electroporation. Clones scored positive for homologous recombination at 3'HS1 by Southern blot hybridization were transiently transfected with a CMV-Cre construct containing a PGK-puromycin selection cassette, followed the next day by a 40 hours selection on medium containing 2 µg/ml of

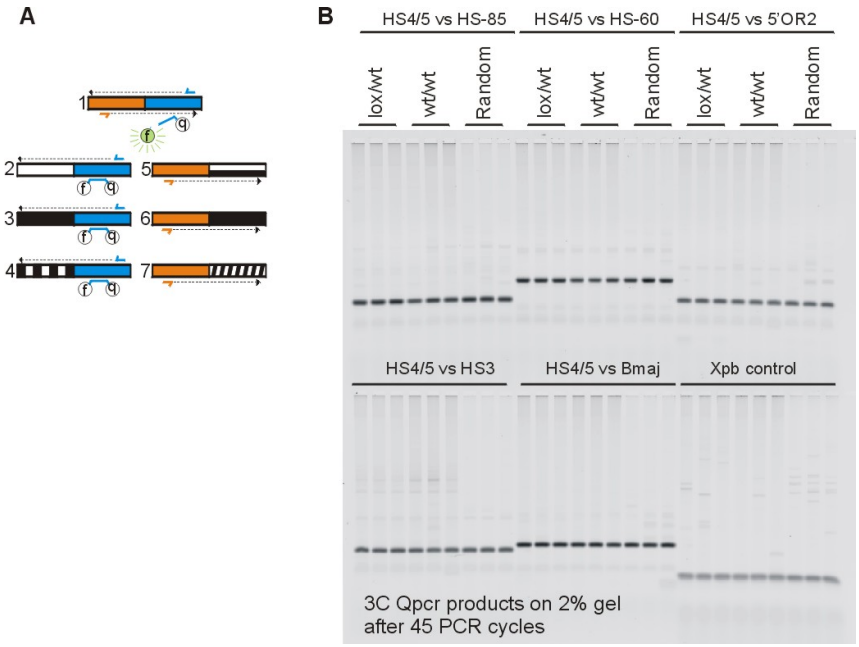
puromycin. Surviving clones were analyzed by Southern blotting for successful Cre-mediated deletion of the neomycin selection cassette and by PCR analysis for the absence of Cre. Positive clones were selected for *in vitro* differentiation into ES-EP cells.

Supplemental Figures



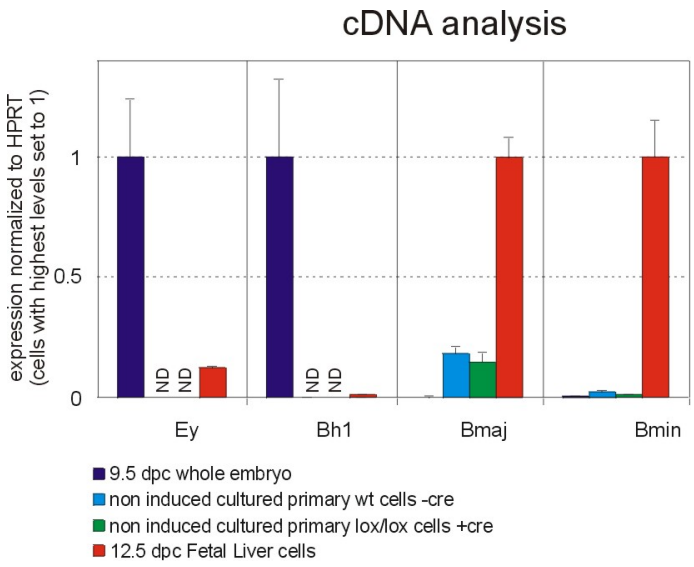
Supplemental Figure 1

ChIP with anti-CTCF antibody on E14.5 fetal liver (black) and brain (white bars).
Data were normalized against input and expressed as enrichment over necdin (highest level set to 1).



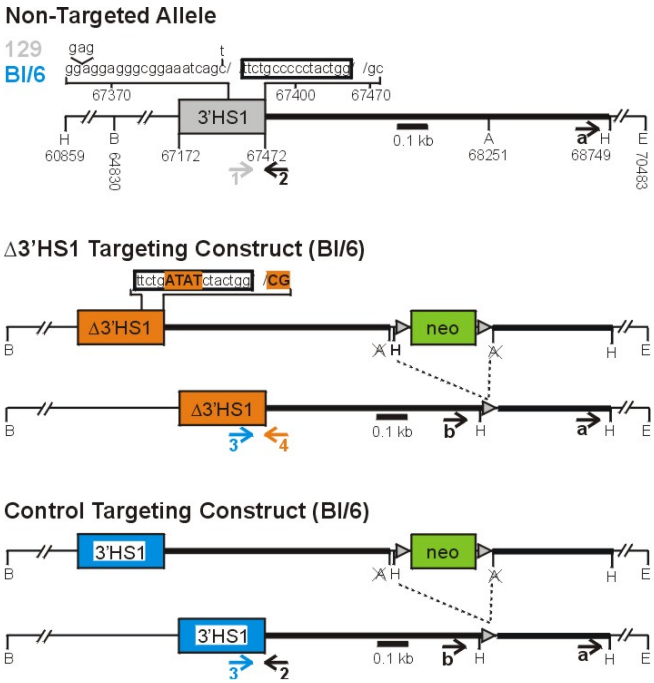
Supplemental Figure 2

(*A*) Q-PCR analysis of ligation frequencies obtained by 3C. The approach entails a primer-probe combination that is specific for a particular restriction fragment (blue), with the probe hybridizing to the opposite strand as compared to the PCR primer. A second PCR primer hybridizes to the fragment (orange) of which one wants to quantify its interaction. The primers/probe configuration guarantees that the probe only signals upon extension of the second primer across the ligated junction (# 1), which is important given the great variety of junctions (e.g. #2-7) formed with each fragment. f (fluorescent group) and q (quencher).
(*B*) Examples of PCR products obtained after 45 cycles of QPCR, analyzed on a 2% agarose gel.



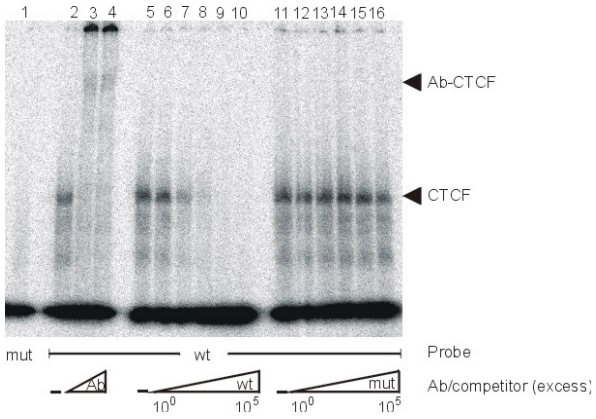
Supplemental Figure 3

Expression of all *β-globin* genes in cultured wild-type versus lox/lox progenitor cells. Messenger RNA levels were normalized to *HPRT* mRNA levels. For comparison, levels in primitive (E9.5) and definitive red blood cells (E12.5 fetal liver) are indicated and set to 1. ND: Not Detectable.



Supplemental Figure 4

Targeting strategy. Top: wild-type (non-targeted) allele. Nucleotide polymorphisms between 129 and B6 used to design allele-specific ChIP-primers #1 (129) and #3 (B6) are shown. Box indicates core CTCF-binding site. Coordinates are relative to the start of most upstream *β-globin* gene (εy). H: *HindIII*, B: *BamHI*, A: *AvrII*, E: *EcoRV*. Sequence between 67172 and 68749 is drawn to scale (see scale bar). 3C primer 'a' (is primer +68(3'HS1)) is used to analyze the 7.9 kb (60859-68749) wild-type HindIII fragment containing 3'HS1. Middle: targeting construct for □3'HS1. Targeted nucleotide changes in core CTCF-binding site (CCCC to ATAT) are indicated in orange. Downstream nucleotide changes (GC to CG) allow the design of ChIP-primer #4 that is specific for this targeted allele. The extra *HindIII* site upstream of the neomycin selection cassette flanked by two loxP sites is maintained after Cre-mediated deletion. This new *HindIII* site creates an allele-specific *HindIII* fragment around 3'HS1 that can be analyzed by primer 'b'. This fragment is 7.4 kb in size, not much smaller than the 7.9 kb fragment analyzed by primer 'a' on the non-targeted allele. On the targeted allele primer 'a' now analyzes a small (0.5kb) *HindIII* fragment downstream of 3'HS1. Bottom: targeting construct for the control ES-EP line (ES-EP(c)). 3'HS1 is untouched, but the extra *HindIII* site is introduced at the same position as before.



Supplemental Figure 5

Nucleotide changes in the core binding site of 3'HS1 effectively disrupt CTCF binding *in vitro*. Gel mobility shift assay with nuclear protein extracts from fetal livers (all lanes) (Wall et al. 1988). ^{32}P-labeled probe used: mutant (lane 1), wild-type (lane 2-16). Lane 3: +1 μl anti-CTCF antibody, lane 4: +2 μl anti-CTCF antibody, lane 6-10: competition with increasing amounts of unlabeled wild-type probe, lane 12-16: competition with increasing amounts of unlabeled mutant probe.

Probe sequences:

3'HS1-wt-S: CGGAAATCAGTGGAACACTTCTGCCCCCTACTGGTATGCAACAGG,

3'HS1-wt-AS: TCCTGTTGCATACCAGTAGGGGGCAGAAGTGTTCCACTGATTTCCG,

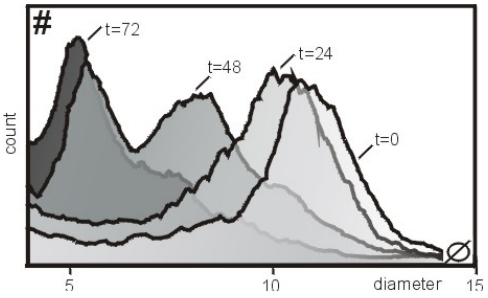
3'HS1-mut-S: CGGAAATCAGTGGAACACTTCTGATATCTACTGGTATGCAACAGG,

3'HS1-mut-AS: TCCTGTTGCATACCAGTAGATATCAGAAGTGTTCCACTGATTTCCG.

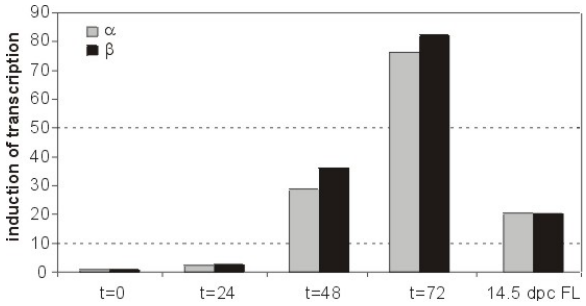
A.

Marker	ES-EP(□3'HS1)	ES-EP(c)
CD117 (%)	74	74
CD71/Ter119 (%)	98	100

B.

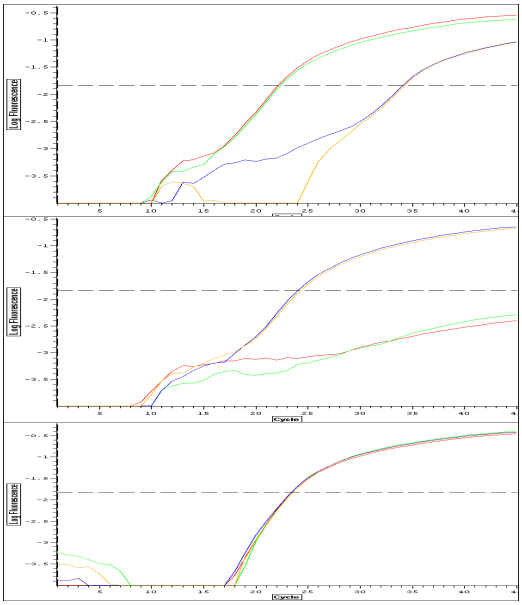


C.



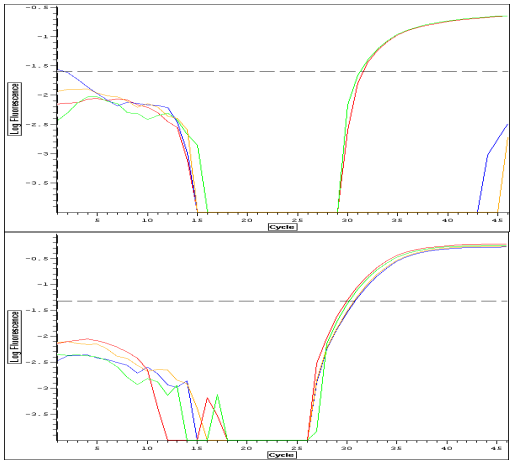
Supplemental Figure 6. Characterization of ES-EP cells

(*A*) ES-EPs express markers characteristic for proerythroblasts.
(*B*) Synchronous *in vitro* differentiation of ES-EP cells. Cell size (μm) was measured at various intervals after induction of differentiation. The plots for the Δ and c lines are identical.
(*C*) Globin mRNA expression during *in vitro* differentiation of ES-EP cells. Quantitative RT-PCR was performed to measure α- and β-globin mRNA levels relative to HPRT. Relative expression at t = 0 hours was set to 1. The fact that globin expression levels are higher in differentiated EP-EPs than in E14.5 fetal liver is attributed to the fetal liver being composed of cells at various stages of differentiation.



Supplemental Figure 7

βmajor intron 2 primers for independent analysis of ongoing transcription from the B6 allele and the 129 allele. Real-time PCR analysis plots showing the log fluorescence (y-axis) versus PCR cycle number (x-axis). Quantitative PCR with SYBR Green was performed on genomic DNA from B6 (red and green curve) and 129 cells (blue and yellow curve). Top: primers GGTTGCAAACAAATGTCTTTA (βmaj-in2-f (B6)) and CCAGGAGCTGTGGGAAG (βmaj-in2-r (B6)) amplify much more efficiently from B6 DNA than from 129 DNA (□Ct = 12, i.e. 129 contribution is ~ 1/4000). Middle: primers GGTTGCAAACAAATGTCTTTG (βmaj-in2-f (129)) and CCAGGAGCTGTGGGAAT (βmaj-in2-r (129)) exclusively amplify βmajor intron 2 from 129 DNA. Bottom: HPRT primers show that equal amounts of 129 and B6 DNA were used as templates.



Supplemental Figure 8

B6 allele-specific Taqman probe for *HindIII* fragment HS4/5. Real-time PCR analysis plots showing the log fluorescence (y-axis) versus PCR cycle number (x-axis). Quantitative PCR was performed on genomic DNA from B6 (red and green curve) and 129 cells (blue and yellow curve).
Top: taqman probe (HS45-B6) AGGAGGAATTTCTTCCGGTTGAATATGCCACAGCC hybridises near the *HindIII* site directly upstream HS5 and signals exclusively from B6 alleles. Primers used: TTCAAGTTCTCATCCTTCACTG (-21(HS4/5)), GCTTTGTGTACACTGTGCAG.
Bottom: A primer/probe combination for the βmajor promoter that works on B6 and 129 DNA shows that approximately equal amounts of DNA were analysed.

Supplemental Table 1

Sequences of primers used for 3C analysis

(names correspond to approximate position (in kb) relative to εy):

-97:	AACCTGGTAACTTTCTGCTGA
-85(HS):	AGGTTGCAGTTGGTGTGA
-79:	CTCATTTGTCAACTCATCCC
-62(HS):	GGGTGTGGGTATTTGTAAGAG
–48:	AGCCTAATCTGGTGCTGG
–42:	ATGAACAAGTTTCATGGGG
-21(HS4/5):	TTCAAGTTCTCATCCTTCACTG
–17(HS3):	CTTGTCCTATGGATGCCA
-10(HS2):	GCAGCTTCCTCATTTAGCA
+11(βh1):	ATAAGTGTAGCTGCCTGGTG
+12(βh1):	GCTTGTGATAGCTGCCTTC
+30(βmaj):	AATCGCTGCTCCCCCTCACT
+52:	TTTAATGTCACGCAAAACATCAG
+68(3'HS1):	TCCTTGCTTTTACTCTTTCTCC (primer 'a')
+88:	CATTCATTGTTCTTCTACCTCG

Primer for extra HindIII site introduced downstream of 3'HS1:

(primer 'b'):	CATCTTTGGACCCTTCTAATCC
XPB-1	GCCCTCCCTGAAAATAAGGA
XPB-2	GACTTCTCACCTGGGCCTACA

Sequences of probes:

3'HS1:	AGCTTCCTTTGAACATAACTTTGCACTTACTTGTTTCTG (with primer 'a');
HS45-B6:	AGGAGGAATTTCTTCCGGTTGAATATGCCACAGCC (with -21(HS4/5))
XPB:	AAAGCTTGCACCCTGCTTTAGTGGCC

Supplemental Table 2. Primer sequences used for ChIP

HS-85-f:	GAGACTAAGTAATTCACCATGGG
HS-85-r:	GGATCTATCTTGATTGTCCTCC
HS-62-f:	GCACATGCCGTAGTTCTC
HS-62-r:	TCTGGAGTTCTCAGTTGTATGAC
HS-62-B6-f:	GAACATGCTGTAGTTCTC
HS-62-B6-r:	TCTGGAGTTCTCGGTTGTATGGC
HS5-f:	ATGAGGCGTTTTCACCAC
HS5-r:	AAGGGGTCTTTTCACCGT
HS3-f:	TTTCAGAACCAGGAGGC
HS3-r:	TTTGCTGTTGTTGTTACTGTTC
HS2-f:	TTCCCTGTGGACTTCCTC
HS2-r:	GTCATGCTGAGGCTTAGG
βH1-f:	TGGAAACCTCTCTTCTGCC
βH1-r:	AAGCCCAAGGATGTCAGC
βmajor-f:	GGGAGAAATATGCTTGTCATC
βmajor-r:	CAACTGATCCTACCTCACCTT
Upstream-3'HS1-f:	TGTGGCTCTGGCTGTATCA
Upstream-3'HS1-r:	GCCAGGGCAACAGATAAGT
3'HS1-f:	AATCAGTGGAACACTTCTGC
3'HS1-r:	GTCTCAGGTTGTCAACTAAAGC
Downstream-3'HS1-f:	CTGTCTGGGAATGCTAACC
Downstream-3'HS1-r:	GCAAGGAGATTCTATGAGCC
Necdin-f:	GGTCCTGCTCTGATCCGAAG
Necdin-r:	GGGTCGCTCAGGTCCTTACTT

Allele-specific primers at 3'HS1 (see Fig. 3A):

1: AGAGGAGGGCGGAAATCAGT

2: GTCTCAGGTTGTCAACTAAAGC

3: GGAGGAGGGCGGAAATCAGC

4: GTCTCAGGTTGTCAACTAAACG

Supplemental Table 3. Primer sequences used for gene expression analysis

HPRT-S:	AGCCTAAGATGAGCGCAAGT
HPRT-AS:	ATGGCCACAGGACTAGAACA
εy-f:	GAACTTGTCCTCTGCCTCT
εy-r:	ATCACCAGCACATTACCCA
βH1f:	TGGACAACCTCAAGGAGAC
βH1r:	AGTAGAAAGGACAATCACCAAC
βmaj-mRNA-S:	ATGCCAAAGTGAAGGCCCAT
βmaj-mRNA-AS:	CCCAGCACAATCACGATCAT
βmaj-in2-f (129):	GGTTGCAAACAAATGTCTTTG
βmaj-in2-r (129):	CCAGGAGCTGTGGGAAT
βmaj-in2-f (B6):	AGGTTGCAAACAAATGTCTTTA
βmaj-in2-r (B6):	CCAGGAGCTGTGGGAAG
βminor-f:	ATCCCAAGGTGAAGGCCCAT
βminor-r:	CCCAGCACAATCACGATCGC
CTCF-f:	GACCACAAATCTAGAACCAAAGAAC
CTCF-r:	GTTGGCTTCGGAGGCTTCATATTACC
MOR3'B1 (Olfr67)-f:	CCTTTGGTAGTTGTGTGTCC
MOR3'B1 (Olfr67)-r:	GTGGTGTATCTCAGTGGGT
MOR3'B2/B3 (Olfr68, 69)-f:	TCAAGTTCAATGGCTCAGTC
MOR3'B2/3 (Olfr68, 69)-r:	ACGATTGAGGAATACAGCCA
MOR3'B4 (Olfr630)-f:	ATGGTGCTGGAGTTTGAGG
MOR3'B4 (Olfr630)-r:	ACTGTGGGATTGGATTGAGC
MOR5'B3 (Olfr64)-f:	CACAGACCTCACAGTTACAC
MOR5'B3(Olfr64)-r:	CCACAATGGAAAGGGAGTG

References

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.

Follenzi, A., Sabatino, G., Lombardo, A., Boccaccio, C., and Naldini, L. 2002. Efficient gene delivery and targeted expression to hepatocytes in vivo by improved lentiviral vectors. *Hum Gene Ther* 13: 243-260.

Splinter, E., Grosveld, F., and de Laat, W. 2004. 3C technology: analyzing the spatial organization of genomic loci in vivo. *Methods Enzymol* 375: 493-507.

Wall, L., deBoer, E., and Grosveld, F. 1988. The human beta-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. *Genes Dev* 2: 1089-1100.

Yu, H., Kessler, J., and Shen, J. 2000. Heterogeneous populations of ES cells in the generation of a floxed Presenilin-1 allele. *Genesis* 26: 5-8.

Zufferey, R., Nagy, D., Mandel, R.J., Naldini, L., and Trono, D. 1997. Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. *Nat Biotechnol* 15: 871-875.

Chapter 5

CTCF controls distinct pathways in T cell development, activation and cytokine production

Submitted

CTCF controls distinct pathways in T cell development, activation and cytokine production

Helen Heath[1+], Claudia Ribeiro de Almeida[2+], Gemma Dingjan[2], Suzanne van de Nobelen[1], Frank Sleutels[1], Kam-Wing Ling[2], Iris Jonkers[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, Justus-Liebig-Universitaet Giessen, Heinrich-Buff-Ring 58-62, D-35392 Giessen, Germany.[4]

[+]Equal contributions

Summary

The conserved transcriptional regulator CTCF is involved in imprinting and long-range chromatin interactions and is thought to be essential for many nuclear events in all cells. We generated mice with a conditional *Ctcf* allele and tested CTCF function in vivo in the T-cell lineage. Deletion of the *Ctcf* gene in early double-negative thymocytes did not hamper T-cell receptor (TCR) gene rearrangement, but impeded the proliferation of β-selected cells and their differentiation into double-positive thymocytes. In contrast, γδ T cell differentiation remained unaffected. CTCF deletion at the double-positive stage allowed development of peripheral T cells, but their activation and proliferation upon anti-CD3 stimulation was severely impaired. Surprisingly, circumventing TCR signaling with phorbol ester and ionomycin, allowed substantial proliferation of CTCF-deficient $CD4^+$ and $CD8^+$ T cells. In Th1/Th2 polarization cultures CTCF-deficient Th1 effector cells produced significant amounts of INF-γ. Expression of IL-4, IL-5 and IL-13 in Th2 cultures was severely reduced, despite the induction of two essential transcription factors, GATA 3 and SATB1. Thus, CTCF plays a major role in GATA 3/SATB1dependent regulation of gene expression within the Th2 cytokine locus. Collectively, our findings demonstrate that CTCF controls distinct pathways in T-cell development and activation, but is not absolutely essential for cell survival or proliferation.

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 Ref. 1, 2). In line with this view, murine CTCF is essential at very early embryonic stages (3), and expression of CTCF in WEHI231 B lymphoma cells results in growth arrest and apoptosis (4). CTCF is the archetypal vertebrate protein that binds insulator sequences, DNA elements that have the ability to protect a gene from outside influences (5). Its methylation-sensitive interaction with the imprinting control

region of the *H19*/*insulin-like growth factor 2* (*Igf2*) genes indeed controls enhancer access (3, 6, 7). CTCF-mediated insulator activity has been predicted at several other sites including the *DM1* locus and boundaries of domains that escape X-chromosome inactivation (8, 9). We have shown that CTCF mediates long-range chromatin interactions and regulates local histone modifications in the *β-globin* locus (10). Evidence has also been presented for a role of CTCF in interchromosomal interactions between Igf2 and other loci (11). During mitosis, CTCF remains bound to mitotic chromosomes, possibly facilitating reformation of higher order chromatin loops after mitosis (12). Genome-wide mapping of CTCF-binding sites revealed ~14,000 sites, the distribution of which correlates with genes but not with transcriptional start sites (13). 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 (14). High-resolution profiling of histone methylation in the human genome showed that CTCF marks boundaries of histone methylation domains (15) consistent with a role for CTCF as an insulator protein. Multiple CTCF-binding sites were identified near genes displaying extensive alternative promoter usage, including *protocadherin γ*, the *immunoglobulin λ* light chain and the *Tcrα/δ* and *β* chains. In mice, CTCF-dependent insulators were found downstream of the *Tcrα/δ* and the *immunoglobulin H* chain loci (16, 17), suggesting a role for CTCF in the regulation of gene transcription or recombination targeting in these complex loci.

To understand how CTCF regulates proliferation and differentiation in vivo, we generated mice with a conditional *Ctcf* allele (*Ctcf* $^{f/f}$), and deleted the gene at sequential stages of 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 Ref. 18). 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. Activation of $CD4^+$ T cells triggers differentiation into effector T cells, classically divided into two distinct subsets: IFN-γ^+ Th1 cells mediating cellular immunity against intracellular pathogens and IL-4, IL-5 and IL-13 producing Th2 cells important for the eradication of parasitic worms, but also implicated in allergic responses (19, 20). When naive $CD4^+$ T cells differentiate into Th1 or Th2 cells, the *Ifn-γ* and Th2 cytokine loci undergo structural and epigenetic changes, thought to be dependent on the transcription factors T-bet and

GATA 3, respectively (21-23). Here we show that CTCF acts as a critical dose-dependent regulator of cellular proliferation and differentiation following β-selection in the thymus. In the absence of the *Ctcf* gene, peripheral $CD4^+$ and $CD8^+$ TCRαβ T lymphocytes survive, but anti-CD3-mediated activation and proliferation is impaired. Remarkably, when TCR signaling is bypassed by phorbol ester and ionomycin, CTCF-deficient T cells show substantial proliferation. However, lack of CTCF severely affects Th2 cytokine production in Th2 cells. Thus, the insulator protein CTCF regulates specific pathways in T cell development and effector function, but is not essential for cell survival or proliferation per se.

Results

Generation of a conditional *Ctcf* allele

We generated a conditional *Ctcf* allele ($Ctcf^{f/f}$) by inserting loxP-sites upstream of exon 3 and downstream of exon 12 (Fig. 1A, B). Normal amounts of CTCF were expressed in $Ctcf^{f/f}$ mice (unpublished data). $Ctcf^{f/f}$ mice were crossed with mice expressing Cre recombinase ubiquitously (24), generating a CTCF-lacZ fusion transcript by removal of *Ctcf* exons 3-12 and inclusion of LacZ sequences (Fig. 1A, B). $Ctcf^{+/-}$ mice appeared normal, but we were unable to generate $Ctcf^{-/-}$ offspring from *Ctcf* +/- crosses (Supplemental Table S1), consistent with an essential role for CTCF in early development (3).

Conditional deletion of the *Ctcf* gene in developing T lymphocytes

CTCF expression was explored using mice carrying a *GFP-Ctcf* knock-in allele. We used flow cytometry with markers for CD3, CD4 and CD8, as well as CD44 and CD25 to subdivide DN cells into DN1-DN4 stages. We detected GFP-CTCF throughout T lymphocyte development, with higher fluorescence levels in subpopulations associated with high proliferation (Supplemental Fig. S1). We subsequently analyzed lacZ expression in $Ctcf^{+/-}$ T cells using fluorescein-di-β-D-galactopyranoside (FDG) as a substrate. LacZ expression was similar to that of GFP-CTCF (unpublished data), indicating that lacZ is a good marker of *Ctcf* gene deletion. T-cell specific deletion of *Ctcf*, was achieved by crossing $Ctcf^{f/f}$ mice with transgenic lines, in which the Cre recombinase was either driven by the proximal Lck promoter (LckCre) or by the CD4 promoter (CD4-Cre) (25, 26). Using lacZ as reporter, we found that the *Ctcf* gene deletion was almost complete from the DN2/DN3 stage onwards for Lck-Cre mice, and from the ISP stage onwards for CD4-Cre mice (Supplemental Fig. S2).

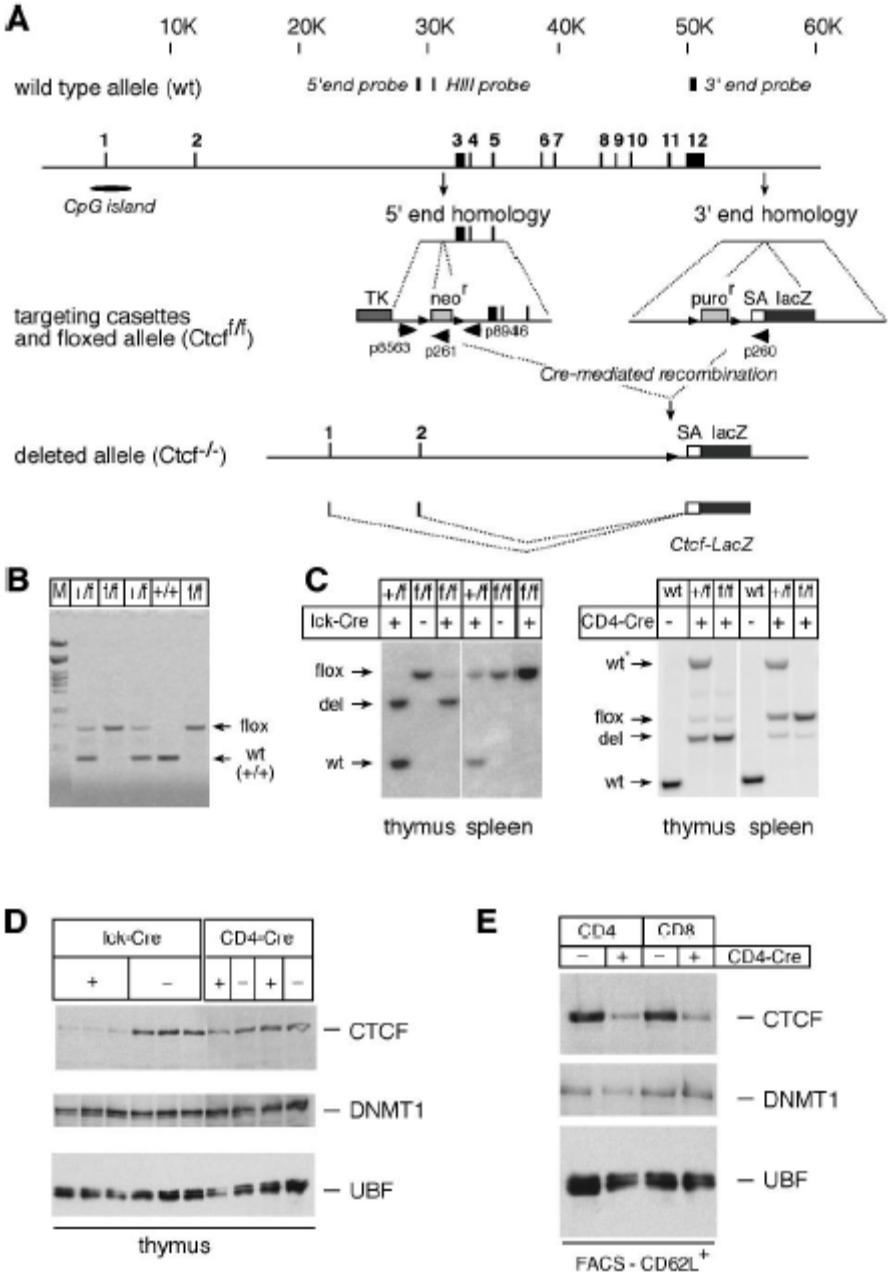

A
10K
20K
30K
40K
50K
60K
wild type allele (wt)
5'end probe
HIII probe
3' end probe
1
2
3 4 5
6 7
8 9 10
11 12
CpG island
5' end homology
3' end homology
targeting casettes and floxed allele (Ctcf f/f)
TK
neo r
puro r
SA lacZ
p8946
p3563
p261
p260
Cre-mediated recombination
deleted allele (Ctcf -/-)
SA lacZ
Ctcf-LacZ
B
M
flox
wt (+/+)
C
lck-Cre
CD4-Cre
flox
del
wt
wt'
thymus
spleen
D
lck-Cre
CD4-Cre
CTCF
DNMT1
UBF
thymus
E
CD4
CD8
CD4-Cre
CTCF
DNMT1
UBF
FACS - CD62L+

Figure 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 kilo base (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 (neor) or a PGK-puromycin cassette (puror), are shown with homologous regions. TK: thymidine kinase gene, SA-LacZ: Splice acceptor-lacZ cassette (41).
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 (*Ctcf* $^{-/-}$), 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) PCR analysis of genomic tail DNA. Genotypes are shown above the lanes (f: targeted allele, flanked by loxP sites; +: WT allele).
(C) Southern blot analysis of Lck-Cre and CD4-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 WT, *Ctcf* $^{f/f}$ (flox) and *Ctcf* $^{-/-}$ (del) alleles are indicated (asterisk indicates a polymorphic WT allele from the FVB background).
(D) Western blot analysis of thymus. Total thymus lysates from Lck-Cre *Ctcf* $^{f/f}$, CD4-Cre *Ctcf* $^{f/f}$ or WT mice (+ indicates presence of Cre transgene; - indicates absence) were analyzed for CTCF, DNMT1 and UBF protein levels.
(E) Western blot analysis of naïve T cells. Total cell lysates from sorted CD62L+ CD4+ and CD8+ T cell fractions from non-transgenic (-) or CD4-Cre transgenic (+) *Ctcf* $^{f/f}$ mice were analyzed for CTCF, DNMT1 and UBF.

