β-Globin Gene Regulation and Nuclear Organisation

Jurgen Kooren

β-Globin Gene Regulation and Nuclear Organisation

β-Globine gen-regulatie en nucleaire organisatie

Proefschrift

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

Prof.dr. S.W.J. Lamberts

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

woensdag 19 december 2007 om 13.45 uur

door

Jurgen Alexander Kooren

geboren te Rotterdam

Promotiecommissie

Promotor:	Prof.dr. F.G. Grosveld
Overige leden:	Dr.ir. D.N. Meijer Prof.dr. J.N.J. Philipsen Dr. M.M. von Lindern
Copromotor:	Dr. W.L. de Laat

Dit proefschrift is tot stand gekomen binnen de vakgroep Celbiologie aan de faculteit der Geneeskunde en Gezondheidswetenschappen van het Erasmus MC te Rotterdam. De vakgroep maakt deel uit van het Medisch Genetisch Centrum Zuid-West Nederland.

Een bijdrage in de drukkosten van dit proefschrift is verkregen van de J.E. Jurriaanse Stichting.

"*Verwondering is het zout der aarde.*"

M.C. Escher

Contents

List of abbreviations

3C	Chromosome Conformation Capture
4C	Chromosome Conformation Capture on Chip
ACH	Active Chromatin Hub
bp	Base pairs
CH	Chromatin Hub
ChIP	Chromatin immuno-precipitation
CTCF	CCCTC-binding factor
DNA	Deoxyribonucleic acid
DNAseI	Deoxyribonuclease I
EKLF	Erythroid Krüppel-like factor
ES-EP	Embryonic stem cell derived erythroid progenitor
FISH	Fluorescent in situ hybridization
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HP1	Heterochromatin protein 1
HS	Hypersensitive site
HSC	Hematopoietic stem cell
ICR	Imprinting control region
LCR	Locus Control Region
MEL	Mouse erythroleukemia
mRNA	messenger RNA
NF-E2	Nuclear factor erythroid derived 2
OR	Olfactory receptor
PCR	Polymerase chain reaction
PEV	Position effect variegation
PIC	Pre-initiation complex
RNA	Ribonucleic acid
RNAPII	RNA polymerase II
TBP	TATA binding protein

1 Introduction

Introduction

The emergence of multicellular life, about one billion years ago, represents one of the key events in the evolution of life on earth. This process initiated a chain of events that eventually led to the biological diversity as we know it today. Given this complexity of life, it is fascinating to realise that the development of every multicellular organism begins with a single cell, e.g. a fertilised oocyte. Throughout development, this single cell divides mitotically to give rise to numerous different cell types, eventually constituting the adult organism. This cellular diversity is remarkable given the fact that every cell harbours the same genetic information that, in cases of higher eukaryotes, is stored in billions of base pairs of deoxyribonucleic acid (DNA). The morphological and functional diversity among cells is the result of the precise regulation of this genetic information at different stages of cellular differentiation. Activation and repression of genes is precisely regulated since abnormal gene expression can lead to defects at the cellular level, which may eventually result in diseases such as cancer. Therefore, different levels of control must exist to safeguard the proper spatiotemporal expression of genes throughout the process of cellular differentiation and this thesis will mainly describe studies concerned with the process of transcriptional regulation.

Regulatory sequences encoded in the primary DNA sequence provide the first level of transcriptional control. These *cis* regulatory elements such as promoters and enhancers are able to bind a diverse set of proteins called transcription factors that can lead to either activation or repression of the genes they control. A second level of control originates from the fact that eukaryotic DNA is packaged into chromatin. Besides serving the compaction of the genome, modifications of this chromatin template can lead to local alterations in its structure thereby facilitating or suppressing transcriptional output. Finally, the three-dimensional topology of the chromatinized DNA template in the spatially confined environment of the nucleus provides another level of regulation. Although less well understood, there is growing evidence that higher order chromatin structure and nuclear organisation provide another level of transcriptional control throughout development and cell differentiation.

Transcriptional regulation on the chromatin template

All hereditary information required for the development and functioning of an organism is stored on long DNA molecules, the chromosomes, that locate inside a eukaryotic nucleus. A DNA molecule consists of two polynucleotide chains composed of different combinations of four types of nucleotide subunits. The entire functional unit controlling a discrete hereditary characteristic (usually corresponding to a single protein or RNA) is called a gene.In higher eukaryotes like humans, the entire genome with a length of up to two meters needs to be compacted in order to fit into a nucleus with a diameter as small as 3-10 µm. This compaction results in the formation

of condensed chromatin fibers that are inherently restrictive to processes (like transcription) requiring access to the DNA sequence.

Packaging of the genome into chromatin

The fundamental packaging unit of chromatin is the nucleosome. Nucleosomes consist of a core histone octamer (2 copies of each core histones H2A, H2B, H3 and H4) around which the DNA double helix is wrapped in 1.7 left-handed superhelical turns spanning 147 bp (Luger et al., 1997). Each nucleosome consists of two functionally different domains. The core globular domain mediates histone-histone and histone-DNA interactions. In addition to this, each of the core histones have a long N-terminal amino acid 'tail', which protrudes from the DNA-histone core. Covalent modifications of these histone tails have a crucial role in regulating chromatin structure by recruiting specific proteins to a marked stretch of chromatin. Furthermore, these modifications affect the charge of the histone proteins resulting in changed functional properties of the chromatin template. Patterns of different modifications are thought to form a specific 'histone code' that is associated with structural changes that occur in chromatin at replication and transcription.

Each nucleosome core particle is separated from the next by a region of linker DNA, which can vary in length from a few nucleotide pairs up to about 80. Nucleosomes together with this linker DNA give rise to the 10-nm fiber ('beads on a string') conformation, which represents the first level of chromosomal DNA packing. In addition to core histones, metazoan chromatin also contains linker histones (such as histone H1), which are not related in sequence to the core histones, but also contain a globular domain flanked by NH_2-terminal and COOH-terminal tail domains (Parseghian and Hamkalo, 2001). Binding of linker histone H1 results in condensation of the nucleosomal array and the second level of chromosomal packing; the more compact 30 nm fiber (Allan et al., 1986) (Figure 1.1). Despite intense efforts, the actual structure of the 30 nm fiber remains unresolved and current biophysical and biochemical studies have led to two principal concepts of fiber architecture underlying the 30 nm fiber: the one-start solenoid model and the zigzag two-start helical model (reviewed in (Tremethick, 2007)). It should be noted however that the 30 nm fiber so far has only been visualized *in vitro* and is not seen as an underlying structure in sections of whole nuclei in most higher eukaryotes cell types (Horowitz-Scherer and Woodcock, 2006).

Further compaction of eukaryotic chromatin into higher order fibers is probably the result of self-association of these 30 nm fibers into a series of chromatin loops and/or coils. It is not surprising that the molecular basis of additional higher order chromatin structure is even less well defined, although several models of large scale chromatin folding exist. The 'giant loop' and 'radial array' models were proposed for interphase chromosomes and are based on the statistical analysis of the mean separation between two chromosomal sites as a function of genomic distance

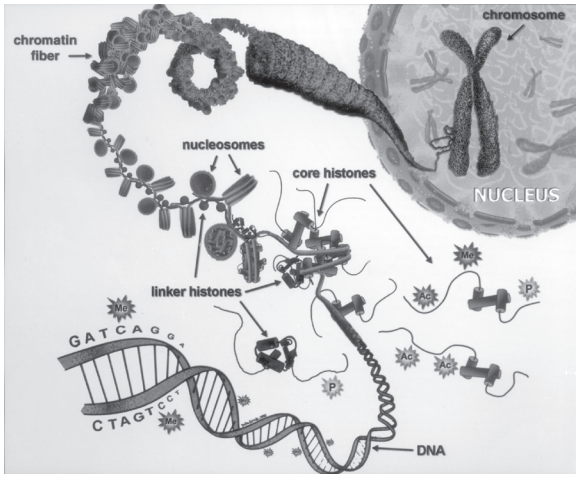

Figure 1.1
DNA is packaged into chromatin. The basic building block of chromatin is DNA wrapped around nucleosomes, which together with linker DNA give rise to a 10-nm 'beads on a string' conformation. The second level of chromosomal packaging is the result of subsequent condensation of this nucleosomal array and the formation of a 30-nm fiber. Further compaction and loop formation results in additional higher-order structure and mitotic chromosome condensation leads to the final level of hierarchy of chromosome packaging.

(Munkel et al., 1999; Sachs et al., 1995). Both models propose that the 30 nm fiber is arranged as DNA loops with variable sizes which extend outwards from a central protein matrix and form a rosette-like conformation. Alternatively, the 'folded chromonema' model is based on light and electron microscopy and predicts that the 30 nm fiber is progressively folded, eventually resulting in a compacted chromonema fiber of 100-300 nm width (Belmont et al., 1989; Li et al., 1998a; Tumbar et al., 1999). The folding of this fiber is postulated to be organized either by nuclear scaffolding proteins or by non-specific interfiber interactions, possibly regulated by different histone variants or tail modifications (Hansen, 2002). However, given the intrinsic artificiality of the experimental setups to study higher order chromatin folding models it is uncertain whether one of these model systems are representative of normal interphase chromatin.

Heterochromatin vs. euchromatin

The interior of the eukaryotic nucleus is clearly non-homogenous and already in 1928 areas of differential compaction were distinguished by studying several species of moss (Heitz, 1928). Subsequent cytological staining and light microscopic studies confirmed this subdivision in two types of chromatin as a general hallmark of eukaryotic interphase nuclei. These observations resulted in the introduction of the terms heterochromatin and euchromatin to describe chromatin fractions showing differences in their degree of condensation. Heterochromatin is the term used for highly condensed, inaccessible chromatin, which is highly ordered in nucleosomal arrays and appears darkly stained throughout the cell cycle. Other regions that appear less densely stained and decondense as the cell progresses from metaphase to interphase are named euchromatin. Because of its higher accessibility, euchromatin is generally more easily transcribed and shows a more disorganised nucleosome array (reviewed in (Dillon and Festenstein, 2002)). An overview of other distinctive features between euchromatin and heterochromatin are presented in Table 1.

Euchromatin

Mammalian euchromatic regions are typically composed of unique, non-repetitive sequences with a high to variable gene density. A distinctive property of euchromatin is its relatively high GC-content (65%) as a result of the presence of unmethylated CpG islands. By contrast, bulk DNA is comparatively GC-poor (40%) and heavily methylated at CpG (Bird et al., 1985; Bird, 1986). CpG islands are associated with approximately 50% of all mammalian gene promoters and often contain multiple binding sites for transcription factors (Somma et al., 1991). Additionally, CpG islands almost completely lack linker histone H1 (which is involved in chromatin condensation) and are often identified as initiation sites for replication (Delgado et al., 1998). The presence of DNAseI hypersensitive sites as a result of its decondensation status is another typical feature of euchromatin. Specific histone modifications associated with euchromatic regions are the hyperacetylated lysines of histone H3 and H4 and the methylation of the lysine 4 residue of histone H3 (H3K4me) (Richards and Elgin, 2002). Taken together, euchromatin is comprised of gene-rich chromosomal regions and forms a relatively open and decondensed chromatin compartment thereby facilitating processes associated with gene activity.

Feature	Euchromatin	Constitutive heterochromatin
Staining/packaging during interphase	Decondensed	Condensed
DNA sequence	Predominantly unique	Predominantly repetitive (satellites; derivatives of viruses, transposons, etc.)
Replication timing	Early/throughout S phase	Late S phase
Chromatin structure	Disorganised nucleosomal array, relative accessible to nucleases, presence of HS sites	Regular nucleosomal array, less accessible to nucleases, loss of HS sites
Presence of genes	High/variable density GC-rich	Low density GC-poor
Activity state		
Euchromatic genes	Genes inducible	Genes silenced (variegated)
Heterochromatic genes	Genes silenced (variegated)	Genes inducible
Characteristic modifications	Histone H3 and H4 hyperacetylation	Histone hypoacetylation
	Histone H3K4me present	Histone H3K9me present
	Cytosine hypomethylation	Cytosine hypermethylation

Table 1. Distinction between Euchromatic and Heterochromatic domains (based on (Dillon and Festenstein, 2002; Richards and Elgin, 2002)).

Heterochromatin

Heterochromatin can be divided in two different populations: constitutive heterochromatin and facultative heterochromatin. Constitutive heterochromatin is mainly found at pericentromeric regions and is located at condensed regions that remain permanently silenced throughout the cell cycle. This type of heterochromatin is generally composed of long stretches of satellite repeats which can vary significantly in length and composition between species. For example, human pericentromeric heterochromatin is mainly composed of a 171-bp α-satellite repeat, whereas in mouse the pancentromeric 234-bp γ-repeats is the most predominant repetitive element present in constitutive heterochromatin. Besides these characteristics present in the primary DNA sequence, the assembly of constitutive heterochromatin involves complex patterns of histone modifications (such as H3K9me, H3K27me and H4K20me), the recruitment of chromodomain proteins such as heterochromatin protein 1 (HP1) and the binding of currently unknown RNA components (Jenuwein and Allis, 2001; Maison and Almouzni, 2004; Maison et al., 2002).

Facultative heterochromatin can be defined as developmentally regulated heterochromatin capable of undergoing a transition between the heterochromatic and euchromatic states. This transition occurs throughout development and differentiation when silencing of a particular subset of genes is needed for proper cell-type specific transcriptional regulation. A prominent example of this type of heterochromatin is the inactivation of either the paternal of maternal X chromosome in female mammals. This inactivated X chromosome remains silent through mitotic cell divisions and can be distinguished from its active counterpart by changes in its chromatin status such as differential histone modifications and DNA methylation (Heard and Disteche, 2006). The occurrence of position effect variegation (PEV) in which genes normally active in a euchromatic domain will typically be silenced when placed adjacent to or within a heterochromatic domain is another example of facultative heterochromatin. This silencing shows a variegating phenotype that is thought to reflect the stochastic nature of heterochromatin assembly at the euchromatin/heterochromatin boundary. Currently there are three well-characterized biochemical markers indicative of heterochromatin: DNA methylation at cytosine residues, hypoacetylation of histone lysine residues and methylation of histone H3 at lysine 9.

The most common form of DNA modification in eukaryotes and an important epigenetic mark that contributes to the stability of pericentromeric heterochromatin is DNA methylation (Okano et al., 1999; Xu et al., 1999). The post-synthetic addition of methyl groups to the 5' position of cytosine alters the structure of the major groove of DNA to which DNA binding proteins bind, thereby preventing transcriptional initiation of for example genes on the inactive X chromosome, imprinted genes and parasitic DNA (Jones and Takai, 2001).

The discovery that many transcription factors are associated with either histone acetyltransferase (HAT) or histone deacetylase (HDAC) activity suggests that histone acetylation

is important for regulating gene expression. Indeed, hypoacetylation of histone lysine residues is associated with both the formation of heterochromatin and subsequent gene silencing, probably by having a direct effect on the stability and higher-order packaging of nucleosomes (Strahl and Allis, 2000; Tse et al., 1998). Acetylation is however not generally associated with activation as examplified by the acetylation of histone H3 at lysine 4 (H3K4ac), which is associated with gene repression. Another covalent modification linked to heterochromatin is the methylation of histone H3 at lysine 9 (H3K9me). Initial experiments on *D. melanogaster* polytene chromosomes showed that the majority of H3K9me is present in pericentromeric regions and repetitive sequences (Jacobs et al., 2001). Methylated histones recruit proteins that either directly or indirectly modify chromatin and methylation of H3K9 results in high-affinity binding of HP1 (Bannister et al., 2001; Lachner et al., 2001). Besides attracting HDACs and other proteins implicated in heterochromatin formation, HP1 has the ability to multimerize through its chromoshadow domain and may thereby promote higher-order structures (Brasher et al., 2000; Cowieson et al., 2000; Yamada et al., 2005). The observation that that the methyltransferase Su(var)3-9 co-immunoprecipitates with HP1 led to a model in which the interaction between the two proteins and methylated H3K9 is involved in the propagation and stabilisation of the epigenetic code present in heterochromatin (Hall et al., 2002; Schotta et al., 2002). This ability to propagate is one of the key features of heterochromatin and influences gene expression of nearby genes in a sequence independent manner. Although this spreading generally causes epigenetic repression of nearby sequences, there are several reports in which formation of heterochromatin is required for activation of gene expression (Lu et al., 2000; Weiler and Wakimoto, 1995; Yasuhara and Wakimoto, 2006). Furthermore, a subset of transcribed genes can be found associated with histone modifications (H3K9me) and proteins (HP1) which typically characterise heterochromatin and possibly influence transcriptional elongation of linked genes (Greil et al., 2003; Piacentini et al., 2003; Vakoc et al., 2005b). These examples indicate that the organisation of the genome is complex and a division between transcriptionally silent heterochromatin and active euchromatin is, although generally correct, an oversimplification.

Modifications of the chromatin template

In general, gene expression is determined by the general accessibility of the DNA template to the many different components of the transcription machinery. Therefore, detailed knowledge about processes influencing the compaction and accessibility of the chromatin template is crucial for understanding how gene expression is regulated. The chromatin template carries numerous modifications of either histones or DNA. Currently, there are two not mutually exclusive proposals for the mechanistic function of these modifications. In the first view, histone modifications directly alter the chromatin packaging (either by changing the electrostatic charge or internucleosomal contacts) resulting in a more open or closed DNA polymer. This effect

directly influences the binding of DNA-binding proteins such as transcription factors. The other view is that attached chemical moieties to nucleosome surfaces or the DNA template promotes the association of chromatin-binding proteins. This binding of effector protein complexes to this so-called 'histone code' influences transcriptional output. Clearly, as the deposition of histone marks occurs at specific sites in the genome, binding of sequence-specific *trans*-acting factors is first required to attract the histone modifying enzymes.

Covalent histone modifications

Both the amino- and carboxy-terminal tails of histones can be covalently modified in several ways and certain modifications seem to correlate with either positive or negative transcriptional states (see Table 2).

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

Table 2. DNA modifications and post translational histone tail modifications (* yeast; *S. cerevisae* / # mammals, based on (Berger, 2007)).

One of the best studied examples of histone tail modification is the addition or removal of acetyl groups. Enzymes called histone acetyl transferases (HATs) function by transferring an acetyl group from acetyl-CoA to the amino group of certain lysine residues (Sterner and Berger, 2000). This neutralises the positive charged lysine residues thereby decreasing the affinity of the nucleosome for the DNA wrapped around it. The resulting hyperacetylation of histones correlates with increased transcription (Peterson, 2002). On the other hand, the opposing histone deacetylases

(HDACs) are often part of protein complexes associated with general repression of gene transcription (Narlikar et al., 2002).

The methylation of lysine at position 4 and 9 of the amino terminus of histone H3 (H3K4 and H3K9) has recently been allocated a central role in the framework of histone modifications (Lachner and Jenuwein, 2002). As indicated before, H3K9 methylation is mediated by the histone methyl transferase (HMTase) Suv39 and provides a high-affinity binding site for heterochromatin protein 1 (HP1). Suv39/HP1-mediated repression is mainly associated with constitutive heterochromatin, but might also be involved in gene repression at euchromatic targets (Nielsen et al., 2001). One possible mechanism for this silencing is the direct recruitment by H3K9 methylated sites of factors involved in DNA methylation (Tamaru and Selker, 2001). Conversely, methylation of H3K4 by specific HMTs like SET-7 and SET-9 appears to render chromatin permissive for transcription. H3K4me3 specifically localised to the 5' ends of ORFs and several complexes associated with transcriptional initiation and elongation (e.g. NURF, Chd1 and NuA3) bind to methylated H3K4 (Berger, 2007).

There is not always a strict division between active and repressive modification states. Complexes like Sin3/Hdac1 and JMJD2A, show binding to H3K4 methylated residues but are associated with transcriptional repression and not gene activation (Huang et al., 2006; Shi et al., 2006). Adding even more to the complexity is the fact that each lysine residue can be mono-, di- and tri-methylated resulting in a different chromatin state and associated protein complexes (Bannister et al., 2002; Iizuka and Smith, 2003). Clearly, the consequences of each modification are often context dependent. Altogether, this leads to the conclusion that regulation by histone modifications is dynamic and the presence of a specific modification does not necessarily indicate a unique regulatory status, making the interpretation of individual modifications more complex than previously thought (Guenther et al., 2007).

ATP dependent chromatin remodelling

Several basal transcription factors, like the TATA binding protein (TBP), can not bind to their target DNA sequence if this site is occluded by a nucleosome. Therefore, specialised classes of chromatin remodelling enzymes exist that are capable of facilitating activator binding by repositioning nucleosomes and adjusting nucleosome spacing. These transcriptional co-activators, which are present in every eukaryotic organism and show a high degree of conservation between species, all belong to the SNF2 family of DNA-dependent ATPases. They all have a helicase-like ATPase domain that use ATP hydrolysis to alter histone-DNA contacts (Cote et al., 1994; Kwon et al., 1994).

Based on the presence of other functional domains, ATP dependent remodellers are grouped into three different subclasses. The SWI/SNF family is characterized by the presence of a bromo domain which binds acetylated histones (Hassan et al., 2002). The family comprises yeast Snf2 and

Sth2, *Drosophila melanogaster* brahma (BRM) and mammalian BRM and brahma-like 1 (BRG1). By contrast, members of the ISWI family have a SANT domain which is thought to be involved in the binding of specifically modified histones (Boyer et al., 2004). This family consists of two ISWI homologues in yeast (Isw1 and Isw2) and mammals (SNF2H and SNF2L). Finally, members of the chromodomain and helicase-like domain (CHD) family are characterized by the presence of two amino-terminal chromodomains, which specifically interact with methylated histone tail (Bannister et al., 2001; Flanagan et al., 2005; Lachner et al., 2001; Sims et al., 2005). The presence of specific histone binding domains raises the possibility that different classes of remodelling enzymes are targeted to regions of specifically modified chromatin, implying functional interplay between ATP dependent remodellers and covalent modifiers. Question remains how these remodellers and modifiers that bind DNA in a sequence-independent manner are targeted to specific loci? In many cases, mammalian differentiation pathways require the combined activity of tissue-specific factors that regulate gene expression and remodelling enzymes (Bultman et al., 2005; Chi et al., 2002; Gresh et al., 2005; Kim et al., 2001; Williams et al., 2004). These sequence-specific transcription factors are able to bind specific DNA regulatory sequences, thereby targeting histone modifying complexes and remodelling complexes to specific loci. Given the different spatio-temporal expression patterns of these sequence specific transcription factors, this generates a mechanistic framework for the execution of different transcription programs at different times and in different cell types. Indeed, many transcriptional regulators such as C/EBPβ (Kowenz-Leutz and Leutz, 1999), MLL (Rozenblatt-Rosen et al., 1998), EKLF (Armstrong et al., 1998; Zhang and Bieker, 1998), NF-E2 (Cheng et al., 1997; Forsberg et al., 1999) and PPARγ (Erickson et al., 2001; Pedersen et al., 2001; Takahashi et al., 2002) can target SWI/SNF enzyme subunits or interact with HATs and HDACs. However, a number of transcription factors, like e.g. GATA-1 in erythroid cells (Rodriguez et al., 2005), have the ability to interact with both activating and repressing complexes, adding another level of complexity. So, the distinction between activation and repression is again not strict and depends on the sequence context and developmental stage.

RNA polymerase II transcription

Eukaryotic protein encoding genes are transcribed into an intermediate messenger RNA (mRNA) by an enzymatic holocomplex containing RNA polymerase II (RNAPII). The initiation of transcription by recruitment of RNAPII and general transcription factors to the promoter of a gene is a highly regulated process. In eukaryotes, this regulation occurs in the context of chromatin, which generally causes repression, necessitating another level of regulation. Consequently, RNAPII transcription also entails recruitment of chromatin remodelling complexes, such as ATP dependent remodellers and histone modifying enzymes. Transcription can be seen as a progression of ordered events where transcription factor interactions and

posttranslational modifications are spatially and temporally coordinated to ensure proper transcriptional progression. The process can be divided into a number of distinct steps consisting of pre-initiation complex assembly, initiation, promoter clearance, elongation and termination. In order to start transcription, a pre-initiation complex (PIC) needs to be assembled on a special DNA recognition sequence called a promoter. Pre-initiation complex assembly proceeds in a stepwise manner where TBP and its associated factors (TFIID) first recognize the core promoter TATA element (Muller and Tora, 2004). Subsequent binding of TFIIA stabilizes TFIID association on the promoter by counteracting inhibitory factors. *In vivo* promoter cross-linking studies in yeast indicate that stable binding of TBP to the TATA-box of an activated gene also requires the function of other general transcription factors like TFIIB and Mediator (Kuras and Struhl, 1999; Li et al., 1999). An early checkpoint for correct promoter complex assembly is the recognition of the TFIID/A/B complex at the TATA box by an RNAPII/TFIIF complex. Subsequent binding of TFIIE leads to the establishment of a proper RNAPII configuration prepared for open complex formation and promoter melting catalysed by TFIIH (Svejstrup, 2004). This indicates that the pathway to assembly of productive pre-initiation complexes should be viewed as a series of rapidly established equilibria between DNA-association and –dissociation of general transcription factors. Only if a complete pre-initiation complex is assembled at the promoter of a gene, the process of transcription is initiated and proceeds to the next step of promoter clearance. The phosphorylation of the C-terminal repeat domain (CTD) of the largest subunit of RNAPII by a kinase component of TFIIH triggers promoter clearance and thereby defines the initiation-to-elongation transition. It results in the phosphorylation of the serine 5 residue of the repeated heptameric Y-S-P-T-S-P-S sequence present in the CTD. Association of RNAPII with the Mediator complex enhances this TFIIH-mediated CTD phosphorylation up to several hundred-fold (Cosma et al., 2001; Kim et al., 1994; Naar et al., 2002). Additional CTD phosphorylation of the serine 2 residue to overcome an early elongation delay is targeted by Cdk9/P-TEFb (Svejstrup, 2004). Both CTD phosphorylation events are thought to disrupt interactions between RNAPII and associated protein complexes like Mediator and other general transcription factors. Simultaneously, new interactions are established with RNA maturation factors such as capping enzymes (Cho et al., 1997; Yue et al., 1997) and other mRNA processing factors (Proudfoot et al., 2002). Furthermore, CTD phosphorylation is also important for interactions between RNAPII and protein complexes related to transcriptional elongation, such as Set1/2 and Elongator (Gerber and Shilatifard, 2003; Hampsey and Reinberg, 2003; Otero et al., 1999). Although many details are still unclear, the phosphorylation mediated disruption of links to the initiation machinery and loading of elongation-specific factors leads to promoter clearance and, ultimately, transcript elongation. Efficient elongation results in the production of nascent RNA (nRNA) and requires that the CTD is kept phosphorylated during the entire length of the run by counteracting the activity of the CTD phosphatase Fcp1 which is responsible for removing CTD phosphates (Kobor et al., 1999; Lin et al., 2002). The identification of a large number of RNAPII associated factors that are

important for the ability to establish efficient transcript elongation underscores the complicated nature of this process (reviewed in (Svejstrup, 2002)). Transcriptional termination by RNAPII differs fundamentally from transcript initiation in that it does not occur at the same position every time, but rather in a zone situated downstream from fairly loosely defined terminator elements on the DNA template. After termination of transcription, the nascent RNA (nRNA) matures into mRNA after removal of intron sequences and covalent modifications of the 5' and 3' end. The 5' end is capped by the addition of a methylated G nucleotide important for initiation of protein synthesis and the 3' end is modified by the addition of a poly-A tail important for the export of mRNA from the nucleus and its stability in the cytoplasm (reviewed in (Proudfoot et al., 2002)).

Cis-regulatory elements and gene expression

As mentioned before, a fundamental process for the survival of every organism is the controlled expression of protein coding genes throughout development and differentiation. Therefore, eukaryotic genomes contain several elements that regulate proper spatio-temporal expression of genes. These so called *cis*-regulatory elements are relatively small DNA fragments (200-300 bp) that can bind both general and sequence specific factors and can be detected as DNAseI hypersensitive sites. Collectively they coordinate the interplay between the genetic information encoded in the primary DNA sequence and the RNA polymerase II machinery. Both *in vitro* and *in vivo* studies have identified several distinct *cis*-regulatory elements, including promoters, enhancers, silencers, Locus Control Regions (LCRs) and insulators.