Southern blotting showed efficient deletion of the *Ctcf* gene in thymus, while in spleen deletion was less evident, reflecting the abundance of non-T lineage cells (Fig. 1C). Western blotting showed that in thymic nuclear extracts from Lck-Cre *Ctcf* $^{f/f}$ mice CTCF protein levels were reduced to ~8 % of control (Fig. 1D). Despite efficient *Ctcf* gene deletion in the thymus of CD4-Cre *Ctcf* $^{f/f}$ mice, close to normal amounts of CTCF protein (~73 %) were detected in this tissue (Fig. 1D), suggesting that CTCF is remarkably stable. Consistent with this notion, residual CTCF protein was still detectable in highly purified fractions of naive peripheral $CD4^+$ and $CD8^+$ T cells (~25 % of control; Fig. 1E). To examine whether CTCF deletion affects global nuclear processes we investigated DNA methylation and nucleolar organization. Neither DNMT1, a maintenance methyltransferase with an important role in T cell development (25), nor UBF, a major accessory factor of RNA polymerase I (27), were significantly affected by deletion of CTCF (Fig. 1D, E). Consistently, the methylation status of ribosomal DNA (rDNA) repeats was normal in *Ctcf* knockout cells (Supplemental Fig. S3A) and the appearance, amount, and organization of rDNA repeats, analyzed by fluorescent *in situ* hybridization (FISH) with an rDNA probe, were similar in *Ctcf* knockout and wild-type (WT) $CD4^+$ and

$CD8^+$ T cells (Supplemental Fig. S3B and Table S2). The paralogue of CTCF, named CTCF-L or BORIS, can bind the same DNA sequences as CTCF (28) and might potentially substitute for the loss of CTCF in T cells. However, CTCF-L was not expressed in normal T cells and was not induced in the absence of CTCF (Supplemental Fig. S3C), eliminating a possible functional substitution.

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

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 (Fig. 2A, B). CD4 -CD8+ cells were mainly of the CD3loCD69lo type, indicative for ISP cells (Fig. 2A). $\alpha\beta$ T cell development was partially arrested at the ISP to DP transition, causing accumulation of DN3, DN4 and ISP cells (Fig. 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 CD4-Cre *Ctcf* $^{f/f}$ mice, thymic cellularity was only modestly reduced and no accumulation of ISP cells was detected (Fig. 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, heterozygous Lck-Cre *Ctcf* $^{+/f}$ and CD4-Cre *Ctcf* $^{f/f}$ mice were significantly reduced (Fig. 2A, B and unpublished data).

CTCF-deficiency had no adverse effect on $\gamma\delta$ T cell development, since the number of $CD3^+$ TCR$\gamma\delta^+$ thymocytes in Lck-Cre *Ctcf* $^{f/f}$ mice were ~2-fold higher than in WT littermates (Fig. 2C, D). The relative proportion of $\gamma\delta$ T cells in the spleens of Lck-Cre *Ctcf* $^{f/f}$ and CD4Cre *Ctcf* $^{f/f}$ mice was markedly increased (Fig. 2C), probably due to impaired $\alpha\beta$ T cell production. Collectively, these findings show a specific role for CTCF in $\alpha\beta$ T cell development, in particular at the ISP to DP transition.

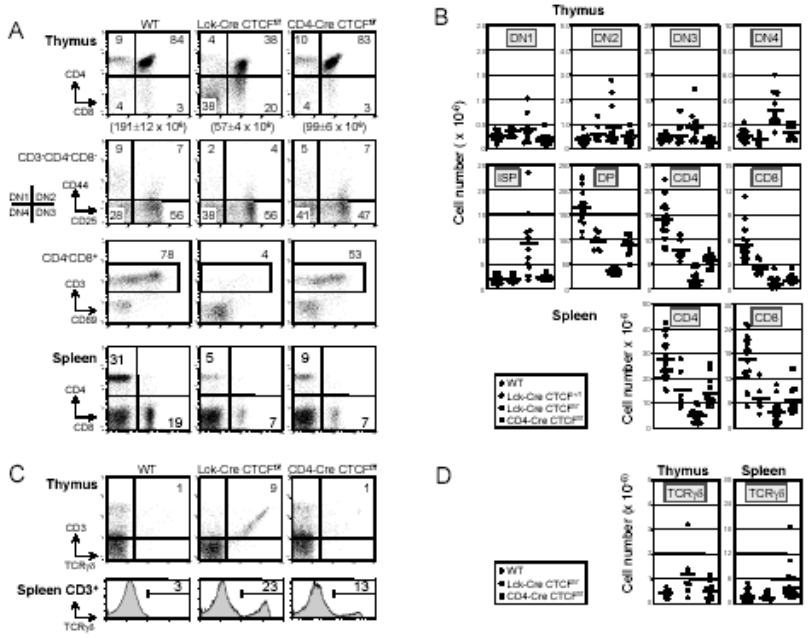


Figure 2. Defective TCRαβ lineage development in CTCF-deficient mice

(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, heterozygous Lck-Cre *Ctcf* $^{f/+}$ mice and CD4-Cre *Ctcf* $^{f/f}$ 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 ($p<0.00001$), in heterozygous Lck-Cre *Ctcf* $^{f/+}$ mice ($p<0.01$) and in CD4-Cre *Ctcf* $^{f/f}$ mice ($p<0.0001$).
(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 ($p<0.05$) and in the spleen of CD4-Cre *Ctcf* $^{f/f}$ mice ($p<0.05$). Data shown are representative of 10-20 animals per group.

Impaired proliferation of β-selected cells in Lck-Cre *Ctcf* $^{f/f}$ mice

TCR≤ rearrangement is generally initiated and completed in DN3. This stage consists of early small cells that have not yet productively rearranged the *Tcrβ* locus and more mature large, proliferating cells expressing TCRβ (29). Remarkably, in Lck-Cre *Ctcf* $^{f/f}$ mice a significant population of large cells containing intracellular TCRβ^{+} was present in DN2 fractions (Fig. 3A). In DN3 cells the proportion of large TCRβ^{+} cells were still elevated. When compared to WT cells, differentiation into DN4 and ISP cells resulted in smaller cells (Fig. 3A; quantified in Supplemental Fig. S4A). The CTCF-deficient ISP population also contained less cycling cells: 29% ± 1 in S/G2/M phase, compared to 53% ± 8 in WT (Fig. 3B). Differentiation was further analyzed by the expression of important T cell factors in sorted CTCF-deficient T cell fractions. C-myc was not affected in CTCF-deficient cells (Fig. 3C), even though CTCF was reported as a negative transcriptional regulator of c-myc (4, 30). GATA3 is critically involved in β-selection and development of CD4 SP cells (31), while SATB1 organizes cell type specific nuclear architecture (32). Expression levels of these transcription factors were somewhat reduced in Lck-Cre *Ctcf* knockout cells. Interestingly, preTα was up-regulated rather than down-regulated in DP cells, in the absence of CTCF (Fig. 3C). The expression of GIMAP4, which is strongly induced by pre-TCR signaling and accelerates T-cell death (33), was also increased in CTCF-deficient T cells (Fig. 3C). Finally, *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.

The accumulation of CTCF-deficient ISP cells could result from a developmental arrest or alternatively 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 (34). To distinguish between these possibilities, we assessed expression of various surface markers on CTCF-deficient ISP cells. CD3 and TCRβ expression was very low, suggesting that these were true ISP cells (Fig. 3D; Supplemental Fig. S4B). Consistently, CTCF-deficient ISP cells expressed low levels of CD5, which is normally present on DN and ISP cells and is up regulated on DP cells (35), and of CD69, which is normally induced in a sub-fraction of DP cells, reflecting TCR-mediated activation (36) (Supplemental Fig. S4). Expression of CD24 (HSA), which is normally high on DN and ISP cells and down-regulated at the ISP to DP transition (34), was reduced in CTCF-deficient cells throughout thymocyte differentiation.

Lck-Cre *Ctcf* $^{f/f}$ DP cells manifested consistently low levels of CD3/TCR expression, but normal levels were present in SP cells (Fig. 3D; Supplemental Fig. S4B). 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-bindings sites in the *Tcrα* gene locus (13, 16), suggested that the arrest of CTCF-deficient thymocytes may result from defective TCRα V(D)J recombination. We therefore crossed Lck-Cre

Ctcf $^{f/f}$ mice with transgenic mice expressing pre-rearranged TCRαβ transgenes that normally drive thymocytes into the CD4 or CD8 lineage, respectively. However the impaired developmental progression of CTCF deficient ISP into DP cells was not rescued (Supplemental Fig. S5). Thus, 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 mimicking pre-TCR signaling (35) was reduced in CTCF-deficient *Rag2* $^{-/-}$ DN cells, compared to CTCF-expressing *Rag2* $^{-/-}$ DN cells (Supplemental Fig. S6).

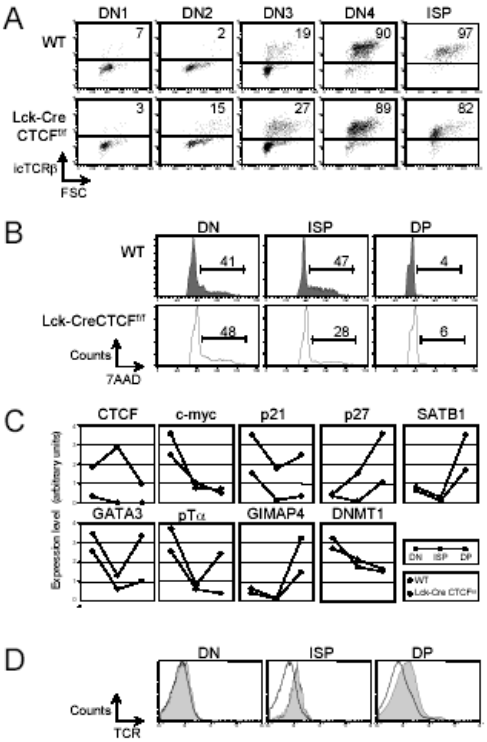


Figure 3. Impaired proliferation and differentiation of β-selected cells in Lck-Cre *Ctcf* $^{f/f}$ mice.

(A) Flow cytometric analyses of intracellular TCRβ protein expression in the indicated thymic subsets from WT and Lck-Cre *Ctcf* $^{f/f}$ mice. TCRβ/forward scatter (FSC) profiles are shown as dot plots and the percentages of TCRβ+ cells are shown.
(B) Cell cycle status of DN, ISP and DP cells, using 7-AAD.
(C) Quantitative RT-PCR analysis in sorted DN, ISP and DP cell fractions from WT and Lck-Cre *Ctcf* $^{f/f}$ mice. The DP fraction also contained CD4SP cells.
(D) Surface TCRβ expression in the indicated thymic subsets. Results are displayed as histogram overlays of Lck-Cre *Ctcf* $^{f/f}$ mice (bold lines) on top of profiles of wild-type mice (gray). Data shown are representative of 6 mice per group.

In summary, in the absence of CTCF, the *Tcrβ* locus can undergo functional V(D)J recombination in DN3, and apparently even in DN2 cells. This results in the initiation of β-selection, as evidenced by cell size increase and differentiation into DN4 and ISP cells. However, CTCF is essential to drive proliferation and maturation of β-selected cells and to efficiently up-regulate CD3/TCR surface expression in DP cells. The arrest of CTCF-deficient DP cells cannot be explained by impaired TCRα gene rearrangement.

Defective TCR/CD3-mediated proliferation in CTCF-deficient TCRαβ cells

The role of CTCF in T cell activation was investigated using in vitro cultures. Remarkably, anti-CD3/CD28 antibody stimulation of purified peripheral T cell fractions from Lck-Cre *Ctcf* $^{f/f}$ mice resulted in a selective outgrowth of TCRγδ T cells (Supplemental Fig. S7A). CTCF is expressed in both TCRαβ and TCRγδ T cells (Supplemental Fig. S1) and was absent in a mixed population of in vitro activated TCRαβ and TCRγδ T cells from Lck-Cre *Ctcf* $^{f/f}$ mice (Supplemental Fig. S7B, C). We therefore conclude that CTCF is essential for TCR mediated activation and proliferative expansion of TCRαβ but not of TCRγδ T cells.

To investigate cellular activation of CTCF-deficient αβ T cells, we performed *in vitro* stimulation experiments with highly purified naive CD62L$^+$ CD4$^+$ and CD8$^+$ T cells from WT and CD4-Cre *Ctcf* $^{f/f}$ spleen and lymph nodes. We evaluated the ability of T cells to go through sequential cell divisions by carboxy-fluorescein succinimidyl ester (CFSE)-labeling and observed severely reduced proliferation of anti-CD3/CD28-activated CTCF-deficient CD4$^+$ T cells at day 3 (Fig. 4A). A similar proliferation defect was also present in CTCF-deficient OTII transgenic CD4$^+$ T cells after a more physiological, antigen-specific, activation by OVA peptide 323-339-pulsed APC (Fig. 4A). In contrast, anti-CD3/CD28-activated CTCF-deficient CD8$^+$ T cells were able to undergo cell division, although they lagged behind WT cells by approximately one cell cycle, and cell recovery was reduced when compared to WT cells.

Phorbol-12-myristate-13-acetate (PMA) bypasses proximal TCR signaling events and directly activates protein kinase C signaling (37). Under conditions where PMA was added as a co-stimulatory signal with anti-CD3, CTCF-deficient cells were defective in proliferation. However, when T cells were stimulated by PMA and the calcium ionophore ionomycin, we observed significant proliferation of CTCF-deficient CD4$^+$ and CD8$^+$ T cells (Fig. 4B). These data indicate that under specific conditions CTCF is not required for cell proliferation.

Anti-CD3/CD28-stimulated CTCF-deficient CD4$^+$ T cell cultures at day 3 showed diminished cell recovery and an almost complete lack of cells in S/G2/M phase of the cell cycle (Fig. 4B). This is consistent with the limited cell division observed in CFSE experiments. Although CTCF-deficient PMA/ionomycin-stimulated CD4$^+$ or CD8$^+$ T cell

cultures exhibited lower expansion rates, their cell cycle profiles at day 3 were similar to WT cells (Fig. 4B).

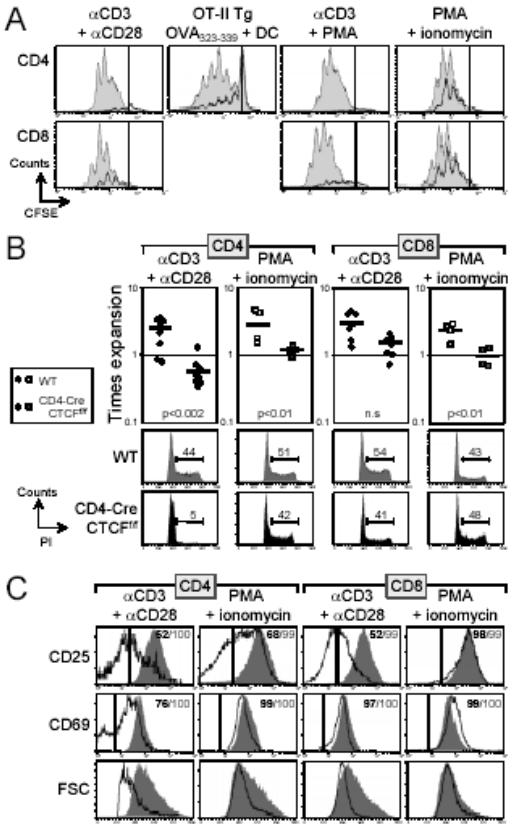


Figure 4. Impaired anti-CD3ε-mediated proliferation of CTCF-deficient CD4+ and CD8+ T cells.

(A) CFSE profiles of T cell cultures of sorted CD62L$^+$ CD4$^+$ and CD8$^+$ T cell populations from WT mice (gray filled histogram) in comparison with cultures from CD4-Cre*Ctcf* $^{f/f}$ mice (black line). Cells were activated by the indicated stimuli and cultured for 3 days (or 4 days for OT-II CD4$^+$ T cells). Dotted lines indicate the fluorescence intensity of un-stimulated cells.
(B) Cellular expansion in 3d cultures upon anti-CD3/CD28 or PMA/ionomycin stimulation (upper part). Symbols represent the expansion values of cultures from individual mice, whereby cell numbers at the start of the culture were set to one; lines indicate average values. The lower part shows the cell cycle status of the indicated cultures, whereby DNA content was examined by propidium iodide (PI) staining. The percentages of cycling cells (S/G2/M phase) are shown.
(C) Phenotypic characteristics of anti-CD3/CD28 or PMA/ionomycin-stimulated CD4$^+$ and CD8$^+$ T cells. CD25, CD69, and FSC profiles are displayed as histogram overlays of WT (gray filled histograms) and CD4-Cre *Ctcf* $^{f/f}$ cultures (bold lines). The percentages shown represent

the fractions of the cells within the indicated marker, in WT (gray) or CD4-Cre *Ctcf* $^{f/f}$ (black, bold) cultures. Data shown are representative of 4-6 animals per group.

The proliferation defect in anti-CD3/CD28-stimulated CTCF-deficient T cells was not due to defective IL-2 production (Supplemental Fig. S8). Nevertheless, induction of the IL-2 receptor CD25 on $CD4^+$ and $CD8^+$ T cells was severely impaired (Fig. 4C). Activation-induced up-regulation of CD69 and cell size increase was particularly affected in $CD4^+$ T cells. When activated by PMA/ionomycin, CTCF-deficient $CD4^+$ T cells displayed a partial defect in CD25 up regulation and cell size increase, while CD69 induction was normal. PMA/ionomycin-stimulated CTCF-deficient $CD8^+$ T cells showed normal CD25, CD69 and cell size up-regulation at day 3 (Fig. 4C), close to WT expansion rates, but reduced production of IFN-γ and granzyme B at day 7 (Supplemental Fig. S9).

Thus, anti-CD3/CD28 treatment does not elicit the proper activation and proliferation of CTCF-deficient peripheral TCR$\alpha\beta$ cells, whereby $CD4^+$ T cells are more affected than $CD8^+$ T cells. CTCF deficiency does not inhibit TCR$\gamma\delta$ T cell proliferation. When TCR signaling was bypassed by PMA and ionomycin, CTCF-deficient $CD4^+$ and $CD8^+$ T cells showed substantial activation and proliferation. These findings point to a specific function of CTCF in the regulation of proximal components of TCR signaling pathways. Unexpectedly, CTCF is not essential for cellular proliferation in vitro.

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

Consistent with the severe reduction of peripheral CD4 + T cell numbers in Lck-Cre *Ctcf* $^{f/f}$ mice, the levels of all serum Ig subclasses, except the T cell-independent isotypes IgM and IgG3, were dramatically reduced (Fig. 5A). Interestingly, total serum Ig levels in CD4-Cre *Ctcf* $^{f/f}$ mice exhibited a different profile, as the levels of the IL-4-dependent isotype IgG1 were more affected than those of INF-γ-dependent IgG2a (~10% and ~60% of wild-type, respectively). We also observed low serum levels of the IL-4-dependent isotype IgE (Fig. 5B). Th2-mediated responses in vivo were tested by injection of TNP-KLH, which boosted serum IgE levels in WT mice, but not in CD4-Cre *Ctcf* $^{f/f}$ animals (Fig. 5B). Thus, in CD4-Cre *Ctcf* $^{f/f}$ mice the Th2-dependent subclasses IgG1 and IgE were severely reduced, while the Th1-dependent isotype IgG2a was only moderately affected.

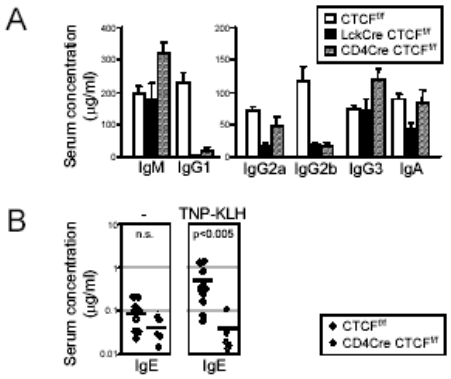


Figure 5. Serum Ig analysis in CTCF-deficient mice

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

Whether CTCF is specifically required for Th2 differentiation was tested by *in vitro* polarization cultures. Sorted naive $CD62L^+$ $CD4^+$ T cells from spleen and lymph nodes were stimulated with PMA/ionomycin under Th0 conditions (without additional cytokines or antibodies), Th1-polarizing conditions (with IL-12 and anti-IL-4) or Th2-polarizing conditions (with IL-4, anti-IL-12 and anti-IFN-γ) for 7 days. CTCF-deficient $CD4^+$ T cells showed substantial expansion (~5-10 times; Fig. 6A), but when compared to WT cells they produced moderately reduced levels of IFN-γ in Th1-polarized cultures and significantly decreased amounts of IL-4 in Th2-polarized cultures, as determined by intracellular FACS (Fig. 6B). Both frequency of $IL\text{-}4^+$ cells and intracellular IL-4 signals per cell were reduced (Fig. 6B). Moreover, quantitative RT-PCR analysis of Th2 cultures showed that in CTCF deficient Th2 cultures transcription of the Th2 locus cytokines *Il-4*, *Il-5* and *Il-13*, as well as of *Il-10* was strongly reduced (Fig. 6C).

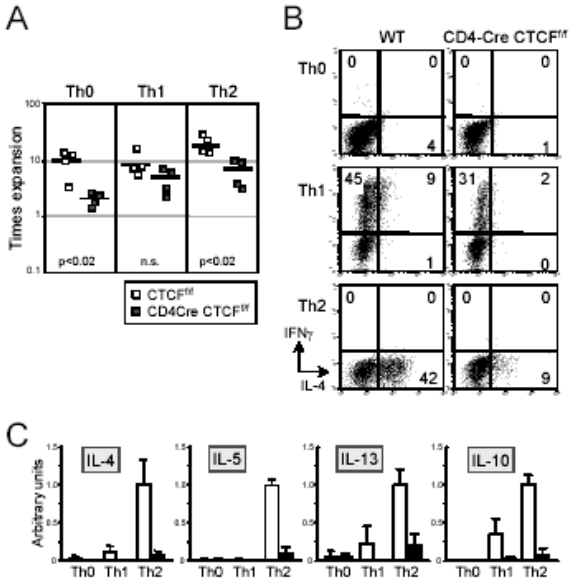


Figure 6. Th2 cytokine production defect in CTCF-deficient mice

(A) Expansion of Th0, Th1 and Th2 cultures 7 days after stimulation with PMA/ionomycin. Symbols represent the expansion values of cultures from individual mice, whereby cell numbers at the start of the culture were set to one. Lines indicate average values.
(B) Flow cytometric analysis for intracellular expression of IFNγ and IL-4 in the indicated T cell cultures after stimulation with PMA/ionomycin. CD4+ T cells were gated and expression profiles are displayed as dot plots. The percentages of cells within the quadrants are given. Mean fluorescence values for IL-4 were 138 and 75 for WT and CD4-Cre *Ctcf* $^{f/f}$ IL-4^{+} Th2 cells, respectively.
(C) Quantitative RT-PCR analysis of expression of the indicated cytokines in different T cell cultures. Expression levels were normalized for GAPDH and are expressed as arbitrary units, whereby the values in WT Th2 cells were set to one. Mean values and SD are given for 4 mice analyzed per group.

As Th2 cytokine production depends on the transcription factor GATA 3 (22), we evaluated its expression in the T cell cultures. GATA 3 expression appeared unaffected in CTCF-deficient Th2 cultures (Fig. 7A), excluding the possibility that Th2 cytokine production was impaired due to defective GATA3 induction. Furthermore, the CTCF deficient Th2 cultures displayed clear features of Th2- polarized cells, including low mRNA levels of *T-bet* and *Stat4* (Fig. 7B). As GATA 3 has the capacity to inhibit *Stat4* transcription (38), the finding of low *Stat4* expression levels suggest that in the absence of CTCF GATA 3 still operated as a negative regulator of *Stat4*. Like GATA 3, SATB1 is important for proper Th2 locus expression (32). SATB1 was specifically induced, both in WT and CTCF-deficient Th2 cultures (Fig. 7C). In CTCF-deficient Th1 cultures, T-bet

expression was reduced when compared to WT (Fig. 7B), which is consistent with the observed moderate reduction in IFN-γ expression (Fig. 6B).

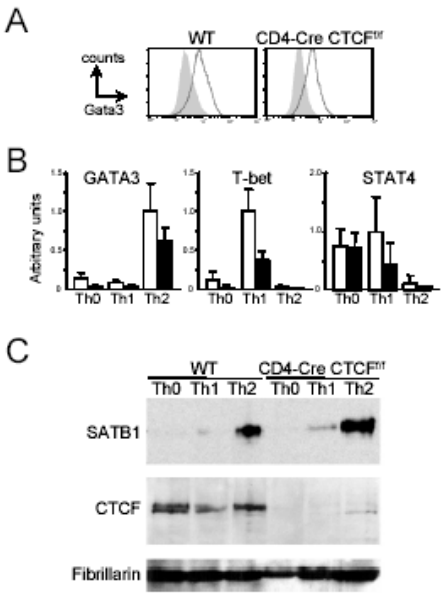


Figure 7. GATA3 and SATB1 are induced in PMA/ionomycin-stimulated CTCF deficient Th2 cultures

(A) Flow cytometric analysis for intracellular GATA3 protein expression in PMA/ionomycin stimulated T cell cultures. CD4+ T cells were gated and for the indicated mice expression profiles are displayed as histograms overlays of Th1 (gray filled histograms) and Th2 cultures (bold lines). (B) Quantitative RT-PCR analysis of GATA3, T-bet and STAT4 expression in different T cell cultures. Expression levels were normalized for GAPDH and are expressed as arbitrary units, whereby expression in WT Th1 cells (T-bet, STAT4) or Th2 cells (GATA 3) was set to 1. Mean values and SD are given for 4 mice analyzed per group.
(C) Western blotting analysis of SATB1 and CTCF protein levels in PMA/ionomycin-stimulated T effector cell cultures at day 7. Fibrillarin was used as a loading control.

As a result of defective TCR/CD3-mediated proliferation, sorted CTCF-deficient $CD62L^+$ $CD4^+$ peripheral T cells did not show expansion in 7-day Th1/Th2 polarization cultures when stimulated with anti-CD3/CD28 (Supplemental Fig. S10A). Moreover, we found a severe Th2-cytokine defect and an absence of GATA 3 induction (Supplemental Fig. S10).

Taken together, our data indicate that CTCF-deficiency affects Th2 cytokine production in multiple ways. CTCF is required for the induction of proliferation and GATA 3 expression upon TCR stimulation. When bypassed with PMA/ionomycin, proliferation is induced and differentiating CTCF-deficient Th2 cells express GATA 3 and activate SATB1. However expression of the Th2 cytokines IL-4, IL-5 and IL-13 remains strongly inhibited.

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 (3), a knockout of *Ctcf* in early embryonic development is lethal, while heterozygous *Ctcf* knockout mice are viable and fertile. No evidence was found for an increased tumor incidence in heterozygous knockout animals, or T lymphoid malignancies in the CTCF-deficient T-cell lineage, arguing against a role for CTCF as a crucial tumor suppressor. Furthermore, 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 mis-expression of genes on the inactive X chromosome.

Within the T-cell lineage, we found that CTCF is not required for V(D)J recombination at the TCRα locus. Importantly, our findings show that CTCF is essential for the efficient proliferation of β-selected cells, 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. The developmental progression 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 $\alpha\beta$ TCR. Therefore, the arrest of CTCF-deficient DP cells cannot result from impaired TCRα gene rearrangement. Because also the TCRβ, δ and γ 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 (15, 17) are not essential for the process of V(D)J recombination, which involves complex long-range DNA interactions.

The impaired generation of SP cells in the thymus of CTCF-deficient mice can be explained by the reduced size of the DP compartment. Thus CTCF appears not required for the substantial epigenetic and regulatory changes (18) that accompany commitment and maturation of CD4 or CD8 SP cells. Like DNMT1 and the RNase III enzyme Dicer (39), CTCF was not essential for $\gamma\delta$ T cell development or anti-CD3-mediated proliferation of these T cells. Apparently, cell division and the regulation of chromatin structure and gene expression in $\gamma\delta$ T cells are very different from $\alpha\beta$ T cells, and do not depend on proteins like CTCF, DNMT1 or Dicer.

We found that CD4-Cre *Ctcf* $^{f/f}$ mice contained substantial numbers of $CD4^+$ and $CD8^+$ T cells in which the expression of CTCF protein was significantly reduced. Either CTCF is not essential for the survival of resting peripheral T cells, or very low amounts of CTCF are enough for cells to survive. When resting naive T cells were activated and driven into cell division in vitro, CTCF was no longer detectable. Since activation by PMA and ionomycin resulted in proliferation of CTCF-deficient αβ T cells, we conclude that CTCF is not essential for cellular proliferation per se. Consistent with this notion, γδ T cells, and to a lesser extent also $CD8^+$ T cells show substantial proliferation in the absence of CTCF. By contrast, CTCF appeared essential for anti-CD3ε-induced activation and proliferation of αβ T cells. It remains to be investigated which pathways are specifically affected in αβ T cells by CTCF-deficiency.

The in vitro Th1/Th2 polarization cultures showed that CTCF-deficient Th1 effector cells produced significant amounts of INF-γ. By contrast, CTCF-deficient Th2 effector cells hardly produced Th2 cytokines or IL-10. In these Th2 effector cells, both GATA3 and SATB1 proteins were present at apparently normal levels and T-bet, the transcription factor that is critical for Th1 differentiation was down regulated. Moreover, levels of Stat4 transcription, which is inhibited by GATA 3 activity (38), were low; indicating that in the absence of CTCF GATA 3 is still functional as a negative regulator of Stat4 transcription. In contrast, transcription of the three Th2 cytokine genes *Il-4*, *Il-5* and *Il-13* was severely reduced.

We therefore propose that CTCF plays a major role in the GATA3- and SATB1-dependent regulation of the expression of genes within the Th2 cytokine locus.

Our findings uncover specific CTCF-dependent pathways in T cell development, activation and effector function. Recently, ~14,000 genome-wide CTCF-binding sites were identified (13, 14). It was proposed that CTCF remains bound to these sites irrespective of cell type (13) and that it marks boundaries of histone methylation domains in human T cells (15). CTCF also mediates long-range chromatin looping in the *β-globin* locus and its deletion alters local histone modifications (10). However deletion of a CTCF binding site in the *β-globin* locus did not affect expression of the globin genes. By contrast, our results in T cells demonstrate an essential role for CTCF specifically in Th2 locus expression. Hence we propose that CTCF has cell type-specific functions; it will be interesting to determine how CTCF performs these specific roles, while remaining bound to its ~14,000 cognate binding sites. Equally interesting are the questions how and which chromosomal interactions, both *in cis* and *in trans*, persist in the absence of CTCF. Importantly, our experiments in mature CTCF-negative T cells show they can proliferate and differentiate under appropriate conditions and it is therefore feasible to address these issues using CTCF knock-down or conditional targeting approaches.

Materials and Methods

Modified *Ctcf* alleles and mouse models

Human CTCF cDNA was used to screen a 129S6/SvevTac mouse PAC library (RPCI-21) (40). 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 (41). 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 (41). Relevant parts of the different constructs were verified by DNA sequencing.

Constructs were targeted into E14 embryonic stem (ES) cells as described (41). DNA from resistant ES cells was analyzed with external radio-labeled probes by Southern blotting. Confirmation of homologous recombination was performed using different 5' end and 3' end probes (Fig. 1A, B) and a PCR-based assay for genotyping. *Ctcf* f/f mice were maintained on a C57BL/6 background. *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 LckCre and CD4-Cre mice (25), kindly provided by Dr. C. Wilson (University of Washington, Seattle, USA). Cre-specific primers were used for genotyping.

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 Fig. 1. Total RNA was prepared using RNA-Bee RNA isolation solvent (TelTest 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 (10) 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. Alternatively, total RNA was extracted using the GeneElute mammalian total RNA miniprep system (Sigma). Primers spanning

at least one intron-exon junction, were designed manually or using the ProbeFinder software (Roche Applied Science, Indianapolis, In). Probes were either chosen from the universal probe library (Roche Applied Science) or designed manually (*Gata3*, *Gapdh*) and purchased from Eurogentec (Seraing, Belgium). Quantitative real-time PCR was performed using the ABI Prism 7700 sequence detection system (Applied Biosystems, Foster City, CA). 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 Ct values were normalized to the Ct 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 as described (10) and analyzed by western blot (41). Primary antibody incubation was done overnight at 4^{o}C 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). CTCF-specific (N3) antibodies, generated as described (42) using a GST-linked chicken CTCF (amino acids 2267) fusion protein, 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 CTCF, UBF and DNMT1 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 (43). All mAbs were purchased from BD Biosciences (San Diego, CA), except for PE-conjugated anti-Granzyme B (GB12; Caltag Laboratories, Burlingame, CA) and anti-GATA 3 (Hg-3-31; Santa Cruz Biotechnology, CA). For intracellular detection of cytokines, cells were re-stimulated with plate-bound antiCD3 (10 μg/ml in PBS; 145-2C11) or phorbol 12-myristate-13-acetate (PMA; 50 ng/ml; Sigma) plus ionomycin (500 ng/ml; Sigma) in the presence of GolgiStop^{TM} (BD PharMingen) for 4h. Cells were harvested and stained extracellularly with PCP-conjugated anti-CD4 or FITC-conjugated anti-CD8 (53-6.7), followed by standard intracellular staining with anti-IL4-PE (BVD4-1D11) or anti-IFN-γ-APC (XMG1.2), using paraformaldehyde and saponin.

For cell cycle profiles of thymic subsets, cells were first stained for surface markers, fixed with 0.25% paraformaldehyde and permeabilized with 0.2% Tween20. 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 0.02mg/ml propidium iodide, 0.1% v/v Triton X-100 and 0.2 mg/ml RNAse. Doublet cells were excluded by measuring peak area and width.

CFSE-labeling of cells was performed as described (44). Samples were acquired on a FACSCaliburTM flow cytometer and analyzed using CellQuestTM (BD, Sunnyvale, CA) or FlowJo™ (Tree Star Inc., Ashland, OR) research software. Statistical evaluations were done by students t-test.

In vitro T cell cultures

For in vitro T cell stimulations and Th1/Th2 polarization cultures, naive $CD62L^+$ $CD4^+$ or $CD8^+$ T cells were purified by cell sorting using a FACSVantage VE equipped with Diva Option and BD FACSDiva software (BD Bioscience). Purity of obtained fractions was >98%. T cells were cultured at a concentration of 1 x 10^6 cells/ml in Iscoves modified Dulbeccos 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), or PMA (50ng/ml) plus ionomycin (300ng/ml).

For Th1-polarizing conditions, anti-IL-4 (10 μg/ml; 11B11) plus IL-12 (10 ng/ml) was added to the medium. Th2-polarizing cultures included anti-IFN-γ (5 μg/ml; R4-6A2,), anti-IL-12 (5 μg /ml; C17.8) and IL-4 (10ng/ml). For Th0 conditions, no cytokines or mAbs were added. For differentiation of $CD8^+$ effector T cells, only recombinant IL-2 (5ng/ml) was added to the medium. On day 3 after activation, T cell cultures were further supplemented with recombinant IL-2 (5ng/ml). Both $CD4^+$ and $CD8^+$ T cells were expanded up to 7 days under the same cytokine conditions as the primary cultures. Cytokines were from R&D Systems Inc., Minneapolis, MN.

Stimulation of OT-II transgenic CD4 + T cells was carried out in the presence of bone marrow derived dendritic cells (BM-DC). Briefly, BM single cell suspensions were prepared from C57BL/6 femurs and seeded at 2 x 10^6 per petri dish in complete IMDM and 200ng/ml murine rGM-CSF (Biosource). On days 3 and 6, 200ng/ml murine recombinant GM-CSF was added in 10 ml of fresh IMDM. On day 8, the non-adherent cells consisting of immature and mature BM-DC were harvested. For in vitro T cell proliferation studies $2x10^6$ CFSE labeled OT-II transgenic naive $CD4^+$ T cells were co-cultured with $2x10^6$ BM-DC previously pulsed with OVA peptide 323-339 (50 μg/ml, manufacturer) in complete IMDM. At day 4, cultured T cells were harvested for proliferation analysis.

Analyses of serum Ig and in vivo immunizations

Total serum Ig levels were determined by subclass-specific sandwich ELISA, as described; IgE was induced by i.p. injection of 10 μg TNP-KLH precipitated on alum (45).

References

1. Lewis, A., and A. Murrell. 2004. Genomic imprinting: CTCF protects the boundaries. Curr Biol 14:R284-286.
2. Ohlsson, R., R. Renkawitz, and V. Lobanenkov. 2001. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. Trends Genet 17:520-527.
3. Fedoriw, A.M., P. Stein, P. Svoboda, R.M. Schultz, and M.S. Bartolomei. 2004. Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting. Science 303:238-240.
4. Qi, C.F., A. Martensson, M. Mattioli, R. Dalla-Favera, V.V. Lobanenkov, and H.C. Morse, 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.
5. Bell, A.C., A.G. West, and G. Felsenfeld. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell 98:387-396.
6. Bell, A.C., and G. Felsenfeld. 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature 405:482-485.
7. Hark, A.T., C.J. Schoenherr, D.J. Katz, R.S. Ingram, J.M. Levorse, and S.M. Tilghman. 2000. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature 405:486-489.
8. Filippova, G.N., M.K. Cheng, J.M. Moore, J.P. Truong, Y.J. Hu, D.K. Nguyen, K.D. Tsuchiya, and C.M. Disteche. 2005. Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development. Dev Cell 8:31-42.
9. Filippova, G.N., C.P. Thienes, B.H. Penn, D.H. Cho, Y.J. Hu, J.M. Moore, T.R. Klesert, V.V. Lobanenkov, and S.J. Tapscott. 2001. CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. Nat Genet 28:335-343.
10. Splinter, E., H. Heath, J. Kooren, R.J. Palstra, P. Klous, F. Grosveld, N. Galjart, and W. de Laat. 2006. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev 20:2349-2354.

11. Ling, J.Q., T. Li, J.F. Hu, T.H. Vu, H.L. Chen, X.W. Qiu, A.M. Cherry, and A.R. Hoffman. 2006. CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. Science 312:269-272.
12. Burke, L.J., R. Zhang, M. Bartkuhn, V.K. Tiwari, G. Tavoosidana, S. Kurukuti, C. Weth, J. Leers, N. Galjart, R. Ohlsson, and R. Renkawitz. 2005. CTCF binding and higher order chromatin structure of the H19 locus are maintained in mitotic chromatin. Embo J 24:3291-3300.
13. Kim, T.H., Z.K. Abdullaev, A.D. Smith, K.A. Ching, D.I. Loukinov, R.D. Green, M.Q. Zhang, V.V. Lobanenkov, and B. Ren. 2007. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell 128:1231-1245.
14. Xie, X., T.S. Mikkelsen, A. Gnirke, K. Lindblad-Toh, M. Kellis, and E.S. Lander. 2007. Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. Proc Natl Acad Sci U S A 104:7145-7150.
15. Barski, A., S. Cuddapah, K. Cui, T.Y. Roh, D.E. Schones, Z. Wang, G. Wei, I. Chepelev, and K. Zhao. 2007. High-resolution profiling of histone methylations in the human genome. Cell 129:823-837.
16. Garrett, F.E., A.V. Emelyanov, M.A. Sepulveda, P. Flanagan, S. Volpi, F. Li, D. Loukinov, L.A. Eckhardt, V.V. Lobanenkov, and B.K. Birshtein. 2005. Chromatin architecture near a potential 3' end of the igh locus involves modular regulation of histone modifications during B-Cell development and in vivo occupancy at CTCF sites. Mol Cell Biol 25:1511-1525.
17. Magdinier, F., T.M. Yusufzai, and G. Felsenfeld. 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.
18. Rothenberg, E.V., and T. Taghon. 2005. Molecular genetics of T cell development. Annu Rev Immunol 23:601-649.
19. Abbas, A.K., K.M. Murphy, and A. Sher. 1996. Functional diversity of helper T lymphocytes. Nature 383:787-793.
20. Szabo, S.J., B.M. Sullivan, S.L. Peng, and L.H. Glimcher. 2003. Molecular mechanisms regulating Th1 immune responses. Annu Rev Immunol 21:713-758.
21. Chang, S., and T.M. Aune. 2007. Dynamic changes in histone-methylation 'marks' across the locus encoding interferon-gamma during the differentiation of T helper type 2 cells. Nat Immunol 8:723-731.
22. Lee, G.R., S.T. Kim, C.G. Spilianakis, P.E. Fields, and R.A. Flavell. 2006. T helper cell differentiation: regulation by cis elements and epigenetics. Immunity 24:369-379.

23. Schoenborn, J.R., M.O. Dorschner, M. Sekimata, D.M. Santer, M. Shnyreva, D.R. Fitzpatrick, J.A. Stamatoyonnapoulos, and C.B. Wilson. 2007. Comprehensive epigenetic profiling identifies multiple distal regulatory elements directing transcription of the gene encoding interferon-gamma. Nat Immunol 8:732-742.
24. Sakai, K., and J. Miyazaki. 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.
25. Lee, P.P., D.R. Fitzpatrick, C. Beard, H.K. Jessup, S. Lehar, K.W. Makar, M. PerezMelgosa, M.T. Sweetser, M.S. Schlissel, S. Nguyen, S.R. Cherry, J.H. Tsai, S.M. Tucker, W.M. Weaver, A. Kelso, R. Jaenisch, and C.B. Wilson. 2001. A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival. Immunity 15:763-774.
26. Wolfer, A., A. Wilson, M. Nemir, H.R. MacDonald, and F. Radtke. 2002. Inactivation of Notch1 impairs VDJbeta rearrangement and allows pre-TCR-independent survival of early alpha beta Lineage Thymocytes. Immunity 16:869-879.
27. Panov, K.I., J.K. Friedrich, J. Russell, and J.C. Zomerdijk. 2006. UBF activates RNA polymerase I transcription by stimulating promoter escape. Embo J 25:3310-3322.
28. Loukinov, D.I., E. Pugacheva, S. Vatolin, S.D. Pack, H. Moon, I. Chernukhin, P. Mannan, E. Larsson, C. Kanduri, A.A. Vostrov, H. Cui, E.L. Niemitz, J.E. Rasko, F.M. Docquier, M. Kistler, J.J. Breen, Z. Zhuang, W.W. Quitschke, R. Renkawitz, E.M. Klenova, A.P. Feinberg, R. Ohlsson, H.C. Morse, 3rd, and V.V. Lobanenkov. 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.
29. Hoffman, E.S., L. Passoni, T. Crompton, T.M. Leu, D.G. Schatz, A. Koff, M.J. Owen, and A.C. Hayday. 1996. Productive T-cell receptor beta-chain gene rearrangement: coincident regulation of cell cycle and clonality during development in vivo. Genes Dev 10:948-962.
30. Lobanenkov, V.V., R.H. Nicolas, V.V. Adler, H. Paterson, E.M. Klenova, A.V. Polotskaja, and G.H. Goodwin. 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.

31. Pai, S.Y., M.L. Truitt, C.N. Ting, J.M. Leiden, L.H. Glimcher, and I.C. Ho. 2003. Critical roles for transcription factor GATA-3 in thymocyte development. Immunity 19:863-875.
32. Cai, S., C.C. Lee, and T. Kohwi-Shigematsu. 2006. SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. Nat Genet 38:1278-1288.
33. Schnell, S., C. Demolliere, P. van den Berk, and H. Jacobs. 2006. Gimap4 accelerates T-cell death. Blood 108:591-599.
34. Williams, C.J., T. Naito, P.G. Arco, J.R. Seavitt, S.M. Cashman, B. De Souza, X. Qi, P. Keables, U.H. Von Andrian, and K. Georgopoulos. 2004. The chromatin remodeler Mi-2beta is required for CD4 expression and T cell development. Immunity 20:719733.
35. Azzam, H.S., A. Grinberg, K. Lui, H. Shen, E.W. Shores, and P.E. Love. 1998. CD5 expression is developmentally regulated by T cell receptor (TCR) signals and TCR avidity. J Exp Med 188:2301-2311.
36. Bendelac, A., P. Matzinger, R.A. Seder, W.E. Paul, and R.H. Schwartz. 1992. Activation events during thymic selection. J Exp Med 175:731-742.
37. Baier, G. 2003. The PKC gene module: molecular biosystematics to resolve its T cell functions. Immunol Rev 192:64-79.
38. Usui, T., R. Nishikomori, A. Kitani, and W. Strober. 2003. GATA-3 suppresses Th1 development by downregulation of Stat4 and not through effects on IL-12Rbeta2 chain or T-bet. Immunity 18:415-428.
39. Cobb, B.S., A. Hertweck, J. Smith, E. O'Connor, D. Graf, T. Cook, S.T. Smale, S. Sakaguchi, F.J. Livesey, A.G. Fisher, and M. Merkenschlager. 2006. A role for Dicer in immune regulation. J Exp Med 203:2519-2527.
40. Osoegawa, K., M. Tateno, P.Y. Woon, E. Frengen, A.G. Mammoser, J.J. Catanese, Y. Hayashizaki, and P.J. de Jong. 2000. Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. Genome Res 10:116-128.
41. Hoogenraad, C.C., B. Koekkoek, A. Akhmanova, H. Krugers, B. Dortland, M. Miedema, A.van Alphen, W.M. Kistler, M. Jaegle, M. Koutsourakis, N. Van Camp, M. Verhoye, A. van der Linden, I. Kaverina, F. Grosveld, C.I. De Zeeuw, and N. Galjart. 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.
42. Hoogenraad, C.C., A. Akhmanova, F. Grosveld, C.I. De Zeeuw, and N. Galjart. 2000. Functional analysis of CLIP-115 and its binding to microtubules. J Cell Sci 113:22852297.

43. Hendriks, R.W., M.F. de Bruijn, A. Maas, G.M. Dingjan, A. Karis, and F. Grosveld. 1996. Inactivation of Btk by insertion of lacZ reveals defects in B cell development only past the pre-B cell stage. Embo J 15:4862-4872.
44. Lyons, A.B., and C.R. Parish. 1994. Determination of lymphocyte division by flow cytometry. J Immunol Methods 171:131-137.
45. Nawijn, M.C., G.M. Dingjan, R. Ferreira, B.N. Lambrecht, A. Karis, F. Grosveld, H. Savelkoul, and R.W. Hendriks. 2001. Enforced expression of GATA-3 in transgenic mice inhibits Th1 differentiation and induces the formation of a T1/ST2-expressing Th2-committed T cell compartment in vivo. J Immunol 167:724-732.

Supplementary Data

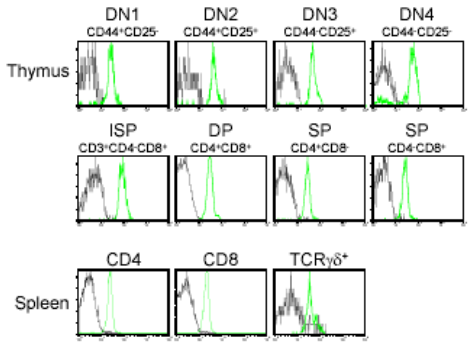


Figure S1. Flow cytometric analysis of GPF-CTCF protein expression

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

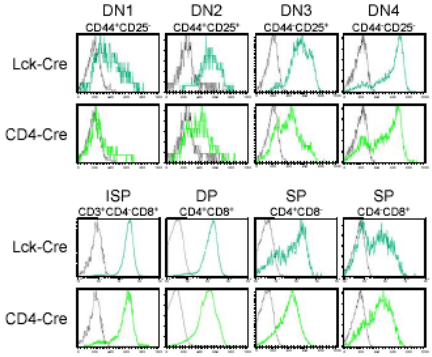


Figure S2. Flow cytometric analysis of lacZ expression in CTCF conditionally deleted mice

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* $^{f/f}$ or CD4-Cre *Ctcf* $^{f/f}$ mice (green) on top of background signals in WT mice (black). High lacZ expression, reflecting almost complete deletion of the *Ctcf* gene is present in DN2/DN3 for Lck-Cre *Ctcf* $^{f/f}$ mice and in ISP for CD4-Cre *Ctcf* $^{f/f}$ mice.

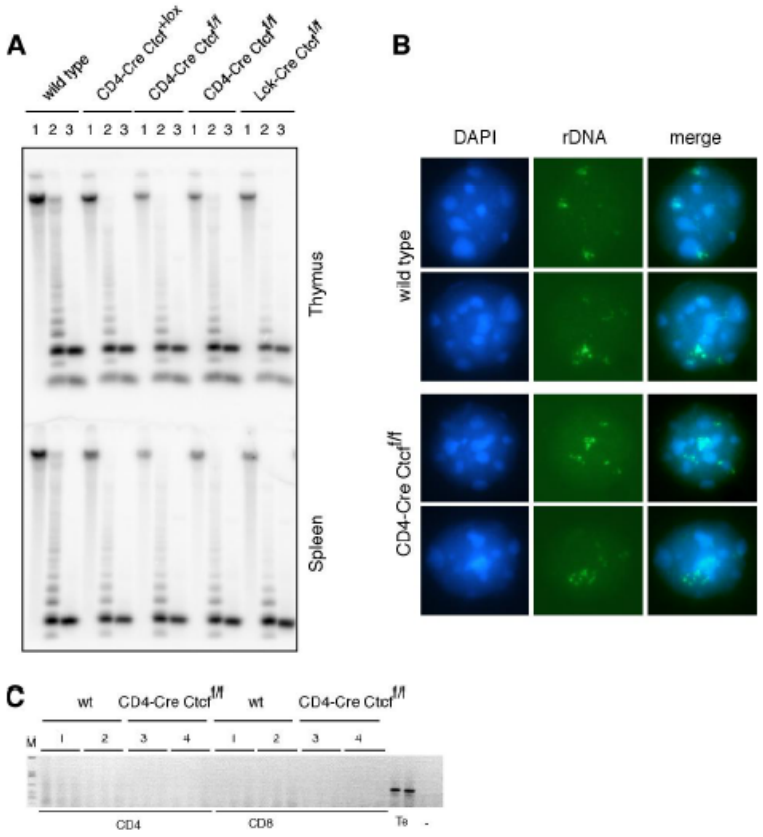


Figure S3. CTCF deletion does not affect DNA methylation or nucleolar organization

A) 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 et al., 2000).
B) 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 et al., 2000). Cells were counterstained with DAPI (blue).
C) RT-PCR analysis for CTCF-L/BORIS expression in sorted naive peripheral $CD62L^+$ $CD4^+$ and $CD8^+$ T cell fractions from wild-type (1,2) and CD4-Cre *Ctcf* $^{f/f}$ (3,4) 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. Products were fractionated by gel electrophoresis and detected with ethidium bromide. M=molecular weight marker.

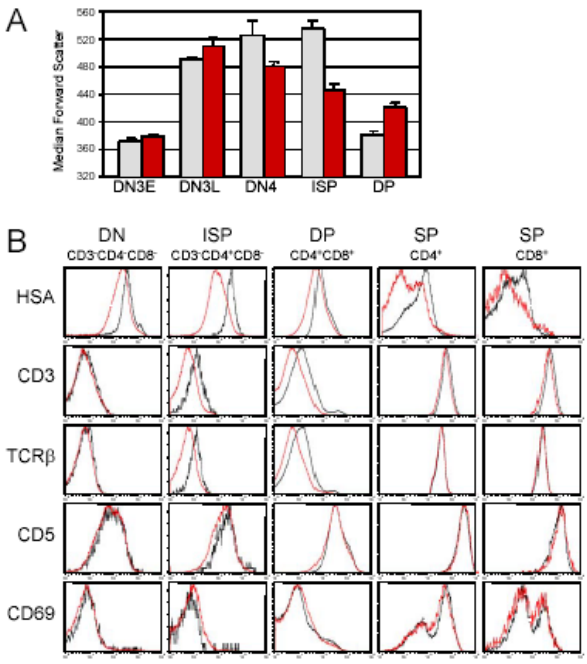


Figure S4. Arrest at the ISP to DP transition in Lck-Cre Ctcf $^{f/f}$ mice

A) Quantification of forward scatter values of the indicated thymocyte subpopulations in WT (gray bars) and Lck-Cre *Ctcf* $^{f/f}$ mice (red bars). Data are average values ± SEM from 5-8 mice per group.
B) Flow cytometric analyses of HSA, CD3, TCR, CD5 and CD69 in the indicated thymocyte sub-populations, displayed as overlays of WT mice (black histograms) and Lck-*Cre Ctcf* $^{f/f}$ mice
(red histograms). Data shown are representative of 5-8 mice per group.

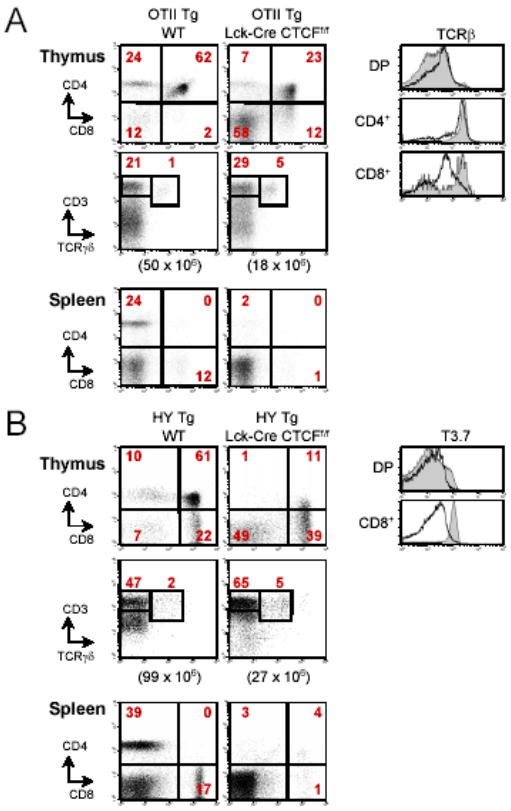


Figure S5. The arrest of CTCF-deficient thymocytes is independent of TCR α rearrangement

Providing Lck-*Cre Ctcf* $^{f/f}$ mice with a pre-rearranged TCRαβ transgene does not correct the developmental arrest of DP cells.
A) The OTII TCRαβ recognizes the OVA323-339 peptide in the context of C57BL/6 MHC class II. OTII TCR transgenic thymocytes are positively selected towards the CD4 lineage (Barnden et al., 1998).
B) The MHC class I-restricted HY TCRαβ recognizes a male-specific HY antigen peptide and in the C57BL/6 H-2b class I female background (Kisielow et al., 1988); HY-specific thymocytes are positively selected towards the CD8 lineage. (A, B) 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). 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).

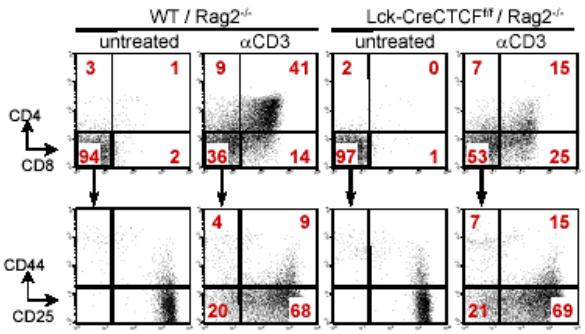


Figure S6. The arrest at the ISP to DP transition is also present in anti-CD3-treated CTCF/RAG2-deficient mice

Flow cytometric analyses of the thymus of the indicated mice, which were either untreated or injected with 50 µg of rat anti-CD3ε 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.

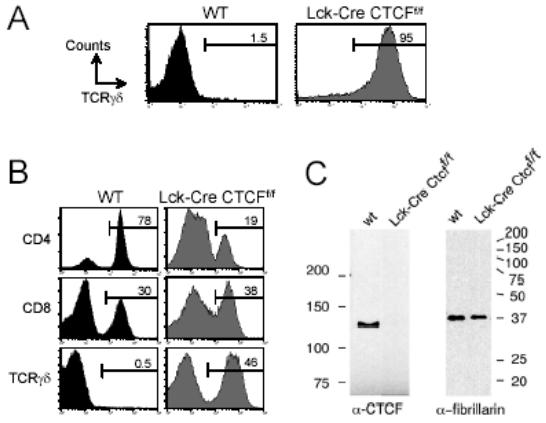


Figure S7. Selective advantage of TCR γδ cells in anti-CD3/CD28-stimulated cultures of Lck-Cre *Ctcf* $^{f/f}$ mice

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.

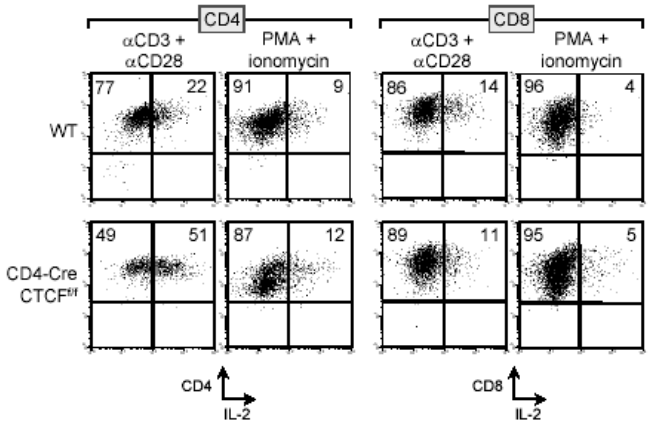


Figure S8. CTCF-deficient CD4$^+$ or CD8$^+$ T cells do not have a defect in IL-2 production

Analysis of IL-2 expression in anti-CD3/CD28 and PMA/ionomycin-stimulated cultures of sorted CD62L$^+$ CD4$^+$ and CD8$^+$ T cell fractions from wild-type and CD4-Cre *Ctcf* $^{f/f}$ mice. At day 3, cells were re-stimulated for 4 hours prior to intracellular flow cytometric analysis. Total living cells were gated and CD4/IL-2 and CD8/IL-2 profiles are displayed as dot plots and the percentages of cells within quadrants are given. Data shown are representative of 4 mice per group.

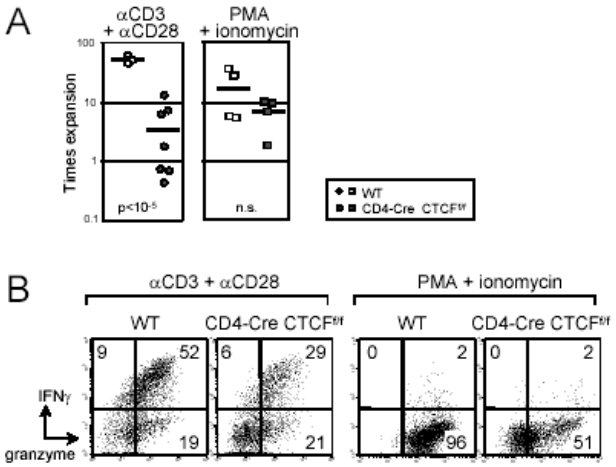


Figure S9. Analysis of defects in the CD8 compartment of CD4-Cre *Ctcf* $^{f/f}$ mice
A) Expansion of CD8+ T cell cultures 7 days after stimulation with αCD3/αCD28 or PMA/ionomycin. Symbols represent expansion values of cultures from individual mice, whereby cell numbers at the start of the culture were set to one. Lines indicate average values.
B) Flow cytometric analysis for intracellular expression of IFNγ and granzyme B in the indicated $CD8^+$ T cell cultures. $CD8^+$ T cells were gated and expression profiles are displayed as dot plots.
The percentages of cells within the quadrants are given.

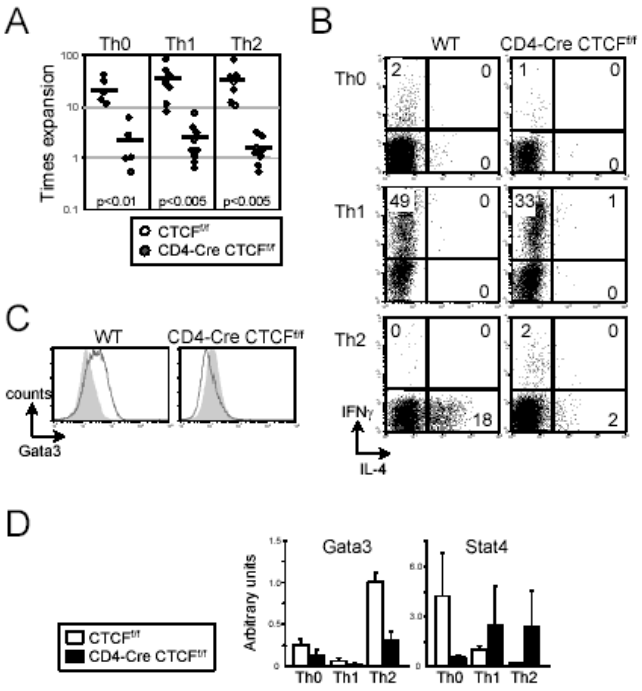


Figure S10. Defective expansion and cytokine production in CTCF-deficient αCD3/ αCD28 stimulated Th cells

A) Expansion of Th0, Th1 and Th2 cultures 7 days after stimulation with αCD3/αCD28. Symbols represent the expansion values of cultures from individual mice, whereby cell numbers at the start of the culture were set to one. Lines indicate average values.
B) Flow cytometric analysis for intracellular expression of IFNγ and IL-4 in the indicated T cell cultures after αCD3/αCD28 stimulation. $CD4^+$ T cells were gated and expression profiles are displayed as dot plots. The percentages of cells within the quadrants are given. Mean fluorescence values for IL-4 were 134 and 70 for wild-type and CD4-Cre *Ctcf* $^{f/f}$ IL-4^+ Th2 cells, respectively.
C) Flow cytometric analysis for intracellular GATA3 protein expression in αCD3/αCD28-stimulated T cell cultures. $CD4^+$ T cells were gated and for the indicated mice expression profiles are displayed as histograms overlays of Th1 (gray filled histograms) and Th2 cultures (bold lines).
D) Quantitative RT-PCR analysis of GATA 3 and STAT4 expression in different T effector cell cultures. Expression levels were normalized for GAPDH and are expressed as arbitrary units, whereby expression in wild-type Th1 cells (STAT4) or Th2 cells (GATA 3) was set to 1. Mean values and SD are given for 4 mice analyzed per group.

Supplementary Tables

Table S1. Genotype of Ctcf$^{+/-}$ x Ctcf$^{+/-}$ offspring.

Stage in development	Number of mice in genotype		
	wild-type	*Ctcf*$^{+/-}$	*Ctcf*$^{-/-}$
E 9.5	13	14	0
E 3.5	10	7	0

Table S2. Number of nucleoli in wild-type and CTCF-negative T cells.