Promoters

A promoter element can be defined as a stretch of DNA found directly upstream of the transcription start site of genes that directs the initiation of transcription. Promoters recruit, position and stabilise the RNA polymerase II transcription machinery to facilitate proper transcriptional initiation of linked genes. RNA polymerase II by itself is incapable of promoter recognition and therefore requires additional factors commonly known as GTFs (general transcription factors) to initiate transcription. One of the key factors capable of binding to the core promoter is the multisubunit TFIID complex that contains TBP (TATA-box binding protein) as well as over ten TAFs (TBP-associated factors). Binding of TFIID comprises the first step in promoter initiation complex (PIC) assembly and hallmarks the start of transcription initiation (reviewed in (Orphanides et al., 1996)). Early observations identified core promoter regions that are typically located between -35 to +35 nucleotides relative to the transcriptional start site. These elements are important for focused transcriptional initiation at a single nucleotide or within a narrow region of several nucleotides. These core promoter regions contain sequence motifs such as the TATA box, the TFIIB recognition element (BRE), the initiator element (Inr)

and downstream core promoter element (DPE) that are sufficient to drive RNA polymerase II transcription from nucleosome free DNA templates *in vitro* (Albright and Tjian, 2000; Juven-Gershon et al., 2006). All of these elements were originally identified in DNA viruses and highly expressed cellular genes which often contain TATA-boxes and show focused transcriptional initiation. However, this initial modular picture of core promoter structure probably needs to be redefined as subsequent studies of other cellular genes reveal that there are no true universal core promoter elements present in higher eukaryotes (Smale, 2001). For example, recent studies suggest that TATA-box elements are only present in approximately 10-15% of human core promoters (Bajic et al., 2004; Gershenzon and Ioshikhes, 2005; Kim et al., 2005). In fact, many mammalian core promoters even lack all of the above-mentioned core motifs and show transcriptional initiation at multiple weak start sites that are dispersed over a broad region of 50-150 bp (Sandelin et al., 2007). This so-called dispersed initiation occurs typically in GC-rich stretches of DNA located in CpG islands, that are often associated with housekeeping genes (Juven-Gershon et al., 2006; Smale, 2001). So, the exact mechanisms of transcriptional initiation at the promoters of eukaryotic genes are not yet clarified and it is likely that many undiscovered core promoter motifs will be identified in the future.

Enhancers and silencers

Transcriptional control of many mammalian genes not only relies on promoters but also on the presence of *cis*-regulatory elements collectively known as enhancers. Enhancers are capable of enhancing the basal transcription levels of a linked promoter and were originally identified in transient transfection assay as sequences capable of transcriptional activation of a linked promoter over large distances in an orientation independent manner (Banerji et al., 1981; Moreau et al., 1981). They show a modular organisation and resemble promoters without a transcriptional start site, containing a collection of protein binding sites capable of binding different classes of transcription factors. By binding both general transcription factors and developmental or cell type specific factors they provide additional specificity to the gene they activate. One of the first enhancers to be characterised was a tandem SV40 repeat of two identical 72 bp elements located 200 bp upstream of the start point of a transcription unit (Moreau et al., 1981). In yeast, enhancer-like elements called upstream activator sequences (UAS) are found that can function in either orientation at variable distances up to about 1 kbp upstream of the promoter (Guarente, 1988). In higher eukaryotes however, enhancers can be located up to several hundreds of kilobases upstream or downstream of the promoter. In order to upregulate transcription over large distances, these enhancers need to communicate with their linked promoters and several models of long-range enhancer action have been proposed (more detailed discussion in next section). Silencers are another type of regulatory elements involved in negatively controlling the transcriptional activity of genes (reviewed in (Ogbourne and Antalis, 1998)). Like enhancers they

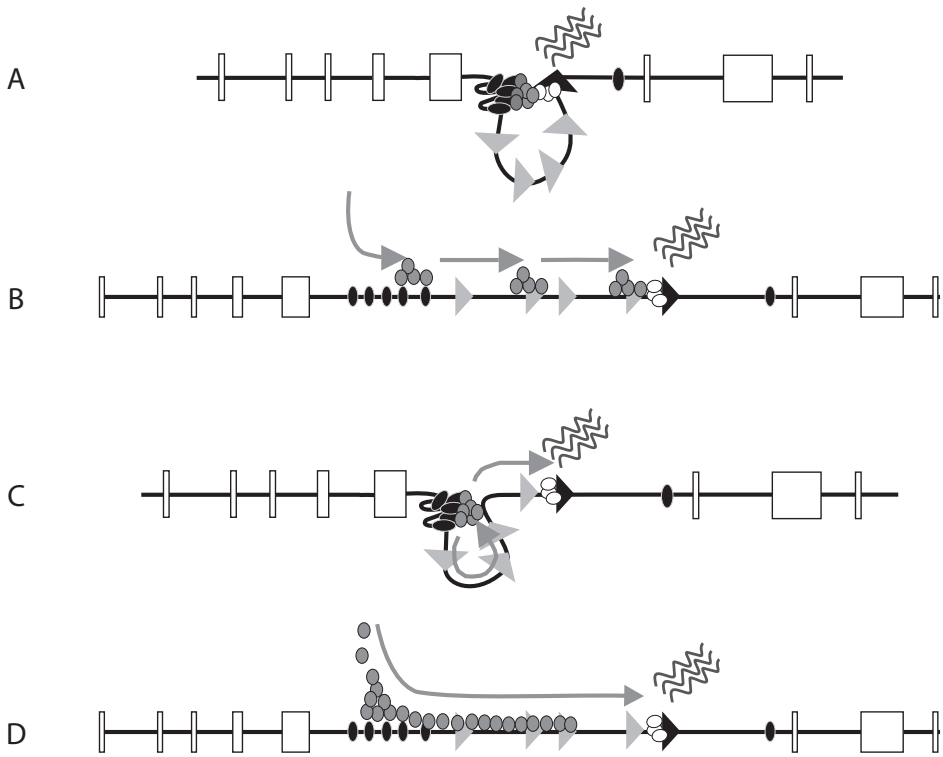

Figure 1.2

Proposed models for long-range enhancer-promoter communication.

Enhancer action over distance is exemplified by communication between the LCR and the globin gene promoter in the β-globin locus. Globin genes are depicted as triangles with actively transcribed gene in black and non-transcribed genes in grey. DNAse I HSs are indicated as black ovals, LCR-bound trans-acting factors as grey ovals and promoter bound factors (e.g. general transcription machinery) are depicted as white ovals. White boxes surrounding the locus represent olfactory receptor (OR) genes.

A. The looping model states that DNA-bound protein-protein interactions result in direct contacts between an enhancer (LCR) and the promoter of a gene. Contacts are established by random collision and are essential for transcription. As a consequence of LCR-interactions the intervening chromatin template is looped out. Note that the looping model, unlike other models, does not depend on the presence of intervening chromatin template for transcription, rather transcription depends on random collision between LCR-bound complexes and promoter-bound complexes.

B. The tracking model proposes that factors nucleate at the enhancer (LCR) and subsequently track along the intervening chromatin fiber towards the promoter of the gene (grey arrows indicate the movement of nucleated complexes along the chromatin template). Once a tracking complex reaches the promoter of a potentiated gene, transcription initiates.

C. The facilitated tracking model unites looping and tracking. This model states that complexes nucleate at the enhancer and form an LCR-protein complex, which tracks along the intervening chromatin template towards a potentially active gene (grey arrows indicate movement). Upon direct contact between the LCR-bound protein complex and the promoter, transcription initiates.

D. According to the linking model, LCR-promoter communication is established by the transmission of a nucleoprotein structure along the intervening chromatin template. This complex gradually polymerises and transcription initiates when it reaches the promoter of a potentiated gene (grey arrow).

contain multiple binding sites for regulatory proteins and act on a promoter in an orientation and position independent manner. The first silencing elements were identified in the mating type loci of yeast (Brand et al., 1985) (reviewed in (Dillon and Festenstein, 2002)) and subsequently in many more loci, including one in the human β-globin locus silencing the ε globin gene (Cao et al., 1989) and one in the chicken lysozyme gene (Baniahmad et al., 1987). Although the exact mechanism of silencing needs to be resolved it is evident that proper spatiotemporal regulation of gene expression depends on accurate cooperation between enhancer and silencer elements.

Mechanisms of long-range enhancer-promoter communication

The basic principle of gene activation by promoter-proximal DNA-binding proteins is primarily based on the concept of recruitment (Ptashne and Gann, 1997). According to this view, various protein factors participating in transcription initiation (GTFs, HATs and chromatin remodelers) are recruited in an ordered manner to activated promoters via protein-protein interactions with the enhancer-bound activator (Cosma, 2002). The recruitment of these proteins to the enhancer eventually results in an increased local concentration of the transcription machinery near the promoter. However, this view is mainly based on studies of enhancer elements in bacteria and eukaryotic promoters and upstream activating sequences (UASs) in yeast. A distinctive feature of these organisms is that most enhancers work on relatively short distances of less than a kilobase from their linked promoters (Guarente, 1988). In higher eukaryotes, enhancers need to communicate over considerable larger distances, often tens of kilobases up to one megabase away from their linked promoter (Kleinjan and van Heyningen, 2005; Lettice et al., 2003). However, recruitment does not work if an enhancer and promoter are separated in space, because in this case recruitment of proteins to the enhancer does not necessarily increase their concentration at the promoter. Therefore, other facilitating mechanisms are required for long-range communication between metazoan enhancer and promoter elements and several models have been proposed to explain enhancer action over distance (Figure 1.2). It is important to realise that most models are not mutually exclusive and combinations of different models have been proposed to explain experimental data generated at different loci and in different organisms.

The looping model

According to the DNA looping model, interaction of an enhancer bound activator protein with a protein at the promoter is accompanied by bringing them in close spatial proximity while the intervening DNA loops out (Ptashne, 1986). This mechanism results in an increase in the local concentration of the transcription machinery near the promoter via protein recruitment mediated by an activator bound to a distally located enhancer. In prokaryotes, DNA looping is a common way of communication among distantly positioned DNA sequences. *Lac* and *gal* repressors can make

stable DNA loops (reviewed in (Matthews and Nichols, 1998)) and the interaction of the AraC activator with RNA polymerase is accompanied by formation of a DNA loop (Lee and Schleif, 1989). Furthermore, bacterial σ^{54}-dependent enhancers activate their linked σ^{54}-dependent promoters via a DNA looping mechanism that also shares several key properties with eukaryotic enhancers such as orientation independent transcriptional activation over a large distance (Buck et al., 2000). Higher eukaryotes have more complex gene clusters with regulatory elements functioning over much greater distance and for a long time direct evidence for eukaryotic chromatin looping between enhancers and promoters was absent. However, a number of observations in eukaryotic systems could only be explained in a satisfactory manner by the looping model. Transvection is a naturally occurring phenomenon in *Drosophila*, where an enhancer on one chromosome activates a promoter *in trans* on the other, paired, homologous allele (Bickel and Pirrotta, 1990). This process, together with other *in vitro* experiments that show that an enhancer on one molecule can activate a promoter *in trans* on another DNA molecule (Dunaway and Droge, 1989; Mahmoudi et al., 2002; Mueller-Storm et al., 1989), demonstrates that a *cis* configuration of a promoter and enhancer is not an absolute prerequisite for interaction, as is only predicted by the looping model. Another observation most easily explained by the looping model is related to gene competition. Transfection assays using plasmids containing different number of genes and enhancers showed that multiple genes can compete for a single enhancer (de Villiers et al., 1983; Wasylyk et al., 1983). This observation was confirmed in the β-globin locus where two developmentally regulated chicken β-globin genes compete for a shared enhancer located between the genes (Choi and Engel, 1988). Subsequent transgenic experiments, in which the order of fetal γ-globin and adult β-globin relative to their shared upstream regulatory element (LCR) was changed, indicated that correct developmental expression depends on both gene order and relative distance to an enhancer element (Hanscombe et al., 1991). The authors proposed that the β-globin genes compete for contacting the LCR for their activation, with proximal genes having a competitive advantage over more distal genes. Indeed, as predicted by the looping model, the competitive advantage of an enhancer proximal gene is lost when genes are more closely spaced at further distance from the enhancer (Dillon et al., 1997; Hanscombe et al., 1991). Furthermore, fluorescence in situ hybridization (FISH) studies analysing ongoing transcription of the γ- and β-globin gene in a single cell show that either the one, or the other, but not both globin genes were active at a given time (Wijgerde et al., 1995). Subsequent studies confirmed these results (Gribnau et al., 1998; Trimborn et al., 1999), indicating the presence of alternate transcription caused by a stochastic 'flip-flop' mechanism of LCR action. Although above mentioned experiments are most easily explained by direct interactions of an enhancer with the promoter, none of them showed directly that two distal elements linked *in cis* come in close spatial proximity. Strong evidence in favour of the looping model was obtained for the mouse β-globin locus using two different biochemical approaches, the recovery of associated protein (RNA TRAP) assay (Carter et al., 2002) and the chromosome conformation capture (3C)

technology (Tolhuis et al., 2002). Both studies show that the endogenous mouse β-globin LCR is in close spatial proximity to the active β-globin gene promoter located 50 kbp away on the linear template while the intervening DNA loops out. Additional transgenic experiments analysing the human β-globin locus using 3C also showed proximity between the LCR and the active β-globin promoter (Palstra et al., 2003). 3C technology has subsequently been applied to many other gene loci and cell types and these studies show that a distantly encoded enhancer is located close to an active promoter. For example, it was shown that a lineage-restricted chromatin loop is formed between the Th2 LCR and its cognate genes encoding interleukin IL-4, IL-5 and IL-13 in CD4+ T-cells and natural killer cells. Moreover, both the transcription factors GATA-3 and STAT6 were shown to be required for these interactions (Spilianakis and Flavell, 2004). In B-cells, actively transcribed Igκ alleles exhibit mutual interactions over 22 kbp between three enhancers and Vκ gene promoters (Liu and Garrard, 2005). In addition to enhancer-promoter loops, parent-specific loop formation was observed between the differentially methylated regions of the imprinted insulin-like growth factor 2 and H19 locus (Murrell et al., 2004). Binding of the transcription factor CTCF to the H19 imprinting control region (ICR) prevented loop formation between the enhancers and the *Igf2* gene on the maternal allele (Kurukuti et al., 2006).

A recent study of the imprinted *Dlx5-Dlx6* locus shows that DNA looping can also play a role in gene silencing. It was shown that the methyl DNA-binding protein MeCP2 was required for the formation of a silent chromatin-loop. In the absence of MeCP2 the loop disappeared, new gene contacts were made with distant activating sequences and expression of *Dlx5* and *Dlx6* was upregulated (Horike et al., 2005). All these examples show that long-range DNA contacts are established by chromatin looping and are important for proper regulation of gene expression by remote control elements.

The tracking model

In the tracking model an enhancer functions as a landing platform for the assembly of a nucleoprotein complex which travels along the chromatin fiber and accumulates in the vicinity of the promoter (Herendeen et al., 1992). The best studied example of a transcriptional enhancer operating by the tracking mechanism is the enhancer activating the late genes of bacteriophage T4 (reviewed in (Kolesky et al., 2002)). There is however no conclusive evidence for enhancer action by a tracking mechanism in eukaryotes and no known example of an activator that has to leave the enhancer to activate transcription exists. A tracking mechanism is consistent with the enhancer-blocking properties of boundary or insulator elements by suggesting that the tracking protein complex is blocked by an insulator bound protein. On the other hand, tracking can not explain how an enhancer on one chromosome could activate transcription from an allelic promoter on another paired chromosome as in transvection (Dunaway and Droge, 1989; Mahmoudi et al., 2002; Mueller-Storm et al., 1989) as

well as the observation of alternate transcription caused by a stochastic 'flip-flop' mechanism of LCR action (Gribnau et al., 1998; Trimborn et al., 1999; Wijgerde et al., 1995). The detection of RNA polymerase II-dependent intergenic transcripts originating from enhancers (Gribnau et al., 1998; Kong et al., 1997; Tuan et al., 1992) has led to the proposal that RNA polymerase II might be the tracking protein. Experimental results are however not conclusive as one study showed that introducing a transcriptional terminator between the promoter and enhancer does not influence enhancer action (Muller et al., 1990), whereas a more recent study reports reduced enhancer function (Ling et al., 2004).

A 'facilitated tracking' mechanism for enhancer function was suggested that incorporates elements from both the looping and tracking model (Blackwood and Kadonaga, 1998). In this model, an enhancer bound complex containing DNA-binding factors and co-activators 'tracks' along the intervening chromatin fiber via small steps (and perhaps scanning) until it encounters the cognate promoter at which a stable looped structure is formed. Experimental evidence has been obtained from a study analysing the mechanism of action of the *HNF-4α* enhancer. A ChIP approach was used to analyse the distribution of the enhancer binding proteins C/EBPα and HNF-3β along different DNA regions *in vivo*. It was shown that these proteins could be cross-linked to the spacer DNA separating the enhancer and the promoter as well as to the enhancer itself. The cross-linking to the 6.5 kbp spacer DNA is only detected after activation of the enhancer, but before actual transcription of the gene (Hatzis and Talianidis, 2002). However, the enhancer-promoter communication is extremely slow (transcription starts 80 hours after induction of the enhancer) and therefore it seem likely that other mechanisms of communication are to be used at larger loci or when fast transcriptional activation is required. Again, it is hard to envision how this mechanism can explain activation of a linked promoter *in trans* or the observation of alternate transcription by a 'flip-flop' mechanism.

The linking model

In the linking model, enhancers act as landing platforms for DNA binding proteins that facilitate polymerisation of the proteins in the direction of the promoter thereby coating the chromatin fiber (Bulger and Groudine, 1999; Dorsett, 1999). If the distance between an enhancer and a promoter is large, it has been proposed that a wave of small protein-stabilized chromatin loops is initiated at the enhancer and moves towards the promoter. The model was proposed to explain the properties of the *Drosophila* CHIP protein. CHIP can not bind to DNA directly, but can interact with numerous transcription factors and facilitate their action over a distance *in vivo* (Morcillo et al., 1997; Torigoi et al., 2000). It was suggested that CHIP is recruited by an activator protein bound at an enhancer and works as a protein 'bridge' between the activator bound at the enhancer and other proteins having multiple weak binding sites between the enhancer and promoter (Dorsett, 1999; Gause et al., 2001). This model makes it difficult to explain the

results of the experiments suggesting a 'flip-flop' mechanism of enhancer action. In particular, it is difficult to see how the enhancer-proximal promoter can be activated after activation of the distal promoter.

Locus Control Regions

The presence of an enhancer as part of a transgenic construct is usually not sufficient to ensure high levels of expression in transfected cells or mice. Expression of the transgene is often low compared with endogenous levels and spatiotemporal regulation is frequently disturbed. These divergent expression patterns are probably caused by the inability of transgenic constructs to overcome the restrictive effects of the chromatin structure at the site of integration. This effect is caused by random integration into the host genome and classical enhancers are not able to shield a transgenic construct from this so-called position effect.

Early transgenic experiments in mice also showed that despite the fact that proximal regulatory elements of the human adult β-globin gene were sufficient for proper spatiotemporal expression, transgenic expression levels were influenced by position effects (Behringer et al., 1987; Kollias et al., 1987; Kollias et al., 1986). Analysis of a β-thalassemia patient resulted in the identification of a deleted region upstream of the globin genes that contained multiple erythroid specific DNaseI HSs (Forrester et al., 1986; Tuan et al., 1985). These data suggested that the deleted DNA segment contained an indispensable *cis*-acting regulatory element required for proper β-globin expression *in vivo*. Indeed, linkage of this region to a β-globin gene resulted in tissue-specific, position-independent and copy-number dependent expression of this transgenic construct in mice (Grosveld et al., 1987) and the identification of the β-globin Locus Control Region (LCR). Based on sequence homology between the human and mouse loci the mouse β-globin LCR was subsequently identified (Moon and Ley, 1990). Originally, LCRs are functionally defined as dominant regulatory elements that can confer position-independent, copy-number dependent and tissue specific activation of a transgene (Grosveld et al., 1987). They are structurally composed of varying numbers of tissue specific DNAseI HSs each harbouring multiple binding sites for transcription factors and typically include enhancer and insulator elements (Bonifer, 2000; Dillon and Sabbattini, 2000; Li et al., 2002a). After the discovery of the β-globin LCR many combinations of *cis*-regulatory DNA elements have been characterized in various vertebrate species that meet the functional definition of an LCR (reviewed in (Li et al., 2002a)). The most prominent property of all LCRs is their strong transcriptional enhancer activity. The absence of an LCR in transgenic mice results in severely reduced transcription of the human β-globin gene to <1% of the endogenous murine β-globin mRNA levels (Kollias et al., 1986; Magram et al., 1985; Townes et al., 1985) and targeted deletion of the LCR from its endogenous position strongly reduces the expression of the murine β-globin genes (Epner et al., 1998). Another property of the LCR is that it can overcome heterochromatin-mediated position

effect variegation (PEV) (Festenstein et al., 1996). Transgenic mice carrying a human CD2 or β-globin minigene linked with a complete LCR show normal levels of expression irrespective of their chromosomal integration site. Partial deletion of the LCR resulted in variegated expression and moreover these transgenic constructs were DNAseI insensitive in the non-transcribed portion of analysed cells (Festenstein et al., 1996; Milot et al., 1996). This shows that both the human CD2 and β-globin LCR are essential for establishing an open chromatin configuration, thereby suppressing PEV. Indicative of open, transcriptionally active chromatin domains is the presence of activating histone marks. Indeed, LCR elements from the human growth hormone (hGH) locus, T-cell receptor γ locus and the murine immunoglobulin heavy chain locus all appear to induce a regional increase in histone acetylation (reviewed in (Li et al., 2002a)). Moreover, studies on the mouse T-cell receptor α/δ indicate also a possible role of the T-cell receptor LCR in tissue-specific DNA demethylation preventing chromatin closure and gene silencing (Santoso et al., 2000). These results indicate that an LCR functions in the establishment and/or maintenance of an open chromatin domain allowing RNA polymerase II transcription in even highly restrictive chromatin environments. However, targeted deletion of the murine LCR shows that the β-globin locus is still DNAseI sensitive and histone marks reflect an open chromatin conformation indicating that in the endogenous mouse β-globin locus the LCR is dispensable for initiating or maintaining an open chromatin conformation (Bender et al., 2000a; Epner et al., 1998; Schubeler et al., 2001). This diversity and complexity regarding LCR function and mode of action highlights the importance to study *in vivo* transcriptional regulation in the context of whole loci, so that essential regulatory elements are not excluded or overlooked.

Insulators

Regulatory elements like enhancers are present genome wide and control target genes in *cis* over considerable distances. This flexibility in enhancer function gives these sequences the possibility to be highly promiscuous. Therefore, mechanisms must exist to ensure that genes are not activated in the wrong place or at the wrong time by enhancers from a neighbouring gene (Eissenberg and Elgin, 1991). Furthermore, a major fraction of vertebrate genomes is composed of repetitive, silenced DNA that exist as large regions of condensed chromatin (Ghirlando et al., 2004; Gilbert et al., 2004)). Chromatin condensation processes are self-propagating and can spread into neighbouring gene loci, potentially affecting their expression (Grewal and Moazed, 2003). Genes must therefore employ mechanisms to avoid influences of their genomic neighbourhood for achieving accurate temporal, spatial and responsive modes of expression. Insulator or boundary elements restrict expression patterns and avoid unwanted activation from *cis* regulatory elements, thereby demarcating domains of autonomous regulated gene expression. They have been experimentally identified based on two defining characteristics. Insulators interfere with interactions between enhancers and promoters and inhibit enhancer-activated

transcription when interposed between the affected enhancer and the promoter. Besides this so-called enhancer blocking activity, insulators also counteract chromosomal position effects by preventing the spreading of nearby heterochromatin, thereby acting as a boundary element (Figure 1.3).

The first characterised insulator element was discovered in *Drosophila* and was found to contain DNAseI HSs located near boundaries of the heat-shock gene locus *hsp70*. These elements were called specialised chromatin structure (scs/scs') and they delimited the region that becomes transcriptionally active in response to heat shock (Kellum and Schedl, 1991). Moreover, it was shown that scs and scs' function as positional enhancer blockers indicating that both insulator activities can be mediated by one kind of DNA element (Kellum and Schedl, 1992). Other *Drosophila* insulators include those found in the *gypsy* transposon and the *Fab-7* and *Fab-8* elements of the bithorax complex (Corces and Geyer, 1991; Hagstrom et al., 1996). Several DNA binding proteins have been identified that bind to these elements. The protein suppressor of hairy wing Su(Hw) is essential to enhancer-blocking properties of *gypsy* (Geyer and Corces, 1992) and zest-white-5 (Zw5) and BEAF-32 have been shown to bind to scs and scs' respectively (Gaszner et al., 1999; Hart et al., 1997). Insulator elements have been described in many organisms but surprisingly no significant similarity is evident among any of the insulator proteins found in *Drosophila*, yeast and vertebrates (Bell et al., 2001). The first vertebrate insulator described was the HS4 element of the chicken β-globin locus which showed enhancer blocking activity when placed between a reporter gene containing the β-globin promoter and an LCR (Chung et al., 1993). The minimal core required for insulation contains binding sites for

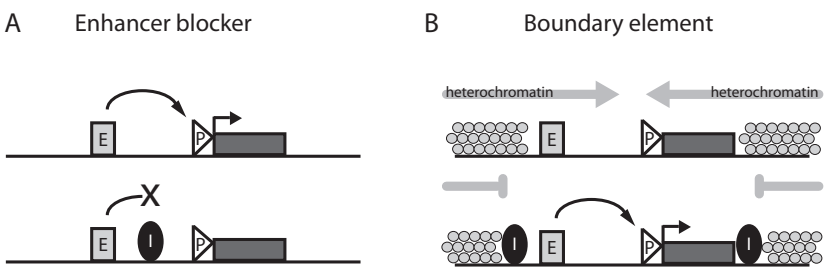

Figure 1.3

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

Insulators have been experimentally identified based on two defining characteristics: enhancer blocking activity and boundary function. Note that the relative position of an insulator is essential for its proper functioning.

A. Enhancer-blocker: an enhancer (E, grey box) can activate transcription (arrow) from a promoter (P, white rectangle) of a *cis*-linked gene (dark-grey rectangle). This enhancer activity is blocked when an insulator element (I, black oval) is positioned in between the enhancer and promoter.

B. Boundary element: spreading of nearby heterochromatin (grey arrows) into the locus prevents the transcriptional activation by an enhancer of a *cis*-linked gene. Insertion of two insulator elements flanking the locus prevents the spread of condensed chromatin, thereby precluding transcriptional inhibition.

the transcription factor CTCF and these sites are both necessary and sufficient for positional enhancer blocking activity (Bell et al., 1999; Bell et al., 2001). In addition, cHS4 can protect a transgene from position effects in *Drosophila* and early-erythroid chicken cell lines (Chung et al., 1993; Pikaart et al., 1998). The CTCF site appears unnecessary to protect against position effects, indicating that the enhancer blocking activity (mediated by CTCF) and boundary function of cHS4 are separable (Recillas-Targa et al., 2002). Additional experiments identified the proteins upstream transcription factor 1 and 2 (USF1/USF2) as likely candidates underlying boundary activity of cHS4. These proteins interact with footprint IV of cHS4, recruit PCAF and CBP/p300 to acetylate lysine 9 and 14 of histone H3, and SET7/9 to methylate lysine 4 of histone H3, thereby probably preventing the assembly of heterochromatin (West et al., 2004).