CD4+	number of nucleoli (%)								
	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 nucleoli (%)								
	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)

Experimental Procedures

Fluorescent in situ hybridization (FISH)

FACS sorted CD4$^+$ and CD8$^+$ T cells were allowed to attach to glass slides for 30 min and fixed for 10 min with 4% PFA/PBS. Slides were stored in 70% EtOH until further use. 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 11.8kb 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.

Magnetic Bead Purification of TCR αβ T cells

$CD4^+$ and $CD8^+$ TCRαβ T cells were enriched (Figure S9) from spleen and lymph node cell suspensions through incubation with biotinylated mAbs to CD11b (M1/70), Gr-1 (RB6-8C5), Ter119 (Ly-76), TCRγδ (GL3), B220 (RA3-6B2), NK1.1 (PK136), followed by streptavidin-conjugated microbeads and autoMACS depletion according to the manufacturer's instruction (Milteny Biotec, Bergisch Gladbach, Germany). The purity of T cells was confirmed to be >90%.

References

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.*, et al.* (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, 4463-4474.

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.

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.

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.

Chapter 6

Dynamic behaviour of GFP tagged CTCF

Manuscript in preparation

Chapter 6

Dynamic behaviour of GFP tagged CTCF

Helen Heath[1], Frank Sleutels[1], Katharina Draegestein[1], Wiggert van Cappellen[2], Anton Grootegoed[2], Frank Grosveld[1], and Niels Galjart[1]

1) Department of Cell Biology and Genetics
2) Department of Reproduction and Development

Erasmus Medical Center, 3000 CA Rotterdam, The Netherlands

Summary

CTCF, an 11 Zn-finger protein, is involved in long-range chromatin interactions, imprinting and transcriptional regulation. Recent genome-wide analyses have revealed ~14000 CTCF-binding sites, which mark boundaries of histone modifications and might partition the genome in expression domains, indicating that CTCF acts as an insulator protein. Analyses using antibodies against CTCF mostly show CTCF in a speckled pattern in the interphase nucleus. Furthermore, CTCF remains associated with mitotic chromosomes. Here, we analyse the dynamic behaviour and localization of GFP-tagged CTCF. We show that this fusion protein is functional since: 1) it binds to cognate sites in the genome, 2) it localizes similar to endogenous CTCF, and 3) it functionally substitutes for CTCF in ES cells. We have generated a knock-in allele (*Ctcf*^{ki}) encoding GFP-CTCF instead of CTCF. Homozygous *Ctcf*$^{ki/ki}$ mice die at birth, due to the reduced expression of GFP-CTCF. Heterozygous *Ctcf*$^{ki/+}$ mice are viable and fertile. We show that GFP-CTCF is expressed in spermatogonia, spermatocytes and Sertoli cells of intact seminiferous tubules, but it is not detectable in spermatids. Fluorescence-based analysis demonstrates that ~200,000 molecules of CTCF are present in a fibroblast nucleus, which are partitioned into mobile and highly immobile fractions. A similar dynamic behaviour of GFP-CTCF is observed in fibroblasts, ES cells and male germ cells. Thus, CTCF behaves independently of the epigenetic state of a cell. As the number of immobile CTCF molecules within a nucleus is about five-fold higher than the number of reported binding sites, we propose that CTCF is has additional roles besides acting as an insulator protein.

Introduction

The highly conserved DNA binding protein CTCF (CCCTC-binding factor) was first identified as a transcriptional regulator of the chicken c-myc gene (Lobanenkov et al., 1990) and, independently, as the Negative Protein 1(NeP1), a chicken lysozyme gene silencer binding protein (Baniahmad et al., 1990). Although the name might suggest differently, CTCF is known to bind to sequences with a loose consensus. The flexibility in binding is caused by differential use of the 11 zinc fingers present in CTCF (Ohlsson et al., 2001).

A major breakthrough in the functional analysis of CTCF came with the identification of a CTCF-binding site within the DNA hypersensitive site 4 (5'HS4) of the chicken beta-globin insulator (Bell et al., 1999). An insulator is defined as a DNA element that prevents inappropriate gene activation or repression by neighbouring chromatin, by blocking access of flanking elements (such as enhancers or repressors) to a given promoter. The CTCF binding site of the chicken beta globin insulator is essential and sufficient to function as an enhancer blocker. Using chromatin-conformation-capture (3C) technology to investigate the three dimensional conformation of the chromatin fibre, it was discovered that the beta-globin locus forms an active chromatin hub (ACH) in which three CTCF binding elements (5' and 3' HS and the LCR HS) aggregate (Palstra et al., 2003). We have shown that CTCF is important for the 3D conformation of the hub and that absence of CTCF causes local histone modifications (Splinter et al., 2006). Interestingly, we have found that CTCF remains bound to specific binding sites in mitosis, suggesting that higher order chromatin structures need to be maintained during cell division (Burke et al., 2005).

The discovery that CTCF could function as an insulator protein, when present between enhancer and promoter, prompted groups to search for CTCF binding sites within the imprinted H19/IGF2 locus. This locus, which contains the paternally transcribed Igf2 gene and the maternally transcribed H19 gene, has been used as a paradigm for enhancer blocking. An element upstream of H19, known as the Imprinting Control Region (ICR), is essential for blocking enhancers, located downstream of the H19 gene, from activating the Igf2 gene. The ICR can only inhibit Igf2 transcription in its maternally inherited form, free of DNA methylation. Several groups simultaneously identified multiple CTCF binding sites within the ICR and showed that on the maternally inherited chromosome CTCF binds the ICR, yielding an explanation for the enhancer blocked and silent Igf2 gene (Bell and Felsenfeld, 2000; Hark et al., 2000; Kanduri et al., 2000; Szabo et al., 2000). Methylation of the paternal ICR inhibits CTCF binding, thereby allowing the enhancers to activate Igf2. Using 3C analysis the enhancer blocking activity of the CTCF-bound ICR has been explained by invoking differential types of looping in the maternal and paternal loci (Murrell et al., 2004). This looping model shares

characteristics with the ACH of the globin locus, as in both models CTCF binds to the elements that aggregate into a hub. However, in the beta-globin locus removal of a CTCF binding site destabilizes long-range interactions and causes local loss of histone acetylation and gain of histone methylation, but it does not lead to loss of globin expression (Splinter et al., 2006). Together the data positively link CTCF to chromatin architecture. This is supported, among others, by the observation that the nuclear matrix protein nucleophosmin is a CTCF binding partner and can tether CTCF-bound elements to the nucleolus (Yusufzai et al., 2004).

A recent screen for CTCF binding sites in the human genome has yielded 13,804 sites in primary human fibroblasts (Kim et al., 2007). All binding sites were postulated to be potential insulators. Notably, most sequences are located far from transcriptional start sites, yet their distribution is strongly correlated with the presence of genes on a given chromosome. CTCF localization was suggested to be similar in different cell types (Kim et al., 2007). Strikingly, the 20-bp consensus motif found in the majority of the sites (Kim et al., 2007) 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). All these data are consistent with a role for CTCF as an insulator protein.

In 2002 a paralogue of CTCF, named CTCF-L, or BORIS (Brother Of the Regulator of Imprinted Sites), was identified (Loukinov et al., 2002). Both proteins contain 11 zinc fingers (71% identical at the amino acid level) and can bind the same DNA. Interestingly, CTCF-L expression is restricted to testis. Antibody staining studies indicated that CTCF-L-positive spermatocytes stain negative for CTCF (Loukinov et al., 2002), raising the hypothesis that CTCF-L could fulfil an essential function in the absence of CTCF. It was hypothesized that the switch from CTCF to CTCF-L expression corresponded with the erasure of the bulk of DNA methylation (Loukinov et al., 2002). Thus, CTCF and, in testis, CTCF-L, play a key role as transcriptional silencing and activating factors, as well as organizers of (epigenetic) chromatin domains.

CTCF has been shown to localise to the nucleus in a speckled pattern. In specific cell types and/or under unusual conditions, CTCF accumulates on centrosomes and midbody (Zhang et al., 2000), in nucleoli (Torrano et al., 2006), adjacent to nucleoli (Yusufzai et al., 2004), or in other sub-nuclear structures (Kantidze et al., 2007). However, no dynamic analysis of CTCF has been reported yet.

To examine the dynamic behaviour of CTCF *in vivo* we have generated mice and cell lines in which GFP-CTCF is expressed instead of CTCF. Contrary to published data we show that CTCF is expressed in spermatocytes. Furthermore, we show that

CTCF partitions into two fractions: a mobile pool (~30% of the protein) and an immobile pool (~70 %). Significantly, our data supports the localisation of CTCF on mitotic chromosomes (Burke et al., 2005) and reveals a proportion non-bound CTCF protein is retained in mitotic cells and divided between the two daughter cells during cytokinesis. Our data indicate that CTCF is a stable protein. The number of CTCF molecules, as measured by FCS and the dynamics of the protein imply the number of proposed CTCF binding sites is underestimated.

Results

Characterization of the *Ctcf*^{ki} allele

To analyse the dynamic behaviour of CTCF we generated a *Ctcf*^{ki} allele in which GFP-CTCF is expressed instead of CTCF (Fig. 1A). We have used a similar strategy before to generate knock-in mice expressing GFP-CLIP170 (Akhmanova et al., 2005). Southern blot (not shown) and PCR analysis (Fig. 1B and data not shown) demonstrated correct targeting of the *Ctcf* gene and, after the action of Cre recombinase, correct removal of the neomycin resistance gene and one loxP site. We derived mouse embryonic fibroblasts (MEFs) from wild type, heterozygous and homozygous *Ctcf*^{ki} E13.5 day embryos. Western blot analysis on nuclear extracts from these cells demonstrated that GFP-CTCF is expressed, albeit at reduced (~20 %) levels compared to endogenous CTCF (Fig. 1C). Similar results were obtained in *Ctcf*$^{ki/+}$embryonic stem (ES) cells, as well as in kidney, testis and brain extracts from adult *Ctcf*$^{ki/+}$ mice (data not shown). We conclude that we have replaced endogenous CTCF with GFP-CTCF, but that this results in reduced expression of GFP-CTCF from the modified *Ctcf* allele.

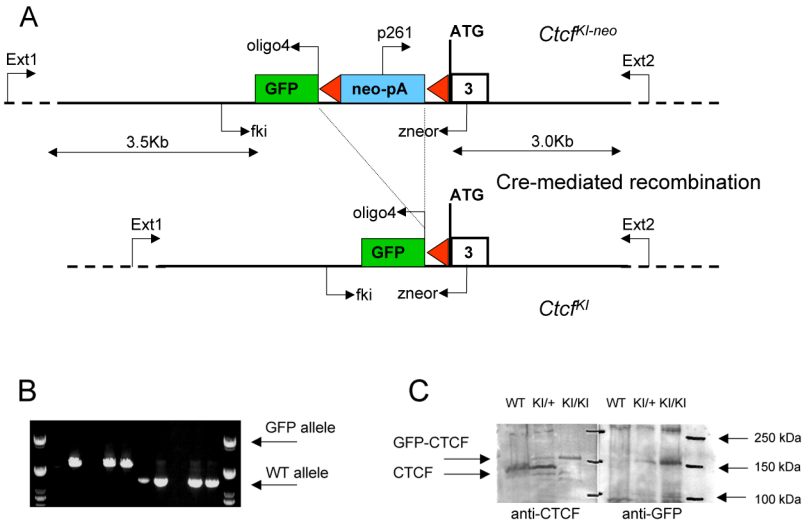


Figure 1. Characterization of the *Ctcf*^{ki} knock-in allele
A) Scheme of the *Ctcf*^{ki} knock-in allele. The top line represents the targeting construct, with *Ctcf* exon 3 (which contains the translation initiation codon (ATG)) indicated by a box. Double-headed arrows indicate the 5'end (3.5 kb) and 3'end (3.0 kb) homologous arms of the targeting construct. The insertion cassette contains GFP, loxP sites (red) and a neomycin resistance gene driven by the pMC1 promoter and containing a polyadenylation sequence (neo-pA, green). The *neo* gene is transcribed antisense to the *Ctcf* gene. Homologous recombination generates a *Ctcf*$^{KI\text{-}neo}$ allele. Cre-mediated excision leaves GFP and one loxP site. As these are in-frame with CTCF encoding sequences, the *Ctcf*^{ki} allele is generated. Positions and direction of PCR primers are indicated.
B) PCR analysis of genomic tail DNA. PCR demonstrates the presence of a homologously targeted *Ctcf*^{ki} allele in ES cells (indicated with GFP allele).
C) Western blot analysis. Nuclear extracts of wild type (WT), heterozygous (KI/+) and homozygous (KI/KI) *Ctcf*^{ki} MEFs were analysed with antibodies against CTCF (left panel) and GFP (right panel). In the heterozygous MEFs both CTCF and GFP-CTCF are present, but the latter protein is expressed at reduced levels.

Ctcf$^{ki/ki}$ mice are not viable

Attempts to generate homozygous *Ctcf*$^{ki/ki}$ mice failed: only 1 homozygous *Ctcf*$^{ki/ki}$ animal was obtained out of 132 offspring from *Ctcf*$^{ki/+}$ x *Ctcf*$^{ki/+}$ matings (Fig. 2A). This homozygous knock-in animal was much smaller than an age-matched wild type mouse (Fig. 2B), and it died after 3 weeks. These data show that *Ctcf*$^{ki/ki}$ mice are not viable. In fact, the *Ctcf*$^{ki/+}$ x *Ctcf*$^{ki/+}$ crosses yielded equal percentages of wild type and heterozygous *Ctcf*$^{ki/+}$mice (Fig. 2A), whereas this ratio should have been 1:2. These data indicate that heterozygous *Ctcf*$^{ki/+}$ mice are also present in less than expected numbers. A similar result was obtained after crossing wild type and heterozygous *Ctcf*$^{ki/+}$ mice (Fig. 2A), i.e. twice as many wild type animals were generated than *Ctcf*$^{ki/+}$ mice (expected ratio 1:1). These data raise the question of whether GFP-CTCF is a functional protein. However, when we analysed the genotypes of embryos at mid- to late gestation (Fig. 2A) the percentage of wild type, *Ctcf*$^{ki/+}$ and *Ctcf*$^{ki/ki}$ embryos was ~21 % (n=20), ~60 % (n=58) and ~19 % (n=18), respectively. Combined these data indicate that the *Ctcf*^{ki} allele is transmitted normally and is not deleterious for embryonic development, but that once mice are born the *Ctcf*^{ki} allele causes a selective disadvantage.

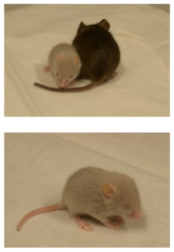
We next tested the proliferation capacity of wild type, heterozygous and homozygous knock-in MEFs, to examine whether presence of GFP-CTCF adversely affects cell growth. As shown in Fig. 2C wild type MEFs grow faster than *Ctcf*$^{ki/ki}$ fibroblasts. The fact that *Ctcf*$^{ki/+}$ MEFs grow faster than *Ctcf*$^{ki/ki}$ fibroblasts, but slower than wild type MEFs, indicates that a proper amount of CTCF is required for cell proliferation, but argues against GFP-CTCF having a dominant negative function.

A

Ctcf$^{ki/+}$ x wild type (n=157)		*Ctcf*$^{ki/+}$ x *Ctcf*$^{ki/+}$ (n=132)		
wt	*Ctcf*$^{ki/+}$	wt	*Ctcf*$^{ki/+}$	*Ctcf*$^{ki/ki}$
101	56	66	65	1

	E12.5 (33)	E13.5 (13)	E14.5 (21)	E15.5 (22)	E17.5 (7)
wt	6	2	5	7	0
Ctcf$^{ki/+}$	19	9	13	11	6
Ctcf$^{ki/ki}$	8	2	3	4	1

B



C

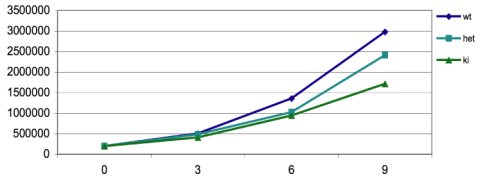


Figure 2. Inheritance and properties of *Ctcf*^{ki} alelle
(A) Inheritance of *Ctcf*^{ki} allele. Heterozygous *Ctcf*$^{ki/+}$ mice were either interbred or crossed to wild type animals. Offspring were counted after birth (upper panel), or at mid-gestation (lower panel).
(B) Impaired development of homozygous *Ctcf*$^{ki/ki}$ mice. The only homozygous *Ctcf*$^{ki/ki}$ mouse surviving to 3 weeks is shown. It is smaller than an age-matched mouse and had apparent neuronal defects (shaking).
(C) Proliferation assay. The proliferation capacity of homozygous *Ctcf*$^{ki/ki}$ and heterozygous *Ctcf*$^{ki/+}$ MEFs was compared to that of wild type MEFs. Homozygous Ctcf^{ki} MEFs have a reduced rate of proliferation.

GFP-CTCF is a functional protein

To examine whether GFP-CTCF is functional, we performed three experiments. First, in MEFs we showed complete colocalization of CTCF and GFP-CTCF (data not shown), indicating that the fusion protein distributes normally. Second, we generated embryonic stem (ES) cells carrying conditional *Ctcf* knock-out alleles (Heath et al., 2007). When treated with Cre recombinase these cells will delete CTCF and, as a consequence, they die (data not shown). We co-transfected these ES cells with GFP-CTCF and CTCF-GFP and picked clones that survived Cre treatment. We examined three independent ES cell lines for CTCF and GFP-CTCF expression. Clearly in all three clones only the fusion protein is expressed (Fig. 3A). These data demonstrate that GFP-CTCF can functionally substitute for CTCF in a cell survival assay. Interestingly, in all three clones the level of GFP-CTCF is highly similar to that of CTCF, strongly suggesting that cells can only grow well when appropriate amounts of CTCF are present.

The GFP-CTCF expressing ES cells were allowed to form small aggregated colonies. Under such conditions the normally round ES cell nucleus becomes highly irregular in shape (Fig. 3B). In most cells GFP-CTCF is localized in a speckled pattern throughout the nucleus and expression is reduced in the nucleolus. This distribution is similar to that reported for other cell types. Interestingly, we also captured images of an ES cell going through mitosis and could show GFP-CTCF on mitotic chromosomes, as reported before (Burke et al., 2005).

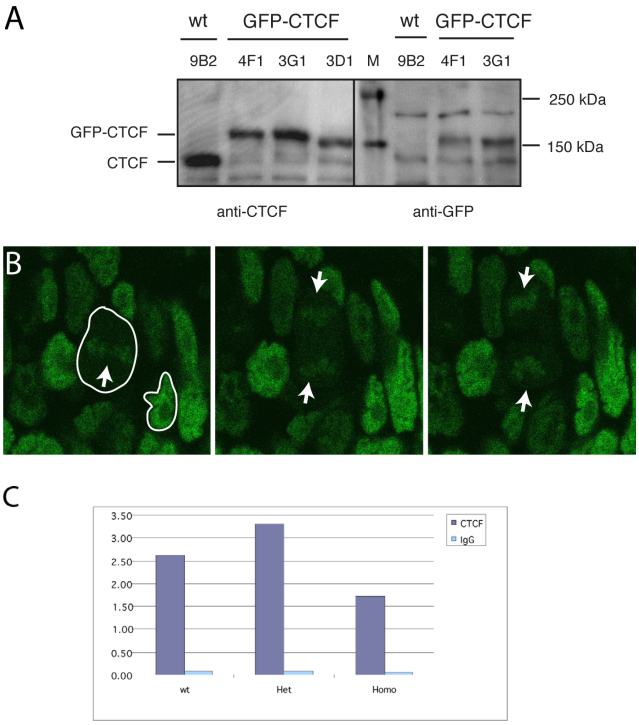


Figure 3. Functionality of GFP-CTCF

Western blot analysis of ES cells.
(A) Wild type ES cells expressing endogenous CTCF (wt, clone 9B2) are compared to three independent ES cell lines that express GFP-CTCF instead of endogenous CTCF (4F1, 3G1, 3D1). Notice that GFP-CTCF migrates at a higher position. Expression levels of GFP-CTCF and endogenous CTCF are similar.
The distribution of GFP-CTCF in ES cells.
(B) GFP-CTCF expression was imaged using confocal microscopy (three still images of a time-lapse movie are shown) (3B). For clarity the nuclei of two ES cells are outlined. One cell of the colony is going through mitosis. GFP-CTCF is seen on the mitotic chromosomes at the metaphase plate (arrow in left image), and moves with the chromosomes at telophase (arrows in middle and right images) to the daughter cells.

C) Chromatin immunoprecipitation (ChIP) analysis. The binding of CTCF and GFP-CTCF in wild type, heterozygous *Ctcf*$^{ki/+}$, and homozygous *Ctcf*$^{ki/ki}$ fetal liver cells to the beta-globin-locus was analysed by ChIP using anti-CTCF antiserum.

Our third approach to demonstrate functionality of GFP-CTCF was to analyse binding of the fusion protein to established sites in the mouse beta-globin locus (Splinter et al., 2006). Chromatin immunoprecipiation (ChIP) indeed showed binding of CTCF and/or GFP-CTCF to the 3' hypersensitive site (3'HS1) site in fetal livers derived from wild type, heterozygous *Ctcf*$^{ki/+}$, and homozygous *Ctcf*$^{ki/ki}$ embryos (Fig. 3C). GFP-CTCF not only binds to this cognate site but was also shown to ChIP other sites in the globin locus (data not shown). We conclude that GFP-CTCF is a functional protein.

In vivo distribution of CTCF

It has been reported that CTCF and CTCF-L are expressed in a mutually exclusive manner in male germ cells (Loukinov et al., 2002). To visualize GFP-CTCF in live germ cells, we dissected testis tubules (with Hoechst to mark nuclei) and examined these at 33^{o}C with a confocal/multiphoton set up. Previously we used this set-up to document expression of GFP-CLIP170 in the testis (Akhmanova et al., 2005). The distribution of GFP-CTCF was examined, tubules were scanned longitudinally, and optical sections acquired through the tubule. We observed GFP-CTCF-positive cells throughout the seminiferous tubule (Fig. 4A-F). In combination with Hoechst nuclear staining we identified cell types. Our data show that GFP-CTCF is expressed in spermatogonia, spermatocytes and Sertoli cells. By contrast, we could not detect GFP signals closer to the lumen of the tubule, in round and elongating spermatids. These results are in stark contrast to immunocytochemistry-based data (Loukinov et al., 2002). We conclude that CTCF localizes in all cells of the testis, except at later stages of spermatogenesis.

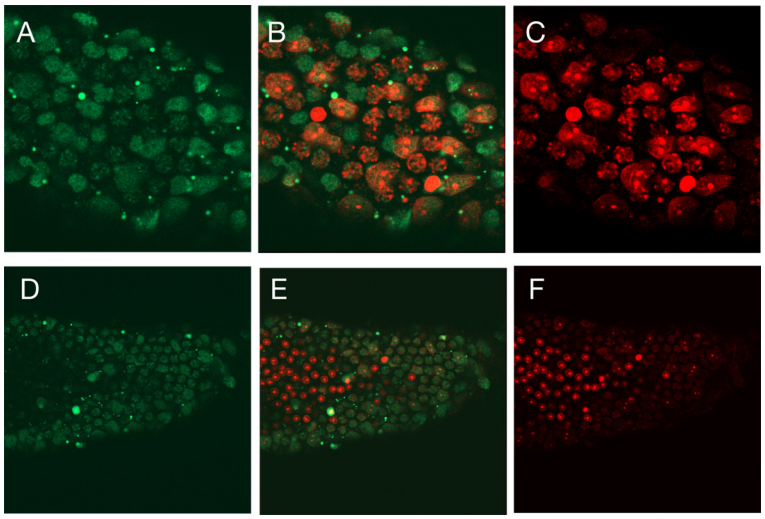


Figure 4. In vivo distribution of GFP-CTCF, GFP-CTCF expression in seminiferous tubules of the testis
GFP-CTCF expression (green) is shown in conjunction with Hoechst (red), a vital DNA stain that was injected into the testis and which stains nuclei on the inside of the blood-testis barrier. In panels A-C a surface staining of CTCF is shown. In panels C-F a cross section of cells within the tubule is shown. Notice that GFP-CTCF protein is expressed in all nuclei, including those of Sertoli cells (recognizable in panels A-C by the large Hoechst-positive dot inside the nucleus). However, GFP-CTCF is notably absent from round spermatids (panels D-F, small nuclei, again recognizable by the large Hoechst-positive dot).

Dynamic behaviour of CTCF

Using fluorescence recovery after photobleaching (FRAP) and fluorescence correlation spectroscopy (FCS) approaches (Bacia et al., 2006), we next analysed the dynamic behaviour of GFP-CTCF in cultured cells (Fig. 5) and in intact seminiferous tubules of the testis (data not shown). Both approaches revealed that a large percentage of CTCF (60-70 %) is immobile. Using FCS we calculated a concentration of ~200nM of GFP-CTCF inside a fibroblast nucleus. Given the fact that GFP-CTCF is expressed at ~20% of endogenous CTCF, this corresponds to about 200,000 molecules of CTCF inside a fibroblast nucleus.

With FRAP similar patterns of recovery were observed in homozygous and heterozygous knock-in MEFs (Fig. 5B), suggesting that the presence of endogenous CTCF does not affect the dynamic behaviour of GFP-CTCF. Furthermore, a similar behaviour was detected in MEFs and ES cells (Fig. 5) and in cells of the seminiferous tubule (data not shown). These data suggest that the dynamic behaviour of CTCF is not affected by epigenetic state of the cells.

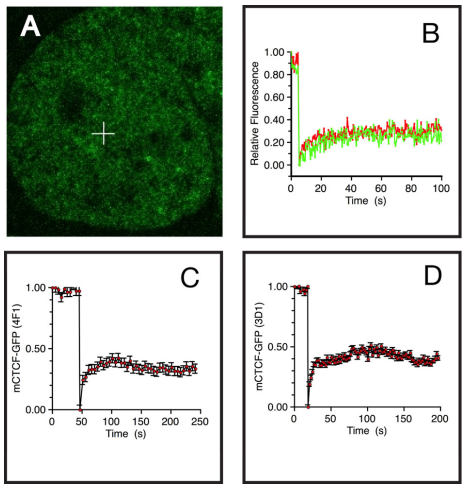


Figure 5. FRAP analysis
FRAP in MEFs. Heterozygous *Ctcf*$^{ki/+}$ and homozygous *Ctcf*$^{ki/ki}$ MEFs were analysed by FRAP. In (A) the GFP-CTCF distribution is in MEFs shown, notice the granular staining pattern and the apparent exclusion from the nucleolus. The crosshair indicates the position of the bleach.
In (B) fluorescent recovery of GFP-CTCF is shown in heterozygous *Ctcf*$^{ki/+}$ (red) and homozygous *Ctcf*$^{ki/ki}$ (green) MEFs. Approximately 70% of the total pool of GFP-CTCF does not recover within the time frame of the experiment. FRAP in ES cells.
(C, D) Two ES cell lines expressing GFP-CTCF were analysed by FRAP. Approximately 60 % of the total pool of GFP-CTCF remains immobile for the duration of the experiment.

Discussion

Here, we have analysed the dynamic behaviour and in vivo distribution of CTCF, tagged with GFP. We wanted to address three questions. First, CTCF was reported to be absent from spermatocytes and was not detected in Sertoli cells (Loukinov et al., 2002). However, the latter is very odd given the essential function of CTCF in virtually all cell types. Using the GFP-CTCF knock-in mice we wanted to address this issue. Second, recent reports emphasize a role for CTCF as an insulator (Barski et al., 2007; Kim et al., 2007; Xie et al., 2007), however, nothing is known about its dynamic behaviour. Third, we wished to calculate the number of CTCF molecules present in a nucleus.

Based on the successful analysis of GFP-CLIP170 knock-in mice (Akhmanova et al., 2005), we decided to use a similar strategy to generate GFP-CTCF expressing knock-in animals. However, we found that the amount of GFP-CTCF in knock-in cells and tissues was reduced to approximately 20 % of the level of endogenous CTCF based on western blot data. Western blot analysis was performed using a CTCF antibody with an N-terminus epitope. Although recognition of GFP-CTCF protein could be sub-optimal due to the antibody used, and therefore indicate lower expression than actual levels, we show other phenotypes that suggest low expression of CTCF occurs. CTCF has been

described to function in a dose-dependent manner in T-cells *in vivo* (Heath et al, manuscript in preparation). Heterozygous GFP-CTCF MEFs have reduced rates of proliferation compared to WT, but undergo greater expansion than homozygous GFP-CTCF MEFs. Interestingly, the inheritance of GFP-CTCF alleles in mice is strikingly similar to that of a CTCF knock out allele (Fig 2a, H Heath unpublished observations). Northern blot data suggest reduced levels of GFP-CTCF protein is not due to a reduction in the level of GFP-CTCF encoding mRNA (not shown). Thus, either GFP-CTCF is less stable than CTCF, or the fusion protein is produced less efficiently. Further research is required to distinguish between these options.

Ctcf$^{ki/ki}$ mice are not viable. We think this is largely due to the fact that GFP-CTCF is present in reduced amounts. Alternatively, GFP-CTCF may have a dominant negative effect. However, the fact that GFP-CTCF can fully substitute for CTCF in ES cells argues against such a scenario. Furthermore, out of 24 backcrosses of heterozygous *Ctcf*$^{+/-}$ knockout mice (Heath et al., 2007) to C57Bl6 wild type mice we obtained 101 wild type offspring and 74 *Ctcf*$^{+/-}$ knockout offspring (H. Heath, unpublished observations). Thus, also when the *Ctcf* gene is deleted heterozygous animals are apparently at a selective disadvantage. These data strongly suggest that the amount of CTCF is important for normal animal survival.

Using heterozygous *Ctcf*$^{ki/+}$ knock-in mice we analysed expression of GFP-CTCF in seminiferous tubules of the testis. We detected GFP-CTCF in all cells of the tubule, except in germ cells undergoing spermiogenesis (i.e. from the round spermatid stage onward). We believe this localization pattern reflects the true distribution of CTCF, as we did not fix tissues and did not have to employ antibodies to examine localization. A distribution of CTCF as described here is logical: CTCF is necessary in all cells of the body, including Sertoli cells and spermatogonia. However, in cells that start to actively compact their DNA and undergo replacement of their histones first with transition proteins and, subsequently, with protamines, CTCF is not required. Thus, the mutually exclusive expression of CTCF and CTCF-L is incorrect – however, the hypothesis that these proteins compete for binding sites on the DNA of male germ cells may still be correct.

Our live analysis in ES cells showed that CTCF is indeed localized on mitotic chromosomes, as previously reported (Burke et al., 2005). However, we did not detect GFP-CTCF on centrosomes (Zhang et al., 2004); this may be a special localization of CTCF in immortal cell lines. Furthermore, unlike in K562 cells (Torrano et al., 2005), we did not detect enrichment of CTCF on nucleoli of ES cells. Instead, in all cells tested, CTCF distributed in a speckled pattern. We could distinguish two fractions: ~30 % of CTCF was mobile, whereas ~70 % of the protein was immobile. The latter behavior explains the difficulties investigators might have in isolating the protein from nuclear extracts.

The number of CTCF molecules inside a fibroblast nucleus is ~200.000. Recent genome-wide analyses have revealed ~20,000 binding sites for CTCF within the human and mouse genomes (Kim et al., 2007; Xie et al., 2007). Thus, the number of CTCF molecules largely exceeds the number of binding sites (~40,000 in a diploid cell). In future experiments we will specifically investigate two possible explanations for this discrepancy. First, it is possible that the 20,000 binding sites that were reported are only a subset of the total number of CTCF binding sites. However, it appears unlikely that there are five times more sites. The second possibility is that CTCF has additional roles besides acting as an insulator. It will be interesting to determine whether the mobile and immobile pools of CTCF have different roles or that these pools do exchange but at very low rates. It is striking that the dynamic behavior of CTCF is so independent of epigenetic state of the cell. This observation also requires further research.

Materials and Methods

Molecular biology and antibodies

DNA, RNA and protein isolations were essentially performed according to standard procedures (Sambrook et al., 1989). Southern, northern and western blot analyses were performed as described (Hoogenraad et al., 2002). PCR analysis was using oligos as described in figure 1. Chromatin immunoprecipiation (ChIP) protocols have been described (Splinter et al., 2006). Antibodies against CTCF are published elsewhere (Heath et al., 2007). Anti-GFP antibodies were purchased from Abcam. Secondary antisera were alkaline phosphatase-labeled anti-rabbit and anti-mouse antibodies (Sigma).