One key question that remains to be answered is what exactly underlies the molecular basis of insulator activity? Models explaining the enhancer blocking activity of insulators have been hampered by the lack of understanding how enhancers activate transcription in higher eukaryotes. The transcriptional model assumes that insulators have a direct effect on transcription and different mechanisms have been proposed. If it is assumed that a signal is propagated along the chromatin fiber from the enhancer to the promoter (tracking model) then insulators might block the propagation of the enhancer signal along the DNA (Dorsett, 1993). This signal could be, for example, a nucleosome modifying helicase complex or RNA polymerase II itself launched from the enhancer (Gaszner and Felsenfeld, 2006). However, this model can not explain the neutralisation of enhancer blocking activity when two tandem copies of a Su(Hw) insulator were introduced between an enhancer and a promoter (Cai and Shen, 2001; Muravyova et al., 2001). Alternatively, insulators could act as a promoter decoy, confusing the enhancer-bound transcription factor into interacting with the insulator instead of the transcription complex at the gene promoter (Geyer, 1997). Supporting this model is the fact that a promoter element has been detected within the scs and scs' insulator sequence (Glover et al., 1995) and enhancer blocking function strongly depends on enhancer and promoter strength in *Drosophila* (Cai and Shen, 2001). However, the decoy model fails to explain how an enhancer blocked on one side by an insulator can still activate a promoter on the other side. The structural model envisions insulators as sequences that organize the chromatin fiber within the nuclear space by creating transcriptionally independent domains via tethering to other elements or fixed structures in the nucleus. Consequently, enhancer blocking would result from the formation of different chromatin loops, in which the enhancer and promoter are separated. An assumption of this model is that the frequency of intra-loop enhancer-promoter interactions is higher than that of inter-loop interactions as a result of steric interference at the base of the loop. Evidence that the *gypsy* insulator establishes chromatin domains comes from observations that Su(Hw) and Mod(mdg)4 associate with approximately 500 sites in the *Drosophila* genome, but coalesce into 25 large structures named insulator bodies (Gerasimova et al., 2000). Furthermore, *in vivo* chromatin loop formation between opposite ends of the *hsp70*

locus is mediated via interaction between the insulator proteins BEAF-32 and Zw5 bound to scs/scs' insulator sequences (Blanton et al., 2003). Even interactions between two heterologous insulator sequences may occur in the genome, as indicated by the pairing of the GAGA factor and the *gypsy* insulator, resulting in bypassing of insulator activity (Melnikova et al., 2004). Chromatin loop formation might also explain the enhancer blocking activity of vertebrate insulator proteins. The ubiquitously expressed zinc-finger protein CTCF can form homodimers to generate clusters of loop domains either directly or via tethering the chromatin fiber to physical nuclear structures such as nucleolar surfaces or the nuclear matrix (Dunn et al., 2003; Yusufzai and Felsenfeld, 2004; Yusufzai et al., 2004).

Other studies have led to the development of several models explaining insulator boundary function. All models link boundary activity to the localized disruption of the polymerization-like reaction cycle underlying heterochromatin spreading. In these models barriers function as chain terminators by either modifying the nucleosomal substrate or by formation of looped structures via anchoring of the chromatin fiber (Gaszner and Felsenfeld, 2006).

One mechanism involves the creation of a nucleosomal gap, thereby disrupting the spread of chromatin-mediated silencing (Bi and Broach, 2001; Bi et al., 2004). Other forms of nucleosome modification in yeast involve the targeted recruitment of histone acetyltransferase (HAT) and ATP-dependent nucleosome-remodelling complexes (Oki et al., 2004). Additionally, the cHS4 vertebrate insulator recruits histone acetyltransferase and methyltransferase activity leading to local histone modifications that probably terminate the spread of heterochromatin formation (West et al., 2004). Although histone modifications appear to be necessary for boundary activity at the cHS4 insulator, they are clearly not sufficient to block heterochromatin encroachment. The same study shows that other protein binding sites in cHS4 are also required for boundary activity despite the fact that no histone modifications are seen at these sites (West et al., 2004). Finally, the tethering of barrier elements to fixed structures like the nuclear pore complex or anchoring via homotypic protein-protein interactions may create a steric hindrance preventing the propagation of heterochromatin (Ishii et al., 2002; Ishii and Laemmli, 2003).

Transcriptional regulation of the β-globin locus

Erythropoiesis

Erythrocytes residing in the circulating blood are of vital importance for the survival of vertebrates, taking up oxygen in the lungs or gills and delivering it to internal organs and tissues. The formation of red blood cells (erythropoiesis) starts from a multipotent hematopoietic stem cell (HSC) residing in the adult bone marrow. HSCs are capable of self-renewal, by producing daughter cells that retain stem cell characteristics, thereby expanding the HSC compartment. In addition, HSCs are pluripotent and can undergo multilineage differentiation under the influence of hematopoietic growth factors to give rise to every type of mature blood cell (Figure 1.4). The erythroid differentiation program is controlled by the expression of a combinatorial set of general and cell lineage specific transcription factors and is characterised by a series of distinct cell intermediates that progressively gain erythroid features and gradually lose proliferative capacities (Perry and Soreq, 2002). In the embryonic stage, the formation of red blood cells is initiated in the yolk sac in a process defined as primitive erythropoiesis. Definitive erythropoiesis starts in the liver during the fetal stage of development and finally resides in the adult bone

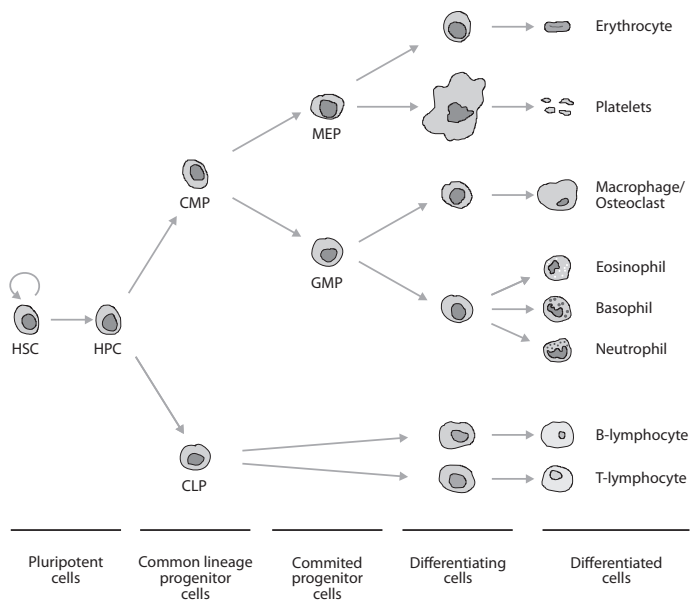

Figure 1.4

Schematic presentation of multi-lineage hematopoietic differentiation.

The hematopoietic stem cell is pluripotent and gives rise to all blood cells. Throughout differentiation, cells loose their proliferative capacities and become more restricted to one of the blood cell lineages. HSC= hematopoietic stem cell; HPC= hematopoietic progenitor cell; CMP= common myeloid progenitor; CLP= common lymphoid progenitor; MEP= myeloid and erythroid progenitor; GMP= granulocyte and monocyte progenitor.

marrow. The first stage of definitive erythroid differentiation following the HSC involves the formation of CFU-S (colony forming unit – spleen) which are still pluripotent but have lost their long-term repopulation capacity. The next progenitor is CFU-GEMM (colony forming unit – granulocyte, erythrocyte, macrophage and megakaryocyte) which can differentiate *in vitro* to all erythroid-myeloid specific cell types but has already lost its lymphoid potential (Johnson and Metcalf, 1977). The first erythroid lineage restricted progenitor is the BFU-E (burst forming unit – erythroid) followed by the formation of the CFU-E (colony forming unit – erythroid). This CFU-E stage is the last intermediate which has still considerable proliferative capacity and is characterised by its ability to form small colonies in semisolid medium (Wong et al., 1986). Next, the first morphologically recognizable differentiated member of the erythroid lineage, the pro-erythroblast, is formed. Around this stage, cells become positive for the TER-119 antigen which is used as a cell surface marker for the identification of cells late in the erythroid lineage (Kina et al., 2000). The completion of the erythroid differentiation program from the pro-erythroblastic stage takes place within 48 to 72 hours and results in the formation of reticulocytes. It is characterised by a number of rapid cell divisions followed by a decreased size of the nucleus and finally gives rise to fully matured, enucleated erythrocytes. The average lifespan of an erythrocyte is approximately 60 days in mice and up to 120 days in humans, which implies that red blood cells have to be replenished continuously from a pool of HSCs. Terminal erythroid differentiation is accompanied by strong chromatin condensation and the expression of a limited number of genes. Most notably, the α- and β-globin genes are highly expressed throughout differentiation and this results in the accumulation of high amounts of hemoglobin in erythrocytes. Hemoglobin is synthesized as a heterotetrameric protein, consisting of two α-like and two β-like globin chains, which can individually bind one heme group that binds oxygen and carbon dioxide in a reversible manner. The specific α- and β-globin chains synthesized differ between species and depend on the developmental stage of the differentiating red blood cell (see next section). Hemoglobin and its mRNA are the most abundant molecules in blood and this contributed to the fact that globin genes and associated diseases were among the first to be studied by biochemical and molecular biological methods. The erythroid compartment is well-characterised and accessible, making it an ideal model system to study developmental gene regulation and tissue specific transcription. As a result, the globin genes have played a major role in the identification and functional mechanisms of *cis*-regulatory DNA elements involved in proper transcriptional regulation.

Genomic organisation of the human and mouse β-globin locus

The human and the mouse β-globin locus are highly conserved and contain a number of β-globin like genes with some well-characterised *cis* regulatory elements like enhancers, promoters and a locus control region (LCR) (Figure 1.5). The genes are aligned in the order of their developmental expression and all encode for a β-globin like protein, which together with two

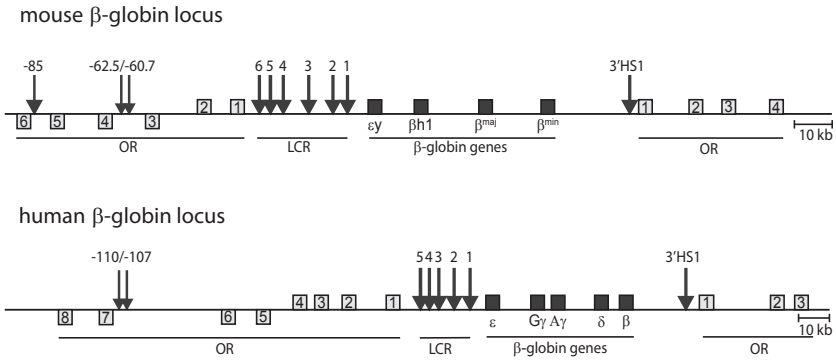

Figure 1.5

The mouse and human β-globin locus.

Schematic presentation of both the mouse and human β-globin locus. Boxes above the horizontal line represent genes transcribing from left to right; boxes below represent genes transcribed in the opposite direction. The β-globin genes, aligned in the order of their developmental expression, are indicated by black boxes. The β-globin LCR is underlined and hypersensitive sites are indicated by black arrows. Additional erythroid specific DNAseI HSs located upstream of the LCR and downstream of the β-globin genes are also indicated by black arrows. Olfactory receptor genes (OR, grey boxes) flank and partially overlap both loci.

α-globin like proteins and four heme groups form a functional hemoglobin molecule. All globin genes are derived from one ancestral gene which duplicated about 450 million years ago and gave rise to separate α- and β-globin genes. Both α- and β-globin genes duplicated several times, evolved independently and ended up on different chromosomes in birds and mammals (Gillemans et al., 2003; Hardison, 1998). The mouse β-globin locus resides on chromosome 7 and contains four functional genes starting 5' with εγ followed by β_{h1}, β_{maj} and β_{min} at the 3' side of the locus. During primitive erythropoiesis, yolk-sac derived erythroid cells express primarily εγ and β_{h1}, while the expression of β_{maj} and β_{min} is low. The generation of definitive erythroid cells commences in the fetal liver around E11 and is characterised by high level transcription of the two adult (β_{maj}/β_{min}) β-globin genes and inactive embryonic (εγ/β_{h1}) genes (Wawrzyniak and Popp, 1987; Whitelaw et al., 1990). Upstream of the genes lies a cluster of erythroid specific DNAseI hypersensitive sites collectively known as the locus control region (LCR), a dominant regulatory element that confers position-independent and copy number dependent expression to linked transgenes (Grosveld et al., 1987). Additional erythroid specific *cis* regulatory DNA elements can be found both upstream of the LCR and downstream of the genes. 3'HS1 is located 68 kb downstream of the εγ cap site (Tuan et al., 1985) and two other hypersensitive sites are located 85/84 kb (5' HS-84/-85) and 62/60 kb (5' HS-60/-62) upstream of the εγ cap site (Bulger et al., 2003; Farrell et al., 2000). The entire mouse β-globin locus is embedded in a cluster of olfactory receptor (OR) genes that are not expressed in erythroid cells (Bulger et al., 2000). The genomic organisation of the human β-globin locus is very similar to the mouse β-globin gene cluster and both the genes and *cis*-regulatory DNA elements show high levels of conservation

between the species (Bulger et al., 1999; Hardison et al., 1997) (Figure 1.5). The human locus is located on chromosome 11 and comprises five functional β-globin genes 5'-ε-Gγ-Aγ-δ-β-3'. The ε gene is expressed early in development when erythropoiesis is located in the blood islands of the yolk-sac. When the site of red blood cell formation switches to the fetal liver after approximately 10 weeks of development, ε expression is silenced with a concomitant increased expression of both γ genes. A second switch from γ to δ and β globin expression takes place around birth when the bone marrow becomes the main site of erythropoiesis. In addition to the LCR and the erythroid specific 3' HS1 hypersensitive site, a homologue of the mouse 5' HS-60/-62 *cis* regulatory element can be found 110/107 kb upstream of the ε gene. However, the *cis* regulatory element corresponding to the 5' HS-84/-85 seems to be absent in the human β-globin locus (Bulger et al., 2003).

The β-globin genes are relatively small genes of about 1500 bp encompassing three coding regions (exons) and two intervening sequences (introns) whose transcription results in the production of a protein of 146 amino acids. Like the gene structures, promoter sequences are very well conserved between the different globin genes. They can be detected as DNAseI hypersensitive sites in erythroid cells and are located within 200 bp upstream of their corresponding transcriptional start site. All β-globin promoters contain recognition sequences for both general and erythroid specific transcription factors thereby providing a mechanism for tissue specific promoter activation and gene regulation. For example, all globin promoters contain TATA, CCAAT and CAC box sequences, although minor variations exist. The functional significance of these sequences is illustrated by several naturally occurring mutations in for example the proximal CACC of the β-globin promoter resulting in β thalassemia (Kulozik et al., 1991) and in abolished γ gene silencing as a result of mutations in the γ-globin CCAAT box region (Collins et al., 1985). Like the promoters, β-globin enhancers are also detected as erythroid specific DNAseI hypersensitive sites (Groudine et al., 1983) and are located in close proximity or even within the β-globin genes (Behringer et al., 1987; Bodine and Ley, 1987; Kollias et al., 1987). Their importance is illustrated by the deletion of the 3' β-globin enhancer sequences which results in decreased transcription of the β-globin gene (Liu et al., 1997).

The β-globin Locus Control Region

Analysis of a Dutch γδβ thalassemia patient indicated the presence of an important distally located regulatory element upstream of the globin genes now known as the β-globin Locus Control Region (LCR). This patient carried a 100 kb deletion on one chromosome which resulted in the deletion of the ε and γ genes, but did not affect the β-globin gene and its proximal *cis* regulatory DNA elements. The unaffected chromosome showed normal globin expression and therefore it was concluded that a deletion far from the β-globin gene resulted in the suppression of its activity rather than a defect in *trans*-acting factors (Kioussis et al., 1983). Subsequent analysis of

this deleted region showed the presence of multiple erythroid specific DNAseI HSs (Forrester et al., 1986; Tuan et al., 1985) and linkage of this region to a β-globin gene resulted in tissue-specific, position independent and copy-number dependent expression in transgenic mice (Grosveld et al., 1987). Hence, the human LCR was identified and based on sequence homology the mouse LCR was discovered some years later (Moon and Ley, 1990).

The human LCR consists of five erythroid specific DNAseI hypersensitive sites located between 6-25 kb upstream of the ε gene. These HSs are 200-300 bp in size and contain binding sites for several transcription factors including GATA-1, NF-E2, EKLF and Sp1 and binding of these factors to their recognition sites is required for LCR hypersensitivity (Goodwin et al., 2001). Formation of the HSs precedes β-globin transcription (Blom van Assendelft et al., 1989), but the LCR needs to be linked to an active promoter to stay hypersensitive (Guy et al., 1996; Reitman et al., 1993; Tewari et al., 1996). Deletion of any individual HS from the β-globin LCR abolishes position independent expression of transgenes, suggesting that in a functional LCR all hypersensitive sites act together as a holocomplex (Ellis et al., 1996). Single cell β-globin mRNA analysis support the notion of an LCR holocomplex by demonstrating that both human (Gribnau et al., 1998; Wijgerde et al., 1995) and mouse (Trimborn et al., 1999) β-globin genes are alternately transcribed, i.e. only one gene is transcribed at any given moment. A fully functional LCR requires the presence of all HSs, but studying them individually shows that the contributions of single hypersensitive sites are not equivalent. Linking of the human 5' HS1 directly to a transgene does not result in expression (Fraser et al., 1990; Fraser et al., 1993), although this site does contribute to LCR function in the context of the complete locus (Milot et al., 1996). The main enhancer activity of the β-globin LCR resides in 5' HS2 (Tuan et al., 1989). This element behaves as a classical enhancer in transient transfection assays and its activity is mediated by the presence of a tandem NF-E2 binding site (Ney et al., 1990a; Ney et al., 1990b; Talbot and Grosveld, 1991). 5' HS2 contributes equally to overall expression throughout development (Fraser et al., 1993) and multi-copy integration of 5'HS2 linked transgenes is sufficient to drive position-independent and copy-number dependent expression (Ellis et al., 1993). 5' HS3 is the only element capable of controlling transcription as a single copy integration in transgenic studies, which suggests a chromatin opening or remodelling function (Ellis et al., 1996). Indeed, the activity of this HS largely depends on the transcription factor EKLF (Gillemans et al., 1998), which has been shown to interact with the acetyltransferases CBP/p300 (Zhang and Bieker, 1998) and is associated with the SWI/SNF chromatin remodelling complex (Zhang et al., 2001). HS3 is the most active site during the embryonic period and the only site capable of conferring high-level expression of γ- and β-globin during fetal erythropoiesis (Fraser et al., 1993). A core deletion of this element resulted in reduced expression of embryonic/fetal genes during primitive erythropoiesis while definitive erythrocytes showed loss of position independent β-globin expression (Navas et al., 1998). The highest activity of 5' HS4 is observed during the adult stage of development (Fraser et al., 1993), but this element is unable to drive expression of

a linked transgene as a single copy (Ellis et al., 1996). Finally, 5' HS5 was originally identified as a constitutive HS with insulator activity (Li and Stamatoyannopoulos, 1994; Li et al., 2002b; Tuan et al., 1985) and contains binding sites for the enhancer-blocking transcription factor CTCF (Farrell et al., 2002). The results mentioned above are mainly based on transgenic experiments and indicate that the human LCR is composed of multiple HSs with separable functions. However, the complex nature of these experiments makes interpretation of results sometimes difficult because they are inevitably influenced by experimental variables like transgene copy-number and position effects. This might explain why targeted deletion of endogenous hypersensitive sites in the mouse β-globin locus shows discrepancies with data obtained for the human locus. In line with the human LCR, mice containing a targeted deletion of the entire LCR show a severe reduction of β-globin gene expression (Epner et al., 1998), but individual properties of murine LCR hypersensitive sites are less well defined and seem to be more redundant. Targeted deletion of 5' HS1 and 5' HS4 results in a 10% decrease of adult β-globin expression (Bender et al., 2001), whereas deletion of 5' HS3 results in a small reduction of the embryonic εγ and β_{h1} genes, but a 30% reduction of adult genes (Hug et al., 1996). Likewise, targeted deletion of 5' HS2 results in approximately 70% residual adult gene expression, although the expression of embryonic genes is not influenced by this deletion (Fiering et al., 1995). Deletion of the 5' HS5 and 5' HS6, the latter not present in the human locus, has minimal effects on expression of the globin genes (Bender et al., 1998). Importantly, none of the hypersensitive site deletions affected the formation of the remaining sites, implicating that there is no dominant or initiating site whose formation must precede the formation of other HSs (Bender et al., 2000b). None of the individual HSs is necessary for proper developmental β-globin gene expression and a mechanistic framework emerges of an LCR holocomplex in which each site independently contributes to LCR function and has an additive effect on transcription (Bender et al., 2001; Fraser et al., 1993).

A domain of erythroid specific DNAseI sensitivity extends from approximately 10 kb upstream of the 5' HS-60/-62 to a few kb downstream of 3'HS1 in the endogenous mouse β-globin locus (Bulger et al., 2003). In erythroid cells, the human locus is also more sensitive to DNAseI than 'bulk' DNA (Weintraub and Groudine, 1976), which implies that both the endogenous mouse and human β-globin locus create a relatively open chromatin domain. The exact role of the LCR in the establishment or maintenance of this active chromatin environment is unclear and contradictory data exist between the human and mouse β-globin locus.

A naturally occurring 35 kb deletion in a Hispanic thalassemia patient results in the removal of 5'HS2-5 of the LCR, but leaves the distally located HSs (5' HS-107/-110 and 3'HS1) and all the β-globin genes intact. This deletion of almost the complete LCR results in complete abrogation of β-globin gene transcription, the locus becomes DNAseI resistant and the remaining HSs are not formed (Driscoll et al., 1989; Forrester et al., 1990). This data supports the notion that the LCR is required for chromatin opening of the endogenous human β-globin locus. However, deleting the human endogenous LCR by targeted recombination in an erythroid background

did completely abolish β-globin gene transcription, but did not influence general DNAseI sensitivity and histone acetylation levels throughout the human β-globin locus, whereas the locus on a chromosome from a Hispanic thalassemia patient remained DNaseI-insensitive and hypo-acetylated. The authors interpreted this as the LCR being necessary to open, but not to maintain an open chromatin formation in the human locus (Reik et al., 1998; Schubeler et al., 2000). These conclusions are in apparent contradiction with the targeted deletion of the LCR in the endogenous mouse locus (Epner et al., 1998). While none of the β-globin genes express beyond basal levels in these mice, they appear to have similar chromatin properties as their wildtype counterparts that express 25-100 fold more efficiently. The β-globin locus in these mice is still DNAseI sensitive and the remaining HSs were present, even after germline passage of the deletion. This suggests that elements elsewhere in the mouse β-globin locus are sufficient to establish and maintain an open chromatin conformation (Bender et al., 2000a). Recent data from erythroid progenitor cells show that there are intrinsic differences in chromatin modifications between the human and the mouse β-globin locus, which might give an explanation for observed functional differences. Both loci are subject to different epigenetic control mechanisms in hematopoietic progenitor cells and this difference is maintained when the human locus is introduced in transgenic mice, suggesting that the primary DNA sequence rather than the organism determines the difference in this epigenetic code (Bottardi et al., 2003).

Erythroid transcription factors

Evidently, an LCR is a *cis*-regulatory element important for the proper transcriptional regulation of linked genes. In order to execute its function, the β-globin LCR contains several binding sites for general and erythroid specific transcription factors (TFs). Transcription factors can be defined as proteins that recognize and bind specific sequences and regulate transcription through interactions with other factors. They are usually comprised of a DNA-binding domain and one or more regulatory domains and based on these functional domains they are classified into families. The DNA-binding domain physically interacts with the DNA template and typically contains structural motifs like e.g. helix-loop-helix, helix-turn-helix, zinc-fingers and leucine zippers (Pabo and Sauer, 1992). Transcription factors contain additional regulatory domains that either activate or repress gene transcription by interacting with other proteins or protein complexes (Buratowski, 1995). For instance, they can bind to proteins of the basal transcription machinery, thereby recruiting this complex to transcription initiation sites. Alternatively they can associate with proteins complexes involved in histone tail modifications or chromatin remodelling. In addition to DNA binding and activation or repression, most transcription factors have sites that are subject to post-translational modifications, which can influence protein stability, protein-protein interactions, subcellular localisation and DNA binding efficiency. Proper spatio-temporal regulation of the β-globin locus requires the presence of the appropriate

transcription factors, the correct fine-tuning of posttranslational modification and an orchestrated cooperation between multiple general- and erythroid specific transcription factors functioning in different complexes (Lemon and Tjian, 2000). Several hematopoietic transcription factors that regulate β-globin gene expression have been identified of which NF-E2, GATA-1 and ELKF are best studied and will be discussed in more detail.

NF-E2

NF-E2 was initially found to bind AP-1 sites in the promoter of the human porphobilinogen deaminase (PBGD) gene (Mignotte et al., 1989). Subsequent studies showed the presence of a tandem AP-1 site in 5'HS2 of the β-globin LCR which enhanced expression of reporter constructs in transfected cells (Ney et al., 1990b) and transgenic mice (Caterina et al., 1994; Talbot et al., 1990). Important insights regarding NF-E2 function have come from studies using the mouse erythroleukemia (MEL) cell line from which NF-E2 was initially purified (Andrews et al., 1993a). The transcription factor is a heterodimer composed of two basic leucine zipper proteins; a conserved hematopoietic specific subunit of 45 kD (p45 NF-E2) that contains a transactivation domain and a smaller, more widely expressed subunit of 18 kD (p18 NF-E2 or MafK) (Andrews et al., 1993a; Andrews et al., 1993b; Ney et al., 1993). Inactivation of the p45 NF-E2 gene by proviral integration in MEL cells shows that this transcription factor is important for β-globin gene expression. In the absence of p45 NF-E2, β-globin gene expression is drastically reduced, but expression levels could be restored upon reintroduction of p45 NF-E2 (Lu et al., 1994). Multiple regions of exogenous p45 NF-E2 are required to restore β-globin gene expression, including two discrete proline rich regions in the transactivation domain indicating that protein-protein interactions are important for p45 NF-E2 function (Bean and Ney, 1997). Indeed, the activation domain of p45 NF-E2 interacts with basal transcription apparatus component TATA-binding protein-associated factor $TAF_{II}130$ in erythroid cells and absence of this domain interferes with β-globin gene expression (Amrolia et al., 1997). Other studies demonstrated interaction between p45 NF-E2 and CREB binding protein CBP/p300 which results in increased nucleosomal HAT activity and acetylation of p45 NF-E2 (Chen et al., 2001; Cheng et al., 1997) and recent data suggests that p45 NF-E2 is involved in binding the H3K4 methyltransferase complex MLL2 to the β-globin LCR (Demers et al., 2007).

Tandem NF-E2 binding sites found within HS2 of both the human and mouse LCR are required for high levels of transcription in both MEL cells and transgenic mice (Caterina et al., 1994; Ney et al., 1990a; Ney et al., 1990b; Talbot and Grosveld, 1991) and NF-E2 directly binds to HS2 of the β-globin LCR (Forsberg et al., 2000). p45 NF-E2 itself interacts only very weakly with the NF-E2 consensus site as a monomer or homodimer and binding is mediated by the small-Maf (p18 NF-E2) subunit which interacts with the TGCTGA(C/G)TCA(T/C) consensus sequence, sometimes referred to as a Maf recognition element or MARE (Andrews et al., 1993b;

Igarashi et al., 1994; Motohashi et al., 1997). In addition to binding p45 NF-E2, Maf proteins can bind to MAREs as homodimers, or as heterodimers with non-erythroid transcription factors e.g. Fos/Jun (Kataoka et al., 1995; Kataoka et al., 1994), NF-E2 related factors (Nrf-1/Nrf-2/Nrf-3) (Caterina et al., 1994; Kobayashi et al., 1999; Moi et al., 1994) and Bach1 and Bach2 (Oyake et al., 1996). It was proposed that competition between different transcription factors for interaction with specific Maf subunits may be involved in erythroid cell differentiation (Kataoka et al., 1995). A first step in the switch from the repressed state to the activated state of the β-globin locus in MEL cells possibly involves the heme-induced displacement of the MafK dimerisation partner Bach1 (Sun et al., 2004). Upon erythroid differentiation an exchange of MafK-binding partners, from Bach1 to p45 NF-E2, is associated with the formation of a MafK/p45 NF-E2 heterodimer activator complex which results in β-globin gene expression (Brand et al., 2004). One study suggested that sequestration of both NF-E2 subunits in different nuclear compartments in non-induced MEL cells may be a rate-limiting step in NF-E2 complex formation, thereby inhibiting high level β-globin gene expression. Upon induction of erythroid differentiation, the β-globin gene loci relocate away from heterochromatin compartments and this relocation correlates with both transcriptional activation of the β-globin genes and relocation of MafK away from heterochromatin to euchromatic compartments in which p45 NF-E2 resides (Francastel et al., 2001). Studying differentiation associated recruitment in MEL cells by chromatin immunoprecipitation (ChIP) showed that activation of β_{major} globin gene transcription is associated with the recruitment of p45 NF-E2 to both the LCR and the active gene promoter (Sawado et al., 2001). However, the functional significance of NF-E2 binding to these regulatory sites is not understood completely. Several studies have linked p45 NF-E2 with the establishment of specific histone modifications associated with activation at the β-globin promoter (Im et al., 2003; Kiekhaefer et al., 2002). Moreover, RNA polymerase II recruitment to the β_{major} promoter in MEL cells requires the presence of p45 NF-E2, while association with the LCR is p45 NF-E2 independent (Johnson et al., 2001; Johnson et al., 2002). *In vitro*, p45 NF-E2 is able to facilitate RNA polymerase loading from the LCR to the β-globin gene (Vieira et al., 2004). These observations have led to the idea that p45 NF-E2 is required for the transfer of RNA polymerase II from the LCR to the promoter of the β-globin gene. Despite the importance of p45 NF-E2 for proper β-globin gene expression in MEL cells, mice lacking p45 NF-E2 show only a mild erythroid phenotype (Shivdasani and Orkin, 1995). The most remarkable aspect of mice with a targeted deletion of p45 NF-E2 is not seen in developing erythrocytes but is related to megakaryocyte differentiation. These mice are characterised by a profound loss of circulating blood platelets, which results in >90% neonatal lethality due to internal bleedings (Shivdasani and Orkin, 1995; Shivdasani et al., 1995). The discrepancy between findings in MEL cells and p45 NF-E2 knockout mice may reflect the possibility that NF-E2 function during development is provided by other redundant factors that are either absent or incompetent in MEL cells. Obvious candidates to functionally replace p45 NF-E2 in these mice are other members known to associate with Maf subunits. The factors Bach1 and

Bach2 lack a canonical transactivation domain and act as repressors in reporter assays (Yoshida et al., 1999), but factors of the AP-1 complex (Fos/Jun) and NF-E2 related factors (Nrf-1/Nrf-2/Nrf-3) are all potential candidates to functionally compensate for the loss of p45 NF-E2. The early embryonic lethality of Nrf-1 deficient mice prevents studying the combined deficiency of p45 NF-E2 and Nrf-1 (Chan et al., 1998). However, the combined absence of p45 NF-E2 and c-jun, p45 NF-E2 and Nrf-2 or p45 NF-E2 and Nrf-3, in double homozygote mice does not result in erythroid maturation defects beyond those seen with loss of p45 NF-E2 alone (Derjuga et al., 2004; Kuroha et al., 1998; Martin et al., 1998; Shivdasani and Orkin, 1995). Based on these results it was concluded that c-jun, Nrf-2 and Nrf-3 do not complement for p45 NF-E2 function *in vivo*. It should be mentioned however, that any functional compensation *in vivo*, if present at all, is incomplete given the consistent presence of erythroid abnormalities in p45 NF-E2 knockout mice. In this respect it should be mentioned that complete deletion of HS2 from the endogenous mouse β-globin locus, which harbours the most prominent NF-E2 binding sites, caused a drop in β-globin gene expression similar to that seen in p45 NF-E2 knockout mice (Fiering et al., 1995).

GATA-1

GATA-1 is the founder member of the GATA family of transcription factors, all binding to the DNA consensus sequence (A/T)GATA(A/G) (Ko and Engel, 1993; Martin and Orkin, 1990; Whyatt et al., 1993; Yamamoto et al., 1990). Initially identified as a protein that binds the β-globin 3' enhancer (Evans et al., 1988; Wall et al., 1988), GATA-1 was subsequently cloned from MEL cells (Tsai et al., 1989). GATA-1 is expressed in primitive and definitive erythroid cells (Fujiwara et al., 1996; Leonard et al., 1993), megakaryocytes (Martin et al., 1990; Romeo et al., 1990), eosinophils (Zon et al., 1993), mast cells (Martin et al., 1990), Sertoli cells (Ito et al., 1993) and dendritic cells (Gutierrez et al., 2007). It is an essential transcription factor for normal erythropoiesis as GATA-1 deficient ES-cells are able to contribute to all tissues in chimeric mice, with the exception of mature red blood cells (Pevny et al., 1991). GATA-1 knockout embryos die from severe anemia between E10.5 and E11.5 of development (Fujiwara et al., 1996) and production of erythroid precursors is arrested at the proerythroblastic stage due to apoptotic cell death (Pevny et al., 1995; Weiss et al., 1994; Weiss and Orkin, 1995). GATA-1 knockdown mice, in which the level of GATA-1 is 5% of the physiological level, lack primitive erythropoiesis and die from anemia between E11.5 and E12.5 (Takahashi et al., 1997), whereas knockdown mice that expresses about 20% of wild-type GATA-1 levels show a somewhat milder phenotype (McDevitt et al., 1997). The majority of these mice die between E13.5 and E14.5 with signs of defective primitive and definitive erythroid differentiation, although some of these anemic mice survive to adulthood. This shows that there is a direct relationship between GATA-1 expression levels and severity of the phenotype. Although a very low amount of GATA-1 in erythroid cells can be sufficient to prevent apoptosis, it is insufficient to promote proper erythroid differentiation

(Pan et al., 2005). GATA-1 binding sites are found in the globin gene promoters and in the core of HSs in the β-globin LCR. GATA-1 functions as either an activator or a repressor of gene expression, depending on the context of the binding sequence and its interaction with other proteins. For example, GATA-1 can activate ε-globin gene expression (Li et al., 1998b), but it can also function as a repressor when bound to the ε-globin gene in the presence of the ubiquitous transcription factor YY1 (Raich et al., 1995). In agreement with the role of GATA-1 as a transcriptional activator of globin expression, the recruitment of RNA polymerase II to the LCR and the β_{major} promoter is GATA-1 dependent (Johnson et al., 2002). Furthermore, GATA-1 is associated with the establishment of specific histone modifications associated with activation in the β-globin locus (Im et al., 2003; Kiekhaefer et al., 2002). GATA-1 controls a large number of genes and a recent study confirmed that GATA-1 can form different protein complexes with distinct transcriptional activity (Rodriguez et al., 2005). GATA-1 can interact via FOG-1 with the repressive MeCP1-complex, which results in the formation of a complex with histone deacetylase activity that binds to repressed early hematopoietic genes and genes of the eosinophilic lineage. The same study showed that GATA-1 can also interact with the essential hematopoietic transcription factor Gfi-1b and bind to repressed proliferative genes, while an activating complex of GATA-1 with TAL-1 and Ldb-1 was found at the enhancer of the active erythroid EKLF gene (Rodriguez et al., 2005). Furthermore, GATA-1 can homodimerise (Crossley et al., 1995) and has been shown to interact with many other transcription factors like EKLF/Sp1 (Merika and Orkin, 1995), FOG-1 (Tsang et al., 1997), PU.1 (Rekhtman et al., 1999), Rb (Whyatt et al., 1997) and CBP/p300 (Blobel et al., 1998). In conclusion, GATA-1 is important for the transcriptional regulation of multiple target genes and is essential for normal erythropoiesis. In the β-globin locus it can act either as a repressor or activator dependent on the protein complexes with which it associates and is thereby able to control globin gene expression at different developmental time points.

EKLF

The erythroid Krüppel-like transcription factor EKLF was initially discovered in a cDNA substraction assay between lymphoid and erythroid transcripts (Miller and Bieker, 1993). EKLF expression was subsequently found in the bone marrow and spleen of the mouse and further studies confirmed that this factor is indeed an erythroid-specific member of the SP/XKLF family of transcription factors (Nuez et al., 1995; Perkins et al., 1995; Southwood et al., 1996). Interestingly, the promoter of the EKLF gene contains a functional binding site for GATA-1, suggesting that the expression of EKLF is dependent on and downstream of GATA-1 (Crossley et al., 1994). Many promoters, including those of the β-globin genes, contain CACC boxes and are bound by SP1 and other Krüppel related proteins (Raich and Romeo, 1993). EKLF binds specifically to the CCACACCCT motif, which is found in the promoter of

the β_{major} gene and 5'HS3 of the LCR (Feng et al., 1994; Gillemans et al., 1998). 5' HS3 has been implicated in chromatin opening of the β-globin locus (Ellis et al., 1996) and EKLF has been shown to play a direct role in LCR function via this site by means of its association with CBP/p300 (Zhang and Bieker, 1998) and the SWI/SNF chromatin remodelling complex (Zhang et al., 2001). *In vitro*, EKLF interacts with the chromatin remodelling complex E-RC1 (Armstrong et al., 1998) and EKLF is required for the recruitment of E-RC1 subunits near the transcription initiation site of the β-globin promoter, suggesting that the complex uses EKLF for specific targeting (Lee et al., 1999). The importance of EKLF binding is illustrated by the strongly reduced expression of β-globin in β-thalassemia patients which carry mutations in the EKLF binding site in the β-globin promoter (Faustino et al., 1996; Feng et al., 1994; Orkin et al., 1984). Mice deficient for EKLF die from anemia around day E14, although the primitive blood cells function sufficiently for normal survival during embryonic development up to approximately day E12 (Nuez et al., 1995; Perkins et al., 1995). This led to the conclusion that EKLF is essential for the final steps of definitive erythropoiesis and may facilitate the completion of the fetal to adult switch in humans. Indeed, EKLF seems to play an important role in γ- to β-globin gene competition. This is illustrated by transgenic mice that contain the human β-globin locus and have a reduced level of EKLF, which results in reduced β-globin expression with a concomitant increase in γ-globin expression (Perkins et al., 1996; Wijgerde et al., 1996), while overexpression of EKLF results in an earlier switch from γ- to β-globin gene expression (Tewari et al., 1998). However, an EKLF-dependent reporter gene has been used to demonstrate that this factor also acts as a transcriptional activator in embryonic erythropoiesis (Tewari et al., 1998) and although the expression of the embryonic εγ and βh1 are not influenced by the absence of EKLF, primitive erythroid cells deficient for EKLF already have other defects. Most likely, this is a result of disregulated EKLF-dependent genes involved in hemoglobin metabolism and membrane stability (Drissen et al., 2005), indicating that EKLF regulates other essential genes in erythropoiesis in addition to the β-globin genes.

β-globin LCR mode of action

The most prominent property of the β-globin LCR is its strong transcriptional-enhancing activity. In order to upregulate transcription over distances up to 50 kb, the LCR needs to communicate with linked promoters of the β-globin genes. But what is the exact mechanism underlying this LCR mediated long-range gene activation in the β-globin locus? How does an LCR communicate with the β-globin genes and which factors or processes are involved in this communication? Evidently, understanding the *in vivo* function of the LCR is intimately associated with our knowledge of the process of β-globin gene activation. Therefore, a good point of departure for answering fundamental questions about the LCR mode of action are studies investigating the developmental regulation of β-globin gene expression.

Developmental regulation and gene switching

The developmental expression pattern of the globin genes has been studied extensively and is characterised by the switching of expression of one globin gene to another. This switch in globin gene expression coincides with changes in morphology of the erythroid cell, the site of erythropoiesis and hemoglobin composition (Stamatoyannopoulos and Grosveld, 2001). Expression of the embryonic genes occurs in primitive cells derived from the yolk sac. In contrast, the fetal and adult genes are expressed in definitive cells that originate from stem cells from the AGM region of the developing embryo (Ling and Dzierzak, 2002). In humans, developmental switching occurs twice, from embryonic (ε) to fetal (γ) and from fetal (γ) to adult (δ/β) globin gene expression. In mice, there is only one developmental switch from embryonic ($\varepsilon\gamma/\beta_{h1}$) to adult ($\beta_{maj}/\beta_{min}$) gene expression (Figure 1.6). Hence, the order of the mouse and human β-globin genes on the linear chromatin template coincides with the order of their developmental expression (Figure 1.5). Several studies have led to the proposal of a dual molecular mechanism underlying β-globin gene switching: autonomous gene control and gene competition for direct interaction with the LCR (Hanscombe et al., 1991; Peterson and Stamatoyannopoulos, 1993). The concept of autonomous gene regulation is largely based on studies of the human β-globin genes in transgenic mice. Integration of human transgenes with only proximal *cis*-regulatory DNA elements resulted in very low levels of globin expression, but this expression was tissue-specific and showed correct developmental timing, albeit in a position dependent manner (Behringer et al., 1987; Chada et al., 1986; Kollias et al., 1986; Magram et al., 1985). Transgenes of the human ε and γ genes that included their proximal regulatory elements together with an LCR, show normal expression in early developmental stages and proper silencing in the adult stage of development (Dillon and Grosveld, 1991; Raich et al., 1990). These silencing elements reside in their promoter region. Moreover, deletions and mutations of a 5' ε silencer element lead to improper developmental timing of the ε-globin gene (Raich et al., 1992; Shih et al., 1993) and point mutations in the promoters of the γ-genes give rise to non-deletion hereditary persistence of fetal hemoglobin (HPFH), which is characterised by continued expression of the fetal γ-genes in adult life (Stamatoyannopoulos and Grosveld, 2001). These observations indicate that elements responsible for developmental-stage specific expression of the human ε- and γ-globin genes lie within their proximal *cis*-regulatory DNA elements and accurate silencing of these genes is independent of the LCR.

Unlike the embryonic genes, transgenic studies with the adult β-globin gene directly linked to the LCR show high expression at all developmental stages (Enver et al., 1990; Grosveld et al., 1987). Correct developmental timing can be restored by introducing a human γ- or α-globin gene in-between the LCR and the β-globin gene, but not when the β-gene was the globin gene closest to the LCR (Hanscombe et al., 1991). Transgenic mice in which the endogenous γ-globin promoter is replaced with the β-spectrin promoter lack developmental switching, indicating that

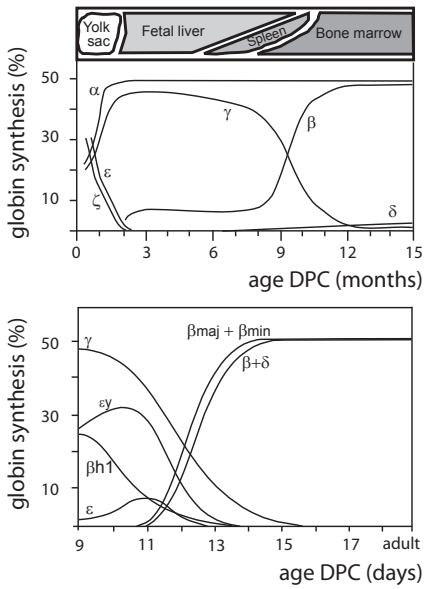

Figure 1.6
Developmental regulation of globin synthesis. Top panel shows the expression levels and site of expression of the individual human globin genes during development. The ζ- and α-globin genes are located in the α-globin locus. The α-like and β-like globin proteins expressed at the same stage of development can form different types of hemoglobin molecules. Bottom panel shows the expression levels of the individual mouse β-globin genes and human β-globin transgenes when expressed in mice.

the γ-globin promoter is necessary and sufficient for the suppression of the β-globin gene at early developmental stages (Sabatino et al., 1998). So, the adult β-globin gene can be expressed at early developmental stages, but the presence of a competing γ-gene more proximal to the LCR prevents its activation. This led to the idea that β-globin developmental gene expression is regulated through competition of the genes for LCR function based on gene order or relative distance. Indeed, this competition model was confirmed in the context of the complete globin locus. The importance of proximity to the LCR was demonstrated by introducing a 'marked' β-globin gene (β_m) that resulted in a competitive advantage of this marked gene over the endogenous β-globin gene when it was located more proximal to the LCR. Upon increasing the distance between the LCR and the proximal β_m gene, and thereby decreasing the distance between β_m and the endogenous β-globin gene, the competitive advantage of the more proximal gene diminished (Dillon et al., 1997). Additionally, inverting the gene order of the locus activated the now more proximal β-globin gene at early stages of development, whereas expression of the more distally located γ- or ε-gene was abolished (Hanscombe et al., 1991; Tanimoto et al., 1999). All these experiments suggest that competition of the genes for LCR function is based on gene order and relative distances. The competition mechanism was further supported by FISH studies analysing pre-mRNA transcripts of both the γ- and β-globin gene in a single cell. This showed that either the one, or the other, but not both globin genes were active at a given time, indicating that switching between globin genes is continuously dynamic and both genes can be alternately transcribed until the γ-genes are autonomously silenced (Gribnau et al., 1998; Wijgerde et al., 1995). A similar mechanism was found for the mouse β-globin locus (Trimborn et al., 1999), indicating the presence of alternate transcription caused by a stochastic 'flip-flop' mechanism of LCR action.

Regulation by *trans*-acting factors plays a crucial role in autonomous gene silencing and hemoglobin gene switching and they can favour activation or silencing of β-globin genes regardless of gene order (Peterson and Stamatoyannopoulos, 1993). For example, a transgenic human β-globin locus expressed the γ-genes at much higher levels than expected, suggesting that in these mice the *trans*-acting environment in embryonic erythroid cells favours LCR-γ interactions, possibly as a result of multiple changes in DNA sequences and transcription factor balance (Strouboulis et al., 1992). A protein complex called direct repeat erythroid definitive (DRED) binds the ε-promoter *in vitro* and is required for the autonomous silencing of the ε gene (Tanimoto et al., 2000). The core of this complex consists of two nuclear orphan receptors (TR2 and TR4) that can bind to direct repeat sites in the ε- and γ-promoters and it was suggested to act as a repressor of embryonic and fetal transcription in definitive erythroid cells (Tanabe et al., 2002). In addition, the erythroid Krüppel-like transcription factor (EKLF) plays a key role in β- and γ-globin gene competition during hemoglobin gene switching. Mice with reduced levels of EKLF display a reduced number of transcriptionally active β-genes, with a reciprocal increase in the number of transcriptionally active γ-genes, although the timing of γ-globin silencing remains the same. Total absence of EKLF results in a further increase in the number of transcriptionally active γ-genes, while β-gene transcription is abolished (Wijgerde et al., 1996). Additional studies on EKLF showed that this factor is required for initiating or stabilising contacts between the LCR and the actively transcribed β_{major} gene in definitive erythroid cells (Drissen et al., 2004). Like EKLF, the erythroid transcription factor GATA-1 and its cofactor FOG-1 are also required for the physical interaction between the β-globin locus LCR and the active β_{major} promoter (Vakoc et al., 2005a). Together, this work highlights the importance of *trans*-interacting factors and indicates that changing the *trans*-acting environment is fundamental for proper developmental regulation and gene switching in the β-globin locus. The adult β-globin genes have the potential to be active at all stages of development, but are transcriptionally inactive due to the presence of the more proximal embryonic and fetal genes. By changing the *trans*-acting factors the embryonic genes are silenced and the competitive advantage shifts to the adult genes that become activated. Altogether, the β-globin locus is characterised by competition among the different β-globin genes for LCR function. The transcriptional outcome of this competition is mainly determined by gene order, relative distance to the LCR and the presence of different *trans*-acting factors that are able to silence or activate individual β-globin genes.

Communication between the LCR and the β-globin genes

Several mechanisms have been proposed that may explain the communication between the β-globin genes and the LCR. However, certain aspects of β-globin developmental transcriptional regulation can only be explained satisfactory by the DNA looping model. According to this model, interaction of an enhancer bound activator with a protein at the promoter is accompanied by

bringing them in close spatial proximity, while the intervening DNA loops out (Ptashne, 1986). This mechanism eventually results in an increase in the local concentration of the transcription machinery near the promoter via protein recruitment mediated by an activator bound to a distal enhancer. The model assumes that the chromatin fiber is flexible which enables dispersed DNA fragments to interact with each other via random collision. The frequency of collision between two sites is inversely correlated with genomic site separation and depends on the flexibility of the chromatin fiber.

All other models explaining promoter-gene communication, e.g. tracking, facilitated tracking and linking, assume that enhancers somehow emit a signal that travels along the intervening chromatin fiber towards the promoter of a gene that gets activated (Figure 1.2). Unlike these models, the looping model does not depend on the presence of intervening chromatin for transcription. A key observation in support of the DNA looping model in the β-globin locus is the presence of gene competition among the different β-globin genes for LCR function, with proximal genes having a competitive advantage over more distal genes (Hanscombe et al., 1991). This competitive advantage of an enhancer-proximal gene is lost when genes are more closely spaced at further distance from the β-globin LCR (Dillon et al., 1997), which is only predicted by the looping model, but not by any other model of long-range enhancer-promoter communication. Another aspect of β-globin gene regulation in strong agreement with a looping mechanism is the alternating transcription caused by a stochastic 'flip-flop' mechanism of LCR action (Gribnau et al., 1998; Trimborn et al., 1999; Wijgerde et al., 1995).

This suggests that the LCR can alternate between the fetal and adult genes for their activation and, most importantly, that the process of β-globin gene transcription is intermittent. This implies that the multiple regulatory sites of the β-globin LCR cluster to form a holocomplex that can only contact one β-globin gene at any given time. All these experiments support the idea that regulatory DNA elements in the LCR of the β-globin locus physically contact downstream genes via a looping mechanism to regulate their expression. However, none of them directly showed *in vivo* that two distal elements linked in *cis* are in close spatial proximity to each other, with the intervening DNA looping out. The recent development of two unrelated technologies made it possible to study the spatial conformation of the β-globin locus in unprecedented detail and confirmed that in erythroid cells the LCR is in close proximity to the promoter of active β-globin genes.

Looping in the β-globin locus: the Active Chromatin Hub

In 2002, two different biochemical techniques were developed that demonstrated that the β-globin LCR is in close spatial proximity to the actively transcribed gene: RNA TRAP (Carter et al., 2002) and Chromosome Conformation Capture (3C) technology (Dekker et al., 2002).

RNA TRAP (RNA Tagging and Recovery of Associated Proteins) involves the targeting of horseradish peroxidase (HRP)-labeled probes against nascent RNA transcripts to the site of active transcription. After addition of biotin-tyramide, the localised peroxidase becomes a highly reactive radical intermediate that covalently deposits biotin on nearby chromatin. After fragmentation of the chromatin, the biotin labelled chromatin is enriched and using PCR, DNA elements are identified that were originally in close proximity to the site of active transcription. After application of this technique to the actively transcribed mouse β-globin, gene, it was shown that a peak of biotin deposition was observed at 5'HS2 of the LCR (Carter et al., 2002). This implies that the LCR is indeed in close spatial proximity to the actively transcribed β-globin gene as predicted by the DNA looping model for enhancer-promoter communication. However, RNA TRAP can only be directed to actively transcribed genes, meaning that results only represent spatial proximity between elements in the transcribed section of a population of cells. Therefore it is impossible to determine if interactions identified by RNA TRAP are present in non-transcribed loci as well. Moreover, the dependence on nascent transcripts may hamper the analysis of interactions with genes that are less efficiently transcribed and precludes analysis between non-genic elements. Chromosome Conformation Capture (3C) is a powerful technique which enables the mapping of *in vivo* interactions between chromosomal regions irrespective of their transcriptional status and without the necessity to modify the locus of interest. Originally developed to study the spatial

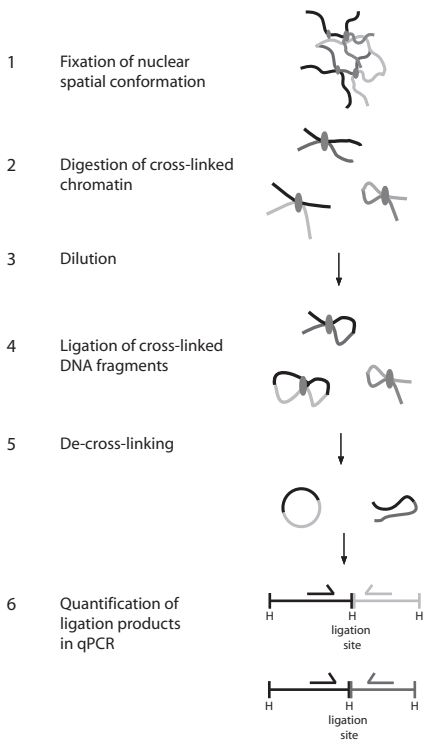

Figure 1.7
3C technology.
Schematic diagram showing the principles of the Chromosome Conformation Capture (3C) technology. The initial step in 3C involves the fixation of whole cells using formaldehyde (1). This step forms cross-links between DNA fragments that are close together in nuclear space. Next, the cross-linked chromatin is digested with an excess of restriction enzymes (2). This separates cross-linked from non-cross-linked DNA fragments. After dilution, cross-linked DNA fragments are ligated (3/4). Note that intramolecular ligation is favoured as a consequence of the heavy dilution. The chromatin crosslinks are subsequently reversed (5) and ligation events between selected pairs of restriction fragments are quantified by qPCR, using primers specific for the given fragments (6).

conformation of chromosomes in yeast (Dekker et al., 2002), this technique was adapted to study the spatial conformation of a 200 kb region encompassing the entire mouse β-globin locus (Tolhuis et al., 2002). An outline of the 3C procedure is given in Figure 1.7. In short, 3C uses formaldehyde on whole cells to trap interactions between chromatin segments by cross-linking proteins to proteins and proteins to DNA. Next, the cross-linked chromatin is digested using a restriction enzyme, followed by ligation under heavily diluted conditions, thereby favouring intramolecular ligation events between DNA fragments that were originally in close spatial proximity. Finally, quantitative PCR across ligation sites with primers diagnostic for DNA fragments of interest gives a measure for proximity frequencies between selected fragments (Dekker et al., 2002; Splinter et al., 2004). Using 3C on the endogenous β-globin locus in expressing fetal liver cells, it was found that the actively transcribed adult β-globin genes contacted the HSs of the LCR, while the intervening DNA harbouring the non-expressed embryonic genes looped out. Importantly, no such interactions were found in non-expressing brain tissue, where the β-globin locus adapted a more or less linear conformation (Tolhuis et al., 2002). Thus, two independent techniques provided direct evidence that regulatory elements in the β-globin LCR loop towards their downstream target genes in cells where the β-globin genes are actively transcribed.

The spatial clustering with the active β-globin genes is not confined to *cis*-regulatory DNA elements residing in the LCR only. In addition, 3C technology also revealed erythroid specific long-range interactions with two other sets of hypersensitive sites upstream (HS-85 and HS-62/60) and downstream (3'HS1) of the β-globin locus. The spatial clustering of these *cis* regulatory elements together with the LCR and the active genes in the β-globin locus of mature erythroid cells is called an Active Chromatin Hub (ACH). Importantly, the transcriptionally silent olfactory receptor (OR) genes located between the upstream HSs and the LCR, like the inactive embryonic genes, do not participate in this clustering, but loop out (Tolhuis et al., 2002) (Figure 1.8). As mentioned before, erythroid differentiation is characterised by a strong increase in β-globin gene transcription. In erythroid progenitor cells, the transcription levels of the β-globin genes are already comparable to those seen for most housekeeping genes. During later stages of erythroid differentiation, this relatively high basal level of expression increases to extremely high levels of transcription. These high transcription rates are dependent on the presence of the LCR as illustrated by the deletion of the LCR from its endogenous locus in the mouse which results in only 1-4% of β-globin gene expression when compared to wild-type littermates (Epner et al., 1998; Schubeler et al., 2001). When 3C was applied to erythroid progenitor cells that express the β-globin genes at basal levels, a pre-structure called the Chromatin Hub (CH) was found (Palstra et al., 2003). This β-globin CH consists of interactions between the outer hypersensitive sites and the 5' side of the LCR, while the β-globin genes and HS1-3 of the LCR loop out. Interestingly, upon induction of erythroid differentiation, the locus changes its conformation and adopts an ACH spatial conformation in which contacts between the entire LCR and the active genes are established (Figure 1.8). These experiments suggest that ACH formation coincides with, and is

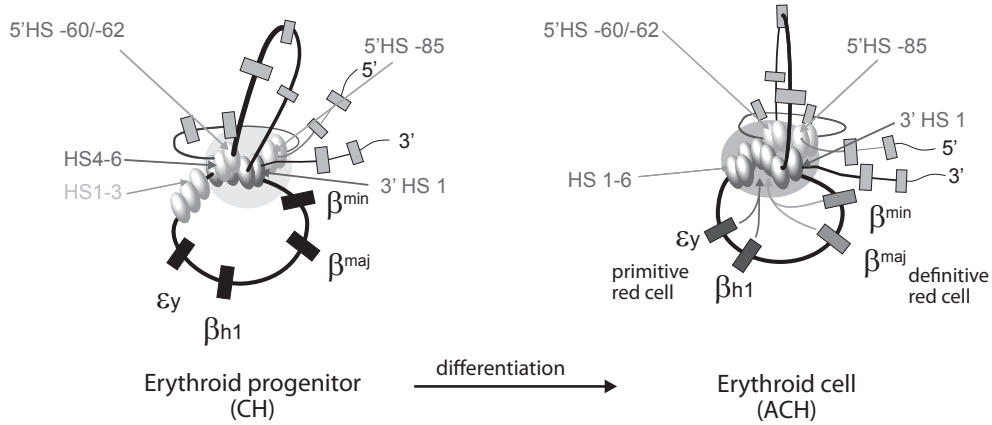

Figure 1.8

Clustering of cis-regulatory elements of the β-globin locus result in an erythroid specific spatial conformation: the Active Chromatin Hub (ACH).