Generation of GFP-CTCF knock-in allele, mice and mouse embryonic fibroblasts

Based on a previously published strategy (Akhmanova et al., 2005) we generated the *Ctcf*^{ki} allele, in which GFP-CTCF is expressed instead of CTCF. The same mouse PAC clones from female 129S6/SvevTac mouse spleen genomic DNA (Osoegawa et al., 2000) as previously used to generate the *Ctcf* knockout constructs (Heath et al., 2007) were now used to insert a GFP-lox-neo-lox casette into the ATG translation initiation codon on *Ctcf* exon 3. Targeting of embryonic stem (ES) cells, and selection and identification of the positive clones was performed as described previously (Hoogenraad et al., 2002). One targeted ES cell clone with the correct karyotype was chosen to convert the knockout allele into a GFP-CTCF knock-in allele by Cre-mediated recombination in ES cells. Heterozygous GFP-CTCF knock-in ES cells were injected into C57Bl/6 blastocysts. Chimeric males were mated to C57Bl/6 females to obtain germ line transmission of the modified allele. Mouse genotyping was routinely performed by PCR.

Generation of GFP-CTCF expressing ES cells

ES cells with a floxed *Ctcf* allele were derived from female conditional *Ctcf* knockout mice (Heath et al., 2007). These cells were infected with a lentivral construct expressing Cre recombinase in combination with GFP-CTCF (F. Sleutels, manuscript in preparation). Thus, while the endogenous *Ctcf* allele is deleted cells can be rescued by expressing "transgenic" GFP-CTCF. Several ES cell lines were isolated and analyzed.

Time lapse imaging

The fluorescence live imaging of cultured somatic cells and of testis tubules was performed as described (Akhmanova et al., 2005). When required, 5 μg/ml Hoechst 33342 (Molecular Probes) was added. In this incubation system, the Hoechst stain is taken up only by nuclei on the basal side of the Sertoli cell barrier (also called "blood-testis barrier"). This allows for identification of all cells on the basal side of this barrier: peritubular cells, Sertoli cells, spermatogonia, and preleptotene spermatocytes. When applicable, the testis was injected through the rete testis with Hoechst 33342 and Trypan blue (Sigma) in 3-5 μl PBS, one hour prior to testis dissection, to allow spreading of the vital DNA stain throughout the adluminal compartment of the testis tubules and uptake by nuclei on the adluminal side of the Sertoli cell barrier.This method makes it possible to identify the germ cell types that have migrated through the Sertoli cell barrier. The testis tubules were examined at 33^0C, using a Zeiss LSM510NLO confocal/multiphoton set up, to allow simultaneous acquisition of phase-contrast, GFP and Hoechst images.

Fluorescence-based biophysical measurements

Fluorescence recovery after photobleaching (FRAP) was carried out with the 63x planapochromat (1.4 NA) oil immersion lens as described (Akhmanova et al., 2005). Bleaching of an outlined region of interest (ROI) was done with several iterations of the 488 nm laser at full transmission (2.6 mW). Prior to and after bleaching the laser was set at 2% transmission (55 μW). The Zeiss LSM software was used to measure pixel intensities inside different ROIs. Recovery values were normalized to the prebleach values for each ROI. Values were imported into Aabel (Gigawiz) for graphical representation and statistical analysis.

Fluorescence correlation spectroscopy (FCS) measurements were conducted as described previously (Drabek et al., 2006) in MEFs expressing GFP-CTCF. For these FCS experiments we used the LSM 510-Confocor II (Zeiss) and the 488nm Ar-laser (beampath: HFT488-Mirror-BP505-550). The system was calibrated with rhodamine 6G (diffusion coefficient: $28x10^{-10}$ m^2/s at 20 $^{\circ}$C).

Using FCS we calculated that approximately 20 molecules of GFP-CTCF are present in the FCS measurement volume of 0.25 fl. To calculate the total number of

CTCF molecules in a nucleus we assumed that GFP-CTCF is expressed at about 20 % of the level of CTCF and that a nucleus has a radius of 5 micrometeres.

References

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*., et al.* (2005). The microtubule plus-end-tracking protein CLIP-170 associates with the spermatid manchette and is essential for spermatogenesis. Genes Dev *19*, 2501-2515.

Bacia, K, Kim, S.A, Schwille, P. (2006). Fluorescence cross-correlation spectroscopy in living cells. Nature Methods. *3* , 83-89.

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

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

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

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.

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

Drabek, K., M. van Ham, T. Stepanova, K. Draegestein, R. van Horssen, C.L. Sayas, A. Akhmanova, T. Ten Hagen, R. Smits, R. Fodde, F. Grosveld, and N. Galjart. 2006. Role of CLASP2 in Microtubule Stabilization and the Regulation of Persistent Motility. *Curr Biol*. 16:2259-64.

Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M., and Tilghman, S.M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature *405*, 486-489.

Heath, H., Ribeiro de Almeida, C., Dingjan, G., Van de Nobelen, S., Sleutels, F., Ling, K.-W., Jonkers, I., Renkawitz, R., Grosveld, F., Hendriks, R.W*., et al.* (2007). CTCF controls distinct pathways in T cell development, activation and cytokine production. manuscript in preparation.

Hoogenraad, C.C., B. Koekkoek, A. Akhmanova, H. Krugers, B. Dortland, M. Miedema, A. van Alphen, W.M. Kistler, M. Jaegle, M. Koutsourakis, N. Van Camp, M. Verhoye, A van der Linden, I. Kaverina, F. Grosveld, C.I. De Zeeuw, and N. Galjart. 2002. Targeted mutation of Cyln2 in the Williams syndrome critical region links CLIP-115 haploinsufficiency to neurodevelopmental abnormalities in mice. *Nat Genet*. 32:116-27.

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.

Kantidze, O.L., Iarovaia, O.V., Philonenko, E.S., Yakutenko, II, and Razin, S.V. (2007). Unusual compartmentalization of CTCF and other transcription factors in the course of terminal erythroid differentiation. Biochim Biophys Acta *1773*, 924-933.

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.

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.

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

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.

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.

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.

Osoegawa, K., M. Tateno, P.Y. Woon, E. Frengen, A.G. Mammoser, J.J. Catanese, Y. Hayashizaki, and P.J. de Jong. 2000. Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome Res*. 10:116-28.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2 Edition. Cold Spring Harbor Laboratory Press, New York.

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.

Szabo, P., Tang, S.H., Rentsendorj, A., Pfeifer, G.P., and Mann, J.R. (2000). Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function. Curr Biol *10*, 607-610.

Torrano, V., Chernukhin, I., Docquier, F., D'Arcy, V., Leon, J., Klenova, E., and Delgado, M.D. (2005). CTCF regulates growth and erythroid differentiation of human myeloid leukemia cells. J Biol Chem *280*, 28152-28161.

Torrano, V., Navascues, J., Docquier, F., Zhang, R., Burke, L.J., Chernukhin, I., Farrar, D., Leon, J., Berciano, M.T., Renkawitz, R.*, et al.* (2006). Targeting of CTCF to the nucleolus inhibits nucleolar transcription through a poly(ADP-ribosyl)ation-dependent mechanism. J Cell Sci *119*, 1746-1759.

Xie, X., Mikkelsen, T.S., Gnirke, A., Lindblad-Toh, K., Kellis, M., and Lander, E.S. (2007). Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. Proc Natl Acad Sci U S A *104*, 7145-7150.

Yusufzai, T.M., Tagami, H., Nakatani, Y., and Felsenfeld, G. (2004). CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. Mol Cell *13*, 291-298.

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.

Zhang, X., Chen, J., Kaoru, M., Wang, X., Liu, T., and Lu, X. (2000). [The promoter sequence of human nitric-oxide-synthase gene: functional study]. Zhonghua Yi Xue Yi Chuan Xue Za Zhi *17*, 13-15.

Chapter 7

Discussion

Chapter 7: General discussion and future directions

What does CTCF do and what have we learnt?

The CTCF protein is known to be required for embryonic development, cell cycle progression, regulation of transcriptional activity, imprinted gene expression and X-inactivation in conjunction with its role as a mediator of chromatin structure, histone modifications, chromatin interactions *in cis* and chromosome interactions *in trans*. The fact that one single protein has such diverse functional roles makes it difficult to tease out its function in distinct cellular events.

In this thesis we focused upon studies where the Ctcf gene has been deleted in a spatial and temporal manner from two different cell lineages. Such studies have not yet been reported. By conditional deletion analysis, we show that cellular expansion and differentiation *in vivo* is CTCF dependent. Surprisingly, our studies have revealed several key cellular processes that do not appear to require CTCF. We also address questions regarding levels of CTCF protein within the cell and suggest that a critical level of CTCF is required for development and cellular function.

CTCF as a tumour-suppressor gene

The current information regarding the potential tumour-suppressor function of CTCF comes from analysis in human cells. CTCF maps to the long arm of chromosome 16, in a region frequently deleted in sporadic breast and prostate tumours, 16q22.1 (Filippova et al. 1998). The binding of CTCF to promoter regions of tumour suppressor genes *Brac1* and *retinoblastoma* (*Rb*) prevents the acquisition of methylation and repression (De La Rosa-Velazquez et al. 2007, Butcher et al. 2004). However, several studies have failed to ascertain correlations between the loss or aberrant expression of CTCF and tumour progression (reviewed in Recillas-Targa et al. 2006). To date, it is unknown if CTCF has the same potential function in mice. In our studies, we found no evidence of obvious tumours in heterozygous CTCF mice, nor the development of lymphoma in mice with CTCF depleted T-cells. CTCF is expressed in both mouse and human tumour-cell lines (MEL cells and HeLa cells), however it is not known if CTCF still interacts with critical target sites, such as *PIM1*, $p16^{INK4a}$, *Rb*. ChIP experiments to determine CTCF interaction with these loci would be very interesting. Changes in DNA methylation may occur at promoter regions in the mouse genome as have been demonstrated in human cells. Preliminary experiments suggest that global patterns of DNA methylation (rDNA locus, H. Heath unpublished observations) do not change in CTCF deficient murine cells. Bisulphate sequencing of specific loci would allow a more sensitive detection of potential changes in methylation patterns.

In mice, especially in lymphoid cells, the proto-oncogene c-myc requires tight regulation. Transgenic mice over-expressing c-myc are susceptible to both B- and T-cell lymphomas, whereas T-cell specific deletion of c-myc induces developmental arrest at the DN3 stage (Adams et al. 1985, Spanopoulou et al.1989, Stewart et al 1993). CTCF was originally identified as a transcriptional repressor of c-myc (Lobanenkov et al. 1990) and increased expression of CTCF in WEHI 231 B-cell lymphoma cultures induced the down-regulation of c-myc expression (Qi et al, 2003). Significantly, in T-cells where the CTCF gene had been conditionally deleted, we saw no evidence of lymphoma in aged mice (45 weeks) or mis-regulated c-myc expression. CTCF binding sites in c-myc regulatory regions are conserved between human and mice. It is not known if these binding sites acquire methylation in the absence of CTCF, or indeed if CTCF remains bound when c-myc is activated and therefore transcriptional regulation could be mediated by co-activating/repressing complexes together with CTCF. Importantly, inducible activation of *lck-c-mycER* in transgenic mice indicated no significant alterations in the proportions of thymocyte subsets comparing mice treated or not with 4-hydroxy-tamoxifen (4-OHT). This indicates the consequences of c-myc activation in T-lymphocytes may be a reflection of the effects on c-myc target genes following constitutive over-expression of c-myc that manifest as malignancies (Rudolph et al. 2000). Lck-Cre-mediated deletion of CTCF in T-cells induced a developmental block at the ISP stage due to an arrest at G_1, and *in vitro* proliferation of peripheral T-cells was impaired in cells derived from $CTCF^{lox/lox}$CD4Cre mice. This does not correlate with de-repression of c-myc, a growth-promoting factor. On the contrary, inhibition of c-myc is known to prevent lymphocytes entering S-phase (Heikkila et al. 1987). Taken together this suggests that in T-cells at least, the loss of CTCF does not directly activate c-myc mRNA expression, nor induce T-cell lymphomas. It will be interesting to see if c-myc expression is mis-regulated in the absence of CTCF in other conditional systems that by-pass the use of transformed cell lines. Mouse embryonic fibroblasts derived from $CTCF^{lox/lox}$ mice infected with Cre-expressing lenti-virus also fail to proliferate and do not have increased expression of c-myc (F. Sleutels and S. vd Nobelen unpublished observations).

Mutations in p53 and the cyclin-dependent kinase (CDK) inhibitor p27 (*CDKN1B*) are also associated with tumour progression in mice (Philipp-Staheli et al.2004). The down-regulation of nuclear p27 or cytosolic sequestering is frequently associated with tumours, particularly in conjunction with the loss of p53. In the absence of CTCF, no significant effect on p53 mRNA expression was observed in thymocytes and expression of both p27 and p21 increased compared to wt cells. The expression of p27 and p21 is high in normal quiescent T-cells and must be down-regulated to allow transition from G_1 to S-phase (Tsukiyama et al. 2001, Wolfraim et al. 2004). The elevated expression of p27 in CTCF deficient thymocytes might inhibit the development of

tumours, if CTCF was to be involved in T-cell lymphoma progression. It is unknown if p27 or p21 expression is elevated in peripheral T-cells and it remains to be seen if CTCF is a direct transcriptional repressor of p27 in thymocytes. CTCF expression is increased in activated/proliferating cells (Herblot et al. 1999), which could indicate the up-regulation of CTCF potentially controls p27 expression, perhaps via recruitment of repressive chromatin modifying complexes. CTCF has been reported to interact with histone deacetylases, although this has not been verified in T-cells. Alternatively, CTCF could initiate/promote the degradation of p27 by triggering appropriate pathways, however this is merely a suggestion as there is currently no evidence of CTCF-mediated degradation of any proteins. The related CDK-inhibitor p21 is a transcriptional target of p53 also associated with tumour-suppressor activity although to a lesser extent that p27. In contrast to T-cells, p27 expression is not elevated in CTCF deficient MEFs, although p21 expression is up-regulated consistent with a proliferation defect in these cells (S. vd Nobelen unpublished observations). This could indicate differential regulatory pathways in these distinct cell types. Cell-cycle inhibition of p21 and p27 occurs via binding of cyclin/CDKs preventing cell-cycle progression. An increased level of p21 mRNA expression is an effect most often mediated by low levels of c-myc. However, it is not likely that increased expression of p21 reflects decreased c-myc expression in the absence of CTCF as no dramatic effect on c-myc mRNA expression was detected in CTCF deficient thymocytes by RT-PCR. A genome-wide analysis of CTCF binding sites in human cells revealed four sites around the p21 promoter (Barski et al. 2007). The corresponding region in MEFs also binds CTCF as indicated by ChIP (S. vd Nobelen unpublished observations). These data indicate that the p21 gene is a direct target of CTCF.

Cell proliferation, differentiation and locus-specific gene expression in the absence of CTCF

Data presented in this thesis show that CTCF is a key regulator of chromatin structure at the *β-globin* locus and is necessary for cellular expansion, cytokine expression and the functional capacity of T-cells *in vivo*. However a number of key regulatory processes apparently occur in a CTCF-independent manner.

Normal expansion of T-cells in the thymus requires CTCF

CTCF is expressed at every stage of T-cell development, as determined by GFP expression from heterozygous knock-in mice. In $CTCF^{lox/lox}$Lck-Cre mice, defective thymocyte differentiation is evident by a dramatic decrease in the total number of thymocytes, an increased number of cells at the DN-to-ISP stage and a significant

reduction in DP cells. The transition between DN and DP cells proceeds through the ISP intermediate, which requires correct preTCR expression and signaling, functional rearrangement of *Tcrβ* genes (β-selection), rapid proliferation, and - in mice - activation of *Cd8* expression. Each of these regulatory events is mediated by changes in chromatin remodeling and histone modifications, DNA methylation and even dynamic positioning of chromatin in nuclear space (Krangel, 2007). Careful analysis of these processes revealed no apparent defects in the absence of CTCF - determined by RT-PCR - with the exception of proliferative capacity. *Tcr* genes are formed by somatic recombination of different gene segments and expression of the four different *Tcr* types requires developmental coordination. CTCF-dependent and independent insulator sequences are present on the *Tcrα/δ* locus (Magdinier et al. 2004). The expression of neither *Tcrα* nor δ genes is aberrantly regulated in the absence of CTCF, suggesting other unknown proteins are responsible for spatial and temporal coordinated expression of this locus.

Subsets of thymocytes (DN, ISP, DP/CD4^{+}SP) analysed by RT-PCR revealed elevated expression of p21 and p27 mRNA in CTCF-/- ISP cells, consistent with the observed arrest in G_1. Interestingly, a number of other studies have highlighted phenotypes similar to CTCF deficiency in T-cells. The bHLH factor HEB and Lsh1, a member of the SNF2 family of chromatin remodeling factors, are both required for the DN-to-DP transition (Barndt et al. 2000, Geiman et al. 2000). The phenotypic similarities between these mutants and CTCF$^{lox/lox}$Lck-Cre mice are suggestive of a role for CTCF in mediating the expression or functional capacity of these proteins. However there is currently no evidence for this. Interestingly, TCF1, a T-cell specific HMG box protein, has been identified in CTCF-biotin tagged pull-downs using nuclear extracts from thymus (S vd Nobelen unpublished observations). The functional significance of this interaction remains to be verified and analysed *in vivo*. TCF1 is important for expression of *Tcrα* genes (Roberts et al. 1997), however *Tcrα* gene expression is not impaired in the absence of CTCF. These studies indicate that an intricate network of factors function coordinately to regulate T-cell differentiation. Importantly, it also suggests T-cell differentiation and gene regulation in T-cells requires more than CTCF *per se* and CTCF is not able to rescue or compensate for defects in factors such as HEB or Lsh1 assuming CTCF is unaffected in these mutants.

Despite the profound developmental block at the ISP stage, T-cell development can proceed in CTCF$^{lox/lox}$Lck-Cre mice and cells can migrate from the thymus and populate the peripheral lymphoid tissues, albeit in very low numbers. CTCF mRNA was strongly reduced in a pool of all DN stages and absent in CTCF-/- thymocytes by the ISP stage. However we have observed that loss of CTCF protein as detected by western blot occurs only 3-4 days after initiation of recombination, therefore CTCF is a highly stable protein. It is possible that residual CTCF remains in developing thymocytes long enough

to support progression through the various checkpoints in the thymus (chapter 5). Importantly, the presence of mutant mature cells in the periphery indicates CTCF is not required for survival of these cells.

The *Cd4* and *Cd8* loci require coordinated and dynamic changes in chromatin modifications to allow the correct expression of the genes that characterize the different developmental stages of thymocyte development. Significantly, the expression of neither *Cd4* nor *Cd8* appears to be mis-regulated in the absence of CTCF. The *Runx1* and *Runx3* transcription factors are highly implicated in the regulation of *Cd4* expression (Taniuchi et al. 2002). Interestingly, Runx factors mediate silencing of *Cd4* through binding to the nuclear matrix and attracting repressive histone modifying complexes in a manner similar to CTCF. CTCF is not known to interact with or regulate the expression of Runx factors, nor interact with enhancers required for *Cd8* expression (Sato et al. 2005). CTCF binding sites within the *Cd4* and *Cd8* loci have not been identified.

CTCF is required for T_H2-mediated responses *in vivo* and distinct activation pathways and cytokine expression *in vitro*

Recombination of *Ctcf* after β-selection increased the numbers of CTCF-/- cells in the periphery, presumably because CTCF is necessary for the highly proliferative stages of thymocyte development around β-selection, but may not affect the less proliferative later stages. We analysed the functional capacity of CTCF-/- peripheral $CD4^+$ cells by ELISA, measuring the concentration of T-helper dependent immunoglobulins (IgGs) in serum. Levels of the Th-2 dependent IgG1 were significantly reduced and levels of the Th-1 dependent IgG2a was slightly lower in CTCF mutant mice. This strongly suggested Th-2 cells in particular were not functional. Th-2 cells are key mediators of humoral immunity that requires activation of B-cells and regulates the response to allergens. Induction of an allergic response as measured by serum levels of IgE following immunisation with TNP-KLH was defective in the absence of CTCF. Further analysis of CTCF deficient mice is required to ascertain the specific defects presented. It would be of interest to challenge CTCF mutant mice with Th-1 inducing pathogens as IGg2a levels were also slightly reduced in the ELISA, and expression of Th-1 dependent factors STAT4, T-bet and Ifn-γ are also reduced.

The differentiation of naïve T-cells into distinct effector cells is characterized by restricted transcription factor expression profiles, DNA methylation and changes in chromatin structure. Histone modification patterns reveal the restricted expression profile of *Il4* is set in the thymus in conjunction with $CD4^+/CD8^+$ lineage choice. It is not clear if the correct epigenetic modifications are made in CTCF -/- thymocytes, and whether or

not $CD8^+$ cells can express IL-4. T-helper cells are characterized by differential cytokine expression. In CTCF depleted T-helper cells, the expression of signature Th-2 cytokines is almost absent and Th-1 cells express reduced levels of Ifn-γ. Differentiation of naïve $CD4^+$ cells into Th-2 cells is accompanied by replication-dependent demethylation of CpG sites within the *Il4/Il13* locus (Lee et al. 2002). The DNA-binding capacity of CTCF is methylation sensitive, therefore if CTCF interacts directly with the promoters of Th-2 cytokine genes, this would occur as a result of activation. Demethylation may be triggered by or require IL-4R signaling, as DNA methylation is retained in both primary $CD4^+$ and $CD8^+$ cells activated in non-polarising conditions. DNA methylation is a critical epigenetic mediator of cytokine expression. Both $CD4^+$ and $CD8^+$ cells express IL-4 in the absence of Dnmt1 (Makar et al. 2003). In the absence of CTCF both Th-1 and Th-2 cells fail to proliferate efficiently *in vitro* when activated by anti CD3/anti CD28. It is possible that the Il4/Il13 locus in CTCF -/- cells remains methylated due to reduced replication-dependent demethylation. In this case, the activation of *Il4, Il5 and Il13* may not be directly dependent on CTCF. Proliferative signals perhaps mediated by CTCF may trigger demethylation of the locus.

The molecular basis for CTCF-mediated cytokine expression remains elusive. The particular T-helper defects may reflect specific CTCF-protein complexes or target genes not required at earlier stages. GATA3 is a key regulator of both direct transcriptional activation and chromatin structure of the *Il4/Il13* locus. In the absence of CTCF, GATA3 protein was not induced. It is unlikely that CTCF directly activates Gata3 transcription given that expression of Gata3 mRNA was not significantly different comparing WT and knock out cells. When stimulated by αCD3ε/αCD28, which mimics TCR signaling, elevated levels of STAT4 repressed the expression of GATA3 possibly via T-bet. However, when PMA/ionomycin was used the induction of GATA3 recovered. This situation is similar to effects seen when T-bet is ectopically expressed in Th-2 cells. Activation by PMA/ionomycin induced a strong increase in T-bet-mediated Ifn-γ production but Ifn-γ expression was only weakly increased when stimulating with IL-12-Il-18 or OVA/APC (Afkarian et al. 2002). PMA therefore bypasses the effect of a factor in either TCR or cytokine-induced pathways. Antigen-APC or cytokine stimulation induces STAT4 dependent Ifn-γ production, whereas PMA/ionomycin induces STAT4 independent Ifn-γ production. In a similar model, PMA/ionomycin could circumvent the need for STAT6 activation of GATA3.

Importantly, even in the presence of GATA3, the expression of *Il-4*, *-5* and *-13* remained low in CTCF deficient cells. This suggests that factors besides GATA3 are required for cytokine expression in the absence of CTCF. GATA-3 and STAT6 are known to regulate chromatin interactions *in cis* at the *Il4/Il13* locus that mediate cytokine expression. SATB1 also forms chromatin loops along this region, and is required for

cytokine transcriptional activity (Cai et al. 2006, Spilianakis et al. 2004). The study by Cai et al suggest GATA3/STAT6 looping reflects a poised chromatin state that is insufficient to induce expression of cytokines, and instead propose that SATB1-mediated c-Maf expression directs the transcription of *Il4, Il5* and *Il13.* The higher order transcriptionally active chromatin configuration at this locus requires SATB1, c-Maf and RNA pol II. In CTCF -/- PMA/ionomycin stimulated Th-2 cells, SATB1 is expressed as revealed by western blot. However we do not know if SATB1-mediated loops are retained or if SATB1 is functional in CTCF deficient cells. SATB1 is required for c-Maf activity, and although we have not demonstrated this, it is unlikely that c-Maf is not induced in the presence of SATB1. CTCF itself is necessary for the formation of chromatin loops *in cis*, although CTCF is not yet known to be necessary for chromatin loops at this locus. Database analysis of CTCF binding sites at the *Il4/Il13* locus in humans reveals strong binding sites flanking the locus, but only weak if any potential binding sites within the locus, including the Th-2 LCR (figure 7.1). It is reasonable to suggest this pattern of binding sites is also present in mice, but this remains to be verified by chromatin immunoprecipitation.

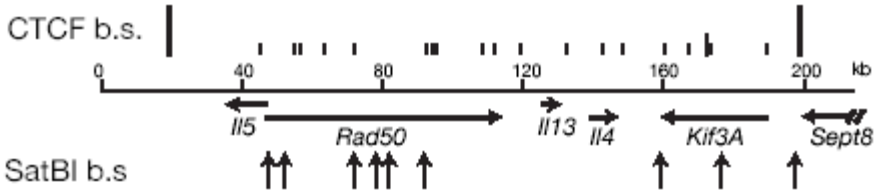


Figure 7.1: A schematic representation of predicted CTCF and SATB1 binding sites in the T_H2 locus.

Transcriptional orientation is indicated by horizontal arrows. Vertical arrows indicate SatB1 binding sites (see text for details). Vertical lines indicate potential CTCF binding sites. The height of vertical lines is indicative of the confidence CTCF will bind at that position (Barski et al. 2007). Position of the binding sites is approximate. Position of CTCF-binding sites was obtained from human T cells (Barski et al. 2007), whereas SATB1-binding sites were obtained from mouse T cells (Cai et al. 2006).

The different effects of stimulation via the TCR or bypassing the TCR with PMA/ionomycin suggests signaling from the TCR is impaired in the absence of CTCF. It is not clear if this is a question of thresholds and whether CTCF -/- cells could respond to TCR-derived signals if the intensity of this was increased. The expression of TCRβ and CD3 is equivalent in WT and CTCF deficient cells as measured by FACS. This suggests there is no obvious discrepancy regarding the number of TCRs present on the cell

surface. RT-PCR analysis of signaling molecules required immediately downstream of the TCR indicated no significant difference in expression between control and CTCF mutant $CD4^+$ cells. Signal cascades are often mediated by phosphorylation, and it remains unknown if this occurs in the absence of CTCF. Western blot analysis of CTCF deficient cells with phosphorylation-specific antibodies would be required to analyse this.

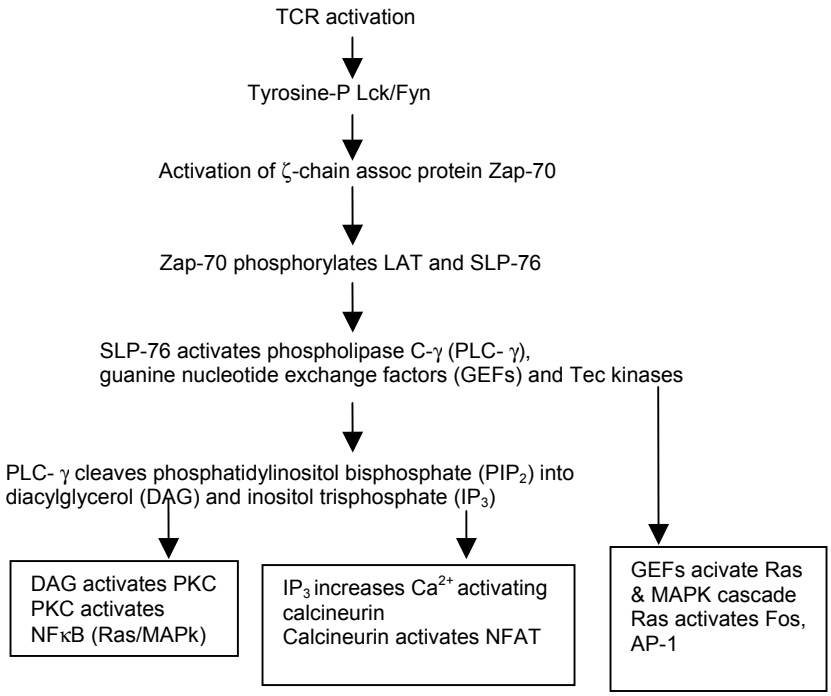


Figure 7.2: A simplified scheme of TCR-mediated intracellular signaling.
Based on (Janeway et al. Immunobiology 6th edition, Garland Science Publishing 2005).

Preliminary experiments suggest a solo effect of ionomycin-derived signals is sufficient to 'rescue' cell division and GATA3 expression in the absence of CTCF. Stimulation of CTCF knock out cells with αCD3ε/PMA did not support proliferation or GATA3 expression. PMA and ionomycin activate the Ras/MAPk pathway and calcium signals respectively (fig 7.2). Importantly, within a cell, pathways as depicted in figure 7.2 will not work independently, but invariably will share regulatory factors and influence each other. Ionomycin induces calcineurin/NF-AT -regulated transcription factors, which are probably responsible for chromatin remodeling and/or GATA3 induction in the absence of CTCF and possibly in the absence of STAT6 signals, although this remains to be verified. Ras-Erk signals are required for the DN-DP transition that is impaired in the absence of

CTCF. However, in human corneal epithelial (HCE) and myeloid cells, CTCF is activated by Erk signals and down-regulated by cellular stress such as UV irradiation or hyper-osmotic stress (Li. 2007). Stress-induced activation of NFκB is proposed to regulate the repression of CTCF as this is reversed in the presence of NFκB inhibitors. The interleukin 1 receptor-associated kinase (IRAK) family activate NFκB via Toll-like receptors. The only lymphoid specific factor known to be a direct target of CTCF is Irak2 (Kuzmin et al. 2005). CTCF was shown to bind to the Irak2 promoter and directly activate Irak2 transcription in human cells. Preliminary analysis by quantitative RT-PCR indicates this function of CTCF is conserved in mice (H. Heath unpublished observations). The IRAK family of proteins are important regulators of immune responses; IRAK4 is a critical regulatory factor of innate and acquired immunity (Suzuki et al. 2006). CTCF binding sites in the IRAK4 promoter remain to be established. The link between CTCF and Irak2 could suggest CTCF may be involved in signaling for the activation of NFκB but not NF-AT and explain why CTCF deficient T-cells respond to ionomycin but not PMA activation. A comparison of NFκB and NF-AT expression in WT and CTCF deficient $CD4^+$ cells by western blot is required in the first instance. The significance of CTCF-mediated down regulation of Irak2 in mice is currently unknown. Irak2 induces pro-apoptotic signals via caspase 8 in response to *Yersinia* infections and LPS. It would be of interest to challenge CTCF mutant mice with inflammatory pathogens.

CTCF is not required for the development and proliferation of γδ-T-cells

CTCF is expressed in γδ T-cells as shown by GFP-CTCF expression and the *Ctcf* gene in γδ T-cells can be recombined in an LCK-Cre-mediated manner as demonstrated by lacZ expression. The LCK-Cre transgene is active at the DN1 stage prior to αβ/γδ fate decisions in thymus (Shimizu et al. 2001). Initial attempts to analyse the proliferation and cytokine expression of $CTCF^{lox/lox}$Lck-Cre -/- $CD4^+$ cells *in vitro* generated cultures producing high levels of Ifn-γ even in Th-2 conditions. FACS analysis of these cultures after 7 days revealed a substantial proportion of γδ T-cells were conditionally deleted as shown by *lacZ* expression and western blot, despite initial MACS purification for $CD4^+$ cells. In the absence of CTCF, γδ T-cells have a proliferative advantage over $CD4^+$ and $CD8^+$ cells *in vitro*. However the molecular mechanism behind this and whether CTCF deficient γδ T-cells are functional *in vivo* is unknown. In humans, γδ T-cells are sensitive to activation-induced cell death, and standard mitogen-activated culturing is not suitable for γδ T-cell expansion (Guo et al. 2002, Lamb et al. 2005). Preferential expression of the IL-15Rα opposed to IL-2Rα alters the responsiveness of γδ-T-cells to IL-2 mediated apoptosis. Interestingly, in mice, IL-2 is an activator of CTCF *in vitro* and *in vivo* (Herblot et al. 1999). Normal activation-induced cell death could be a

response to CTCF rather than IL-2 signalling, and in the absence of CTCF γδ T-cells may be protected. It would be interesting to see if CTCF deficient γδ T-cells express IL-15Rα. The only cell surface marker for γδ T-cells is the TCRγδ itself. This suggests that *Tcr-γ* and -δ loci rearrange in a CTCF-independent manner.