2D-presentation of 3-dimensional interactions that occur between regulatory DNA elements (grey ovals) of the β-globin locus in erythroid progenitor cells (left) and definitive erythroid cells (right). β-globin genes are depicted as rectangles (inactive: black; active: light or dark grey), Olfactory receptor genes (grey squares) are inactive in erythroid cells. In erythroid progenitor cells, that express the β-globin genes at basal levels, a substructure called the Chromatin Hub (CH) is present. This substructure is formed through interactions between the outer hypersensitive sites (HS-85, HS-62/60 and 3'HS1) and the 5' side of the LCR, while the β-globin genes and HS1-3 of the LCR loop out. Upon induction of erythroid differentiation, the β-globin gene that gets activated and the rest of the LCR stably interact with this substructure to form a functional Active Chromatin Hub (ACH). The core of the ACH is erythroid specific and developmentally stable; a developmental switch occurs in globin genes entering this nuclear compartment, as depicted by the arrows.

required for, an increase in β-globin gene expression level going from LCR independent, basal levels, to extremely high levels that are dependent on the LCR (Palstra et al., 2003). Moreover, 3C studies on the β-globin locus at the embryonic stage of development demonstrated that the β-globin genes switch their interaction with the ACH in relation to their transcriptional status. In primitive cells, the active embryonic genes, and not the inactive adult genes, spatially clustered with the LCR and the outer hypersensitive sites and this appeared to be conserved between the mouse and the human β-globin locus (Palstra et al., 2003).

Functional significance and formation of long-range DNA interactions

The experiments mentioned above clearly demonstrate that distant *cis*-regulatory elements in the β-globin locus physically interact to control gene activity, as predicted by the looping model (Tolhuis et al., 2002). Presumably, ACH formation results in a high local concentration of DNA-binding sites for cognate transcription factors and their interacting partners (e.g. HATs and chromatin remodelling complexes), which consequently accumulate at the site. Since efficiency of transcription is proportional to the concentration of transcription factors involved (Droge and Muller-Hill, 2001), DNA contacts formed in the context of the β-globin ACH may therefore be

necessary to drive efficient β-globin gene expression. Analogous to the microscopically visible nucleolus, where nucleolar organising regions from different chromosomes (rDNA) physically interact to form a nuclear entity dedicated to the transcription of the ribosomal genes by RNA polymerase I, the β-globin ACH could represent an example of a structural entity dedicated to efficient RNA polymerase II gene transcription (de Laat and Grosveld, 2003).

But how exactly are contacts established between two sites separated by several kilobases on the linear chromatin template? To explain how a loop can be established between an enhancer and a promoter a number of different models have been proposed. Both the tracking and the linking model assume that enhancers function as a landing platform for a signal that subsequently propagates along the intervening DNA from the enhancer to a promoter. This signal can either be an enhancer bound nucleoprotein complex scanning the DNA (tracking) or a growing chains of proteins that nucleates at the enhancer (linking). Alternatively, the presence of intergenic transcription in the β-globin locus has led to the suggestion that transcription underlies the establishment of lang-range interactions between regulatory elements (Ling et al., 2004; Tuan et al., 1992; Zhao and Dean, 2004). In this model, the LCR associates with an elongating RNA polymerase II complex to the promoter of a gene. However, models that propose a role for the intervening chromatin fiber in establishing LCR-promoter contacts are difficult to reconcile with phenomena like competition of multiple genes for a single regulator (de Villiers et al., 1983; Hanscombe et al., 1991; Wasylyk et al., 1983) and alternating transcription of the β-globin genes by a stochastic 'flip-flop' mechanism (Gribnau et al., 1998; Trimborn et al., 1999; Wijgerde et al., 1995). Moreover, tracking and linking can not explain how an enhancer on one chromosome could activate transcription from an allelic promoter on another paired chromosome in a process called transvection (Dunaway and Droge, 1989; Mahmoudi et al., 2002; Mueller-Storm et al., 1989). These examples imply that interaction between an enhancer and promoter can take place in the absence of intervening DNA. The random collision model assumes that the chromatin fiber shows some degree of flexibility and as a consequence dispersed DNA fragments can randomly collide. These collisions occur in a stochastic manner of which the frequency depends on genomic site separation and the relative affinities between proteins bound to the DNA fragments. However, theoretical calculations and experimental measurements of site-specific recombination suggest that random diffusion alone is not enough to establish contacts between two sites separated by several kilobases on a chromatin fiber (Ringrose et al., 1999; Rippe, 2001). Nevertheless, many nuclear processes like e.g. V(D)J recombination and transcription have been shown to involve DNA looping over vast distances (Sayegh et al., 2005; Skok et al., 2007; Spilianakis and Flavell, 2004). Histone modifications such as acetylation might enhance random diffusion by altering chromatin flexibility (reviewed in (Li et al., 2006)). However, recombination events require short lived, single interactions while transcription of e.g. the β-globin genes, requires longer contact times in the range from 45 to 80 minutes for the β-globin-LCR interaction in definitive erythrocytes (Wijgerde et al., 1995). Therefore, stabilisation of interactions might be

an important determinant of successful loop formation and maintenance. Transcription factors that bind to their cognate sites in the β-globin locus and subsequently (homo-) multimerise could possibly function as stabilising factors. Indeed, structural analysis in cells lacking the erythroid specific transcription factors EKLF or GATA-1 have shown that these factors play a pivotal role in the establishment and/or maintenance of LCR-promoter contacts. Using 3C technology, it was shown that EKLF is indispensable for ACH formation. In the absence of this factor, a CH pre-structure similar to that observed in erythroid precursor cells was found. Progression to, or stabilisation of, a fully functional ACH, which entails the participation of the actively transcribed gene and the HSs 3' of the LCR, requires the presence of EKLF (Drissen et al., 2004). Likewise, it was shown that the transcription factor GATA-1 and its cofactor FOG-1 are required for the physical interaction between the β-globin LCR and the β_{major} globin promoter (Vakoc et al., 2005a). Both EKLF and GATA-1 are known to bind DNA directly, but it is unlikely that these factors are directly responsible for bringing distant DNA elements together. Indeed, EKLF and GATA-1 are already bound to their cognate binding sites in the β-globin locus at a stage of erythroid differentiation that precedes ACH formation (Kooren et al., 2007). Both EKLF and GATA-1 are known to interact with large protein complexes like CBP/p300 and Mediator. These complexes are able to interact simultaneously with factors bound at enhancers and promoters and could thereby stabilise promoter-LCR contacts (Chan and La Thangue, 2001; Kuras et al., 2003). Besides EKLF and GATA-1, p45 NF-E2 is another lineage restricted transcription factor associated with β-globin gene regulation. The role of this transcription factor in ACH formation will be discussed in more detail in Chapter 3 of this thesis. In short, it was shown that in the absence p45 NF-E2 long-range contacts in the β-globin locus were formed normally and hence we conclude that p45 NF-E2 is dispensable for β-globin ACH formation (Kooren et al., 2007).
In addition to transcription factors, the role of individual hypersensitive sites in the spatial conformation of the β-globin locus was investigated. When the β-globin promoter or HS3 of the LCR is deleted from a transgenic construct containing the entire human β-globin locus and the complete LCR, the ACH is maintained. However, deletion of both HS3 and the β-globin promoter resulted in a strong destabilisation of the ACH with a concomitant strong reduction of β-globin gene expression, demonstrating that multiple interactions between the LCR and the β-globin promoter are required to stabilise the spatial conformation of the locus *in vivo* (Patrinos et al., 2004). Interestingly, removal of only the core-regions of HS3 already results in the disruption of long-range interactions within the human β-globin locus without the need of to additionally remove the β-globin promoter (Fang et al., 2007). So, the ACH is a complex entity containing multiple protein complexes and binding sites for transcription factors. The removal of one of these transcription factors may cause ACH destabilisation and the adoption of a structure that can be formed independent of this factor. The data are in strong agreement with the idea that the LCR functions as a holocomplex in which cooperativeness of a minimum amount of *cis* regulatory elements is required for stabilising the spatial conformation of the β-globin locus.

CTCF and long-range DNA interactions

Besides the presence of LCR-gene contacts in the context of the β-globin ACH, the function of the outer hypersensitive sites HS-85, HS-62/-60 and 3'HS1 participating in this structural organisation is unclear (Tolhuis et al., 2002). Together with 5'HS5 of the LCR, these distal HSs are already spatially clustered in a pre-structure called the CH in erythroid progenitor cells prior to high level β-globin gene expression (Palstra et al., 2003). Importantly, no such long-range DNA interactions were detected in non-erythroid cells. This tissue-specificity suggests a role for the clustering of these outer HSs in developing red blood cells, perhaps facilitating interactions between the LCR and the β-globin genes during later stages of erythroid differentiation. Interestingly, all these interacting sites are known to bind the ubiquitous zinc finger transcription factor CTCF (CCCTC-binding factor) and both 5'HS5 and 3'HS1 have been shown in reporter assays to function as CTCF-dependent enhancer-blocker when placed between an enhancer and a promoter (Bulger et al., 2003; Farrell et al., 2002). In the chicken β-globin locus, CTCF dependent insulator sites were also found at both sides of the locus, namely 5'HS4 of the LCR and 3'HS1 downstream of the β-globin genes. Both sites coincide with erythroid specific transitions in DNAseI chromatin sensitivity (Saitoh et al., 2000), which led to the suggestion that CTCF might partition the genome in physically distinct domains of gene expression. Studies in *Drosophila* using another insulator called the suppressor of Hairy-wing Su(Hw) showed that these elements were located at the base of a chromatin loop. Insertion of an additional insulator in the centre of the loop results in the formation of two smaller loops and mutations of insulator-binding proteins lead to the disruption of the loop (Byrd and Corces, 2003). Based on these observations, a similar role for the mammalian insulator CTCF in higher order chromatin structure organisation may be anticipated.

CTCF is a ubiquitously expressed 11 zinc finger protein and is the prototype transcription factor harbouring insulator activity in vertebrates. For example, CTCF acts as an enhancer blocker if bound to the 5'HS4 LCR element in the chicken β-globin locus (Prioleau et al., 1999; Recillas-Targa et al., 1999) and *in vitro* experiments suggested that CTCF functions as a barrier against repressive telomeric heterochromatin (Defossez and Gilson, 2002). Moreover, CTCF binding sites have been found at the imprinting centre that determines choice of X-inactivation (Chao et al., 2002), CTCF mediates the enhancer blocking activity of the c-myc insulator element (Gombert et al., 2003) and CTCF binding sites have been found at boundaries of domains that escape X-inactivation (Filippova et al., 2005).

Important insights regarding CTCF function have come from studies in both the human and mouse Igf2/H19 locus, which showed that this factor regulates the imprinting of both genes in a methylation-sensitive manner. Both genes are located approximately 80 kb apart and are reciprocally imprinted such that Igf2 is expressed from the paternal allele and H19 from the maternal allele. Transcription of both genes is controlled by a set of shared enhancers downstream

of H19 and an imprinting control region (ICR) containing binding sites for CTCF located upstream of H19. On the maternally inherited chromosome, CTCF binds to the unmethylated ICR and thereby prevents the Igf2 promoter to gain access to the downstream enhancers. On the paternally inherited allele, the ICR is methylated which prevents the binding of CTCF and the promoter of the Igf2 gene is able to interact with the downstream enhancers (Bell and Felsenfeld, 2000; Hark et al., 2000). Mutations in the binding site for CTCF in the H19 ICR prevented binding of this protein and resulted in loss of imprinting of the maternal Igf2 gene, indicating that CTCF plays a pivotal role in organising epigenetically controlled expression domains (Pant et al., 2003; Schoenherr et al., 2003). Moreover, two other differentially methylated regions (DMRs) are located upstream of Igf2 promoter 1 (DMR1) and within exon 6 of Igf2 (DMR2), which are a methylation-sensitive silencer (Constancia et al., 2000; Eden et al., 2001) and a methylation-dependent activator respectively (Murrell et al., 2001). Interestingly, the unmethylated maternal H19 ICR is required to protect DMR1 and DMR2 from methylation and DMR1 is required to protect DMR2 from methylation. Since this coordination is not due to linear spreading of methylation it was suggested that long-range interactions between these regulatory elements underlie these epigenetic modifications (Lopes et al., 2003). By applying 3C technology to study the spatial conformation of the locus, it was demonstrated that the H19 ICR indeed is in close spatial proximity to the two DMRs at the Igf2 locus. On the maternal allele, the H19 ICR interacts with DMR1 which renders the Igf2 gene inactive, whereas on the paternal allele the H19 ICR interacts with DMR2 which results in the activation of Igf2 (Murrell et al., 2004). Subsequent studies showed that these long-range interactions are dependent on CTCF. Binding of CTCF to the maternal DMR1 *in vivo* requires intact CTCF binding sites in the H19 ICR and using 3C it was shown that the interaction between the ICR and DMR1 depends on binding of CTCF to the maternal H19 ICR (Kurukuti et al., 2006). Furthermore, the chromatin loop between the CTCF-bound ICR and DMR1 are maintained during mitosis whereas enhancer-promoter loops are absent, providing a possible role of CTCF in epigenetic memory during cell division (Burke et al., 2005). In addition it was shown that the H19 ICR functions by disrupting promoter-enhancer interactions that are associated with transcriptional activation. Instead, the H19 ICR promotes alternative long-range interactions between itself and the blocked enhancer or promoter (Yoon et al., 2007). These results led to the proposal of a chromatin-loop model in which the imprinting of Igf2 is explained by a CTCF dependent partitioning of the paternal and maternal Igf2 alleles into active or repressed chromatin loop domains. Moreover, insertion of the ICR element into heterologous positions in the mouse genome shows that this insulator promotes alternative long-range interactions between itself and blocked enhancers and promoters (Yoon et al., 2007). All these examples indicate that CTCF binding to the H19 ICR in the imprinted Igf2/H19 locus is required for long-range interactions between regulatory elements and hence, CTCF is a factor involved in the organisation of higher order chromatin structure.

The presence of multiple CTCF binding sites surrounding the murine β-globin locus and the participation of these sites in the CH and ACH in erythroid cells prompted us to investigate the role of CTCF in higher order chromatin structure of the mouse β-globin locus (see Chapter 4). The fact that it binds to both sides of the locus has led to the idea that CTCF may serve as an enhancer blocker that prevents inappropriate activation of surrounding olfactory receptor gene by the β-globin LCR in erythroid cells (Farrell et al., 2002) or may prevent spreading of heterochromatin into the β-globin locus by demarcating the boundaries. Using a conditional knockout mouse model, which depleted most CTCF protein, and targeted disruption of the CTCF binding site residing within 3'HS1 it was shown that CTCF is indeed required for long-range interactions between cognate CTCF binding sites in the β-globin locus (Splinter et al., 2006). Surprisingly, disruption of CTCF-binding to the β-globin locus had no measurable effect on the expression of the β-globin genes, nor did it result in inappropriate activation of the olfactory receptor genes surrounding the β-globin locus. Removal of CTCF resulted in a loss of histone acetylation and gain of histone methylation, but only locally at the CTCF binding site and not elsewhere in the locus (Splinter et al., 2006). Expression of the β-globin genes was not influenced either in mouse models that carried combinations of deletions of the outer CTCF binding hypersensitive sites (Bender et al., 2006). These data show that CTCF is required for higher order chromatin structure in the β-globin locus, but CTCF mediated loops do not play an important role in β-globin gene expression. Although CTCF-dependent loops are tissue specific and evolutionary conserved between mouse and man, this data does not support a role for CTCF as a boundary element demarcating the borders of an erythroid specific expression domain. Perhaps, β-globin gene expression benefits from CTCF mediated loops to an extent that is sufficient for evolutionary selection, but too limited to be detected. Alternatively, CTCF is a structural factor that binds to many loci for the sole purpose of folding their chromatin. If this is the case, it is likely that evolutionary selection against sites forming chromatin loops within a gene locus explains why CTCF binding sites are often found outside gene loci such as the β-globin locus, without necessarily having a positive or negative influence on gene expression (Dillon and Sabbattini, 2000). In fact, a recent report analysing genome wide CTCF binding sites in the human genome indeed confirms that the majority of CTCF binding occurs outside gene loci at sites remotely located from gene promoters of known genes (Kim et al., 2007).

Studying nuclear organisation using 4C technology

Long range DNA interactions are established both *in cis* and *in trans*

The development of 3C technology has given researchers the opportunity to investigate the spatial configuration of many gene loci in unprecedented detail and has boosted our understanding of the functional significance of DNA folding. Using this method the importance of chromatin folding at the level of individual gene loci has firmly been established. However, interactions between gene loci separated by tens of megabases also exist and even interactions between regions that reside on different chromosomes have been reported. A classical example of interactions between chromosomal regions that are located on different chromosomes is transvection. This naturally occurring phenomenon in *Drosophila* involves a process in which an enhancer on one chromosome can activate a promoter *in trans* on the other, paired, homologous chromosome (Bickel and Pirrotta, 1990). Such pairing of homologs is not observed in mammals and a topical research question is whether functionally relevant interactions take place between gene loci located far apart on the same chromosome or even present on different chromosomes.

At the level of complete chromosomes it was shown by 2D and 3D FISH that a subset of mouse chromosomes form distinct types of spatial clusters in different tissues. Moreover, a correlation was found between tissue specific spatial proximity and tissue specific translocation prevalence, suggesting that distant interchromosomal loci involved in recurrent translocations are frequently in close spatial proximity in nuclei of relevant cell types (Parada et al., 2004).

At the level of single gene loci it was shown by RNA FISH and DNA immuno-FISH that a number of selected loci co-localise with the active β-globin locus in erythroid cells, despite the fact that these regions were up to 40 Mb apart with respect to the underlying chromosome template. This co-localisation occurred when loci were actively transcribed and took place at so-called transcription factories. Moreover, two of the four genes that showed co-localisation with the β-globin gene were other erythroid specific genes. 3C technology was used to confirm this co-localisation (Osborne et al., 2004). In another study, the nuclear position of the co-ordinately regulated α- and β-globin genes was measured using 3D FISH. In this study it was reported that the human α- and β-globin genes are frequently in close proximity when active in erythroid cells, perhaps suggesting that interchromosomal association is important for their co-regulation (Brown et al., 2006). However, proximity was defined as being within 1 μm and the percentage of truly overlapping signals was not significantly different from those measured between random loci. Moreover, this interaction does not occur between the mouse α- and β-globin genes, demonstrating that the globin genes regulate their expression irrespective of their relative nuclear separation (Brown et al., 2006).

An example of both intrachromosomal and interchromosomal associations that occur within one locus can be found in the T helper type 2 (T_h2) cytokine locus on mouse chromosome 11.

This locus contains the cytokine genes *IL4*, *IL5* and *IL13*, which are expressed in T_h2 cells under the control of an LCR that is located in between the genes. Using 3C technology it was shown that a 'pre-poised' structure was present in various cell types, formed by contacts between the promoters of the different interleukin genes. Only in cells of the T lineage, the interleukin gene promoters also interact with the T_h2 LCR to give rise to a 'poised' conformation, which suggests a phenomenon in which the locus is poised for rapid expression of lineage-specific cytokines after stimulation (Spilianakis and Flavell, 2004). In addition to these intrachromosomal interactions, the same authors used a 3C approach to detect interchromosomal interactions between multiple regulatory regions of the T_h2 cytokine locus on chromosome 11 and the promoter region of the *IFN-γ* gene on chromosome 10 in naïve T-cells. Expression of the interleukin genes on chromosome 11 specifies commitment towards the T_h2 lineage, while *IFN*-γ expression induces T_h1 differentiation. After differentiation of these naïve T-cells into effector T_h1 or T_h2 cells these interchromosomal interactions are greatly reduced in favour of the establishment of different intrachromosomal interactions. 2D FISH experiments confirmed these observations by showing single allele interchromosomal co-localisation in almost 40% of naïve T-cells. Moreover, deletion of the LCR in the T_h2 cytokine locus not only affected expression of the linked cytokine genes, but also had a negative influence on *IFN-γ* activation after stimulation, suggesting a regulatory role for this interchromosomal association (Spilianakis et al., 2005). This is not the only example of a *trans* interaction identified by 3C technology. Another study, using an adapted 3C method, proposed that CTCF mediates interchromosomal interactions between the H19 ICR located on chromosome 7 and the Wsb1/Nf1 gene on chromosome 11 (Ling et al., 2006). It should be noted however that this interchromosomal interaction was one of the only three interactions seen in total after a nested PCR approach, which questions the validity of these results (see also Chapter 2). In contrast, another study showed that the H19 ICR had many, but completely different, *trans* interactions in addition to its abundant *cis* interactions (Zhao et al., 2006). Additional evidence for interchromosomal interactions comes from studies that address the role of the H enhancer element in the olfactory receptor system. This H element resides on chromosome 14 upstream of an olfactory receptor (OR) gene cluster and was shown to be required for the expression of OR transgenes (Serizawa et al., 2003). Like the gene competition studies in the β-globin locus, insertion of the H element more proximal to the OR gene cluster results in a competitive advantage of the nearest gene. This suggests that genes within the cluster compete for interaction with the distal H element, which ensures the stochastic activation of a single gene of this cluster per cell (Serizawa et al., 2004). A modified version of 3C technology using sequencing of ligation products was used to identify DNA fragments that interact with the H enhancer element in sensory neurons (Lomvardas et al., 2006). Apart from interactions with the nearest OR gene cluster located 75 kb from the H element, also sequences from other OR gene clusters residing on other chromosomes were found. Subsequent RNA and DNA FISH experiments showed that these contacts are related to OR gene expression since co-localisation

with the H-element was only found when the given allele was actively expressed. Based on these observations it was suggested that OR genes on different chromosomes compete for an interaction with one active H enhancer and this mechanism may underlie the stochastic activation of only one OR allele in each olfactory neuron (Lomvardas et al., 2006). However, recent experiments in mice carrying a targeted deletion of the H-element only showed abolished expression of genes located within a nearby OR gene cluster and no demonstrable effect was seen on the expression of all other OR genes analysed (Fuss et al., 2007). Obviously, these results question the relevance of the above mentioned interchromosomal interactions and the role of the H-element as an essential trans-acting enhancer element for genome wide regulation of all OR-genes.

Studies concerning gene positioning and nuclear organisation have predominantly been carried out using FISH technology, although recently 3C technology has also been used and adapted in order to identify or confirm a (limited) number of *trans* interactions. A clear disadvantage of FISH studies is that they are biased towards the loci selected for that study and allow analysis of only a limited number of loci simultaneously. Results obtained by this technique therefore often are non-general and rather anecdotal. Moreover, the criteria for truly overlapping FISH-signals often differ greatly between different studies, which make the results obtained by this technique difficult to interpret. 3C technology has been developed to identify interacting elements between selected parts of the genome and requires the design of individual primers for all restriction fragments analysed. This means 3C is also a biased method, since it requires prior knowledge of the loci that need to be analysed and the identification of unpredicted, novel interacting regions outside the analysed locus is therefore difficult. Moreover, the mere detection of a given ligation product does not necessarily reveal a genuine interaction between the analysed sites and a lack of proper controls and quantification often undermines the reliability of a 3C analysis (Dekker, 2006) (see Chapter 2). Sites separated over large genomic distances, including interchromosomal interactions, often do not form enough ligation products for accurate quantification by 3C, even if microscopy studies suggest these sites come together in a significant proportion of cells. Therefore 3C is well suited to study the spatial conformation of loci up to several hundreds of kilobases in size, but not to discover or verify interactions over larger genomic distances. Recently, a number of new strategies have been developed that allow the screening of the entire genome in an unbiased manner for DNA segments that physically interact with a DNA fragment of choice (Dostie et al., 2006; Simonis et al., 2006; Wurtele and Chartrand, 2006; Zhao et al., 2006). All these methods are based on 3C technology. One of these novel methods developed is referred to as Chromosome Conformation Capture on Chip (4C) and will be discussed in more detail in the next section.

4C: 3C goes genome-wide

4C technology combines 3C technology with either micro-array analysis (Simonis et al., 2006) or sequencing (Ling et al., 2006; Wurtele and Chartrand, 2006; Zhao et al., 2006) to allow for the unbiased identification of DNA elements that interact with a target sequence in the nuclear space. An outline of the 4C procedure, developed recently in our lab, is given in Figure 1.9 (Simonis et al., 2006). Just like 3C, 4C involves PCR amplification of DNA fragments cross-linked and ligated to a DNA restriction fragment of choice (the 'bait'). First, the standard 3C procedure is followed, using a six-cutter as a restriction enzyme which yields a regular 3C template that contains de-cross-linked ligation products. Next, a four-cutter restriction enzyme is used to reduce the size of the captured fragments followed by re-ligation of trimmed ligation products under conditions that favour the formation of self-ligated circles. Finally, all the DNA fragments captured by the bait sequence in the population of cells are simultaneously and linearly amplified via inverse PCR, using two bait-specific primers that amplify from the circularised ligation products (Simonis et al., 2006). This amplified material, representing the genomic environment of the bait fragment, is labelled and subsequently hybridised to a tailored micro-array containing unique probes located directly adjacent to each recognition site of the used six-cutter. Our original probe design represented seven complete mouse chromosomes to be spotted on a single array, whereas more recent designs cover the entire mouse or human genome on a single

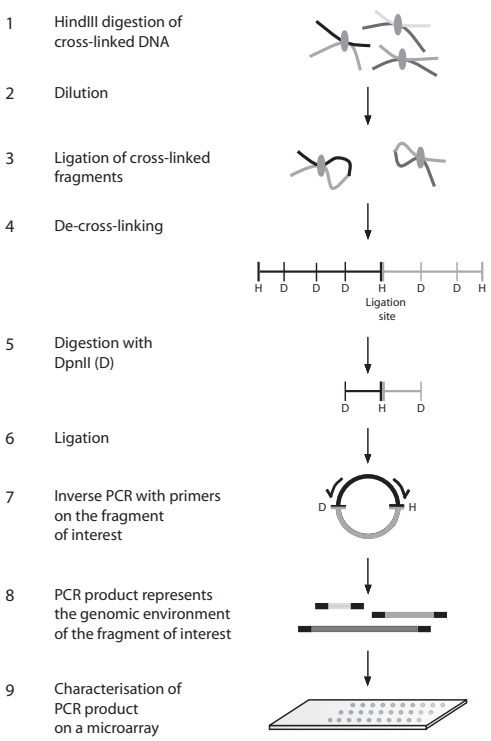

Figure 1.9
4C technology.
Schematic diagram showing the principles of Chromosome Conformation Capture on Chip (4C). The initial steps of 4C involve the standard 3C procedure, using a six-cutter (HindIII) as a restriction enzyme yielding a regular 3C template containing de-cross-linked ligation products (1-4). In 4C, the captured fragments are subsequently reduced in size by using a frequently cutting secondary restriction enzyme (DpnII)(5), followed by re-ligation of trimmed ligation products to form small, self-ligated DNA circles (6). Next, all the DNA fragments captured by the bait sequence in the population of cells are simultaneously and linearly amplified via inverse PCR, using two bait-specific primers that amplify from the circularised ligation products (7). This amplified material, representing the genomic environment of the bait fragment (8), is labelled and subsequently hybridised to a tailored micro-array containing unique probes located directly adjacent to each recognition site of the used six-cutter (9).

Nimblegen microarray (400.000 probes), enabling the identification of interactions at a resolution of approximately 7 kb. 4C technology analyses interactions with many thousands of genomic sites simultaneously and does not rely on the detection of a single ligation product but on the identification of clusters of neighbouring restriction fragments, each showing interactions with the bait sequence. The design of the array is such that each probe analyses one independent ligation event. Since only two fragments can be captured per cell, the clustering of interacting DNA fragments strongly indicates that this genomic region contacts the bait sequence in multiple cells. Thus, 4C technology is the first method that allows for an unbiased genome-wide screen for DNA fragments that interact with a locus of choice. As such, it is expected to contribute importantly to a more comprehensive understanding of nuclear architecture. In the next section, the application of 4C on both the β-globin locus and HoxB1 gene cluster will be discussed in more detail as two examples of how this novel unbiased, high throughput technology can be used to study long-range DNA interactions at higher organisational levels.