γδ T-cells in humans and mice are important for anti-tumour immunosurveillance. Intradermal injections of chemical mutagens or tumour cells, or application of chemical carcinogens to the skin induce the development of tumours more rapidly in mice without γδ T-cells (Girardi et al. 2001). It is unknown whether CTCF deficient γδ T-cells can still mediate the progression of tumour development. In the future this would be an interesting experiment incorporating the functional capacity of CTCF deficient γδ T-cells and the tumour-suppressor function of CTCF.

The characteristic cytokines of γδ T-cells are Ifn-γ and Il-17. Il-17 is also the characteristic cytokine of the proinflammatory Th-17 cells (Laurence & O'Shea. 2007). Preliminary experiments suggest Il-17 can be expressed in γδ T-cells in the absence of CTCF, but this needs to be verified and checked with intracellular FACS. γδ T-cells were the first cells identified in this study to proliferate in a CTCF-independent manner. Understanding why γδ T-cells can proliferate in the absence of CTCF would be very important. Microarray analysis of γδ T-cells would be useful in terms of identifying CTCF target genes. However, analysis of the potential functional capacity of γδ T-cells that lack CTCF would be of most interest. Equally so is the question of whether Th-17 cells can be derived *in vitro* from CTCF deficient naïve $CD4^+$ cells.

CTCF and chromatin architecture

The chromatin structure of the *β-globin* locus is regulated by CTCF (Splinter et al. 2006, this thesis). CTCF binding sites flank the locus and interact by looping out the intervening chromatin. In erythroid precursor cells, a transcriptionally poised but inactive structure is established (chromatin hub) (Palstra et al. 2003). Differentiation into globin expressing erythroid cells requires the incorporation of the globin genes into the hub via interactions between the LCR and gene promoters. Significantly, we have shown that CTCF-mediated ACH formation at the *β-globin* locus is not necessary for transcriptional activity. This raises the question of why this loop is formed and what is the purpose of it?

ES cells carrying a targeted mutation of the 3'HS1 CTCF binding site could efficiently differentiate into erythroid progenitors (ES-EP), suggesting chromatin loops as defined and analysed in this study were not required for ensuring commitment to the erythroid lineage. Interestingly, GATA-1 expression was unaffected by the loss of CTCF in lenti-viral cre transfected fetal liver cells, indicating that GATA1 is not a target of CTCF. GATA1 in particular is necessary for erythroid differentiation from multi-potent

progenitor cells. The expression of GATA1 and other erythroid specific transcription factors such as EKLF were not analysed in ES-EPs. However deletion of 3'HS1 is known to have little effect on globin gene transcription (Strouboulis et al. 1992). Accordingly mutant ES-EPs did express globin and therefore must also express the required transcription factors. CTCF is not known to bind *β-globin* gene promoters, interact with or regulate GATA1, EKLF or other erythroid transcription factors. It would be of interest to analyse the capacity of CTCF deficient ES-cells to become ES-EPs and to measure the transcriptional activation of globin genes in conditionally deleted fetal liver cells. Importantly, LCR-gene contacts were retained in mutated ES-EPs indicating transcription factor recruitment is the main regulatory feature of this locus. This is consistent with the finding that EKLF was necessary for LCR-gene interactions (Drissen et al. 2004). How specific transcription factors are recruited to the *β-globin* locus is currently unclear. It was previously thought that establishment of chromatin loops would cluster regulatory regions close by in nuclear space, and that this clustering would attract transcription factor complexes. Our data suggest that the chromatin loops analysed in this study were not the determining factor. Additional flanking CTCF binding sites are predicted more external to –85 and 3'HS1 (Sleutels & vd Nobelen). It is possible that the main CTCF-mediated regulatory loop of the *β-globin* locus has not yet been identified. Verification of CTCF binding at these external sites followed by mutational analysis will be required to ascertain the significance of potentially super-β-globin loops.

Insulators are characterised by two separable functions; enhancer blocking and protection against chromatin position effects. The latter is CTCF-independent (Recillas-Targa et al. 2002). The *β-globin* locus is embedded in a region of olfactory receptor genes (*MOR*). CTCF binding sites flanking the *β-globin* locus are thought to function as insulator sequences, preventing activation of *MOR* genes by the *β-globin* LCR. The lack of expression of the surrounding *MOR* genes in CTCF mutant cells indicate either additional sequences are required for enhancer (LCR)-blocking or the transcription factor environment in erythroid cells does not support MOR expression.

The loss of CTCF binding from all *β-globin* target sites is accompanied by a local gain of di-meH3K9/27. It is not known if a DNA methylation gain also occurs. The change from H3 acetylation to repressive modifications is restricted to the CTCF binding site. No spreading of repressive chromatin modifications into the locus is observed. The association of CTCF with HAT complexes is not verified, and it is not clear why repressive histone modifications replace CTCF binding at this locus, as it has no impact on the transcriptional activity of *β-globin* genes.

GFP-CTCF knock-in mice; the dynamic behaviour of CTCF

Studies mapping CTCF binding sites acknowledge the omission of repetitive sequences from the analysis. Consequently current data regarding CTCF binding site location and number does not fully represent all CTCF interaction sites and functions. Most analyses also use fixed chromatin material and therefore cannot accurately address the concentration of CTCF protein in terms of molecule number present in a given nucleus or give any information regarding dynamics of the CTCF protein. The generation of mice in which GFP is incorporated in-frame 5' to the endogenous CTCF translation initiation site has circumvented many of the issues regarding fixed cells or chromatin. However this modification of CTCF strongly affects the survival of mice around birth. Increased incidences of wt mice are born, suggesting heterozygous GFP-CTCF mice are developmentally impaired. In contrast to the complete CTCF knockout, GFP-CTCF homozygous embryos develop until a relatively late stage in embryogenesis. It is currently unclear why homozygosity of the GFP-CTCF fusion protein is detrimental to post-natal development; a detailed pathology of mice just before and after birth would be of great interest.

The development of late-stage embryos permitted the isolation of MEFs for various analyses; proliferative capacity, the dynamic localisation of GFP-CTCF throughout the cell cycle and expression levels of GFP-CTCF. We show by western blot that the protein level of the GFP-modified allele is significantly lower than endogenous CTCF. Antibodies against CTCF were derived using the N-terminus domain. It is possible that the antibody used does not recognise GFP-CTCF as efficiently as a C-terminus antibody might. Preliminary analysis of N- and C-terminus CTCF antibodies detecting COS-1 cells transiently transfected with GFP-CTCF and CTCF-GFP did not suggest either antibody detected the GFP-tagged protein with more or less efficiency. Furthermore, independent data reveals CTCF functions in a dose-dependent manner in T-cells *in vivo*. The effects of reduced GFP-CTCF protein levels also manifest as impeding the rate of cellular proliferation, which is affected in MEFs in a GFP-CTCF dose-dependent manner. Together these data imply that a critical amount of CTCF is necessary for development, but also suggests that the loss of CTCF binding to some sites – as is inevitable with decreased amounts of protein in the nucleus – is enough to impair but not block proliferation. It will be of interest to detect which binding sites retain CTCF in the homozygous GFP-CTCF cells. ChIP revealed GFP-CTCF binds to CTCF target sites within the *β-globin* locus in fetal liver cells. The expression of *β-globin* genes is not required for cellular proliferation. It is currently unknown if GFP-CTCF can still bind CTCF target sites in proliferation dependent gene loci such as c-myc.

A large fraction (~70 %) of GFP-CTCF is immobile as determined by FRAP analysis in the interphase nucleus, apparently remaining bound either to chromatin, the nuclear matrix or both. This is a particularly unusual feature for a transcription factor, and probably reflects the structural role of CTCF regarding chromatin organisation, although a minority mobile fraction (~30 %) of GFP-CTCF does exist. The functional distinction between these two pools of GFP-CTCF remains to be elucidated. It may reflect modifications of the CTCF protein. The treatment of GFP-CTCF cells with HDAC inhibitors (TSA) or transcriptional inhibitors (DRB) will be useful in understanding the regulation of CTCF dynamics within the nucleus.

FRAP and FCS analyses of GFP-CTCF knock-in fibroblasts and ES-cells with GFP-CTCF rescue constructs both demonstrate similar dynamics and protein concentration within the nucleus. The concentration of GFP-CTCF as determined by FCS indicates there are approximately 200,000 CTCF molecules in a given nucleus. FCS software takes the presence of dimeric/multimeric complexes into account. When considering the dynamics of the protein, this equates to around 130,000 CTCF binding sites in the mouse genome. This is significantly more than current estimates (Vetchinova et al. 2006, Kim et al. 2007, Xie et al. 2007), but not necessarily an unrealistic figure considering there are ~22, 000 genes -averaging nearly 6 CTCF target sites per gene in the genome that may require regulation in one form or other by CTCF. Clear technical distinctions can be drawn between the current knowledge of numbers of CTCF binding sites and the data revealed in this thesis. The GFP-CTCF model is a direct *in vivo* analysis that is not susceptible to limiting conditions such as interpretation of arrays, antibodies or fixation techniques that may impede the detection of CTCF binding sites. Significantly, the GFP-CTCF model revealed for the first time the dynamics of the CTCF protein and the proportion of CTCF that is immobile compared to mobile. The fact that some CTCF molecules are not constitutively bound to chromatin further suggests the number of potential CTCF binding sites within the genome may be underestimated when using fixed cells. Current analyses of CTCF binding sites within the human genome have revealed ~ 14 000 CTCF-associated insulator sequences (Kim et al. 2007). The data presented in this thesis suggests more insulator sequences exist than current estimates. Alternatively, CTCF binds a substantial number of sites that have no insulator function.

An accumulation of CTCF at the nucleoli is observed by immunofluorescent analysis of CTCF in K562 cells (Torrano et al. 2005), however in primary GFP-CTCF MEFs or indirect immunofluorescence in primary cells, CTCF is absent from the nucleolus. This may highlight a difference between primary and transformed cell lines. Interestingly, CTCF is an interacting protein partner of two distinct nucleolar proteins, nucleophosmin and UBF (Yusufzai et al. 2004, S. vd Nobelen unpublished observations), although the functional significance of this is unknown and it remains to

be seen if CTCF mediates the shuttling of either of these factors between the nucleus and nucleoli. Live imaging of GFP-CTCF ES cells showed that CTCF is indeed localized on mitotic chromosomes, as previously reported (Burke et al. 2005). However, we did not detect GFP-CTCF on centrosomes or the midbody (Zhang et al. 2004); this may be a special localization of CTCF in transformed cell lines in contrast to immortal non-differentiated ES cells. The dynamic and specific localization of GFP-CTCF on mitotic chromosomes suggests CTCF is perhaps required for regulating the condensation of chromatin at mitosis, although mechanisms for this are unclear. Alternatively, CTCF interaction with mitotic chromosomes may function as an epigenetic mark, a mechanism of cellular memory. Interestingly, not all CTCF binding sites are occupied by CTCF during mitosis (Burke et al. 2005, Komura et al. 2007). It will be interesting to investigate if there is any correlation between the specific loci CTCF remains bound to. Importantly, if CTCF is required for mitosis, the localization of GFP-CTCF on mitotic chromosomes raises questions of how T-cells can proliferate in the absence of CTCF. It would be of interest to analyse the dynamic distribution of GFP-CTCF in T-cells to confirm the same mitotic distribution is observed. Results presented in this thesis show a notable proportion of GFP-CTCF protein does not bind to mitotic chromosomes, but is retained and divided between the two daughter cells. The significance of this 'free' CTCF remains to be elucidated.

We have used GFP-CTCF mice to analyse the expression pattern of CTCF during spermatogenesis by the live imaging of seminiferous tubules. Both male and female germ cells undergo the erasure and re-establishment of methylation marks. Loukinov et al proposed that during spermatogenesis, this stage coincided with a loss of CTCF expression and a concomitant up-regulation of the CTCF paralogue CTCF-L (BORIS) (Loukinov et al. 2002). Our live cell imaging does support this idea. We observe GFP-CTCF expression in all stages of spermatogenesis with the exception of mature spermatids, which have an altered compact chromatin structure consisting of protamines. The apparent conflict of data is most likely due to technicalities; indirect immunofluorescence of CTCF and BORIS was used to ascertain mutually exclusive patterns of expression in testis (Loukinov et al. 2002). GFP-CTCF imaging reflects the live *in vivo* dynamics of the protein and is not susceptible to fixation conditions or protein conformations. Importantly, studies by Szabo et al have eliminated the suggestion of a CTCF-BORIS reciprocal expression pattern in male and female germ cells to regulate methylation imprints (Szabo et al. 2004). Although global methylation in germ cells may still require BORIS, we suggest this does not happen independently of CTCF.

It is clear that studies presented in this thesis have revealed more surprises and raised more questions than first anticipated. The generation of mice carrying conditional *Ctcf* alleles has allowed the first *in vivo* analysis of CTCF and is a very powerful model

for investigating the functional capacity of cells and cellular systems in the absence of CTCF. The precise function of CTCF remains elusive, and I do not agree with a single function existence. CTCF continues to present as a multi-faceted protein and the consequences of deletion as a function of the cell type, specific locus and stage in development investigated. The generation of GFP-CTCF knock in mice has been a useful model to approach the idea of function from a different perspective, and has presented novel insights to both the activity of CTCF and the expression levels of CTCF required to support development and proliferation. Several key processes can still occur in the absence of CTCF. Hence the data generated in these studies and presented in this thesis have enhanced our understanding of the complexities behind gene regulation and cellular survival and the contribution of CTCF in these areas. The data presented in this thesis demonstrate that the original characterization of CTCF as a transcription factor is valid, as long as we conclude transcription factors also have other functions within the nucleus.

References

Abbas, A.K., Murphy, K.M., and Sher, A. 1996. *Nature* **383**(6603): 787-793.
Abkowitz, J.L., Golinelli, D., Harrison, D.E., and Guttorp, P. 2000. *Blood* **96**(10): 3399-3405.
Adams, J.M., Harris, A.W., Pinkert, C.A., Corcoran, L.M., Alexander, W.S., Cory, S., Palmiter, R.D., and Brinster, R.L. 1985. *Nature* **318**(6046): 533-538.
Affar el, B., Gay, F., Shi, Y., Liu, H., Huarte, M., Wu, S., Collins, T., Li, E., and Shi, Y. 2006. *Mol Cell Biol* **26**(9): 3565-3581.
Afkarian, M., Sedy, J.R., Yang, J., Jacobson, N.G., Cereb, N., Yang, S.Y., Murphy, T.L., and Murphy, K.M. 2002. *Nat Immunol* **3**(6): 549-557.
Agarwal, S. and Rao, A. 1998. *Immunity* **9**(6): 765-775.
Agnello, D., Lankford, C.S., Bream, J., Morinobu, A., Gadina, M., O'Shea, J.J., and Frucht, D.M. 2003. *J Clin Immunol* **23**(3): 147-161.
Akashi, K., Traver, D., Miyamoto, T., and Weissman, I.L. 2000. *Nature* **404**(6774): 193-197.
Akey, C.W. and Luger, K. 2003. *Curr Opin Struct Biol* **13**(1): 6-14.
Akhmanova, A., Miedema, K., and Hennig, W. 1996. *FEBS Lett* **388**(2-3): 219-222.
Allen, R.D III., Bender, T.P., Siu, G. 1999. *Genes & Development.* **13** 1073-1078.
Allfrey, V.G., Faulkner, R., and Mirsky, A.E. 1964. *Proc Natl Acad Sci U S A* **51**: 786-794.
Alvarez, J.D., Yasui, D.H., Niida, H., Joh, T., Loh, D.Y., and Kohwi-Shigematsu, T. 2000. *Genes Dev* **14**(5): 521-535.
Alvarez-Silva, M., Belo-Diabangouaya, P., Salaun, J., and Dieterlen-Lievre, F. 2003*Development* **130**(22): 5437-5444.
Ananiev, E.V. and Gvozdev, V.A. 1974. *Chromosoma* **45**(2): 173-191.
Anastassova-Kristeva, M. 1977. *J Cell Sci* **25**: 103-110.
Anderson, M.K., Weiss, A.H., Hernandez-Hoyos, G., Dionne, C.J., and Rothenberg, E.V. 2002. *Immunity* **16**(2): 285-296.
Anguita, E., Hughes, J., Heyworth, C., Blobel, G.A., Wood, W.G., and Higgs, D.R. 2004. *Embo J* **23**(14): 2841-2852.
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. *Nat Cell Biol* **7**(3): 303-310.
Arents, G., Burlingame, R.W., Wang, B.C., Love, W.E., and Moudrianakis, E.N. 1991. *Proc Natl Acad Sci U S A* **88**(22): 10148-10152.
Asnagli, H. and Murphy, K.M. 2001. *Curr Opin Immunol* **13**(2): 242-247.
Aulmann, S., Blaker, H., Penzel, R., Rieker, R.J., Otto, H.F., and Sinn, H.P. 2003. *Breast Cancer Res Treat* **80**(3): 347-352.
Awad, T.A., Bigler, J., Ulmer, J.E., Hu, Y.J., Moore, J.M., Lutz, M., Neiman, P.E., Collins, S.J., Renkawitz, R., Lobanenkov, V.V., and Filippova, G.N. 1999.. *J Biol Chem* **274**(38): 27092-27098.
Bacher, C.P., Guggiari, M., Brors, B., Augui, S., Clerc, P., Avner, P., Eils, R., and Heard, E. 2006. *Nat Cell Biol* **8**(3): 293-299.
Bacia, K., Kim, S.A., Schwille, P. (2006). *Nature Methods* **3** (2): 83-89.
Baguet, A. and Bix, M. 2004. *Proc Natl Acad Sci U S A* **101**(31): 11410-11415.
Bain, G., Robanus Maandag, E.C., te Riele, H.P., Feeney, A.J., Sheehy, A., Schlissel, M., Shinton, S.A., Hardy, R.R., and Murre, C. 1997. *Immunity* **6**(2): 145-154.
Bain, G., Romanow, W.J., Albers, K., Havran, W.L., and Murre, C. 1999. *J Exp Med* **189**(2): 289-300.
Baniahmad, A., Steiner, C., Kohne, A.C., and Renkawitz, R. 1990. *Cell* **61**(3): 505-514.
Bank, A. 2005. *J Clin Invest* **115**(6): 1470-1473.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. 2001. *Nature* **410**(6824): 120-124.
Barndt, R., Dai, M.F., and Zhuang, Y. 1999. *J Immunol* **163**(6): 3331-3343.
Barndt, R.J., Dai, M., and Zhuang, Y. 2000. *Mol Cell Biol* **20**(18): 6677-6685.
Becker, M., Baumann, C., John, S., Walker, D.A., Vigneron, M., McNally, J.G., and Hager, G.L. 2002. *EMBO Rep* **3**(12): 1188-1194.
Bell, A.C., West, A.G., and Felsenfeld, G. 1999. *Cell* **98**(3): 387-396.
Belmont, A.S., Braunfeld, M.B., Sedat, J.W., and Agard, D.A. 1989. *Chromosoma* **98**(2): 129-143.
Bergstrom, R., Whitehead, J., Kurukuti, S., and Ohlsson, R. 2007. *Cell Cycle* **6**(4): 450-454.
Berry, M., Grosveld, F., and Dillon, N. 1992. *Nature* **358**(6386): 499-502.
Bestor, T.H. 1992. *Embo J* **11**(7): 2611-2617.
Bhandoola, A., von Boehmer, H., Petrie, H.T., and Zuniga-Pflucker, J.C. 2007. *Immunity* **26**(6): 678-689.
Blackwood, E.M. and Kadonaga, J.T. 1998. *Science* **281**(5373): 60-63.
Blander, G. and Guarente, L. 2004. *Annu Rev Biochem* **73**: 417-435.
Blanton, J., Gaszner, M., and Schedl, P. 2003. *Genes Dev* **17**(5): 664-675.
Blower, M.D., Sullivan, B.A., and Karpen, G.H. 2002. *Dev Cell* **2**(3): 319-330.
Bode, J., Goetze, S., Heng, H., Krawetz, S.A., and Benham, C. 2003. *Chromosome Res* **11**(5): 435-445.
Bode, J., Kohwi, Y., Dickinson, L., Joh, T., Klehr, D., Mielke, C., and Kohwi-Shigematsu, T. 1992. *Science* **255**(5041): 195-197.
Boeger, H., Bushnell, D.A., Davis, R., Griesenbeck, J., Lorch, Y., Strattan, J.S., *FEBS Lett* **579**(4): 899-903.
Boeger, H., Griesenbeck, J., Strattan, J.S., and Kornberg, R.D. 2003. *Mol Cell* **11**(6): 1587-1598.
Boumil, R.M., Ogawa, Y., Sun, B.K., Huynh, K.D., and Lee, J.T. 2006. *Mol Cell Biol* **26**(6): 2109-2117.
Branco, M.R. and Pombo, A. 2006. *PLoS Biol* **4**(5): e138.
Brown, K.E., Amoils, S., Horn, J.M., Buckle, V.J., Higgs, D.R., Merkenschlager, M., and Fisher, A.G. 2001. *Nat Cell Biol* **3**(6): 602-606.
Brown, K.E., Guest, S.S., Smale, S.T., Hahm, K., Merkenschlager, M., and Fisher, A.G. 1997. *Cell* **91**(6): 845-854.
Bulger, M., Schubeler, D., Bender, M.A., Hamilton, J., Farrell, C.M., Hardison, R.C., and Groudine, M. 2003. *Mol Cell Biol* **23**(15): 5234-5244.
Burcin, M., Arnold, R., Lutz, M., Kaiser, B., Runge, D., Lottspeich, F., Filippova, G.N., Lobanenkov, V.V., and Renkawitz, R. 1997. *Mol Cell Biol* **17**(3): 1281-1288.
Burke, L.J., Hollemann, T., Pieler, T., and Renkawitz, R. 2002a. *Mech Dev* **113**(1): 95-98.
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. *Embo J* **24**(18): 3291-3300.
Burke, L.J., Zhang, R., Lutz, M., and Renkawitz, R. 2002b. *J Steroid Biochem Mol Biol* **83**(1-5): 49-57.
Butcher, D.T., Mancini-DiNardo, D.N., Archer, T.K., and Rodenhiser, D.I. 2004. *Int J Cancer* **111**(5): 669-678.
Cai, S., Han, H.J., and Kohwi-Shigematsu, T. 2003. *Nat Genet* **34**(1): 42-51.
Cai, S., Lee, C.C., and Kohwi-Shigematsu, T. 2006. *Nat Genet* **38**(11): 1278-1288.

Capone, M., Hockett, R.D., Jr., and Zlotnik, A. 1998. *Proc Natl Acad Sci U S A* **95**(21): 12522-12527.

Caron, H., van Schaik, B., van der Mee, M., Baas, F., Riggins, G., van Sluis, P., Hermus, M.C., van Asperen, R., Boon, K., Voute, P.A., Heisterkamp, S., van Kampen, A., and Versteeg, R. 2001. *Science* **291**(5507): 1289-1292.

Catez, F., Yang, H., Tracey, K.J., Reeves, R., Misteli, T., and Bustin, M. 2004. *Mol Cell Biol* **24**(10): 4321-4328.

Cerrato, F., Dean, W., Davies, K., Kagotani, K., Mitsuya, K., Okumura, K., Riccio, A., and Reik, W. 2003. *Hum Mol Genet* **12**(23): 3123-3132.

Cerwenka, A., Carter, L.L., Reome, J.B., Swain, S.L., and Dutton, R.W. 1998. *J Immunol* **161**(1): 97-105.

Chakalova, L., Osborne, C.S., Dai, Y.F., Goyenechea, B., Metaxotou-Mavromati, A., Kattamis, A., Kattamis, C., and Fraser, P. 2005. *Blood* **105**(5): 2154-2160.

Chao, W., Huynh, K.D., Spencer, R.J., Davidow, L.S., and Lee, J.T. 2002. *Science* **295**(5553): 345-347.

Chen, D., Dundr, M., Wang, C., Leung, A., Lamond, A., Misteli, T., and Huang, S. 2005. *J Cell Biol* **168**(1): 41-54.

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. *Mol Cell Biol* **27**(5): 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. *J Biol Chem* **275**(38): 29915-29921.

Chess, A., Simon, I., Cedar, H., and Axel, R. 1994. *Cell* **78**(5): 823-834.

Chi, T.H., Wan, M., Zhao, K., Taniuchi, I., Chen, L., Littman, D.R., and Crabtree, G.R. 2002. *Nature* **418**(6894): 195-199.

Chubb, J.R., Boyle, S., Perry, P., and Bickmore, W.A. 2002. *Curr Biol* **12**(6): 439-445.

Ciavatta, D., Kalantry, S., Magnuson, T., and Smithies, O. 2006. *Proc Natl Acad Sci U S A* **103**(26): 9958-9963.

Clayton, A.L., Hazzalin, C.A., Mahadevan, L.C. 2006. *Molecular Cell* **23** 289-296.

Cobb, B.S., Nesterova, T.B., Thompson, E., Hertweck, A., O'Connor, E., Godwin, J., Wilson, C.B., Brockdorff, N., Fisher, A.G., Smale, S.T., and Merkenschlager, M. 2005. *J Exp Med* **201**(9): 1367-1373.

Cohen, D.E. and Lee, J.T. 2002. *Curr Opin Genet Dev* **12**(2): 219-224.

Collins, F.S. and Weissman, S.M. 1984. *Prog Nucleic Acid Res Mol Biol* **31**: 315-462.

Conaway, J.W., Florens, L., Sato, S., Tomomori-Sato, C., Parmely, T.J., Yao, T., Swanson, S.K., Banks, C.A., Washburn, M.P., and Conaway, R.C. 2005. *FEBS Lett* **579**(4): 904-908.

Conaway, R.C. and Conaway, J.W. 1997. *Prog Nucleic Acid Res Mol Biol* **56**: 327-346.

Cosgrove, M.S., Boeke, J.D., and Wolberger, C. 2004. *Nat Struct Mol Biol* **11**(11): 1037-1043.

Costanzi, C. and Pehrson, J.R. 1998. *Nature* **393**(6685): 599-601.

Cote, J., Quinn, J., Workman, J.L., and Peterson, C.L. 1994. *Science* **265**(5168): 53-60.

Cote-Sierra, J., Foucras, G., Guo, L., Chiodetti, L., Young, H.A., Hu-Li, J., Zhu, J., and Paul, W.E. 2004. *Proc Natl Acad Sci U S A* **101**(11): 3880-3885.

Cremer, T. and Cremer, C. 2001. *Nat Rev Genet* **2**(4): 292-301.

Cremer, T., Cremer, M., Dietzel, S., Muller, S., Solovei, I., and Fakan, S. 2006. *Curr Opin Cell Biol* **18**(3): 307-316.

Cui, H., Niemitz, E.L., Ravenel, J.D., Onyango, P., Brandenburg, S.A., Lobanenkov, V.V., and Feinberg, A.P. 2001. *Cancer Res* **61**(13): 4947-4950.
Cunningham, J.M., Purucker, M.E., Jane, S.M., Safer, B., Vanin, E.F., Ney, P.A., Lowrey, C.H., and Nienhuis, A.W. 1994. *Blood* **84**(4): 1298-1308.
Dahmus, M.E. 1995. *Biochim Biophys Acta* **1261**(2): 171-182.
Dave, V.P., Allman, D., Keefe, R., Hardy, R.R., and Kappes, D.J. 1998. *Proc Natl Acad Sci U S A* **95**(14): 8187-8192.
de la Cruz, X., Lois, S., Sanchez-Molina, S., and Martinez-Balbas, M.A. 2005. *Bioessays* **27**(2): 164-175.
De La Rosa-Velazquez, I.A., Rincon-Arano, H., Benitez-Bribiesca, L., and Recillas-Targa, F. 2007. *Cancer Res* **67**(6): 2577-2585.
de Laat, W. and Grosveld, F. 2003. *Chromosome Res* **11**(5): 447-459.
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. *J Biol Chem* **280**(52): 43017-43023.
Dessypris, E.N. 1998. Erythropoeisis. In *Wintrobe's Clinical Hematology* (ed. J.F.G.R. Lee., Lukens, J., Rodgers, G.M., Grer, J.P., Paraskevas, F. 169-192. Williams and Wilkins.
Devlin, R.H., Holm, D.G., Morin, K.R., and Honda, B.M. 1990. *Genome* **33**(3): 405-415.
Dickinson, L.A., Joh, T., Kohwi, Y., and Kohwi-Shigematsu, T. 1992. *Cell* **70**(4): 631-645.
Dickinson, L.A. and Kohwi-Shigematsu, T. 1995. *Mol Cell Biol* **15**(1): 456-465.
Dietzel, S., Schiebel, K., Little, G., Edelmann, P., Rappold, G.A., Eils, R., Cremer, C., and Cremer, T. 1999. *Exp Cell Res* **252**(2): 363-375.
Dimitri, P., Junakovic, N., Arca, B. 2003. *Mol Biol Evol* **20** 503-512.
Donohoe, M.E., Zhang, L.F., Xu, N., Shi, Y., and Lee, J.T. 2007. *Mol Cell* **25**(1): 43-56.
Drissen, R., Palstra, R.J., Gillemans, N., Splinter, E., Grosveld, F., Philipsen, S., and de Laat, W. 2004. *Genes Dev* **18**(20): 2485-2490.
Drueppel, L., Pfleiderer, K., Schmidt, A., Hillen, W., and Berens, C. 2004. *FEBS Lett* **572**(1-3): 154-158.
Dunn, K.L., Zhao, H., and Davie, J.R. 2003. *Exp Cell Res* **288**(1): 218-223.
Earnshaw, W.C. 1988*Bioessays* **9**(5): 147-150.
Ehrenhofer-Murray, A.E. 2004. *Eur J Biochem* **271**(12): 2335-2349.
Ellmeier, W., Sawada, S., and Littman, D.R. 1999. *Annu Rev Immunol* **17**: 523-554.
Engel, I., Johns, C., Bain, G., Rivera, R.R., and Murre, C. 2001. *J Exp Med* **194**(6): 733-745.
Fan, J.Y., Gordon, F., Luger, K., Hansen, J.C., and Tremethick, D.J. 2002. *Nat Struct Biol* **9**(3): 172-176.
Fan, X., Shi, H., and Lis, J.T. 2005a. *Nucleic Acids Res* **33**(3): 838-845.
Fan, Y., Nikitina, T., Zhao, J., Fleury, T.J., Bhattacharyya, R., Bouhassira, E.E., Stein, A., Woodcock, C.L., and Skoultchi, A.I. 2005b. *Cell* **123**(7): 1199-1212.
Farrell, C.M., Grinberg, A., Huang, S.P., Chen, D., Pichel, J.G., Westphal, H., and Felsenfeld, G. 2000. *Proc Natl Acad Sci U S A* **97**(26): 14554-14559.
Farrell, C.M., West, A.G., and Felsenfeld, G. 2002. *Mol Cell Biol* **22**(11): 3820-3831.
Fearon, D.T. and Locksley, R.M. 1996*Science* **272**(5258): 50-53.
Fedoriw, A.M., Stein, P., Svoboda, P., Schultz, R.M., and Bartolomei, M.S. 2004. *Science* **303**(5655): 238-240.
Felsenfeld, G. and Groudine, M. 2003. *Nature* **421**(6921): 448-453.
Feng, D. and Kan, Y.W. 2005. *Proc Natl Acad Sci U S A* **102**(28): 9896-9900.