Studying nuclear architecture using 4C

A recent 4C study performed in our lab focussed on the murine β-globin locus and provided for the first time a comprehensive overview of all long-range DNA interactions of different selected active and inactive gene loci (Simonis et al., 2006). Using a bait sequence containing HS2 of the mouse β-globin locus, interacting DNA regions were defined in both transcriptionally active erythroid cells and transcriptionally inactive brain tissue. Using tailored micro-arrays able to analyse possible interactions across 7 complete mouse chromosomes, it was demonstrated in different tissues that both the active and inactive β-globin locus are engaged in many long-range intrachromosomal interactions over tens of megabases. Multiple regions of approximately 200 kb were identified to interact with the β-globin locus, so interactions were not confined to single genes or promoters, but rather to larger chromosomal regions. Strikingly, the active β-globin locus in erythroid cells contacted a completely different set of loci than the inactive locus in brain cells. When active, the β-globin locus predominantly contacted other transcriptionally active regions located mostly towards the telomere of the acrocentric mouse chromosome, whereas the inactive β-globin locus preferentially interacted with regions that were not transcribed, which located mostly towards the centromere of chromosome 7. The same study applied 4C to the housekeeping gene *Rad23a*, which is present in an active, gene-dense region and is expressed ubiquitously. This gene was, like β-globin, found to contact many regions *in cis*, but also many interchromosomal interactions were identified. Moreover, these interactions were essentially the same between erythroid cells and brain tissue and were mainly established with other transcriptionally active, gene dense regions. Subsequent analysis by cryo-FISH confirmed multiple long-range interactions (Branco and Pombo, 2006), and showed that individual interactions were present in a significant but limited percentage of cells, suggesting that the

collection of different long-range interactions identified mostly reflected cell-to-cell differences in chromatin architecture. Based on these results it was concluded that the difference in genomic environment observed for the β-globin locus between erythroid and brain cells was directly related to the difference in transcriptional status of the locus in the two tissues (Simonis et al., 2006). Altogether, the data strongly suggests that active and inactive genomic regions each have their own preferred set of long-range DNA interactions and the characteristics of both intra- and interchromosomal interactions is mainly determined by the transcriptional status of the locus.

Another recent study used a sequencing based 4C method to characterise the spatial genomic environment of the HoxB1 gene cluster during the induction of its expression in mouse embryonic stem (ES) cells (Wurtele and Chartrand, 2006). In concordance with the 4C analysis of the β-globin locus, this study also found that sequences located close on the linear chromosome template from the 4C bait region were largely overrepresented, irrespective of the expression status of the HoxB1 gene. Moreover, the spatial genomic environment of the HoxB1 gene changes during induction of expression. The proportion of distal intrachromosomal interactions were significantly lower after HoxB1 induction and this diminution is translated into a higher proportion of inter-chromosomal interactions (Wurtele and Chartrand, 2006). These results are in agreement with the reported relocalisation of the HoxB1 gene outside of its chromosome territory during ES cell differentiation and mouse embryonic development (Chambeyron and Bickmore, 2004; Chambeyron et al., 2005). This study indicates that the genomic environment of the HoxB1 gene is dynamic and a change in the number of intra- and interchromosomal interactions is concomitant with a change in the transcriptional status of the gene. No matter how informative, it needs to be emphasised that in this study only a limited number of all captured interactions were sequenced. Moreover, the characteristics of the reported intra- and interchromosomal associations are not addressed in detail, which makes the exact relevance of these interactions difficult to interpret.

Application of the 4C method is not restricted to analyse only the nuclear environment of genes or their linked regulatory elements. Being an unbiased method, it provides a unique possibility to study the nuclear environment of other genomic elements. Indeed, the genome is comprised of many non-genic sequences that will likely shape and influence the spatial organisation of the genome at a higher organisational level. Chapter 5 of this thesis describes a 4C approach to study the nuclear organisation of both peri-centromeric and peri-telomeric regions on acrocentric mouse chromosomes. Our analysis focussed on unique sequences near the ends of three chromosomes with distinct transcriptional activities. In line with other 4C studies, the results show that there is a strong correlation between the transcriptional status of a chromosomal region and their interacting loci, providing additional evidence that active and inactive chromatin domains separate in the nucleus. Furthermore we show that, next to the transcriptional status, the proximity to defined repetitive sequences strongly influences the positioning of a locus in the interphase nucleus. This data together with other 4C studies provides

evidence for self-organising principles underlying chromatin folding, where the preferred nuclear environment of a locus is not only determined by its own functional properties, but also by the properties of neighbouring DNA segments and, by extrapolation, of the entire chromosome.

Scope of this thesis

Proper spatiotemporal expression of genes is essential for the function and survival of every eukaryotic cell and therefore requires meticulous regulation. Often gene expression is controlled by regulatory DNA elements positioned away from the gene on the linear chromosome template. This thesis describes experiments in which this process of transcriptional regulation is addressed in more detail. The next chapters will mainly focus on the role of different transcription factors in the erythroid specific spatial conformation of the mouse β-globin locus and will discuss novel technologies that provide insight into DNA interactions inside the living cell nucleus.

Chromosome Conformation Capture (3C) technology has become a standard research tool for studying the relationship between nuclear organisation and transcriptional regulation *in vivo*. Based on 3C technology, a number of new 4C methods have recently been developed. Both 3C and 4C based methods have important limitations and preconditions that need to be recognised and addressed properly by each researcher who wants to apply these methods. **Chapter 2** of this thesis will therefore discuss and evaluate potentials and pitfalls of currently used 3C-based methods.

Expression of the β-globin genes proceeds from basal to exceptionally high levels during erythroid differentiation *in vivo*. This high transcription rate at later stages of erythroid differentiation is dependent on the β-globin Locus Control Region (LCR) and coincides with the formation of a three-dimensional structure of the locus known as the Active Chromatin Hub (ACH). **Chapter 3** describes experiments in which recently established I/11 cells are used as an erythroid model system to study the molecular events that accompany and underlie ACH formation. Moreover, the role of the erythroid transcription factor p45 NF-E2 in β-globin ACH formation is addressed in more detail.

In **Chapter 4** the role of the prototype vertebrate insulator protein CTCF in the spatial conformation of the mouse β-globin locus is described in more detail. CTCF-binding sites flank both the human and mouse β-globin locus and in erythroid cells these sites were found to participate in spatial interactions that are involved in ACH formation. A conditional CTCF knock-out mouse and a cell line containing a targeted disruption of a CTCF-binding site were used to investigate the involvement of CTCF in loop formation in the mouse β-globin.

Furthermore, novel 4C technology was applied on unique chromosomal end sequences and **Chapter 5** describes experiments in which the nuclear organisation of both peri-centromeric and peri-telomeric chromosomal regions is examined. These experiments attempt to define and describe important parameters influencing the positioning of a locus in the nucleus. Finally, an overall discussion of this thesis is presented in **Chapter 6** that describes the general implications of this work and future directions.

2

An evaluation of 3C-based methods capturing DNA interactions

Based on: Nature Methods 4, 895-901 (2007)

Summary

The shape of the genome is thought to play an important role in the coordination of transcription and other DNA-metabolic processes. Chromosome conformation capture (3C) technology is a recently developed biochemical technique that allows analysing the folding of chromatin *in vivo* at a resolution beyond that provided by current microscopy techniques. It has been used for example to demonstrate that regulatory DNA elements communicate with distant target genes via direct physical interactions that loop out the intervening chromatin fiber. Here, we will discuss the intricacies of 3C technology and new 3C-based methods including 4C, 5C and ChIP-loop assay.

3C technology

3C technology was originally developed to study the conformation of a complete chromosome in yeast (Dekker et al., 2002) and was subsequently adapted to investigate the intricate folding of complex gene loci in mammalian cells (Tolhuis et al., 2002). 3C technology has now become a standard research tool for studying the relationship between nuclear organisation and transcription *in vivo*. Detailed 3C technology protocols that should help researchers setting up the technology in their own laboratory have been published previously (Miele et al., 2006; Splinter et al., 2004). In short, the 3C procedure involves the following experimental steps: 1. Cells are fixed with formaldehyde, which forms cross-links between DNA segments that are close together in the nuclear space. 2. Cross-linked chromatin is digested with an excess of restriction enzyme, separating cross-linked from non-cross-linked DNA fragments. 3. DNA ends are ligated under conditions that favour junctions between cross-linked DNA fragments. 4. Cross-links are reversed. 5. Ligation events between selected pairs of restriction fragments are quantified by PCR, using primers specific for the given fragments (Figure 2.1).

3C technology is particularly suited to study the conformation of genomic regions that are roughly between five to several hundreds of kilobases in size. For example, it has been used extensively to demonstrate that transcriptional regulatory DNA elements communicate with distant target genes via direct physical interactions that loop out the intervening chromatin fiber (Liu and Garrard, 2005; Murrell et al., 2004; Palstra et al., 2003; Spilianakis and Flavell, 2004; Tolhuis et al., 2002). To our knowledge, the smallest region studied so far by 3C technology spans 6,7 kb (O'Sullivan et al., 2004), while the largest region analysed spans ~600 kb (Skok et al., 2007). The technique enables the identification of physical interactions between distant DNA segments and of chromatin loops that are formed as a consequence of these interactions. Since physically linked DNA segments are engaged in random collisions due to flexibility of the chromatin fiber, with a frequency that is inversely proportional to their genomic site separation, the mere detection of a given ligation product does not necessarily reveal a specific interaction

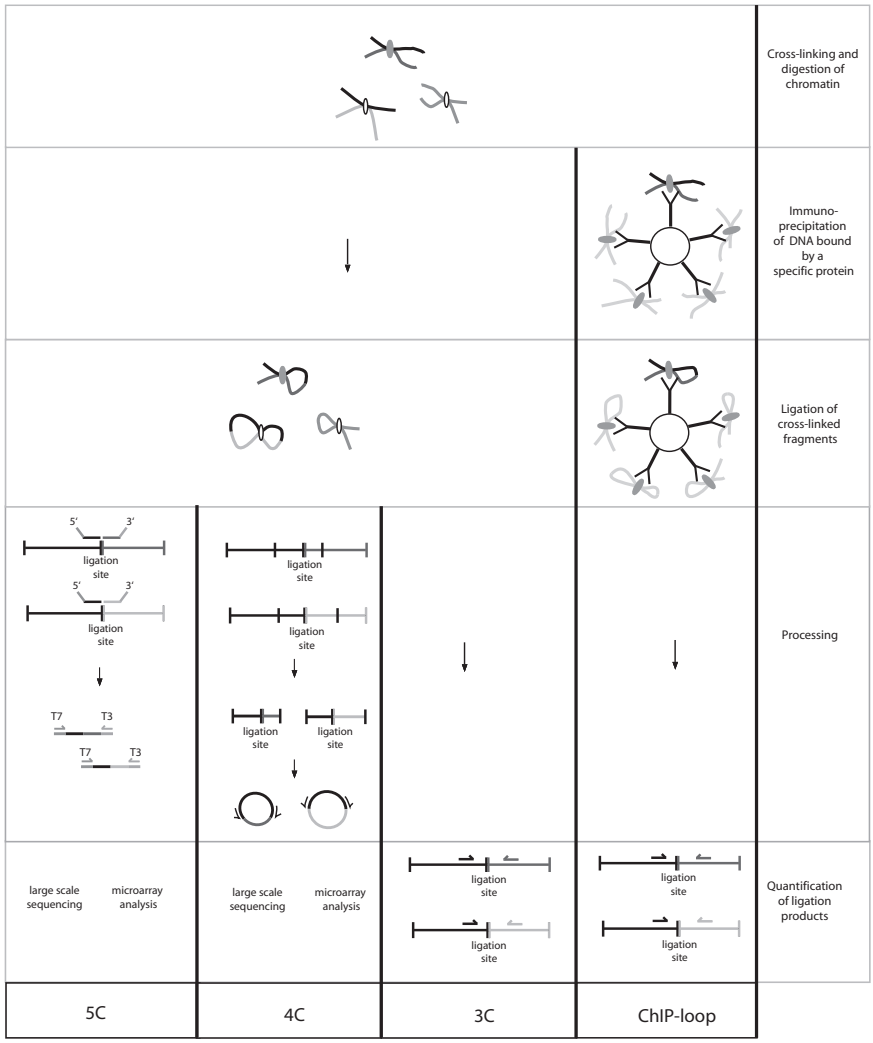

Figure 2.1

Schematic representation of 3C based methods.

In all 3C-based methods the conformation of the chromatin is captured by formaldehyde treatment followed by DNA digestion with a restriction enzyme. Cross-linked fragments are ligated to each other and ligation frequencies are measured. In the ChIP-loop assay, immunoprecipition enriches the sample for fragments bound by a specific protein and restriction fragments are ligated to each other on the beads. In ChIP-loop and 3C, ligation frequencies are measured by quantitative PCR. This requires a specific PCR reaction for each ligation combination. In 5C, oligonucleotides are annealed and ligated in a multiplex setting. The specific oligonucleotides contain either a 5' T7 primer extension or a 3' T3 primer extension, such that multiple ligation events can be amplified in one PCR reaction and analysed by large-scale sequencing or microarray analysis. In 4C, ligation junctions are first trimmed by digestion with an enzyme that cuts more frequently than the first enzyme used. The shortened products are circularized, such that all fragments ligated to a fragment of choice can be amplified in one PCR reaction. If a frequent cutting enzyme is used in the first digestion the second digestion can be omitted (see Figure 2.5). The 4C PCR product is analysed by large-scale sequencing or microarray analysis.

between the analysed sites. This requires the demonstration that two DNA sites interact more frequently with each other than with neighbouring DNA sequences. Thus, 3C technology is a quantitative assay and a meaningful analysis critically relies on an accurate comparison of interaction frequencies between multiple DNA segments. Indeed, the smaller the region studied, the more difficult it becomes to discriminate between specific DNA interactions that potentially have functional relevance and random collisions that occur non-specifically due to the flexibility of the chromatin fiber. 3C and 3C-based technologies provide information about the frequency, but not the functionality, of DNA interactions. Thus, a specific interaction may be identified (i.e. two DNA sites looping towards each other) but additional, often genetic, experiments are required to address whether such interaction is functionally meaningful or just the consequence of general chromosome folding properties. For example, many of the long-range intra- and interchromosomal interactions between genic and non-genic regions that can be identified by 4C technology (Simonis et al., 2006) are expected to be non-functional but to reflect the general folding patterns of chromosomes (de Laat, 2007).

During most of the cell cycle a single mammalian cell provides maximally two events for 3C analysis, as it contains only two copies of a given restriction fragment, each end of which can be ligated to maximally one other restriction site during the 3C procedure. This implies that a meaningful (i.e. quantitative) 3C PCR analysis must be performed on a DNA template that represents many genome equivalents. It also implies that DNA interactions can only be quantified accurately if they occur in a significant proportion of the cells. Sites separated over large genomic distances (i.e. hundreds of kilobases or more) often form not enough ligation products for accurate quantification, even if microscopy studies suggest that they come together in a significant proportion of cells. To study such long-range interactions, we recommend using high-throughput 4C technology. 4C technology puts individual interactions in the context of many thousands of other interactions along the chromosome and relies on the identification of clusters of neighbouring restriction fragments each showing interactions with the site of interest. 4C technology also allows identifying inter-chromosomal interactions and there is mounting evidence in mammalian cells that such *trans*-interactions are more abundant than previously anticipated. For example, a housekeeping gene, *Rad23a*, was found to contact many other genes on unrelated chromosomes (Simonis et al., 2006). Again, many of these interactions may not be functionally relevant but just the consequence of the general folding patterns of the genome. Unfortunately, it is difficult to identify interactions between sister chromatids and homologous chromosomes by 3C-based methods, as these usually cannot be separated from intra-chromosomal interactions. Below, a more detailed outline of the experimental steps involved in all 3C-based technologies will be presented in order to allow a better appreciation of the potentials and limitations of these methods.

Step 1: formaldehyde cross-linking

The method uses formaldehyde to crosslink protein-protein and protein-DNA interactions via their amino and imino groups. Advantage of this cross-linking agent is that it works over a relatively short distance (2 Å) and that cross-links can be reversed at higher temperatures (Jackson, 1999; Orlando et al., 1997; Solomon and Varshavsky, 1985). Although cross-linking is sometimes performed on isolated nuclei, it is preferentially done on living cells, since this better guarantees taking a faithful snapshot of the chromatin conformation *in vivo*. Routinely, cells are cross-linked at room temperature for ten minutes, using a formaldehyde concentration of 1-2%, but optimal fixation conditions depend on the frequency and stability of the interactions analysed and need to be redefined for every new 3C experiment. It is important to keep in mind that more stringent fixation conditions will lower the subsequent restriction enzyme digestion efficiency (Splinter et al., 2004).

Many 3C experiments demonstrate preferential interactions between transcription regulatory DNA elements. These sites are known to carry transcription factors and often contain less histone proteins, hence their hypersensitivity to nuclease digestion. A concern often raised is that the 3C assay may be biased due to better cross-link ability of these sites. Evidence that the contrary may be true comes from recently developed FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) (Giresi et al., 2007; Hogan et al., 2006). FAIRE involves phenol-chloroform extraction of formaldehyde cross-linked and sonicated chromatin and isolates regulatory DNA sequences based on the fact that they tend to end up in the aqueous phase more than other genomic regions. This indicates that the bias is against these regulatory DNA elements for being cross-linked to proteins by formaldehyde.

Formaldehyde is also used under similar experimental conditions in chromatin immunoprecipitation (ChIP) experiments as the cross-linking agent that captures protein-DNA interactions *in vivo*. It is conceivable that formaldehyde often produces complex aggregates containing more than two DNA fragments. In support of this, it was found that a single restriction fragment frequently captures two or more other restriction fragments together in a 4C experiment (Zhao et al., 2006). This notion would imply that both ChIP and 3C-like technologies also pick up indirect interactions.

Step 2: Restriction enzyme digestion

After cross-linking, nuclei are isolated and digested with a restriction enzyme of choice. The choice of restriction enzyme will mainly depend on the locus to be analysed. The restriction enzyme should dissect the locus such that it allows for the separate analysis of the relevant regulatory elements (gene bodies, promoters, enhancers, insulators, etc.). Analysing the topology of small loci (< 10-20 kb) requires the use of frequently cutting restriction enzymes such as *Dpn*II or *Nla*III (4-base cutters). When analysing larger loci, 6-base cutters can also be used. Not all enzymes

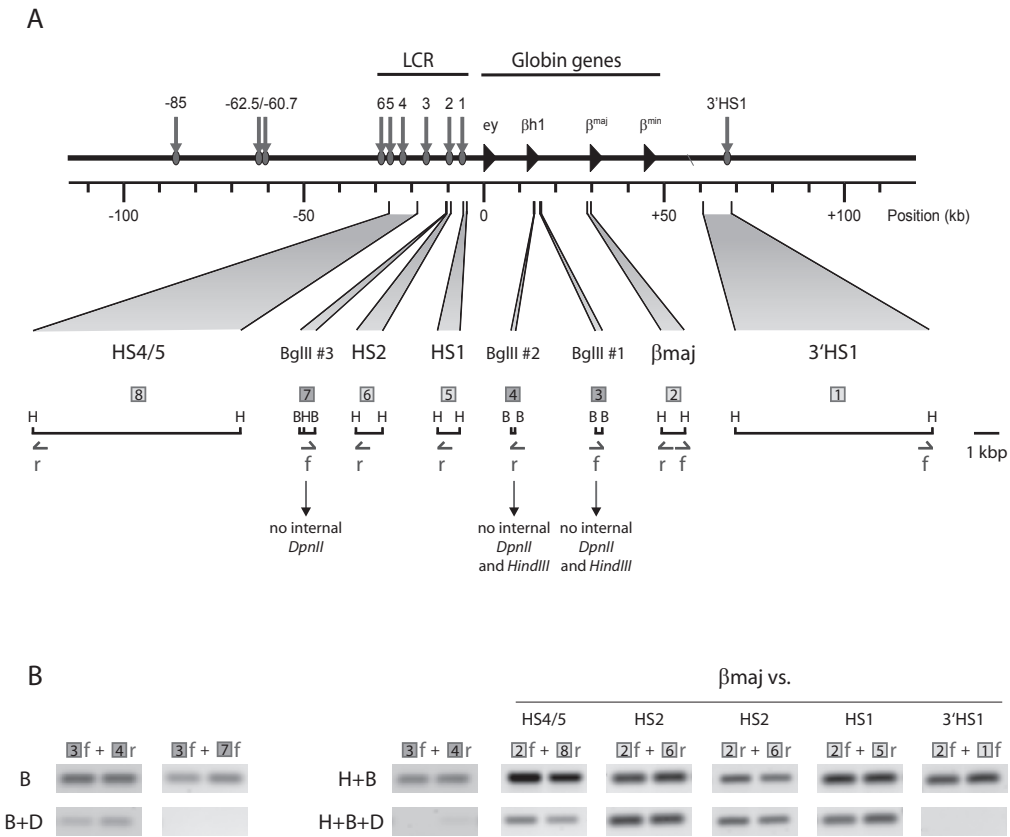

Figure 2.2

3C measurements and the use of frequently cutting restriction enzymes.

A. Schematic presentation of the mouse β-globin locus. Analysed *BglII* (B) and *HindIII* (H) restriction fragments are shown. Primers used are plotted as arrows below the restriction fragments.

B. Additional digestion with the frequently cutting restriction enzyme *DpnII* (D) reduces the number of junctions formed between unrelated fragments not processed by this restriction enzyme (fragment 3, 4 and 7). This shows that formaldehyde forms cross-linked aggregates containing more than two DNA fragments. Most likely, *DpnII* separates DNA fragments not directly cross-linked to each other. Interestingly, additional digestion with *DpnII* has no effect on junctions that analyse specific interactions between β_{major} and HS2, or β_{major} and HS1, suggesting that a more frequently cutting restriction enzyme reduces background while leaving specific, functionally relevant interactions more intact. Data shown are representative for experiments performed on three 3C different templates obtained from E14.5 liver cells.

digest cross-linked DNA equally well and we prefer to use *Eco*RI, *Bgl*II or *Hind*III (Splinter et al., 2004). When we digest overnight with a large excess of one of these restriction enzymes, we do not observe significant preferential digestion of e.g. open over closed chromatin. This may be different with different enzymes and conditions and we recommend for each new 3C experiment to exclude a bias in the assay due to preferential digestion of one site over the other. Digestion efficiency drops with increased cross-linking stringency. We recommend that at least

60-70% of the DNA, but preferably 80% or more, is digested before continuing with the ligation step.

The efficiency of ligation between two physically linked but functionally unrelated fragments not only depends on their genomic site separation but also on the number of restriction enzyme recognition sites located in between them. This is illustrated by the following experiment that analyses interactions in the β-globin locus in fetal liver cells that express the β-globin genes (Figure 2.2). When a cross-linked chromatin sample is first digested by BglII and subsequently split in two parts whereby one is receiving additional BglII and the other the more frequently cutting restriction enzyme DpnII, ligation products between such unrelated BglII fragments are lost only upon additional DpnII treatment (Figure 2.2B). Since DpnII treatment leaves the analysed BglII fragments intact, the observed drop in their ligation frequency strongly indicates that they were not cross-linked directly to each other but via other DNA sequences. Thus, formaldehyde can form chromatin aggregates that contain more than two restriction fragments: in these aggregates, multiple restriction ends will compete for ligation to each other during the next step of the 3C procedure. Interestingly, additional DpnII digestion had no effect on certain specific, functionally relevant, interactions between hypersensitive site 2 (HS2) or HS1 of the β-globin locus control region (LCR) and the active β_{major} gene, no matter which side of the corresponding fragments were analysed (Figure 2.2B). These data strongly suggest that a more frequently cutting restriction enzyme trims the cross-linked chromatin, thereby reducing background interactions while leaving the specific, functionally relevant interactions more intact.

Step 3: Ligation

A critical selective step in the procedure is the ligation step carried out under conditions that favour intra-molecular ligation events between cross-linked DNA fragments. This step creates the actual 3C library that is enriched for ligated junctions between DNA fragments that originally were close together in the nuclear space. A single cell can only provide maximally two junctions per restriction site for analysis. An accurate quantitative comparison of ligation events by means of PCR requires that sufficient copies of each junction are included in the PCR reaction. It is therefore relevant to know how frequent a given restriction end is ligated to a selected other restriction site. We have carefully quantified the abundance of the most recurrent ligation products (Figure 2.3). Independent of the restriction site analysed, two junctions are always over-represented. The first most abundant junction is with the neighbouring DNA sequence. This junction is the result of incomplete restriction enzyme digestion and can constitute up to 20-30% of all the junctions; this number drops when less stringent fixation conditions are used. The second most abundant junction is with the other end of the same restriction fragment, as a consequence of restriction fragment circularisation. This product can be formed independent of the cross-linking step and can account for up to 5-10% of all the junctions formed. Interestingly,

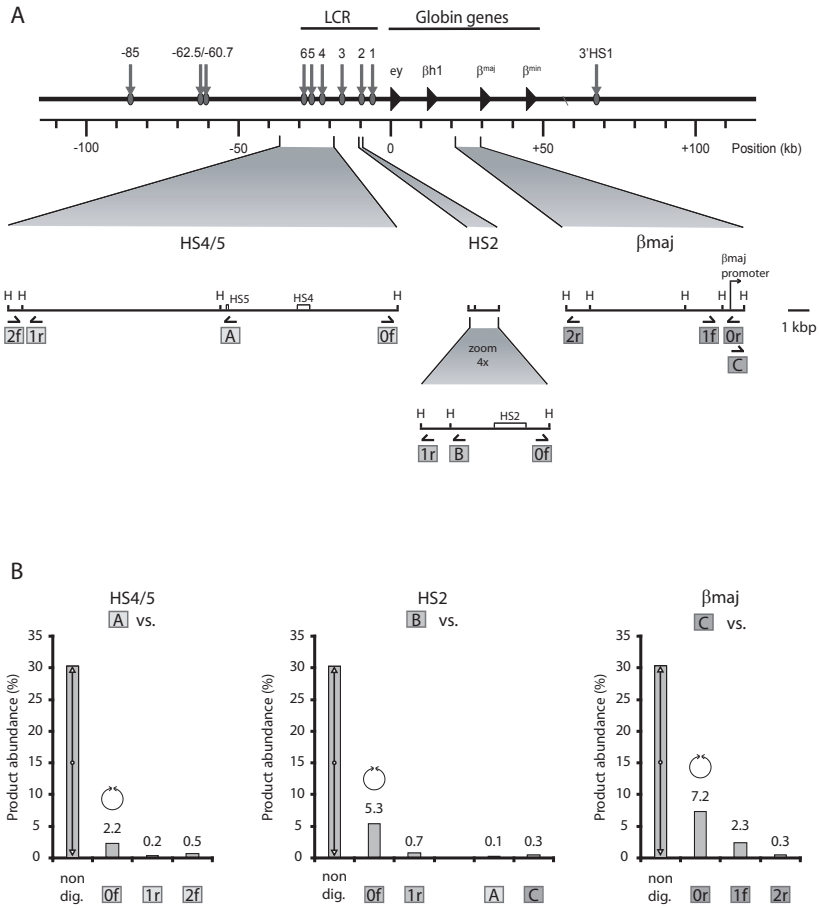

Figure 2.3

Relative abundance of frequently formed 3C ligation products.

A. Schematic presentation of the mouse β-globin locus. Analysed *HindIII* (H) restriction fragments are shown below and the positions of HS5, 4 and 2 and the β_{major} promoter are indicated. 3C primers used are plotted as arrows below the restriction fragments.

B. Plots show the percentage of junctions that are formed with a given end of a fixed fragment (A, B or C) as a consequence of non-digested DNA, ligation-mediated self-circularisation (analysed in combination with primer '0') or ligation to an end of a neighbouring restriction fragment (with primer '1' or '2'). Also shown is the relative abundance of junctions formed between the fixed fragment ends B and A, and B and C. These junctions analyse interactions between HS2 and HS4/5, and HS2 and the β_{major} promoter, respectively, which are thought to contact each other frequently in the tissue analysed (E14.5 liver). Data were based on 3C-qPCR analysis, comparing the abundance of each junction between a 3C template obtained from fetal liver and a control template containing all analysed junctions and several internal fragments in equimolar amounts. In order to generate this control template, specific junctions were PCR amplified from a 3C template and subsequently gel purified. The percentages plotted are representative for experiments performed on three different 3C templates, each obtained from fetal liver cells cross-linked for 10 minutes in 2% formaldehyde at room temperature. Note that the relative abundance of these junctions may be different when other fixation conditions are used, but junctions present as a consequence of incomplete DNA digestion and formed via self-circularisation are expected to always be the most abundant products. The data show that even frequent interactions are captured in less than 1% of the cells and that most ligation products are formed in less than 1 in 500 cells.

this percentage goes up when less stringent cross-linking conditions are used (data not shown), suggesting that under such conditions less restriction fragments are cross-linked together. The formation of other junctions is much less efficient. For example, ligation to ends of directly neighbouring restriction fragments (which will always be close together in the nuclear space and therefore should also ligate relatively efficiently) already only occurs 0.2-0.5% of the time. This percentage quickly drops down to <0.1% with increasing genomic site separation, unless two sites are engaged in a specific interaction. However, even sites thought to frequently interact with each other, such as sites in the β-globin locus control region and the active β-globin genes 30-50 kb away, only account for 0.2-0.5% of the junctions formed between them. It should be clear that in order to accurately quantify such rare events that often occur in less than 1/1000 cells, many genome equivalents need to be included in a PCR reaction. A schematic overview of the relative abundance of frequently formed ligation products at the mouse β-globin locus is presented in Figure 2.4.

PCR

After reversal of the cross-links, ligation frequencies of restriction fragments are analysed by PCR, using primers specific for the restriction fragments of interest. We routinely use 50-200 ng of 3C template, or ~$8x10^4$ – $3x10^5$ genome equivalents, per PCR reaction. A meaningful 3C analysis critically relies on the accurate quantification of the different ligation products and measurements therefore need to be taken when each DNA amplification reaction is in the linear range. In standard 3C experiments, a PCR protocol, which uses a standard number of PCR cycles and a standard amount of DNA template, is applied to the analysis of all different ligation products. Amounts are then estimated by measuring the intensity of ethidium bromide-stained PCR products separated by gel electrophoresis. Disadvantage of this semi-quantitative method is that it is prone to provide inaccurate data, as measurements will not always be taken when the DNA amplification reaction is in the linear range. To overcome this limitation and thus provide more accurate quantitative measurements, a real-time PCR approach using TaqMan® probes, called 3C-qPCR, was developed for the 3C analysis of ligation products (Splinter et al., 2006; Wurtele and Chartrand, 2006). The TaqMan® probe works together with a constant PCR primer, which both hybridise to the restriction fragment of interest. In combination with test primers hybridising to other restriction fragments, they can analyse junctions of choice formed with this constant fragment. The TaqMan® probe is designed such that it hybridises in between the restriction site of interest and the constant PCR primer, with probe and primer hybridising to opposite strands (Splinter et al., 2006). This configuration ensures that the fluorescent signal provided by the probe is strictly specific to the amplification of the ligation product selected for analysis. A detailed protocol for this approach was published recently (Hagege et al., 2007).
Different primer pairs will have different amplification efficiencies and in order to account for

this, these efficiencies need to be assessed. This is done on a control template containing all ligation products in equimolar amounts (Dekker, 2006; Dekker et al., 2002; Palstra et al., 2003; Splinter et al., 2004; Tolhuis et al., 2002), mixed with the same amount of genomic DNA as is present in the 3C PCR reaction. To be able to compare results obtained with different 3C templates, one needs to account for possible differences in quality and quantity between these templates. This is done by analysing interaction frequencies between segments in a control locus that is expected to adopt a similar conformation in the different cell-types of interest, as has been described previously (Dekker, 2006; Splinter et al., 2004; Tolhuis et al., 2002).

The ChIP-loop assay

It is often found by 3C technology that a single DNA site interacts with multiple other sites in the population of cells analysed. In many cases, this is likely to reflect cell-to-cell differences in chromatin conformation and it is very well possible that in different subpopulations of cells distinct proteins bind to such a given site and mediate the different DNA interactions. The chromatin immunoprecipitation-combined loop (ChIP-loop) assay was developed to investigate this (Cai et al., 2006; Horike et al., 2005; Kumar et al., 2007). The method involves formaldehyde cross-linking of cells, restriction enzyme digestion and urea gradient purification of cross-linked chromatin, immunoprecipitation using an antibody against the protein of interest, ligation of precipitated DNA fragments (still coupled to the beads) and PCR analysis of the junctions (Figure 2.1).

In our opinion, a number of technical aspects complicate the analysis of results obtained by ChIP-loop. First, it is not clear why current protocols ligate the fragments when they are bound and concentrated to the beads. Concentrating the DNA on the beads prior to ligation is expected to facilitate the formation of junctions between bead-associated, but not necessarily formaldehyde-crosslinked, DNA fragments, hence also producing results that reflect loops formed on the beads rather than in nuclear space. Unless it is demonstrated that such undesired events do not take place, we would argue it is better to carry out the precipitation after the ligation step, which needs to be performed under conditions as described in the 3C procedure. Second, accurate quantification of ligation products is already very challenging in standard 3C and will even be more difficult in ChIP-loop assays, because it must take into account the relative enrichment of each site on the beads. For example, we would argue that ChIP-loop assays should only be directed to the analysis of fragments that are both enriched by ChIP. Indeed, we tend to question the relevance of analysing, via ChIP-loop, interactions between DNA segments that are not bound by the protein of interest, or between DNA segments of which only one is enriched by the antibody. After all, if a sequence had been co-precipitated because it was cross-linked to a target sequence of the protein of interest, it should also be found enriched in the ChIP assay. It may be possible to obtain unique information, not obtainable from ChIP or

3C only, when studying loops formed between sites that are both precipitated via the protein-antibody interaction. However, like in 3C, the mere detection of a ligation product will not tell whether they randomly collide or interact specifically. Interaction frequencies need to be quantified accurately and compared to other interactions between sequences enriched in the same experiment. To decide whether the ChIP captured a specific DNA-DNA interaction, one must then also take into account the genomic site separation between each pair of segments and the relative enrichment of each site on the beads. This seems to make the interpretation of ChIP-loop results very difficult. ChIP-loop assays can be useful though to identify proteins participating in interactions between sites far apart on the chromosome (i.e. hundreds of kilobases or more) or sites located on different chromosomes, as in this case the interpretation of results will not be complicated by frequent random collisions.

5C technology

Large-scale mapping of for example several hundreds of chromatin interactions using standard 3C is time consuming and difficult. The introduction of the 3C-Carbon Copy (5C) method generates the possibility of such large-scale locus-wide analysis (Dostie and Dekker, 2007; Dostie et al., 2006). The method uses a multiplex ligation-mediated amplification (LMA) step to detect and amplify selected ligation junctions, thereby generating a quantitative carbon copy of a part of the initial 3C library, which is subsequently analysed via microarray detection or high throughput sequencing (Figure 2.1). LMA involves using a combination of test and fixed 5C primers that hybridise to the sense and antisense strand, respectively, of the restriction fragment ends analysed. Fixed and test primers will be directly juxtaposed when a ligation junction is formed between the corresponding restriction sites, allowing subsequent primer-primer ligation. Universal tails protruding from the test and fixed primers, such as T7 and complementary T3 promoter sequences, subsequently enable massive parallel quantitative amplification of all ligation products formed between selected the fragments. 5C technology has the opportunity of analysing a locus from a single or multiple fixed points. It can generate a complex matrix of interaction frequencies for a given genomic region, which can be used to reconstruct the intricate topology of this region. The size of the genomic region that can be studied is limited by the number of 5C primers that can be used simultaneously. Scanning hundreds of megabases of the genome will require using tens of thousands of 5C primers. 5C technology is therefore not very suitable to scan the entire genome for DNA interactions (Dostie and Dekker, 2007).

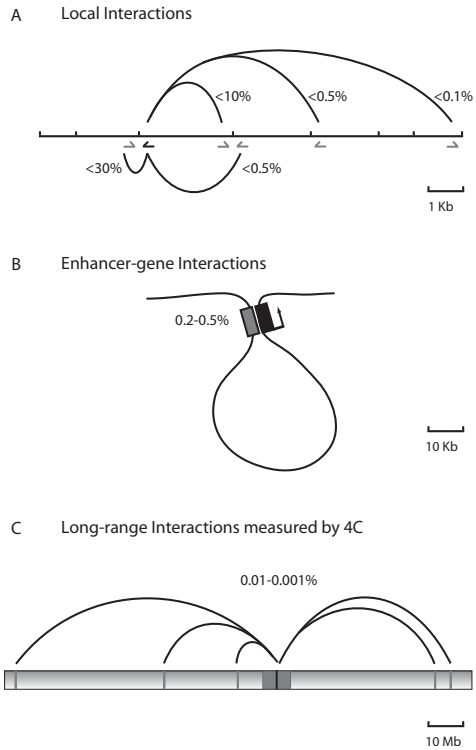

Figure 2.4
Ligation events measured at the β-globin locus.
Schematic presentation of the relative abundance of frequently formed ligation products at the mouse β-globin locus. Typical values for ligation events (in % alleles) are indicated for **A**. Local interactions with directly neighbouring restriction fragments; data measured using 3C-qPCR. Arrows below the restriction fragments indicate the location and direction of 3C primers ('bait' primer indicated in black) **B**. Enhancer-gene interactions over 30-100 Kb; data measured using 3C-qPCR **C**. Long-range interactions *in cis* and *in trans* (>1Mb); data measured by 4C **D**. Table summarizing the relative abundance of different ligation products formed at the mouse β-globin locus (in % alleles).

4C technology

3C and 5C technology have been developed to identify interacting elements between selected parts of the genome and both techniques require the design of primers for all restriction fragments analysed. Recently, new strategies have been developed that allow screening the entire genome in an unbiased manner for DNA segments that physically interact with a DNA fragment of choice (Lomvardas et al., 2006; Simonis et al., 2006; Wurtele and Chartrand, 2006; Zhao et al., 2006). They are all based on 3C technology and very similar in principle and we therefore collectively refer to these techniques as '4C technology'.

An outline of 4C technology is provided in (Figure 2.1). Just like 3C, 4C technology depends on the selective ligation of cross-linked DNA fragments to a restriction fragment of choice (the 'bait'). In 4C technology, all the DNA fragments captured by the bait in the population of cells

are simultaneously amplified via inverse PCR, using two bait-specific primers that amplify from circularized ligation products. Essentially two strategies can be pursued to obtain these DNA circles (Figure 2.5). One strategy relies on the formation of circles during the standard 3C ligation step, i.e. while the DNA is still cross-linked (Lomvardas et al., 2006; Zhao et al., 2006). Here, circle formation requires both ends of the bait fragment to be ligated to both ends of a captured restriction fragment. If multiple restriction fragments are cross-linked together (see above), circles may still be formed but they can contain more than one captured fragment and will therefore be larger (Figure 2.5). After de-crosslinking, captured DNA fragments are directly amplified by inverse PCR, using bait-specific primers facing outwards. Restriction enzymes recognizing four or six basepairs can be used in this set up. Four-cutters are preferred in this method though (Zhao et al., 2006), since they produce smaller restriction fragments (average size 256 bp, versus ~4 kb for six-cutters) and linear PCR amplification of the captured DNA fragments requires that the average product size is small.

The second strategy relies on the formation of DNA circles after the chromatin has been de-cross-linked (Figure 2.5). Here, the standard 3C procedure is followed, using a six-cutter as the restriction enzyme, which yields a 3C template that contains de-cross-linked ligation products. These junctions are then further processed to reduce the size of the captured fragments, which is done by a second round of digestion, this time with a four-cutter restriction enzyme. The trimmed ligation junctions are subsequently re-ligated under conditions that favor the formation of self-ligated circles and inverse PCR primers hybridizing to the bait are used to linearly amplify (the small outer ends of) captured DNA fragments (Simonis et al., 2006; Wurtele and Chartrand, 2006).

Since the two strategies have not been worked out in similar detail yet, it is currently difficult to compare them. Theoretically, each strategy will have its own advantages and disadvantages. A clear advantage of the first approach is that it requires less processing steps. Additionally, the use of four-cutters will provide a higher resolution (256 bp versus 4 kb for a four-cutter versus six-cutter, respectively). This should better allow defining the site of interaction, which is expected to be particularly useful for identifying *cis*-regulatory DNA elements that locate away from a gene of interest (see below). A potential issue of concern exists if formaldehyde cross-links multiple DNA fragments together (see above). As a consequence of this, circles formed between cross-linked DNA fragments may contain more than two captured fragments, which will often be too large to be amplified in a linear fashion. This, in turn, may affect the reproducibility of the approach. It is also not clear how efficient circle formation is between DNA fragments that reside in cross-linked chromatin aggregates. Analysis of fragments captured by this approach so far has been limited to the sequencing of relatively small numbers of clones, 114 (Zhao et al., 2006) and 320 (Lomvardas et al., 2006), respectively. Although these studies identified interesting DNA fragments, it is not clear whether such small number of clones provides a fair representation of the complex library of ligation junctions.

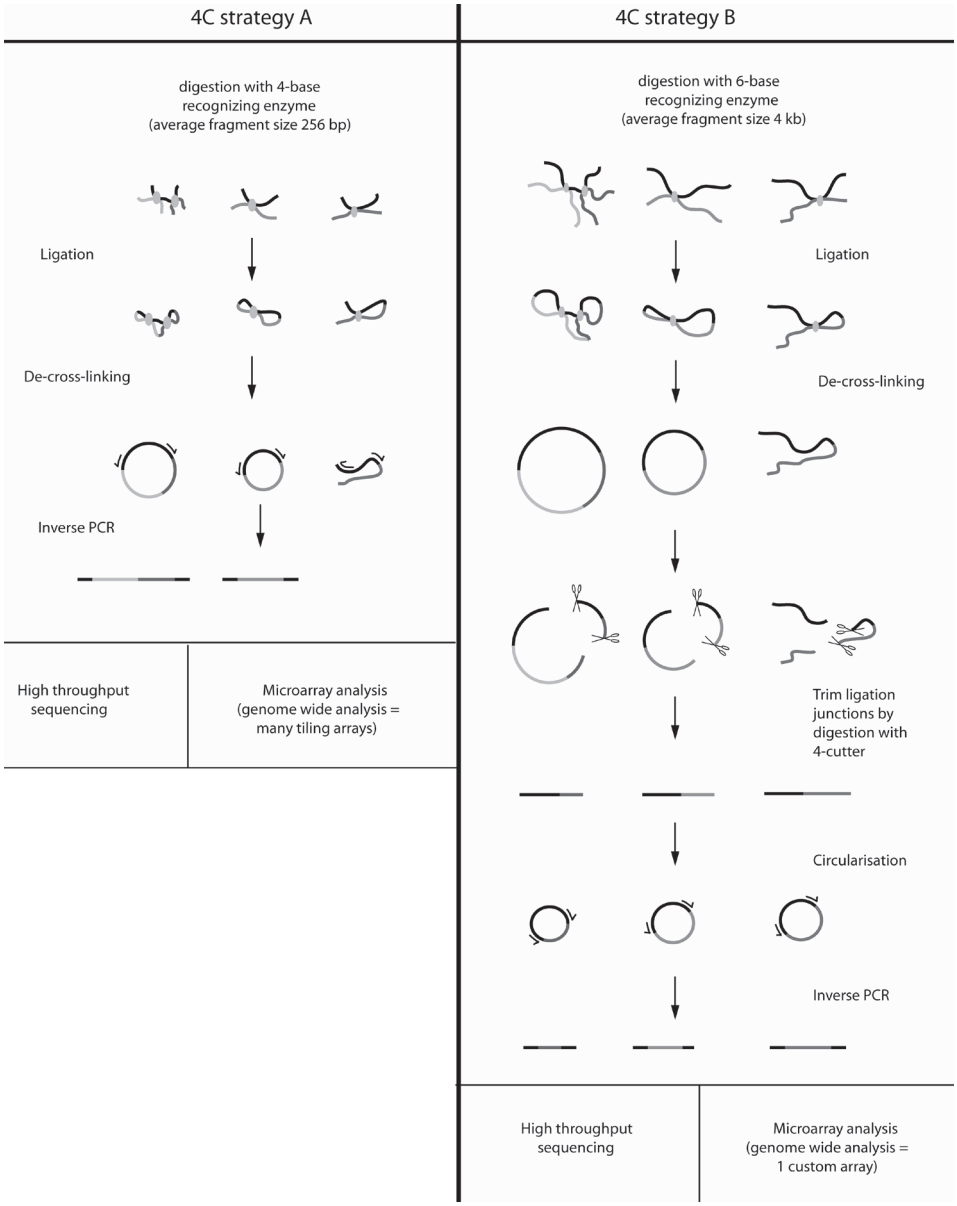

Figure 2.5

Outline of the two basic 4C strategies.

4C is based on 3C, but here all fragments ligated to a "bait" fragment (black) are amplified in one inverse PCR on circularized ligation products, using primers that amplify from the "bait". There are two 4C strategies. Strategy A entails the use of a 4-cutter enzyme, resulting in a resolution of 256 bp, and relies on the circular ligation of DNA fragments while they are still cross-linked. In strategy B, the cross-linked material is digested with a 6-cutter enzyme resulting in a resolution of 4 kb. The ligation junctions are trimmed and circularized after de-cross-linking. In both strategies PCR products can be analysed by either large-scale sequencing or microarray analysis, in which strategy A requires (many) tilling arrays and strategy B one custom designed array, for a whole genome analysis.

The advantage of the second approach, which uses six-cutters and relies on the formation of circles after the DNA has been de-cross-linked, is that it depends on the ligation of only one end of the bait to one end of a cross-linked DNA fragment, which will be more efficient than forming a circle between cross-linked DNA fragments. Circle formation takes place when the DNA is naked, which will also be more efficient than when the DNA is cross-linked. Products to be amplified will generally be smaller, as the circles will not contain more than one captured fragment, and therefore easier to amplify in a linear fashion. Since the strategy selectively amplifies the ligated outer ends of the restriction fragments created by the six-cutter, the complexity of the genomic library to be analysed is strongly reduced. One can take advantage of this by designing tailored microarrays containing only probes located directly adjacent (within 100 bp) to each recognition site of a given six-cutter (e.g. *HindIII*) in the genome to analyse the captured DNA fragments (Simonis et al., 2006). This design allows for a large representation of the genome to be spotted on a single array. Original designs represented seven complete mouse chromosomes, while current designs cover the complete human or mouse genome on a single Nimblegen microarray (400.000 probes), enabling the identification of interactions at a resolution of ~7 kb (unpublished data).

Tailored microarrays were used to simultaneously analyse hundred thousands of fragments captured by the second approach. Replicate experiments performed on biologically independent samples demonstrated this strategy to be highly reproducible. Independent of the bait chosen for analysis (we have now analysed interactions with more than 15 different baits), it is always found that sequences physically close on the linear chromosome template are largely over-represented (Figure 2.6A). In fact, restriction fragments within 5 to 10 Mb from the bait are always captured so efficiently that they saturate every corresponding probe present on the array, precluding a quantitative analysis of local signal intensities. Further away from the bait and on other chromosomes, clusters of 20-50 neighbouring restriction fragments can be identified that all show increased hybridisation signals (Figure 2.6B). Since each probe analyses an independent ligation event and only two fragments can be captured per cell, such clustering of interacting DNA fragments strongly indicates that this genomic region contacts the bait in multiple cells. Importantly, high-resolution cryo-FISH confirmed independently for more than 20 of these regions that they truly represent interacting regions in *cis* and in *trans* (Simonis et al., 2006). These experiments also showed that 4C technology identifies interacting regions in *trans* even if they are together in only 4% of the cells (cryo-FISH background in *trans*: 0-2%) and in *cis* even if they are together in only 6% of the cells (cryo-FISH background in *cis*: 2-4%). Compared to local signals surrounding the target sequence, the amount and intensity of non-specific signals found on unrelated chromosomes are minimal, showing that the background (caused e.g. by random ligation) of this technology is very low (Simonis et al., 2006).

Potential pitfalls of 4C technology

Number of cells

Whichever strategy is followed, several critical steps need to be considered. First, analysis must be performed on a relatively large population of cells. Even frequent interactions between fragments close on the linear chromosome template are captured often in less than 1/500 cells and we think that the *trans* and long-range *cis*-interactions that we identify are captured in only 1/10.000 or even 1/100.000 cells. We routinely process 10 million cells and perform 16 inverse PCR reactions on 200ng template, which we subsequently pool and label for microarray hybridisation. Hence, we analyse an equivalent of approximately one million ligation events on a single microarray.

PCR

The advantage of 4C (and 5C) over 3C is that only 2 primers are required to amplify all products, circumventing the problem of differences in primer pair efficiencies. All PCR based methods suffer from the fact that different amplicons amplify with a different efficiency. By performing the same PCR on a control template containing all ligation products in equimolar amounts one can correct for these differences in 3C and 5C, but not in 4C. It is absolutely critical to optimise the 4C-PCR step, as this step will select the DNA fragments for analysis, which need to correctly represent the fragments captured by the bait. Typically, 80% of the DNA fragments are smaller than 600 bp when samples are processed first with a six cutter and then with a four-cutter, but one also wants larger fragments to be amplified in a linear fashion. Different polymerases will perform this task with different levels of success (data not shown). One can use 3C primers and real-time PCR to test if the abundance of different sized products is similar before and after the inverse PCR step in 4C. We have used this strategy to define conditions that allow fragments of at least 1.2 kb to be amplified at very similar efficiencies. An important test to check if the PCR is set up correctly is to perform reactions on independently processed samples. When separated by gel electrophoresis, they all should give a similar smear of PCR products and a number of more prominent bands that are reproducible between the samples (Figure 2.6C). One should also check if the theoretically most abundant products that originate from the non-digested template and from the self-ligated circle are prominently present, which also confirms that the inverse PCR works (Figure 2.6C).

High-throughput analysis

While sequencing of even hundreds of clones may reveal potentially interesting DNA fragments, we strongly recommend high-throughput analysis of captured DNA fragments, using either microarrays or large-scale sequencing, to exclude that analysis is focused on a misrepresentation of the actual library of captured fragments. Indeed, whatever the bait chosen for analysis

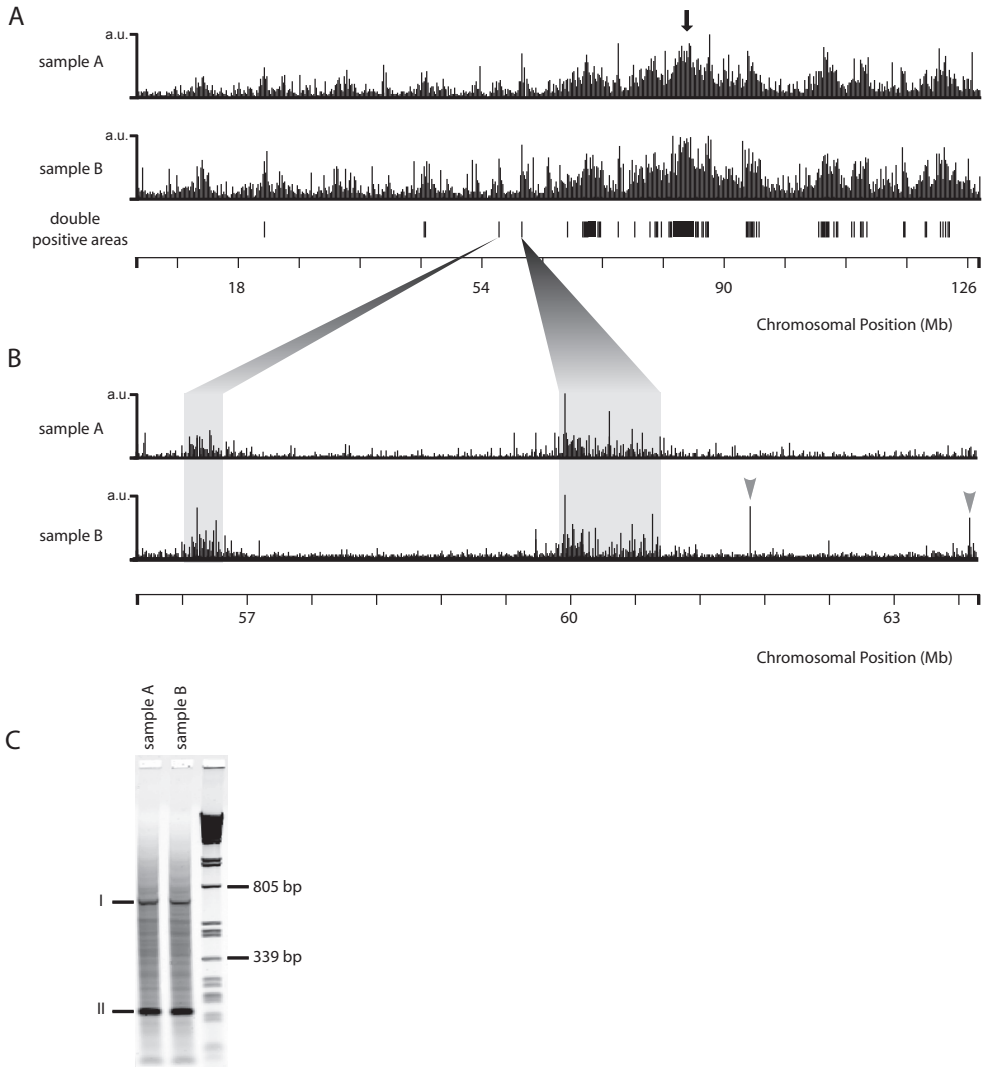

Figure 2.6

Results obtained by 4C technology are highly reproducible.

Example of 4C data analyzing murine *Rad23a*. **A**. Raw data on the cis chromosome of two independent experiments on biological replicates. The arrow indicates the position of the 4C primers. **B**. Raw data of two independent experiments showing reproducible clustering of high signals. Arrowheads indicate irreproducible, isolated high signals, representing random ligation events. **C**. Two 4C PCR products of biological replicates analysed with gel electrophoresis. The appearance is highly reproducible. I- indicates the self-circularized fragment. II -indicates the non-digested product (see text for explanation).

and whatever the 4C strategy used, the great majority of captured fragments will always be located close to the bait on the linear chromosome template (Simonis et al., 2006; Wurtele and Chartrand, 2006).

Analyzing 4C data

The capture of a single restriction fragment away from the bait in *cis* or on another chromosome does not necessarily reveal a specific interaction. High-throughput microarray analysis shows that probes with high signals are found across the chromosome and to a lesser extent also on other chromosomes. Many of these captures are random though, as they are not reproducible between independent duplicate experiments. Thus, highly specific long-range intra- and interchromosomal interactions with single restriction fragments may exist, but it is very difficult to discriminate them from random captures. The presence of genomic clusters of restriction fragments that show increased hybridisation signals in biologically replicate experiments reveal interacting regions, as explained above (Figure 2.6B). These regions can be identified by the application of a sliding window approach that provides a measure for the relative abundance of ligated fragments per genomic area (Simonis et al., 2006).

Verification of 4C data

3C technology may be used as a first verification of data obtained by 4C technology. However, they are not independent technologies and long-range interactions identified by 4C technology should therefore always be verified by completely independent methods such as FISH. Preferably this is done by high-resolution FISH studies, such as 3D-FISH or cryo-FISH (Branco and Pombo, 2006) that use fixation conditions, which preserve the nuclear ultra-structure well. It needs to be demonstrated that two regions identified to interact by 4C technology indeed come together more frequently in the population of cells than two randomly chosen loci.

Concluding remarks and perspectives

The development of 3C technology has contributed enormously to our understanding of the intricate folding of gene loci and revealed for example that transcriptional regulatory DNA elements loop towards their target genes to regulate the expression. Based on 3C technology, a number of new approaches have recently been developed. The ChIP-loop assay may direct structure analysis to specific protein-bound DNA sequences, but correct interpretation is currently still complicated, as it requires a quantitative comparison between ChIP-loop, ChIP and 3C data. 5C technology is expected to provide unprecedented insight into the conformational fine-structure of selected regions in the genome. Like 4C, it may help screening a genomic region for DNA elements that interact with a DNA segment of choice, being either a gene (promoter), an insulator sequence, an enhancer, an origin of replication, etc. 4C technology is expected to contribute importantly to a comprehensive understanding of nuclear architecture (de Laat, 2007), picking up interactions not previously anticipated and putting the relative frequency of interactions in perspective. Current 4C microarray studies allow identifying long-range interactions in cis, over tens of megabases and in trans, between chromosomes.

The large over-representation of fragments closer to the bait precludes a quantitative analysis of local interactions, but it is to be expected that 4C can be modified to also identify loops formed in smaller genomic regions. In the near future more novel 3C-based methods may be expected. Their potential should be evaluated not so much on the possibly exciting nature of the interactions identified but on independent evidence, obtained for example by FISH, that is provided to demonstrate that interactions are real.

Acknowledgements

We thank Frank Grosveld and the members of our lab for discussion. This work was supported by grants from the Dutch Scientific Organisation (NWO) (016-006-026) and (912-04-082) to W.d.L.