Ferrick, D.A., Schrenzel, M.D., Mulvania, T., Hsieh, B., Ferlin, W.G., and Lepper, H. 1995. *Nature* **373**(6511): 255-257.
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. *Dev Cell* **8**(1): 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. *Mol Cell Biol* **16**(6): 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. *Genes Chromosomes Cancer* **22**(1): 26-36.
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. *Nat Genet* **28**(4): 335-343.
Fillingham, J., Keogh, M.C., and Krogan, N.J. 2006. *Biochem Cell Biol* **84**(4): 568-577.
Fitzpatrick, D.R. and Wilson, C.B. 2003. *Clin Immunol* **109**(1): 37-45.
Fraga, M.F., Ballestar, E., Villar-Garea, A., Boix-Chornet, M., Espada, J., Schotta, G., Bonaldi, T., Haydon, C., Ropero, S., Petrie, K., Iyer, N.G., Perez-Rosado, A., Calvo, E., Lopez, J.A., Cano, A., Calasanz, M.J., Colomer, D., Piris, M.A., Ahn, N., Imhof, A., Caldas, C., Jenuwein, T., and Esteller, M. 2005. *Nat Genet* **37**(4): 391-400.
Gebuhr, T.C., Kovalev, G.I., Bultman, S., Godfrey, V., Su, L., and Magnuson, T. 2003. *J Exp Med* **198**(12): 1937-1949.
Geiman, T.M. and Robertson, K.D. 2002. *J Cell Biochem* **87**(2): 117-125.
Georgopoulos, K., Bigby, M., Wang, J.H., Molnar, A., Wu, P., Winandy, S., and Sharpe, A. 1994. *Cell* **79**(1): 143-156.
Germain, R.N. 1994. *Cell* **76**(2): 287-299.
Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N.P., and Bickmore, W.A. 2004. *Cell* **118**(5): 555-566.
Girardi, M. 2006. *J Invest Dermatol* **126**(1): 25-31.
Girardi, M., Oppenheim, D.E., Steele, C.R., Lewis, J.M., Glusac, E., Filler, R., Hobby, P., Sutton, B., Tigelaar, R.E., and Hayday, A.C. 2001. *Science* **294**(5542): 605-609.
Goldmit, M. and Bergman, Y. 2004. *Immunol Rev* **200**: 197-214.
Gombert, W.M., Farris, S.D., Rubio, E.D., Morey-Rosler, K.M., Schubach, W.H., and Krumm, A. 2003. *Mol Cell Biol* **23**(24): 9338-9348.
Gray, C.E. and Coates, C.J. 2005. *BMC Mol Biol* **6**(1): 16.
Grosveld, F., van Assendelft, G.B., Greaves, D.R., and Kollias, G. 1987. *Cell* **51**(6): 975-985.
Grunstein, M. 1997. *Nature* **389**(6649): 349-352.
Guo, B.L., Hollmig, K.A., and Lopez, R.D. 2002. *Cancer Immunol Immunother* **50**(11): 625-637.
Hancock, A.L., Brown, K.W., Moorwood, K., Moon, H., Holmgren, C., Mardikar, S.H., Dallosso, A.R., Klenova, E., Loukinov, D., Ohlsson, R., Lobanenkov, V.V., and Malik, K. 2007. *Hum Mol Genet* **16**(3): 343-354.
Hansen, J.C. 2002. *Annu Rev Biophys Biomol Struct* **31**: 361-392.
Hardison, R.C. 1996. *Proc Natl Acad Sci U S A* **93**(12): 5675-5679.
Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M., and Tilghman, S.M. 2000. *Nature* **405**(6785): 486-489.
Harman, B.C., Jenkinson, W.E., Parnell, S.M., Rossi, S.W., Jenkinson, E.J., and Anderson, G. 2005. *Blood* **106**(3): 886-892.

Harrington, L.E., Hatton, R.D., Mangan, P.R., Turner, H., Murphy, T.L., Murphy, K.M., and Weaver, C.T. 2005. *Nat Immunol* **6**(11): 1123-1132.
Hashimshony, T., Zhang, J., Keshet, I., Bustin, M., and Cedar, H. 2003. *Nat Genet* **34**(2): 187-192.
Havran, W.L. and Boismenu, R. 1994. *Curr Opin Immunol* **6**(3): 442-446.
Hawwari, A. and Krangel, M.S. 2005. *J Exp Med* **202**(4): 467-472.
Hayes, J.J. 2002. Changing chromatin from the inside. *Nat Struct Biol* **9**(3): 161-163.
Heath, V.L., Showe, L., Crain, C., Barrat, F.J., Trinchieri, G., and O'Garra, A. 2000. *J Immunol* **164**(6): 2861-2865.
Heikkila, R., Schwab, G., Wickstrom, E., Loke, S.L., Pluznik, D.H., Watt, R., and Neckers, L.M. 1987. *Nature* **328**(6129): 445-449.
Heng, H.H., Krawetz, S.A., Lu, W., Bremer, S., Liu, G., and Ye, C.J. 2001. *Cytogenet Cell Genet* **93**(3-4): 155-161.
Herblot, S., Chastagner, P., Samady, L., Moreau, J.L., Demaison, C., Froussard, P., Liu, X., Bonnet, J., and Theze, J. 1999. *J Immunol* **162**(6): 3280-3288.
Hernandez-Hoyos, G., Sohn, S.J., Rothenberg, E.V., and Alberola-Ila, J. 2000. *Immunity* **12**(3): 313-322.
Herrscher, R.F., Kaplan, M.H., Lelsz, D.L., Das, C., Scheuermann, R., and Tucker, P.W. 1995. *Genes Dev* **9**(24): 3067-3082.
Hewitt, S.L., High, F.A., Reiner, S.L., Fisher, A.G., and Merkenschlager, M. 2004. *Eur J Immunol* **34**(12): 3604-3613.
Hikichi, T., Kohda, T., Kaneko-Ishino, T., and Ishino, F. 2003. *Nucleic Acids Res* **31**(5): 1398-1406.
Hino, S., Akasaka, K., and Matsuoka, M. 2006. *J Mol Biol* **357**(1): 18-27.
Holohan, E.E., Kwong, C., Adryan, B., Bartkuhn, M., Herold, M., Renkawitz, R., Russell, S., and White, R. 2007. *PLoS Genet* **3**(7): e112.
Holmquist, G.P & Ashley, T. 2006. *Cytogenet. Genome Research.* **114** 96-125.
Holtmeier, W. and Kabelitz, D. 2005. *Chem Immunol Allergy* **86**: 151-183.
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. *Cancer Res* **65**(17): 7763-7774.
Houtsmuller, A.B. and Vermeulen, W. 2001. *Histochem Cell Biol* **115**(1): 13-21.
Howlett, S.K. and Reik, W. 1991. *Development* **113**(1): 119-127.
Huber, T.L., Kouskoff, V., Fehling, H.J., Palis, J., and Keller, G. 2004. *Nature* **432**(7017): 625-630.
Hurst, L.D. and McVean, G.T. 1998. *Curr Opin Genet Dev* **8**(6): 701-708.
Hutchins, A.S., Mullen, A.C., Lee, H.W., Sykes, K.J., High, F.A., Hendrich, B.D., Bird, A.P., and Reiner, S.L. 2002. *Mol Cell* **10**(1): 81-91.
Iavarone, A., King, E.R., Dai, X.M., Leone, G., Stanley, E.R., and Lasorella, A. 2004. *Nature* **432**(7020): 1040-1045.
Iben, S., Tschochner, H., Bier, M., Hoogstraten, D., Hozak, P., Egly, J.M., and Grummt, I. 2002. *Cell* **109**(3): 297-306.
Iizuka, M & Smith, M.M. 2003. *Nature* **178**(4537): 792-794.
Irving, B.A., Alt, F.W., and Killeen, N. 1998. *Science* **280**(5365): 905-908.
Ishihara, K., Oshimura, M., and Nakao, M. 2006. *Mol Cell* **23**(5): 733-742.
Iwasaki, H., Mizuno, S., Wells, R.A., Cantor, A.B., Watanabe, S., and Akashi, K. 2003. *Immunity* **19**(3): 451-462.
Jackson, A.M. and Krangel, M.S. 2006. *Immunol Rev* **209**: 129-141.

Jenuwein, T & Allis, C.D. 2001. *Science* **293** 1074-1080
Kamakaka, R.T. and Biggins, S. 2005. *Genes Dev* **19**(3): 295-310.
Kang, J., Volkmann, A., and Raulet, D.H. 2001. *J Exp Med* **193**(6): 689-698.
Kantidze, O.L., Iarovaia, O.V., Philonenko, E.S., Yakutenko, II, and Razin, S.V. 2007. *Biochim Biophys Acta* **1773**(6): 924-933.
Kappler, J.W., Staerz, U., White, J., and Marrack, P.C. 1988. *Nature* **332**(6159): 35-40.
Katsura, Y. 2002. *Nat Rev Immunol* **2**(2): 127-132.
Kaufmann, S.H. 1996. *Proc Natl Acad Sci U S A* **93**(6): 2272-2279.
Kaul, D.K., Fabry, M.E., Nagel, R.L. 1996. *Blood Rev.***10** 29-44.
Kawamoto, H., Ohmura, K., and Katsura, Y. 1997. *Int Immunol* **9**(7): 1011-1019.
Kee, B.L. and Murre, C. 1998. *J Exp Med* **188**(4): 699-713.
Khorasanizadeh, S. (2004). *Cell* **116** 259-272.
Kim, J., Kollhoff, A., Bergmann, A., and Stubbs, L. 2003. *Hum Mol Genet* **12**(3): 233-245.
Kim, M., Li, D., Cui, Y., Mueller, K., Chears, W.C., and DeJong, J. 2006. *J Biol Chem* **281**(45): 34288-34298.
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. *Cell* **128**(6): 1231-1245.
Kimura, H. and Cook, P.R. 2001. *J Cell Biol* **153**(7): 1341-1353.
Kingsley, P.D., Malik, J., Emerson, R.L., Bushnell, T.P., McGrath, K.E., Bloedorn, L.A., Bulger, M., and Palis, J. 2006. *Blood* **107**(4): 1665-1672.
Kingsley, P.D., Malik, J., Fantauzzo, K.A., and Palis, J. 2004. *Blood* **104**(1): 19-25.
Kioussis, D., Vanin, E., deLange, T., Flavell, R.A., and Grosveld, F.G. 1983. *Nature* **306**(5944): 662-666.
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. *J Neurosci* **24**(26): 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. *Mol Cell Biol* **21**(6): 2221-2234.
Klenova, E.M., Morse, H.C., 3rd, Ohlsson, R., and Lobanenkov, V.V. 2002. *Semin Cancer Biol* **12**(5): 399-414.
Klenova, E.M., Nicolas, R.H., U, S., Carne, A.F., Lee, R.E., Lobanenkov, V.V., and Goodwin, G.H. 1997. *Nucleic Acids Res* **25**(3): 466-474.
Koesters, C., Unger, B., Bilic, I., Schmidt, U., Bluml, S., Lichtenberger, B., Schreiber, M., Stockl, J., and Ellmeier, W. 2007. *Immunol Lett* **109**(2): 165-174.
Kondo, M., Weissman, I.L., and Akashi, K. 1997. *Cell* **91**(5): 661-672.
Kornberg, R.D. and Lorch, Y. 1999. *Cell* **98**(3): 285-294.
Kosaka, Y., Felices, M., and Berg, L.J. 2006. *Trends Immunol* **27**(10): 453-460.
Kouzarides, T. 2007. *Cell* **128**(4): 693-705.
Krangel, M.S. 2007. *Nat Immunol* **8**(7): 687-694.
Kraus, W.L. and Lis, J.T. 2003. *Cell* **113**(6): 677-683.
Kurukuti, S., Tiwari, V.K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., Lobanenkov, V., Reik, W., and Ohlsson, R. 2006. *Proc Natl Acad Sci U S A* **103**(28): 10684-10689.
Kuzmin, I., Geil, L., Gibson, L., Cavinato, T., Loukinov, D., Lobanenkov, V., and Lerman, M.I. 2005. *J Mol Biol* **346**(2): 411-422.
Lamb, L.S., Jr. and Lopez, R.D. 2005. *Biol Blood Marrow Transplant* **11**(3): 161-168.
Langst, G. and Becker, P.B. 2001. *J Cell Sci* **114**(Pt 14): 2561-2568.
Laurence, A. and O'Shea J, J. 2007. *Nat Immunol* **8**(9): 903-905.

Layon, M.E., Ackley, C.J., West, R.J., and Lowrey, C.H. 2007. *J Mol Biol* **366**(3): 737-744.
Lee, D.U., Agarwal, S., and Rao, A. 2002. *Immunity* **16**(5): 649-660.
Lee, G.R., Spilianakis, C.G., and Flavell, R.A. 2005. *Nat Immunol* **6**(1): 42-48.
Lee, M.G. and Van der Ploeg, L.H. 1997. *Annu Rev Microbiol* **51**: 463-489.
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. *Immunity* **15**(5): 763-774.
Leighton, P.A., Ingram, R.S., Eggenschwiler, J., Efstratiadis, A., and Tilghman, S.M. 1995. *Nature* **375**(6526): 34-39.
Lemischka, I.R. 1992. *Cancer Surv* **15**: 3-18.
Lensch, M.W. and Daley, G.Q. 2004. *Curr Top Dev Biol* **60**: 127-196.
Lenschow, D.J., Walunas, T.L., and Bluestone, J.A. 1996. *Annu Rev Immunol* **14**: 233-258.
Li, B., Carey, M., and Workman, J.L. 2007. *Cell* **128**(4): 707-719.
Li, E. 2002. *Nat Rev Genet* **3**(9): 662-673.
Li, E., Beard, C., and Jaenisch, R. 1993. *Nature* **366**(6453): 362-365.
Li, G., Sudlow, G., and Belmont, A.S. 1998. *J Cell Biol* **140**(5): 975-989.
Li, Q., Peterson, K.R., Fang, X., and Stamatoyannopoulos, G. 2002. *Blood* **100**(9): 3077-3086.
Li, T. and Lu, L. 2007. *Exp Cell Res* **313**(14): 3057-3065.
Lighvani, A.A., Frucht, D.M., Jankovic, D., Yamane, H., Aliberti, J., Hissong, B.D., Nguyen, B.V., Gadina, M., Sher, A., Paul, W.E., and O'Shea, J.J. 2001. *Proc Natl Acad Sci U S A* **98**(26): 15137-15142.
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. *Science* **312**(5771): 269-272.
Litt, M.D., Simpson, M., Recillas-Targa, F., Prioleau, M.N., and Felsenfeld, G. 2001. *Embo J* **20**(9): 2224-2235.
Lobanenkov, V.V., Nicolas, R.H., Adler, V.V., Paterson, H., Klenova, E.M., Polotskaja, A.V., and Goodwin, G.H. 1990. *Oncogene* **5**(12): 1743-1753.
Lopez, R.A., Schoetz, S., DeAngelis, K., O'Neill, D., and Bank, A. 2002. *Proc Natl Acad Sci U S A* **99**(2): 602-607.
Lornvardas, S., Barnea, G., Pisapia, D.J., Mendelsohn, M., Kirkland, J., Axel, R. (2006). *Cell* **126** 403-413.
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. *Proc Natl Acad Sci U S A* **99**(10): 6806-6811.
Loyola, A. and Almouzni, G. 2004. *Biochim Biophys Acta* **1677**(1-3): 3-11.
Luderus, M.E., de Graaf, A., Mattia, E., den Blaauwen, J.L., Grande, M.A., de Jong, L., and van Driel, R. 1992. *Cell* **70**(6): 949-959.
Luger, K. 2003. *Curr Opin Genet Dev* **13**(2): 127-135.
Luger, K. 2006. *Chromosome Res* **14**(1): 5-16.
Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. 1997. *Nature* **389**(6648): 251-260.
Luger, K. and Richmond, T.J. 1998. *Curr Opin Struct Biol* **8**(1): 33-40.

Lutz, M., Burke, L.J., Barreto, G., Goeman, F., Greb, H., Arnold, R., Schultheiss, H., Brehm, A., Kouzarides, T., Lobanenkov, V., and Renkawitz, R. 2000. *Nucleic Acids Res* **28**(8): 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. *Embo J* **22**(7): 1579-1587.

Lyon, M.F. 1961. *Nature* **190**: 372-373.

Macara, I.G. 2001. *Microbiol Mol Biol Rev* **65**(4): 570-594, table of contents.

Magdinier, F., Yusufzai, T.M., and Felsenfeld, G. 2004. *J Biol Chem* **279**(24): 25381-25389.

Makar, K.W., Perez-Melgosa, M., Shnyreva, M., Weaver, W.M., Fitzpatrick, D.R., and Wilson, C.B. 2003. *Nat Immunol* **4**(12): 1183-1190.

Manley, N.R. and Blackburn, C.C. 2003. *Curr Opin Immunol* **15**(2): 225-232.

Martinez, E. 2002. *Plant Mol Biol* **50**(6): 925-947.

Mathieu, N., Hempel, W.M., Spicuglia, S., Verthuy, C., and Ferrier, P. 2000. *J Exp Med* **192**(5): 625-636.

Maurice, D., Hooper, J., Lang, G., and Weston, K. 2007. *Embo J* **26**(15): 3629-3640.

Maya-Mendoza, A., Hernandez-Munoz, R., Gariglio, P., and Aranda-Anzaldo, A. 2003. *Nucleic Acids Res* **31**(21): 6168-6179.

McGhee, J.D. and Felsenfeld, G. 1980. *Annu Rev Biochem* **49**: 1115-1156.

McGrath, K.E., Koniski, A.D., Malik, J., and Palis, J. 2003. *Blood* **101**(5): 1669-1676.

McKittrick, E., Gafken, P.R., Ahmad, K., and Henikoff, S. 2004. *Proc Natl Acad Sci U S A* **101**(6): 1525-1530.

McMurry, M.T. and Krangel, M.S. 2000. *Science* **287**(5452): 495-498.

Medvinsky, A. and Dzierzak, E. 1996. *Cell* **86**(6): 897-906.

Miosge, L.A. and Goodnow, C.C. 2005. *Immunol Cell Biol* **83**(4): 318-335.

Moon, A.M. and Ley, T.J. 1990. *Proc Natl Acad Sci U S A* **87**(19): 7693-7697.

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. *EMBO Rep* **6**(2): 165-170.

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. *Genome Res* **14**(8): 1594-1602.

Murre, C. 2005. *Nat Immunol* **6**(11): 1079-1086.

Murrell, A., Heeson, S., and Reik, W. 2004. *Nat Genet* **36**(8): 889-893.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. 1998. *Nature* **393**(6683): 386-389.

Narlikar, G.J., Fan, H.Y., and Kingston, R.E. 2002. *Cell* **108**(4): 475-487.

Navarro, P., Page, D.R., Avner, P., and Rougeulle, C. 2006. *Genes Dev* **20**(20): 2787-2792.

Nawijn, M.C., Ferreira, R., Dingjan, G.M., Kahre, O., Drabek, D., Karis, A., Grosveld, F., and Hendriks, R.W. 2001. *J Immunol* **167**(2): 715-723.

Nickerson, J. 2001. *J Cell Sci* **114**(Pt 3): 463-474.

Noll, M. and Kornberg, R.D. 1977. *J Mol Biol* **109**(3): 393-404.

Norman, M.U. and Hickey, M.J. 2005. *Tissue Antigens* **66**(3): 163-172.

O'Garra, A. and Arai, N. 2000. *Trends Cell Biol* **10**(12): 542-550.

O'Neill, D., Yang, J., Erdjument-Bromage, H., Bornschlegel, K., Tempst, P., and Bank, A. 1999. *Proc Natl Acad Sci U S A* **96**(2): 349-354.

Erratum.

p.21: Superfluous bracket in reference (Howlett and Reik 1991, Mayer, 2000).

p.32: Missing full stop (Recillas-Targa et al. 2002).

p.33: Missing abbreviations, Replication protein A (RPA), Runt-related transcription factors (Runx).

p.38: In chicken, the gene coding for CTCF consists of 10 exons that generate a 728 amino acid protein. Murine CTCF has 12 coding exons that encode a 736 amino acid protein.

p.54: Gap between bracket and reference (Kantidze et al. 2007, Torrano et al. 2006).

p60: Superfluous bracket in reference (McGrath et al. 2003, Kingsley et al. 2004)

p61: Superfluous reference number in figure legend 3.2 (Dessypris 1998).

p69: Missing information in figure 3.4 (5' HS-85).

p72: Superfluous bracket in reference (Harman et al. 2005, Masuda 2005).

p74: Missing information in figure 3.5 (pTα).

p81: Missing information in figure 3.6 (Th-17, iT-reg).

Chapter 4: Figure 2A/C, 3A/B, 4B/C, supplemental fig.4, figure legends (Δ3'HS1). Figure 2B (Δ and c). Supplemental figure 7 (ΔCt).

p.191: Our live cell imaging does not support this idea.

Oestreich, K.J., Cobb, R.M., Pierce, S., Chen, J., Ferrier, P., and Oltz, E.M. 2006. *Immunity* **24**(4): 381-391.
Oganesian, L. and Bryan, T.M. 2007. *Bioessays* **29**(2): 155-165.
Ohhata, T., Tachibana, M., Tada, M., Tada, T., Sasaki, H., Shinkai, Y., and Sado, T. 2004. *Genesis* **40**(3): 151-156.
Ohlsson, R., Renkawitz, R., and Lobanenkov, V. 2001. *Trends Genet* **17**(9): 520-527.
Okano, M., Bell, D.W., Haber, D.A., and Li, E. 1999. *Cell* **99**(3): 247-257.
Otten, A.D. and Tapscott, S.J. 1995. *Proc Natl Acad Sci U S A* **92**(12): 5465-5469.
Oudet, P., Gross-Bellard, M., and Chambon, P. 1975. *Cell* **4**(4): 281-300.
Ouyang, W., Ranganath, S.H., Weindel, K., Bhattacharya, D., Murphy, T.L., Sha, W.C., and Murphy, K.M. 1998. *Immunity* **9**(5): 745-755.
Pai, S.Y et al. 2003. *Immunity* **19** 863-875.
Palis, J., Robertson, S., Kennedy, M., Wall, C., and Keller, G. 1999. *Development* **126**(22): 5073-5084.
Palstra, R.J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., and de Laat, W. 2003. *Nat Genet* **35**(2): 190-194.
Parada, L. and Misteli, T. 2002. *Trends Cell Biol* **12**(9): 425-432.
Park, H., Li, Z., Yang, X., Chang, S.H., Nurieva, R., Wang, Y-H., Wang, Y., Hood, L., Zhu, Z., Tian, Q., Dong, C. 2005. *Nature Immunology*. **6** (11): 1133-1141.
Parseghian, M.H. and Luhrs, K.A. 2006. *Biochem Cell Biol* **84**(4): 589-604.
Patrinos, G.P., de Krom, M., de Boer, E., Langeveld, A., Imam, A.M., Strouboulis, J., de Laat, W., and Grosveld, F.G. 2004. *Genes Dev* **18**(12): 1495-1509.
Paule, M.R. and White, R.J. 2000. *Nucleic Acids Res* **28**(6): 1283-1298.
Pauler, F.M. and Barlow, D.P. 2006. *Genes Dev* **20**(10): 1203-1206.
Peters, A.H., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., Martens, J.H., and Jenuwein, T. 2003. *Mol Cell* **12**(6): 1577-1589.
Petersson, K., Ivars, F., Sigvardsson, M. 2002. *Eur. J. Immunol.* **32** 911-920.
Phelan, M.L., Sif, S., Narlikar, G.J., and Kingston, R.E. 1999. *Mol Cell* **3**(2): 247-253.
Philipp-Staheli, J., Kim, K.H., Liggitt, D., Gurley, K.E., Longton, G., and Kemp, C.J. 2004. *Oncogene* **23**(4): 905-913.
Pikaart, M.J., Recillas-Targa, F., and Felsenfeld, G. 1998. *Genes Dev* **12**(18): 2852-2862.
Pina, B., Truss, M., Ohlenbusch, H., Postma, J., and Beato, M. 1990. *Nucleic Acids Res* **18**(23): 6981-6987.
Polach, K.J., Lowary, P.T., Widom, J. 2000. *J. Mol. Biol.* **298** 211-223.
Prohaska, S.S., Scherer, D.C., Weissman, I.L., and Kondo, M. 2002. *Semin Immunol* **14**(6): 377-384.
Ptashne, M. 1986. *Nature* **322**(6081): 697-701.
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. *Gene* **375**: 26-36.
Pugacheva, E.M., Tiwari, V.K., Abdullaev, Z., Vostrov, A.A., Flanagan, P.T., Quitschke, W.W., Loukinov, D.I., Ohlsson, R., and Lobanenkov, V.V. 2005. *Hum Mol Genet* **14**(7): 953-965.
Qi, C.F., Martensson, A., Mattioli, M., Dalla-Favera, R., Lobanenkov, V.V., and Morse, H.C., 3rd. 2003. *Proc Natl Acad Sci U S A* **100**(2): 633-638.
Quina, A.S., Buschbeck, M., and Di Croce, L. 2006. *Biochem Pharmacol* **72**(11): 1563-1569.

Ragoczy, T., Telling, A., Sawado, T., Groudine, M., and Kosak, S.T. 2003. *Chromosome Res* **11**(5): 513-525.
Raisner, R.M. and Madhani, H.D. 2006. *Curr Opin Genet Dev* **16**(2): 119-124.
Rakha, E.A., Pinder, S.E., Paish, C.E., and Ellis, I.O. 2004. *Br J Cancer* **91**(8): 1591-1596.
Rao, A. and Avni, O. 2000. *Br Med Bull* **56**(4): 969-984.
Rao, S., Gerondakis, S., Woltring, D., and Shannon, M.F. 2003. *J Immunol* **170**(7): 3724-3731.
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. *Cancer Res* **61**(16): 6002-6007.
Razin, S.V., Iarovaia, O.V., Sjakste, N., Sjakste, T., Bagdoniene, L., Rynditch, A.V., Eivazova, E.R., Lipinski, M., and Vassetzky, Y.S. 2007. *J Mol Biol* **369**(3): 597-607.
Recillas-Targa, F., De La Rosa-Velazquez, I.A., Soto-Reyes, E., and Benitez-Bribiesca, L. 2006. *J Cell Mol Med* **10**(3): 554-568.
Recillas-Targa, F., Pikaart, M.J., Burgess-Beusse, B., Bell, A.C., Litt, M.D., West, A.G., Gaszner, M., and Felsenfeld, G. 2002. *Proc Natl Acad Sci U S A* **99**(10): 6883-6888.
Redon, C., Pilch, D., Rogakou, E., Sedelnikova, O., Newrock, K., and Bonner, W. 2002. *Curr Opin Genet Dev* **12**(2): 162-169.
Reese, J.C. 2003. *Curr Opin Genet Dev* **13**(2): 114-118.
Renaud, S., Loukinov, D., Bosman, F.T., Lobanenkov, V., and Benhattar, J. 2005. *Nucleic Acids Res* **33**(21): 6850-6860.
Richards, E.J. and Elgin, S.C. 2002. *Cell* **108**(4): 489-500.
Richmond, T.D., Chohan, M., and Barber, D.L. 2005. *Trends Cell Biol* **15**(3): 146-155.
Roberts, J.L., Lauzurica, P., and Krangel, M.S. 1997. *J Exp Med* **185**(1): 131-140.
Rocha, B., Guy-Grand, D., and Vassalli, P. 1995. *Curr Opin Immunol* **7**(2): 235-242.
Rochette-Egly, C., Adam, S., Rossignol, M., Egly, J.M., and Chambon, P. 1997. *Cell* **90**(1): 97-107.
Rock, E.P., Sibbald, P.R., Davis, M.M., Chien, Y.H. 1994. *J Exp. Med.* **179** 323-328.
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. *Embo J* **24**(13): 2354-2366.
Roh, T-Y., Cuddapah, S., Zhao, K. 2005. *Genes and Development* **19** 542-552.
Rooney, J.W., Hoey, T., and Glimcher, L.H. 1995. *Immunity* **2**(5): 473-483.
Rothenberg, E.V. 2007. *Curr Opin Hematol* **14**(4): 322-329.
Rudolph, B., Hueber, A.O., and Evan, G.I. 2000. *Oncogene* **19**(15): 1891-1900.
Ryu, C.J., Haines, B.B., Lee, H.R., Kang, Y.H., Draganov, D.D., Lee, M., Whitehurst, C.E., Hong, H.J., and Chen, J. 2004. *Mol Cell Biol* **24**(16): 7015-7023.
Saint-Ruf, C., Panigada, M., Azogui, O., Debey, P., von Boehmer, H., and Grassi, F. 2000. *Nature* **406**(6795): 524-527.
Sambandam, A., Maillard, I., Zediak, V.P., Xu, L., Gerstein, R.M., Aster, J.C., Pear, W.S., and Bhandoola, A. 2005. *Nat Immunol* **6**(7): 663-670.
Sanders, V.M. 2006. *Brain Behav Immun* **20**(4): 317-324.
Santos-Rosa, H. and Caldas, C. 2005. *Eur J Cancer* **41**(16): 2381-2402.
Sato, T., Ohno, S., Hayashi, T., Sato, C., Kohu, K., Satake, M., and Habu, S. 2005. *Immunity* **22**(3): 317-328.
Sawada, S., Scarborough, J.D., Killeen, N., and Littman, D.R. 1994. *Cell* **77**(6): 917-929.

Schubeler, D., Francastel, C., Cimbora, D.M., Reik, A., Martin, D.I., and Groudine, M. 2000. *Genes Dev* **14**(8): 940-950.
Schubeler, D., Groudine, M., and Bender, M.A. 2001. *Proc Natl Acad Sci U S A* **98**(20): 11432-11437.
Schwarz, P.M. and Hansen, J.C. 1994. *J Biol Chem* **269**(23): 16284-16289.
Sebzda, E., Mariathasan, S., Ohteki, T., Jones, R., Bachmann, M.F., and Ohashi, P.S. 1999. *Annu Rev Immunol* **17**: 829-874.
Serizawa, S., Miyamichi, K., Nakatani, H., Suzuki, M., Saito, M., Yoshihara, Y., and Sakano, H. 2003. *Science* **302**(5653): 2088-2094.
Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., Shi, Y. 2004. *Cell* **119** 941-953.
Shimizu, C., Kawamoto, H., Yamashita, M., Kimura, M., Kondou, E., Kaneko, Y., Okada, S., Tokuhisa, T., Yokoyama, M., Taniguchi, M., Katsura, Y., and Nakayama, T. 2001. *Int Immunol* **13**(1): 105-117.
Shoemaker, J., Saraiva, M., and O'Garra, A. 2006. *J Immunol* **176**(6): 3470-3479.
Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R., and Peterson, C.L. 2006.. *Science* **311**(5762): 844-847.
Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. 2006. *Nat Genet* **38**(11): 1348-1354.
Sjakste, N., Iarovaia, O.V., Razin, S.V., Linares-Cruz, G., Sjakste, T., Le Gac, V., Zhao, Z., and Scherrer, K. 2000. *Mol Gen Genet* **262**(6): 1012-1021.
Sleutels, F. and Barlow, D.P. 2001. *Mol Cell Biol* **21**(15): 5008-5017.
2002. *Adv Genet* **46**: 119-163.
Smale, S.T. 2001. *Genes Dev* **15**(19): 2503-2508.
Soldani, C. and Scovassi, A.I. 2002. *Apoptosis* **7**(4): 321-328.
Spadafora, C., Bellard, M., Compton, J.L., and Chambon, P. 1976. *FEBS Lett* **69**(1): 281-285.
Spain, L.M., Guerriero, A., Kunjibettu, S., Scott, E.W. 1999. *J. Immunol.* **163** 2681-2687.
Spilianakis, C.G. and Flavell, R.A. 2004. *Nat Immunol* **5**(10): 1017-1027.
Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R., and Flavell, R.A. 2005a. *Nature* **435**(7042): 637-645.
Spilianakis, C.G., Lee, G.R., and Flavell, R.A. 2005b. *Eur J Immunol* **35**(12): 3400-3404.
Splinter, E., Heath, H., Kooren, J., Palstra, R.J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. 2006. *Genes Dev* **20**(17): 2349-2354.
Stamatoyannopoulos, G. 1991. *Science* **252**(5004): 383.
Stein, G.S., Zaidi, S.K., Braastad, C.D., Montecino, M., van Wijnen, A.J., Choi, J.Y., Stein, J.L., Lian, J.B., and Javed, A. 2003. *Trends Cell Biol* **13**(11): 584-592.
Sterner, D.E. and Berger, S.L. 2000. *Microbiol Mol Biol Rev* **64**(2): 435-459.
Strahl, B.D. and Allis, C.D. 2000. *Nature* **403**(6765): 41-45.
Strouboulis, J., Dillon, N., and Grosveld, F. 1992. *Genes Dev* **6**(10): 1857-1864.
Sullivan, B.A. and Karpen, G.H. 2004. *Nat Struct Mol Biol* **11**(11): 1076-1083.
Sun, G., Liu, X., Mercado, P., Jenkinson, S.R., Kypriotou, M., Feigenbaum, L., Galera, P., and Bosselut, R. 2005. *Nat Immunol* **6**(4): 373-381.
Suto, R.K., Clarkson, M.J., Tremethick, D.J., and Luger, K. 2000. *Nat Struct Biol* **7**(12): 1121-1124.
Suzuki, N., Suzuki, S., Millar, D.G., Unno, M., Hara, H., Calzascia, T., Yamasaki, S., Yokosuka, T., Chen, N.J., Elford, A.R., Suzuki, J., Takeuchi, A., Mirtsos, C., Bouchard, D., Ohashi, P.S., Yeh, W.C., and Saito, T. 2006. *Science* **311**(5769): 1927-1932.

Svejstrup, J.Q. 2004.. *Biochim Biophys Acta* **1677**(1-3): 64-73.
Szabo, P.E., Han, L., Hyo-Jung, J., and Mann, J.R. 2006. *Cytogenet Genome Res* **113**(1-4): 238-246.
Szabo, P.E., Tang, S.H., Silva, F.J., Tsark, W.M., and Mann, J.R. 2004. *Mol Cell Biol* **24**(11): 4791-4800.
Szabo, S.J., Dighe, A.S., Gubler, U., and Murphy, K.M. 1997. *J Exp Med* **185**(5): 817-824.
Szabo, S.J., Kim, S.T., Costa, G.L., Zhang, X., Fathman, C.G., and Glimcher, L.H. 2000. *Cell* **100**(6): 655-669.
Szabo, S.J., Sullivan, B.M., Peng, S.L., and Glimcher, L.H. 2003. *Annu Rev Immunol* **21**: 713-758.
Szabo, S.J., Sullivan, B.M., Stemmann, C., Satoskar, A.R., Sleckman, B.P., and Glimcher, L.H. 2002. *Science* **295**(5553): 338-342.
Takatori, H., Nakajima, H., Kagami, S., Hirose, K., Suto, A., Suzuki, K., Kubo, M., Yoshimura, A., Saito, Y., and Iwamoto, I. 2005. *J Immunol* **174**(7): 4105-4112.
Takemoto, N., Koyano-Nakagawa, N., Yokota, T., Arai, N., Miyatake, S., and Arai, K. 1998. *Int Immunol* **10**(12): 1981-1985.
Tamaru, H. and Selker, E.U. 2001. *Nature* **414**(6861): 277-283.
Taniuchi, I., Ellmeier, W., and Littman, D.R. 2004. *Adv Immunol* **83**: 55-89.
Taniuchi, I., Osato, M., Egawa, T., Sunshine, M.J., Bae, S.C., Komori, T., Ito, Y., and Littman, D.R. 2002. *Cell* **111**(5): 621-633.
Ting, C.N., Olson, M.C., Barton, K.P., and Leiden, J.M. 1996. *Nature* **384**(6608): 474-478.
Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and de Laat, W. 2002. *Mol Cell* **10**(6): 1453-1465.
Tomita, K., Hattori, M., Nakamura, E., Nakanishi, S., Minato, N., and Kageyama, R. 1999. *Genes Dev* **13**(9): 1203-1210.
Tonegawa, S. 1983. *Nature* **302**(5909): 575-581.
Torrano, V., Chernukhin, I., Docquier, F., D'Arcy, V., Leon, J., Klenova, E., and Delgado, M.D. 2005. *J Biol Chem* **280**(30): 28152-28161.
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. *J Cell Sci* **119**(Pt 9): 1746-1759.
Tremethick, D.J. 2007. *Cell* **128**(4): 651-654.
Trimborn, T., Gribnau, J., Grosveld, F., Fraser, P. 1999. *Genes and Development* **13** 112-124.
Tse, C., Sera, T., Wolffe, A.P., Hansen, J.C. 1998. *Mol Cell Biol.* **18** 4629-4638.
Tsien, R.Y. 1998. *Annu Rev Biochem* **67**: 509-544.
Tsukiyama, T., Ishida, N., Shirane, M., Minamishima, Y.A., Hatakeyama, S., Kitagawa, M., Nakayama, K., and Nakayama, K. 2001. *J Immunol* **166**(1): 304-312.
Tuan, D., Solomon, W., Li, Q., and London, I.M. 1985. *Proc Natl Acad Sci U S A* **82**(19): 6384-6388.
Turner, B.M. 2002. *Cell* **111**(3): 285-291.
Vakoc, C.R., Letting, D.L., Gheldof, N., Sawado, T., Bender, M.A., Groudine, M., Weiss, M.J., Dekker, J., Blobel, G.A. 2005. *Molecular Cell* **17** 453-462.
Valadez-Graham, V., Razin, S.V., and Recillas-Targa, F. 2004. *Nucleic Acids Res* **32**(4): 1354-1362.
van Daal, A. and Elgin, S.C. 1992. *Mol Biol Cell* **3**(6): 593-602.
van Gent, D.C., Ramsden, D.A., and Gellert, M. 1996.. *Cell* **85**(1): 107-113.

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. *Cancer Res* **65**(17): 7751-7762.
Vetchinova, A.S., Akopov, S.B., Chernov, I.P., Nikolaev, L.G., and Sverdlov, E.D. 2006. *Anal Biochem* **354**(1): 85-93.
Vigneau, S., Augui, S., Navarro, P., Avner, P., and Clerc, P. 2006. *Proc Natl Acad Sci U S A* **103**(19): 7390-7395.
Viney, J.L., Dianda, L., Roberts, S.J., Wen, L., Mallick, C.A., Hayday, A.C., and Owen, M.J. 1994*Proc Natl Acad Sci U S A* **91**(25): 11948-11952.
von Boehmer, H., Aifantis, I., Feinberg, J., Lechner, O., Saint-Ruf, C., Walter, U., Buer, J., and Azogui, O. 1999. *Curr Opin Immunol* **11**(2): 135-142.
Vostrov, A.A. and Quitschke, W.W. 1997. *J Biol Chem* **272**(52): 33353-33359.
Wade, P.A., Gegonne, A., Jones, P.L., Ballestar, E., Aubry, F., and Wolffe, A.P. 1999. *Nat Genet* **23**(1): 62-66.
Wang, J.H., Nichogiannopoulou, A., Wu, L., Sun, L., Sharpe, A.H., Bigby, M., and Georgopoulos, K. 1996. *Adv Immunol* **58**: 29-85.
Weber, M., Hagege, H., Murrell, A., Brunel, C., Reik, W., Cathala, G., and Forne, T. 2003. *Mol Cell Biol* **23**(24): 8953-8959.
Weintraub, H. and Groudine, M. 1976. *Science* **193**(4256): 848-856.
Wen, J., Huang, S., Rogers, H., Dickinson, L.A., Kohwi-Shigematsu, T., and Noguchi, C.T. 2005. *Blood* **105**(8): 3330-3339.
West, A.G. and Fraser, P. 2005. *Hum Mol Genet* **14 Spec No 1**: R101-111.
West, A.G., Gazner, M., Felsenfeld, G. 2002. *Genes and Development* **16** 271-288.
Whitlock, J.P., Jr. and Simpson, R.T. 1977. *J Biol Chem* **252**(18): 6516-6520.
Wiest, D.L., Berger, M.A., and Carleton, M. 1999. *Semin Immunol* **11**(4): 251-262.
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. *Immunity* **20**(6): 719-733.
Willoughby, D.A., Vilalta, A., and Oshima, R.G. 2000. *J Biol Chem* **275**(2): 759-768.
Wilson, C.B., Makar, K.W., and Perez-Melgosa, M. 2002. *J Infect Dis* **185** Suppl 1: S37-45.
Wilson, N.J., Boniface, K., Chan, J.R., McKenzie, B.S., Blumenschein, W.M., Mattson, J.D., Basham, B., Smith, K., Chen, T., Morel, F., Lecron, J.C., Kastelein, R.A., Cua, D.J., McClanahan, T.K., Bowman, E.P., and de Waal Malefyt, R. 2007. *Nat Immunol* **8**(9): 950-957.
Wong, C., Antonarakis, S.E., Goff, S.C., Orkin, S.H., Boehm, C.D., and Kazazian, H.H., Jr. 1986. *Proc Natl Acad Sci U S A* **83**(17): 6529-6532.
Woolf, E., Xiao, C., Fainaru, O., Lotem, J., Rosen, D., Negreanu, V., Bernstein, Y., Goldenberg, D., Brenner, O., Berke, G., Levanon, D., and Groner, Y. 2003. *Proc Natl Acad Sci U S A* **100**(13): 7731-7736.
Xie, X., Mikkelsen, T.S., Gnirke, A., Lindblad-Toh, K., Kellis, M., and Lander, E.S. 2007. *Proc Natl Acad Sci U S A* **104**(17): 7145-7150.
Xu, F., Zhang, K., and Grunstein, M. 2005. *Cell* **121**(3): 375-385.
Xu, N., Donohoe, M.E., Silva, S.S., Lee, J.T. 2007. *Nature Genetics*. doi: 10. 1038/ng.2007.5.
Xue, Y., Wong, J., Moreno, G.T., Young, M.K., Cote, J., and Wang, W. 1998. *Mol Cell* **2**(6): 851-861.
Yan, Z.J. and Qian, R.L. 1998. *Cell Res* **8**(3): 209-218.
Yang, P.K. and Kuroda, M.I. 2007. *Cell* **128**(4): 777-786.

Yang, Y., Contag, C.H., Felsher, D., Shachaf, C.M., Cao, Y., Herzenberg, L.A., Herzenberg, L.A., and Tung, J.W. 2004. *Proc Natl Acad Sci U S A* **101**(11): 3898-3902.

Yao, S., Osborne, C.S., Bharadwaj, R.R., Pasceri, P., Sukonnik, T., Pannell, D., Recillas-Targa, F., West, A.G., and Ellis, J. 2003. *Nucleic Acids Res* **31**(18): 5317-5323.

Yasui, D., Miyano, M., Cai, S., Varga-Weisz, P., and Kohwi-Shigematsu, T. 2002. *Nature* **419**(6907): 641-645.

Yates, A., Callard, R., and Stark, J. 2004. *J Theor Biol* **231**(2): 181-196.

Yeh, A., Wei, M., Golub, S.B., Yamashiro, D.J., Murty, V.V., and Tycko, B. 2002. *Genes Chromosomes Cancer* **35**(2): 156-163.

Yoon, B., Herman, H., Hu, B., Park, Y.J., Lindroth, A., Bell, A., West, A.G., Chang, Y., Stablewski, A., Piel, J.C., Loukinov, D.I., Lobanenkov, V.V., and Soloway, P.D. 2005. *Mol Cell Biol* **25**(24): 11184-11190.

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. *Nat Genet* **36**(10): 1105-1110.

Yusufzai, T.M. and Felsenfeld, G. 2004. *Proc Natl Acad Sci U S A* **101**(23): 8620-8624.

Yusufzai, T.M., Tagami, H., Nakatani, Y., and Felsenfeld, G. 2004. *Mol Cell* **13**(2): 291-298.

Zhang, R., Burke, L.J., Rasko, J.E., Lobanenkov, V., and Renkawitz, R. 2004. *Exp Cell Res* **294**(1): 86-93.

Zhang, W. and Bieker, J.J. 1998. *Proc Natl Acad Sci U S A* **95**(17): 9855-9860.

Zhang, Y., Apilado, R., Coleman, J., Ben-Sasson, S., Tsang, S., Hu-Li, J., Paul, W.E., and Huang, H. 2001. *J Exp Med* **194**(2): 165-172.

Zheng, W. and Flavell, R.A. 1997. *Cell* **89**(4): 587-596.

Zheng, W.P., Zhao, Q., Zhao, X., Li, B., Hubank, M., Schatz, D.G., and Flavell, R.A. 2004. *J Immunol* **172**(1): 114-122.

Zhou, G.L., Xin, L., Song, W., Di, L.J., Liu, G., Wu, X.S., Liu, D.P., and Liang, C.C. 2006. *Mol Cell Biol* **26**(13): 5096-5105.

Zhu, J., Cote-Sierra, J., Guo, L., and Paul, W.E. 2003. *Immunity* **19**(5): 739-748.

Zhu, J. and Emerson, S.G. 2002. *Oncogene* **21**(21): 3295-3313.

Zhu, J., Min, B., Hu-Li, J., Watson, C.J., Grinberg, A., Wang, Q., Killeen, N., Urban, J.F., Jr., Guo, L., and Paul, W.E. 2004. *Nat Immunol* **5**(11): 1157-1165.

Zhu, J., Yamane, H., Cote-Sierra, J., Guo, L., and Paul, W.E. 2006. *Cell Res* **16**(1): 3-10.

Zuniga-Pflucker, J.C. and Lenardo, M.J. 1996. *Curr Opin Immunol* **8**(2): 215-224.

Zuniga-Pflucker, J.C. and Schmitt, T.M. 2005. **35**(7): 2016-2018.

Summary

The genetic information used in each cell type is different. Differential gene expression generates cellular diversity and mediates the development of a multi-cellular body. The molecular mechanisms behind this are of great interest. Substantial advances in the understanding of gene regulation and function have been made based on the approaches of analysing gene expression, knocking individual genes out in various model organisms and trying to ascertain functions based on the obtained phenotype, and by examining the localization and behaviour of proteins encoded by these genes. In doing this we have hoped to understand from the bottom-up fundamental components of aspects of multi cellular existence[*]. Spatial and temporal expression of genes is mediated co-ordinately by direct modifications of chromatin, modifications of the DNA and interaction of the DNA with numerous general and cell type specific proteins, some of which function to remodel the chromatin structure. I have described these processes in chapter 1.

Ascertaining functions of genes (or the proteins they encode) becomes particularly problematic when genes are expressed in many different cells at different stages of an organism's development, and deletion of such genes is detrimental to cell viability. This is the case for the *Ctcf* gene, which encodes an 11-zinc-finger protein. CTCF binds DNA with a loose consensus and has multiple interaction partners. It was the first vertebrate protein identified to have enhancer blocking activity and was proposed to regulate the transition between transcriptionally active and repressed chromatin domains. CTCF is in fact a multi-functional protein shown to regulate transcriptional activation, mediate histone modifications, chromatin looping *in cis* and chromosomal interactions *in trans*. The current knowledge regarding CTCF was detailed in chapter 2.

To better understand the role of this protein *in vivo*, we have generated conditional *Ctcf* knockout mice in order to choose both the time during development and the specific cell type where deletion of the *Ctcf* gene occurs. Embryonic development in mice is severely impaired in the absence of CTCF, which precluded the analysis of a complete knockout of CTCF during late stages of development. However, by conditional deletion of the *Ctcf* gene we have been able to analyse the role of CTCF in hematopoietic tissues. The haematopoietic system consists of many different cell types, each with distinct functional capacities. Erythropoiesis and the development of T-lymphoid cells are well-characterised examples of cellular differentiation pathways that require tight regulation of sequential proliferation and differentiation. These aspects were discussed in chapter 3.

[*] From *The Music of Life; Biology Beyond the Genome. Noble, D. Oxford University Press, 2006.*

The transcription of *β-globin* genes is restricted to erythroid cells. Previous studies have revealed that the chromatin structure of the *β-globin* locus in mice is organised into a chromatin hub in erythroid precursor cells. Activation of *β-globin* genes occurs via the recruitment of specific transcription factors and contacts between the Locus Control Region (LCR) and genes that are mediated by chromatin looping. We have demonstrated that the chromatin configuration of the mouse *β-globin* locus is dependent on CTCF-mediated clustering of DNA hypersensitive sites that flank this locus. The binding of CTCF to the *β-globin* locus also prevents the gain of repressive histone modifications to CTCF binding sites. The significance of defined chromatin domains and looping however is brought into question as we show *β-globin* gene transcription occurs in the absence of a normal ACH formation. The results of this study were presented in chapter 4.

When CTCF protein levels are specifically depleted in T-cells, we show that the expansion of T-cells within the thymus is very sensitive to the level of CTCF. In the absence of CTCF, T-cell differentiation in the thymus is impaired and an accumulation of ISP cells occurs. $CD4^+$ cells leaving the thymus do differentiate into effector T-helper cells in the periphery. However, when tested for their efficacy, Th2 cells are completely incompetent to mediate an allergic response. T-helper cells express characteristic cytokine genes that are arranged in a chromatin configuration similar to the *β-globin* ACH. Our studies indicate that despite the presence of key regulatory proteins such as GATA3 and SatB1, transcriptional activation of cytokine genes is impaired in the absence of CTCF. We also show that the activation of T-cells *in vitro* via signals through the T cell receptor require CTCF. Interestingly, in the presence of the calcium ionophore ionomycin CTCF-deficient cells can divide, indicating that CTCF is not necessarily required for cell proliferation. The data were detailed in chapter 5.

In chapter 6, I have described the generation and analysis of a GFP-CTCF knock-in allele. This allele allowed analysis of the localisation and activity of the CTCF protein *in vivo*. A previous publication indicated that a window of CTCF-negative germ cells occurs during spermatogenesis. Our data actually show that CTCF is ubiquitously expressed during spermatogenesis, with the exception of spermatids when DNA is further compacted. We used intact seminiferous tubules from the GFP-CTCF knock-in mice, mouse embryonic fibroblasts (MEFs) from GFP-CTCF knock-in embryos, and embryonic stem (ES) cells expressing CTCF tagged with GFP, to study the dynamics of the CTCF protein. This data revealed two distinct dynamic populations of CTCF within the nucleus and suggest that the number of binding sites for CTCF in mice exceeds current predictions.

The studies addressed in this thesis have provided novel insight into the significance of CTCF acting in different cells and differentiation pathways. I have

discussed the findings of our experiments and their significance with regard to published observations of CTCF function in the final chapter 7. Two conclusions stand out, namely that CTCF is not absolutely required for T cell proliferation (see chapter 5), and that the protein is only essential for some cell-specific and perhaps locus-specific programs (chapters 4 and 5). Given the ubiquitous expression of CTCF and its presumed essential role, these findings are surprising and warrant further detailed analysis.

Samenvatting

Alhoewel vrijwel elke cel in ons lichaam precies dezelfde genetische informatie bevat, wordt die per celtype verschillend afgelezen. Deze verschillen in genexpressie patronen zijn mede verantwoordelijk voor de diversiteit in cel typen. Het is belangrijk om uit te vinden wat nu de onderliggende moleculaire mechanismen zijn. Er is goede vooruitgang geboekt in ons begrip over hoe genen gereguleerd worden en wat hun functie is door te kijken naar: (1) gen expressie patronen; (2) door genen te inactiveren in verschillende model organismen en door gen functie te destilleren op basis van waargenomen fenotypen; (3) door de intracellulaire distributie en het gedrag te bestuderen van eiwitten die door de desbetreffende genen worden gecodeerd. Door dit soort studies hoopten we de meest fundamentele componenten van het multi-cellulaire bestaan te isoleren en te begrijpen.

Een functionele analyse van een gen in een organisme (ofwel van het eiwit gecodeerd door dat gen) wordt bemoeilijkt als het gen in kwestie in veel cellen en in een groot aantal stadia van de ontwkkeling van het organisme tot expressie komt en als inactiverig van het gen al vroeg in de ontwikkeling zware afwijkingen tot gevolg heeft. Dit is het geval voor het *Ctcf* gen, wat codeert voor een eiwit met 11 zink vingers. CTCF bindt verschillende DNA sequenties en heeft meerdere interactie partners. CTCF was het eerste eiwit in vertebraten waarvan is aangetoond dat het de activiteit van een enhancer op een promoter kan blokkeren. Het eiwit lijkt ook betrokken te zijn bij het scheiden van actief transcriberende chromatine domeinen van inactieve domeinen. Vandaar dat CTCF een “enhancer-blocker” en “insulator” wordt genoemd. CTCF is feitelijk een multifunctioneel eiwit betrokken bij transcriptie activatie, histon modificaties, de vorming van intrachromosomale chromatine lussen, maar ook bij interchromosomale interacties. De huidige kennis omtrent CTCF heb ik beschreven in hoofdstuk 2.

Om de *in vivo* rol van CTCF beter te begrijpen hebben we zogenaamde conditionele *Ctcf* “knockout” muizen gemaakt. Zo konden we zelf de tijd van, en het cel type waarin, *Ctcf* deletie plaatsvond in het zich ontwikkelende organisme bepalen. We hebben laten zien dat de embryonale ontwikkeling van muizen zwaar is aangedaan in de afwezigheid van CTCF. Dit maakte de analyse van volwassen muizen waarin CTCF afwezig is in alle cellen onmogelijk. We konden daarentegen wel de rol van CTCF bestuderen in het hematopoietische (bloedvormend) systeem. Dit bestaat uit veel verschillende celtypen, met een specifieke functie voor elk type cel. Erythropoiese en T-cel lymphopoiese zijn goed gekarakteriseerde voorbeelden van cel differentiatie routes waarin strikt gereguleerde, elkaar opvolgende stappen van proliferatie en differentiatie uiteindelijke zorgen voor, respectievelijk, een volwassen rode bloedcel en een T cel. Deze processen zijn beschreven in hoofdstuk 3.

De transcriptie van de β-globine genen vindt alleen plaats in rode bloedcellen. Eerdere studies hebben laten zien dat al in de erythroïde voorlopers het chromatine in het β-globine locus georganiseerd is in een specifieke structuur van lussen en knopen die een “chromatin hub” wordt genoemd. De activering van de β-globine genen gebeurt doordat specifieke transcriptie factoren worden aangetrokken

en doordat er intensieve contacten komen tussen het Locus Control Region (LCR) en de β-globine genen zelf. Dit alles wordt gecoördineert via de “chromatin hub”. Wij hebben laten zien dat de chromatine structuur van het β-globine locus afhankelijk is van CTCF en dat dit eiwit essentieel is voor de vorming van contacten tussen DNAse gevoelige gebieden aan beide kanten van het β-globine locus. CTCF binding in het β-globine locus voorkomt ook dat histonen gemodificeerd worden op een manier die karakteristiek is voor inactief chromatine. Een verrassend resultaat was dat ondanks de duidelijke rol van CTCF in deze processen de transcriptie van de β-globine genen niet significant verminderd in de afwezigheid van een CTCF bindingsplaats. Dit werpt de vraag op wat nu de functie is van de “chromatin hub” in het β-globine locus. De resultaten van deze studies staan beschreven in hoofdstk 4.

Als we het *Ctcf* gen specifiek inactiveren in T cellen van het immuunsysteem laten we zien dat deze cellen erg gevoelig zijn voor vermindering van de niveaus van dit eiwit. In afwezigheid van CTCF is de differentiatie van T cellen in de thymus geblokkeerd en stapelen de zogenaamde ISP cellen zich op. CD4-positieve T cellen kunnen de thymus verlaten en verder differentiëren naar zogenaamde T-helper of –effektor cellen. Wanneer deze cellen echter getest worden op hun effectiviteit dan blijkt dat in ieder geval de T helper 2 (Th2) cellen niet in staat zijn om een adequate afweer op te bouwen. Th2 cellen expresseren een specifieke set van cytokines, waarvan de coderende genen in een gebied liggen die een chromatine conformatie heeft die veel lijkt op die van het globine locus. Onze resultaten laten zien dat ondanks de aanweigheid van belangrijke T cel transcriptie factoren als GATA3 en SATB1 de activering van de cytokine genen in afwezigheid van CTCF geremd is. We laten ook zien dat gekweekte T cellen CTCF nodig hebben om een goede signaal transductie cascade vanaf de T cel receptor op te zetten. Een interessant resultaat is dat in aanwezigheid van ionomycine (een calcium ionofoor) cellen kunnen delen in afwezigheid van CTCF, wat aantoont dat CTCF niet perse nodig is voor cel proliferatie. Deze resultaten heb ik beschreven in hoofdstuk 5.

In hoofdstuk 6 heb ik de analyse beschreven van ons zogenaamde GFP-CTCF “knock-in” allel. In plaats van CTCF wordt van dit allel het fusie eiwit GFP-CTCF afgeschreven, waardoor we het dynamische gedrag van CTCF *in vivo* kunnen bestuderen. Voorgaande publicaties suggereren dat er gedurende de spermatogenese een periode is dat CTCF niet in mannelijke kiemcellen voorkomt. Onze data laten zien dat dit niet waar is en dat CTCF in alle cellen van de testis tubuli tot expressie komt behalve in spermatiden: kiemcellen waarin het DNA nog verder samengebundeld wordt. We hebben intacte testis tubuli van de GFP-CTCF “knock-in” muizen, embryonale fibroblasten van dezelfde muizen, alsmede embryonale stem cellen die GFP-CTCF expresseren in plaats van CTCF allemaal gebruikt om het dynamische gedrag van CTCF te bestuderen. Het blijkt dat dit eiwit in twee fracties voorkomt, een kleine mobiele populatie en een gote immobiele populatie. Onze resultaten suggereren dat het aantal gepubliceerde bindingsplaatsen voor CTCF veel minder is dan het aantal CTCF moleculen dat aanwezig is in een kern.

De studies die in dit proefschrift zijn beschreven hebben tot een nieuw inzicht geleid omtrent de functie van CTCF in verschillende cel typen en

differentiatie routes. In het laatste hoofdstuk (7) heb ik onze resultaten vergeleken met die van een groot aantal publicaties over CTCF. Voor wat betreft ons werk noem ik nogmaals twee resultaten, namelijk dat CTCF niet echt nodig is voor cel proliferatie (zie hoofdstuk 5) en dat het eiwit soms wel en soms niet noodzakelijk is voor cel-specifieke en locus-specifeke transcriptie programma's (zie hoofdstukken 4 en 5). Gegeven de expressie van CTCF in vrijwel alle cellen van de muis en de gedachte dat dit geconserveerde eiwit essentieel is, zijn dit verrassende resultaten die verdere studie naar dit eiwit noodzakelijk maken.

Vertaling: N. Galjart

Curriculum vitae

Name: Helen Elizabeth Heath

Date of birth: 13 August 1976

Place of birth: Sutton Coldfield, England

Education

1995-1999: School of Biological Sciences, University of Surrey, Guildford, England. BSc (Hons) in Microbiology (Medical).

1997-1998: Janssen Research Foundation, Beerse, Belgium.
Department of Biotechnology.
Project: Molecular cloning and expression of GDNF homologues and analysis of ligand-receptor interactions in the yeast two-hybrid system.

1999-2007: PhD program. Department of Cell Biology, Erasmus Medical Center, Rotterdam, The Netherlands.
PhD project as described in this thesis.
Supervisors: Dr. N. Galjart and Prof. Dr. F. Grosveld.

Publications

Rapid switching of TFIIH between RNA polymerase I and II transcription and DNA repair in vivo. Hoogstraten D, Nigg AL, Heath H, Mullenders LH, van Driel R, Hoeijmakers JH, Vermeulen W, Houtsmuller AB. **Molecular Cell. 2002. 10 (5): 1163-1174.**

Differential contributions of mammalian Rad54 paralogs to recombination, DNA repair, and meiosis. Wesoly J, Agarwal S, Sigurdsson S, Bussen W, Van Komen S, Qin J, van Steeg H, van Benthem J, Wassenaar E, Baarends WM, Ghazvini M, Tafel AA, Heath H, Galjart N, Essers J, Grootegoed JA, Arnheim N, Bezzubova O, Buerstedde JM, Sung P, Kanaar R. **Molecular and Cellular Biology. 2006. 26 (3): 976-989.**

CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus. Splinter E, Heath H, Kooren J, Palstra R.J, Klous P, Grosveld F, Galjart N, de Laat W. **Genes and Development. 2006. 20 (17): 2349-2354.**

Acknowledgements

Thank you

This long and almost impossible pursuit has been made not only possible but became a great experience. For giving me the opportunity to begin, continue and finish, thank you Prof. Grosveld. Your advice on all sorts of topics and support throughout has been invaluable.

Sometimes saying thank you feels remarkably ineffective and Niels, this is certainly true. It is no understatement that I would not be here without you. It has been hard and sometimes very difficult, somehow we still came though and in the end I think we pulled it off (?!). Thank you for the laughs, giving me the space and time required when needed, but above all thank you, really for your courage and unwavering defence. Cheers.

Thank you to all members of my promotion committee who agreed to participate in the final chapter of this story and endure the inconceivably early start to the proceedings.

I had reached the conclusion long ago that the life of a PhD student, particularly one who undertakes the challenge of doing it abroad, is a life of transitions, people come, people go, situations change dramatically and it takes nerve and great friends to hang on. I have been fortunate enough to meet and work with many people over the years, to all those who shared smiles and tears with me, thank you.

To everyone who joined the CTCF cause, especially lab 702, and those that helped along the way, thank you.

Dear Marike and Jasperina, your help and support in all matters both trivial and solemn have often made the difference between things being easy and despairingly difficult, thank you so very much. I would like to extend thanks especially to Ton de Wit and Jon Kong-a-San. Thank you to all the various people of the EDC for looking after the mice, putting up with my demands especially in the beginning and being such a great help.
Casper and Bjorn, without you I would not have survived the early years, cheers.
Dear An, your words of kindness and compassion never failed to impress me, thank you.

Kam-Wing and Jan-Piet, for looking after me at my first T-cell meeting. I wish you both well for the future.
Rudi, you showed me a different understanding, a different way of working, and this manifested in a new belief. For the many spontaneous work discussions and above all your fortitude and patience, thank you.

Laura, this small world allowed our paths to cross for a second time, I am so happy it did. You made a big difference to the lab, I wish you every success and happiness in Spain. Dear Magda, what can I say, you know it all, the ups, downs and the very low downs, throughout you have always been someone I want to stop and talk to and you have always stopped, thank you. Deep breaths and get this PhD done, all fingers crossed for you.

Despite the many changes both personally and professionally I am very lucky to have met friends whose laughter, love and support will stay with me for life. I'll try my best not to be sentimental; I'll save that for the party!
Cristina, you are truly beautiful, courageous, hilarious and fantastic, I am going to miss you!!! Come and visit. Frank, for being the one person who knew what I was thinking at the absolute worst moment of my life, you probably don't remember, but that point and so many others are typical of you. I am happy to have met you and wish you all the strength and good luck required in this job, you know how good you are, don't loose sight of it. Suzanne, your determination and dedication are truly impressive, thank you for caring and for all the conversations and making me laugh, I wish you the best of luck with everything, hope to see you in England one day soon. I just have to mention the eleven magic fingers, yeah, we all know who! Athina and Kris, this is now getting hard….. I have laughed to the point of tears so many times with you and carry so many good memories in my heart, not just of all the fun times, but the downright serious and profound moments as well. I know I am going to see you both for many years to come, so will not embarrass either of you by saying you really did change my life and I can't thank either of you enough for that, those cupid arrows are well dangerous! See you in England.

Ggggrrrrrllluuuggggghhhhugh-ppffffff said the seal!

To my family and my beautiful friends who make life easy, I love you.

I'm coming home.

Erratum.

p.21: Superfluous bracket in reference (Howlett and Reik 1991, Mayer, 2000).

p.32: Missing full stop (Recillas-Targa et al. 2002).

p.33: Missing abbreviations, Replication protein A (RPA), Runt-related transcription factors (Runx).

p.38: In chicken, the gene coding for CTCF consists of 10 exons that generate a 728 amino acid protein. Murine CTCF has 12 coding exons that encode a 736 amino acid protein.

p.54: Gap between bracket and reference (Kantidze et al. 2007, Torrano et al. 2006).

p60: Superfluous bracket in reference (McGrath et al. 2003, Kingsley et al. 2004)

p61: Superfluous reference number in figure legend 3.2 (Dessypris 1998).

p69: Missing information in figure 3.4 (5' HS-85).

p72: Superfluous bracket in reference (Harman et al. 2005, Masuda 2005).

p74: Missing information in figure 3.5 (pTα).

p81: Missing information in figure 3.6 (Th-17, iT-reg).

Chapter 4: Figure 2A/C, 3A/B, 4B/C, supplemental fig.4, figure legends (Δ3'HS1). Figure 2B (Δ and c). Supplemental figure 7 (ΔCt).

p.191: Our live cell imaging does not support this idea.