β-Globin Active Chromatin Hub formation in differentiating erythroid cells and in p45 NF-E2 knockout mice

Based on: Journal of Biological Chemistry 282, 16544-16552 (2007)

Summary

Expression of the β-globin genes proceeds from basal to exceptionally high levels during erythroid differentiation *in vivo*. High expression is dependent on the locus control region (LCR) and coincides with more frequent LCR-gene contacts. These contacts are established in the context of an Active Chromatin Hub (ACH), a spatial chromatin configuration in which the LCR, together with other regulatory sequences, loops towards the active β-globin-like genes. Here, we used recently established I/11 cells as a model system that faithfully recapitulates the *in vivo* erythroid differentiation program to study the molecular events that accompany and underlie ACH formation. Upon I/11 cell induction, histone modifications change, the ACH is formed and the β-globin-like genes are transcribed at rates similar to those observed *in vivo*. The establishment of frequent LCR-gene contacts coincides with a more efficient loading of polymerase onto the β-globin promoter. Binding of the transcription factors GATA-1 and EKLF to the locus, while previously shown to be required, is not sufficient for ACH formation. Moreover, we use knockout mice to show that the erythroid transcription factor p45 NF-E2, which has been implicated in β-globin gene regulation, is dispensable for β-globin ACH formation.

Introduction

The mammalian β-globin gene loci serve as a model system for studying developmental gene regulation. The murine β-globin locus contains four β-like globin genes that are arranged on the DNA in the order of their developmental expression. Expression of the β-like globin genes is restricted to the erythroid lineage. In erythroid progenitor cells, they are expressed at basal transcription levels, comparable to that of most other genes. However, at later stages of erythroid maturation they are expressed at exceptionally high levels. This high transcription rate is dependent on the β-globin locus control region (LCR) (Epner et al., 1998), a *cis*-regulatory DNA element located upstream of the β-like globin genes that contains six erythroid-specific DNaseI hypersensitive sites (HSs) (Figure 3.1A).

Recently developed 3C technology provides insight into the spatial conformation of the β-globin locus. This technique involves quantitative PCR analysis of formaldehyde cross-links made between selected DNA fragments as a measure of their interaction frequency. 3C technology revealed that high expression of the β-globin genes at later stages of differentiation coincides with the formation of an Active Chromatin Hub (ACH), a spatial configuration of the locus in which the LCR, together with additional HSs upstream and downstream of the locus, loops towards the active β-like globin genes (Tolhuis et al., 2002). At early stages of erythroid differentiation a smaller chromatin hub (CH), is present. It is composed of contacts between the outer HSs of the locus and part of the LCR, but does not contain the genes (Palstra et al., 2003). The β-globin genes switch interaction with the ACH in relation to their transcriptional

activity during development. The embryonic genes contact the LCR in embryonic blood, while the adult genes do so in fetal liver and adult bone marrow (Palstra et al., 2003). Based on these observations, ACH formation was proposed to be crucial for the high expression rates of the β-globin genes: spatial clustering of *cis*-regulatory DNA elements ensures a high local concentration of transcription factors, required for efficient gene transcription (de Laat and Grosveld, 2003). While this is likely true for LCR-gene contacts, the functional significance of the outer HSs participating in the hub is still unclear. Deletion of these sites, or depletion of the transcription factor CTCF, which abrogates interactions with the outer HSs, has no measurable effect on β-globin gene expression (Bender et al., 2006; Splinter et al., 2006). Little is known about the molecular events that are involved in the establishment of contacts between the LCR and the genes as the locus proceeds from a CH to an ACH during erythroid differentiation.

Mice containing a targeted deletion of the β-globin LCR have provided insight into LCR-independent modifications of the locus (Epner et al., 1998). While none of the β-globin genes express beyond basal levels in these mice, they appear to have similar chromatin properties as their wildtype counterparts that express 25-100 fold more efficiently. Thus, the promoters are hypersensitive, contain highly acetylated histones and bind both basal transcription factors, such as TATA binding protein (TBP), and tissue-specific transcription factors, such as NF-E2 and GATA-1, albeit at slightly reduced efficiency (Bender et al., 2000; Sawado et al., 2003; Schubeler et al., 2000; Vakoc et al., 2005). RNA polymerase II (RNAP II) binding to the β_{major} promoter was reduced only twofold in the absence of the LCR, while a more dramatic reduction (5-fold) in binding was observed at the third exon of β_{major}, leading to the conclusion that the LCR primarily acts by enhancing the transition from initiation to elongation (Sawado et al., 2003).

GATA-1, EKLF and NF-E2 are the best characterized tissue-specific transcription factors involved in β-globin gene transcription. All three factors bind to elements in the β-globin LCR and to promoters of the β-globin genes (Forsberg et al., 2000; Im et al., 2005; Johnson et al., 2002; Kang et al., 2002). GATA-1 is essential for the development of the erythroid lineage in mice (Pevny et al., 1991) and restoration of GATA-1 activity in a GATA-1 null cell line leads to β-globin gene activation (Weiss et al., 1997). EKLF is essential for adult β-globin gene expression in mice (Nuez et al., 1995; Perkins et al., 1995). EKLF and GATA-1 both function in β-globin ACH formation; frequent contacts between the LCR and the genes are lacking in their absence (Drissen et al., 2004; Vakoc et al., 2005). The role of NF-E2 in β-globin gene regulation is less well understood. NF-E2 is composed of a ubiquitously expressed subunit, MafK (or p18 NF-E2), and an erythroid-specific transactivator subunit, p45 NF-E2 (Andrews et al., 1993). NF-E2 binds most prominently to HS2 in the LCR but also to an element just downstream of the adult β-globin gene promoter (Forsberg et al., 2000; Kang et al., 2002). Based on studies using a mouse erythroleukemia (MEL) cell line lacking the p45 subunit, NF-E2 is thought to be essential for β-globin gene transcription and to mediate the hyperacetylation of histones and the deposition of RNAP II at the β-globin gene promoters (Johnson et al., 2001; Kotkow and Orkin, 1995; Lu et al.,

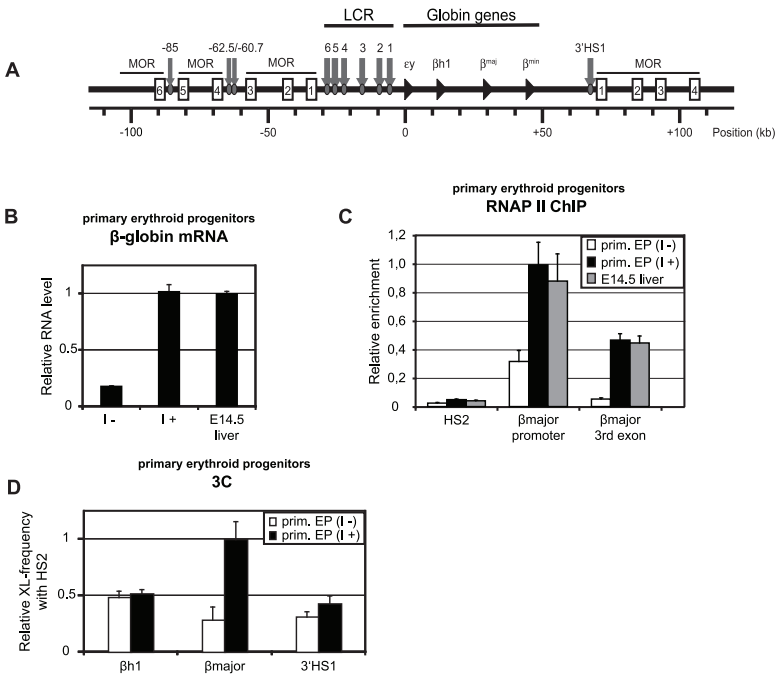

Figure 3.1

Characterization of murine primary erythroid progenitor cells.

A. Schematic presentation of the mouse β-globin locus. MORs (mouse olfactory receptor genes) are indicated by boxes; embryonic (εy, β_{h1}) and adult (β_{major} and β_{minor}) β-globin genes are indicated by triangles. DNaseI HSs are indicated by arrows.

B. Relative β-globin mRNA levels (as determined by quantitative RT-PCR) in non-induced, proliferative primary erythroid progenitor cells (I-) and induced, differentiated primary erythroid cells (I+) and E14.5 liver cells (set to 1).

C. Chromatin immunoprecipitation (ChIP) using an antibody against RNAP II. Plotted are relative enrichments measured for sites in the β-globin locus in non-induced (I-) and induced (I+) primary erythroid progenitor cells and E14.5 liver cells. Highest value set to 1.

D. Quantitative 3C analysis on non-induced (I-) and induced (I+) primary erythroid cells. Shown are relative cross-linking frequencies with HS2. Highest value set to 1. In panel **B**, **C** and **D** the standard error of the mean (SEM) is indicated.

1994; Vieira et al., 2004). However, mice lacking the erythroid-specific subunit p45 show normal erythropoiesis and express the β-globin genes at slightly reduced levels compared to wildtype (Shivdasani and Orkin, 1995). The current view is that p45 NF-E2 is crucial in certain cultured erythroid cell systems, but has redundant factors *in vivo*. However, mice lacking both p45 NF-E2 and the related factors Nrf-2, or Nrf-3, did not show an erythroid phenotype beyond that seen with loss of p45 NF-E2 alone (Derjuga et al., 2004; Kuroha et al., 1998; Martin et al., 1998). It is unknown whether NF-E2 functions in β-globin ACH formation.

Here we analysed the role of erythroid-specific transcription factors and characterize the molecular events that accompany activation of high β-globin gene expression rates at the late stages of erythroid differentiation. Careful analyses of recently established I/11 erythroid cells

show that they provide a good model system for these studies, where upon maturation, β-globin gene expression increases to reach the high transcription levels observed *in vivo*. Increased transcription coincides with stabilized LCR-gene contacts and a more efficient loading of RNAP II onto the β-globin gene promoters. GATA-1 and EKLF bind their target sites in the β-globin locus prior to ACH formation, showing that although required, the mere binding of these factors is not sufficient for the establishment of frequent LCR-gene contacts. Similarly, NF-E2 already binds to the locus in erythroid progenitor cells prior to ACH formation. We used p45 knockout mice to investigate the role of NF-E2 in β-globin gene regulation and demonstrate that while β-globin gene expression is slightly reduced in the absence of this factor, the β-globin ACH is formed normally. Thus, unlike GATA-1 and EKLF, NF-E2 is dispensable for long-range LCR-gene contacts in the β-globin locus.

Results

I/11 cells as a model system to study LCR mediated β-globin gene expression

Primary erythroid cells directly isolated from mouse fetal liver can both be expanded *in vitro* and induced to undergo synchronous differentiation to fully mature, enucleated erythrocytes (Dolznig et al., 2005). We found that they expressed the adult β_{major} gene at basal levels when brought into culture under conditions that enriched for progenitor cells (Figure 3.1B). Differentiation of these cells *in vitro* resulted in an increase in steady-state β-globin messenger RNA levels to amounts similar to that observed *in vivo* (Figure 3.1B). Differentiation was accompanied by an increase in RNAP II-binding to the promoter and third exon of the β_{major} gene (Figure 3.1C) and an increase in interaction frequency between the LCR (HS2) and the β_{major} gene (Figure 3.1D), as determined by an improved 3C strategy that uses Taqman probes for a more quantitative analysis of interaction frequencies (Hagege et al., 2007; Splinter et al., 2006). However, primary progenitor cells often differentiate spontaneously, which precludes the collection of large homogeneous cell populations required for large-scale chromatin immuno-precipitation (ChIP) and chromosome conformation capture (3C) (Dekker et al., 2002) experiments. We therefore compared I/11 erythroid progenitors (Dolznig et al., 2001) to these primary erythroid cells with respect to β-globin gene expression characteristics during erythroid differentiation.

I/11 cells are erythroid progenitor cells established from fetal livers of p53 -/- mice. The cells can be expanded indefinitely *in vitro* and upon exposure to physiologically relevant stimuli such as erythropoietin, they undergo terminal differentiation into enucleated erythrocytes in a synchronized manner (Dolznig et al., 2001; von Lindern et al., 2001). Detailed analysis of β-globin gene expression showed that differentiation of I/11 cells was accompanied by an approximately 20-fold increase in stead-state levels of β_{major} transcripts, reaching levels comparable to those observed *in vivo* (Figure 3.2A). Also, RNAP II-binding to the promoter and third exon of the β_{major} gene increased and reached levels observed *in vivo* (Figure 3.2B). Finally, more frequent

LCR-gene contacts were observed upon differentiation of I/11 cells (Figure 3.2C) (Palstra et al., 2003), in a manner similar to that seen in primary erythroid cells (Figure 3.1D). Therefore, we conclude that I/11 cells faithfully recapitulate the later stages of erythroid differentiation not only morphologically (Dolznig et al., 2001) but also with respect to the transcriptional profile of the β-globin genes. Moreover, we note that non-induced proliferating I/11 progenitor cells express the β-globin genes at levels comparable to the basal β-globin gene expression levels observed in transgenic mice lacking the LCR (Epner et al., 1998). Non-induced cells also lack the

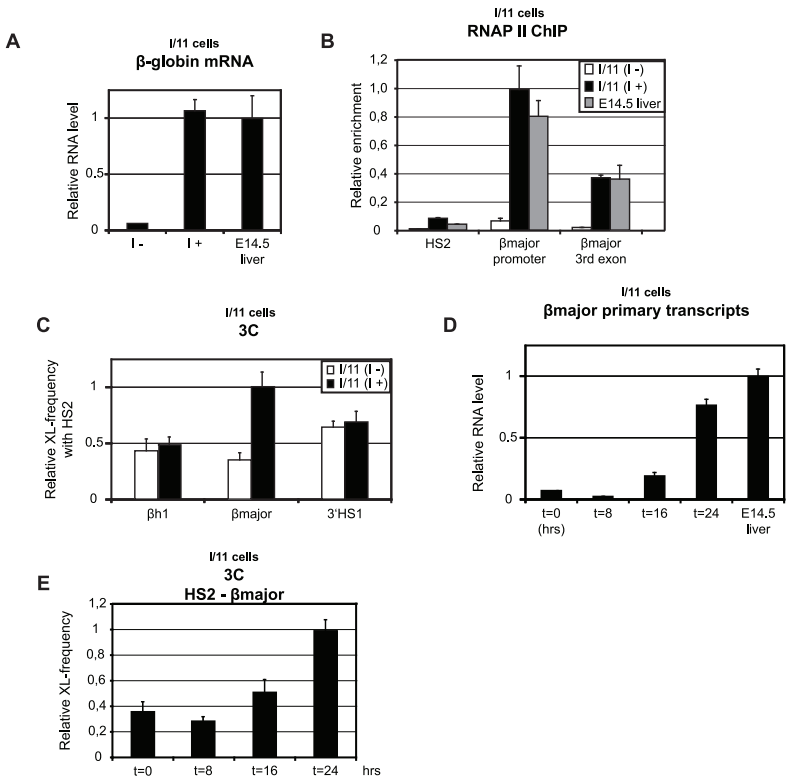

Figure 3.2

I/11 cells as a model system to study LCR-mediated β-globin gene expression.

A. Relative β-globin mRNA levels (as determined by quantitative RT-PCR) in non-induced, proliferative (I-) and induced, differentiated (I+) I/11 cells and E14.5 liver (set to 1).

B. Chromatin immunoprecipitation (ChIP) using an antibody against RNAP II. Plotted are relative enrichments measured for sites in the β-globin locus in non-induced (I-) and induced (I+) I/11 cells and E14.5 liver cells. Highest value set to 1.

C. Quantitative 3C analysis on non-induced (I-) and induced (I+) I/11 cells. Shown are relative cross-linking frequencies with HS2. Highest value set to 1.

D. Relative β-globin primary transcript levels (as determined by quantitative RT-PCR) in I/11 cells at different time-points (hours) after induction of differentiation and in E14.5 liver cells (set to 1).

E. Quantitative 3C analysis on I/11 cells at different time-points (hours) after induction of differentiation. Shown is the relative cross-linking frequency between HS2 and β_{major}. Highest value set to 1. In all panels the standard error of the mean (SEM) is indicated.

frequent β-globin LCR-gene contacts observed in induced I/11 cells, suggesting that proliferating I/11 progenitor cells represent a stage of erythroid differentiation where transcription of the β-globin-like genes is mostly LCR-independent. Indeed, when I/11 cells were analysed at different time intervals after induction of differentiation, increased expression rates of β_{major}, as measured by the amount of nascent transcripts (Figure 3.2D), coincided well with increased HS2/β_{major} interaction frequencies (Figure 3.2E), supporting the idea that the LCR regulates transcription by contacting the gene. We therefore conclude that I/11 cells are a good model system to study the molecular mechanisms behind LCR-mediated activation of high β-globin gene expression levels.

Histone modifications accompanying ACH formation

Next, ChIP experiments were performed on I/11 cells to investigate if ACH formation and transcriptional enhancement is accompanied by changes in histone modifications at the regulatory sites of the β-globin locus. Acetylation of histone H3 (AcH3) is a mark for active chromatin. The murine β-globin locus is characterized by high levels of erythroid-specific AcH3 at the *cis*-regulatory sites present in the LCR and at the promoters of the active genes (Forsberg et al., 2000; Schubeler et al., 2001). We found that in non-induced I/11 cells, histones at the HSs of the LCR and the promoter of the active β_{major} gene, but not at the inactive βh1 gene, were already hyper-acetylated. In differentiated I/11 cells, H3 acetylation levels increased maximally 3.5-fold at sites within the LCR and less than 2-fold at the β_{major} promoter (Figure 3.3A). The observation

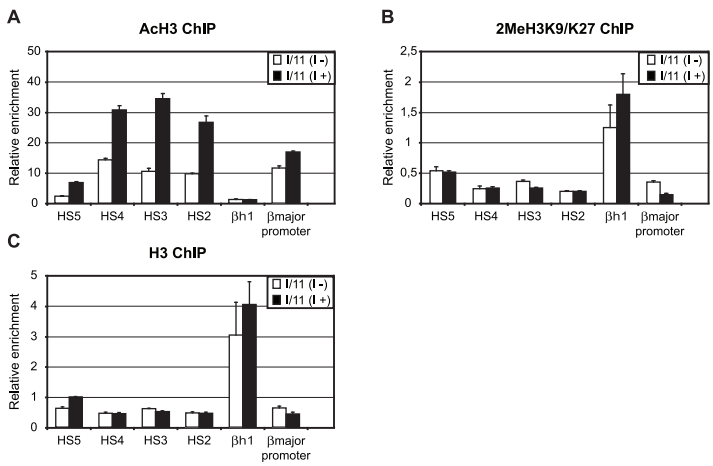

Figure 3.3

Differentiation of I/11 cells is accompanied by an increased ratio of positive versus negative chromatin modifications at the LCR and β_{major} promoter.

A,**B**,**C**. Chromatin immunoprecipitation (ChIP) using an antibody against acetylated histone H3 (AcH3) (**A**), di-methylated lysine-9 and -27 of histone H3 (2MeH3K9/K27) (**B**) and C-terminus of histone H3 (H3) (**C**). Plotted are relative enrichments measured for sites in the β-globin locus in non-induced (I-) and induced (I+) I/11 cells. In all panels the standard error of the mean (SEM) is indicated.

that histone H3 at the β_{major} promoter is already hyper-acetylated prior to ACH formation is in agreement with the fact that promoter hyper-acetylation occurs also in the absence of the LCR (Schubeler et al., 2000).

The data show that a strong increase in β_{major} transcription activity that takes place upon erythroid differentiation is accompanied with a small increase in levels of H3 acetylation at the promoter and we therefore investigated if local abundance of a repressive histone modification also changes upon I/11 differentiation. Using an antibody that recognizes both di-methylated lysine-9 and -27 of histone H3 (2MeH3K9/K27), we found that non-induced I/11 cells already contained low levels of 2MeH3K9/K27 at the HSs of the LCR that did not change substantially when cells underwent differentiation (Figure 3.3B). Levels at the β_{major} promoter were also low prior to induction, but dropped even further (2-fold) upon cellular differentiation, while 2MeH3K9/K27 levels at the inactive β_{h1} promoter were high throughout differentiation. Using an antibody against the C-terminal part of histone H3, we found high amounts of H3 at the inactive βh1 promoter and low amounts of H3 at the regulatory sites of the LCR and the β_{major} promoter (Figure 3.3C), in agreement with their low nucleosome density and hypersensitivity for nuclease digestion. Histone H3 abundance at the sites remained largely the same during differentiation, showing that our ChIP results obtained with antibodies against acetylated and methylated H3 truly reflect differentiation-dependent changes in histone H3 modifications. We conclude that the transition from basal to highly activated β-globin gene expression during erythroid differentiation is accompanied by an increased ratio of positive versus negative chromatin modifications at the β_{major} promoter and the HSs of the LCR.

GATA-1 and EKLF are required, but not sufficient, for ACH formation

Erythroid-specific β-globin gene expression is in part regulated through the action of lineage-restricted transcription factors such as GATA-1 and EKLF that bind to specific motifs found at regulatory sequences throughout the locus. Recent insight into the regulatory role of these transcription factors has come from EKLF knockout mice and GATA-1-null cells containing inducible versions of the corresponding proteins. For both transcription factors it was shown that induction of their activity resulted in more frequent contacts between the LCR and the β_{major} gene and increased β_{major} expression (Drissen et al., 2004; Vakoc et al., 2005). Without EKLF, the β-globin locus adopted a spatial configuration reminiscent of the CH present in erythroid progenitor cells (Drissen et al., 2004). Thus EKLF and GATA-1 appear to be required for the conformation of the β-globin locus to proceed from a CH, containing only the outer HSs of the locus and part of the LCR, to an ACH containing also the entire LCR and the actively transcribed gene. The experimental systems used for these studies precluded investigating if binding of EKLF and/or GATA-1 to the β-globin locus is also sufficient for ACH formation. We therefore analysed if EKLF and GATA-1 are bound to the β-globin locus in non-induced I/11 progenitor cells that do

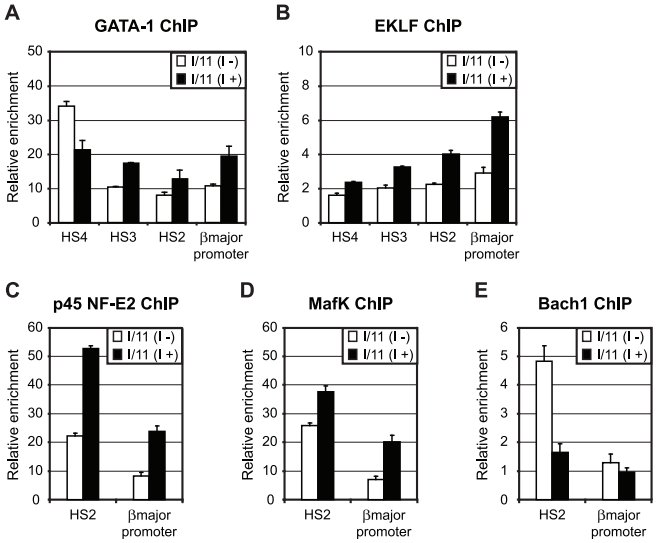

Figure 3.4
GATA-1, EKLF and p45 NF-E2 are already bound to their cognate binding sites prior to ACH formation.
A,**B**,**C**,**D**,**E**. Chromatin immunoprecipitation (ChIP) using an antibody against GATA-1 (**A**), EKLF (**B**), p45 NF-E2 (**C**), MafK (**D**) and Bach1 (**E**). Plotted are relative enrichments measured for sites in the β-globin locus in non-induced (I-) and induced (I+) I/11 cells. In all panels the standard error of the mean (SEM) is indicated. Note that the scale on the Y-axis differs between the panels.

not show ACH formation. ChIP experiments on non-induced I/11 cells revealed strong binding of GATA-1 to its target sites in the β-gobin locus (Figure 3.4A). EKLF was also significantly enriched at the HSs of the LCR and the β_{major} promoter in these cells, albeit not as pronounced, which may be a consequence of the quality of the antibody (Figure 3.4B). These data demonstrate that binding per se of GATA-1 and EKLF to the locus, while required, is not sufficient for ACH formation. We found that at later stages of differentiation, when the β-globin ACH has formed, EKLF-binding to the locus was stronger, as was binding of GATA-1 to most of its β-globin target sites (Figure 3.4A and 3.4B).

Binding of p45 NF-E2 is not sufficient for ACH formation

We then analysed the binding characteristics of NF-E2 to the β-globin locus. First, we focused on the erythroid-specific transactivator subunit p45 NF-E2. Previously, in non-induced MEL cells this protein was shown to be absent from the β_{major} promoter and to bind weakly to HS2 (Brand et al., 2004; Sawado et al., 2001). However, we found that p45 NF-E2 strongly bound to HS2 and the β_{major} promoter in non-induced I/11 cells (Figure 3.4C). The same was true for its heterodimerization partner MafK (Figure 3.4D). Recruitment of p45 NF-E2 to the β_{major} promoter in progenitor cells lacking frequent LCR-gene contacts seems in agreement with the observation in mice lacking the LCR that NF-E2 binding to the promoter is LCR-independent (Sawado et al., 2003). Upon differentiation of I/11 cells, binding of both p45 NF-E2 and MafK to HS2 and the β_{major} promoter increased (up to 2.5-fold) (Figure 3.4C and 3.4D), as was also observed for GATA-1 and EKLF. The small subunit of NF-E2, MafK, which contains a DNA binding domain, can not only form a complex with the transactivator protein p45 NF-E2 but also with the repressive factor Bach1, and the exchange of binding partners from Bach1 to p45 NF-E2 was reported to be

a key step in the activation of β-globin gene transcription during MEL cell differentiation (Brand et al., 2004). Bach1 indeed associated with HS2, but not significantly with the β_{major} promoter, in non-induced I/11 cells and its binding to HS2 was lost upon I/11 differentiation (Figure 3.4E). Thus, Bach1 dissociates from HS2 while p45 NF-E2 binding to this site and the β_{major} promoter increases as cells progress through the later stages of erythroid differentiation. However, the exchange of MafK binding partners that accompanies the induction of high levels of β-globin gene expression during I/11 cell differentiation is not as absolute as observed during MEL cell differentiation, since we find that p45 NF-E2 is already abundantly present at HS2 and the β_{major} gene promoter before induction of differentiation. Based on this result we conclude that, like GATA-1 and EKLF, NF-E2 binding to its target sites in the β-globin locus is not sufficient for ACH formation and the induction of high levels of β-globin gene expression.

p45 NF-E2 is dispensable for β-globin ACH formation

To further investigate the role of p45 NF-E2 in β-globin ACH formation and gene activation we used mice deficient for this transcription factor (Shivdasani et al., 1995). We first performed ChIP experiments to confirm that p45 NF-E2 was absent from the β-globin locus in E14.5 livers from NF-E2 knockout embryos. No recruitment of the protein to HS2 or the β_{major} promoter was found with an antibody recognizing the C-terminus of p45 NF-E2. The same result was obtained using an antibody against the N-terminal part of p45 NF-E2, showing that also no truncated versions of the protein associated with the β-globin locus (Figure 3.5A). We then analysed β-globin gene expression in these mice. Previously this was analysed by measuring steady-state messenger RNA and protein levels (Shivdasani and Orkin, 1995). We used intron-specific primers in a real-time quantitative PCR assay and found that in the absence of p45 NF-E2, the rate of β_{major} transcription was reduced to approximately 65% of wild type levels (Figure 3.5B). In agreement, ChIP analyses showed that RNAP II binding to the gene was also reduced to 60-70% in the absence of p45 NF-E2. This reduction was found both at the promoter and the third exon of the β_{major} gene, but not at HS2 of the LCR (Figure 3.5C). Thus, depletion of p45 NF-E2 in mice has a mild but significant effect on β-globin gene expression, in agreement with the presence of a mild erythroid defect in these mice (Shivdasani and Orkin, 1995).

Next, we asked whether p45 NF-E2 is required for β-globin ACH formation. Locus-wide 3C technology was applied to samples obtained from E14.5 livers and revealed that interactions of the β_{major} gene with regulatory DNA elements elsewhere in the locus were not affected by the loss of p45 NF-E2. Thus, the β-globin ACH, containing the outer HSs (HS-85, HS-62/-60 and 3'HS1), the LCR and the active β_{major} gene, was formed normally in p45 knockout mice (Figure 3.5D). We noticed that β_{major} interaction frequencies with HS2 and HS1, both target sites of NF-E2 (Forsberg et al., 2000; Sawado et al., 2001) increased compared to wildtype. Subsequently, we investigate if p45 NF-E2 depletion influenced the binding of its complexing partner MafK to

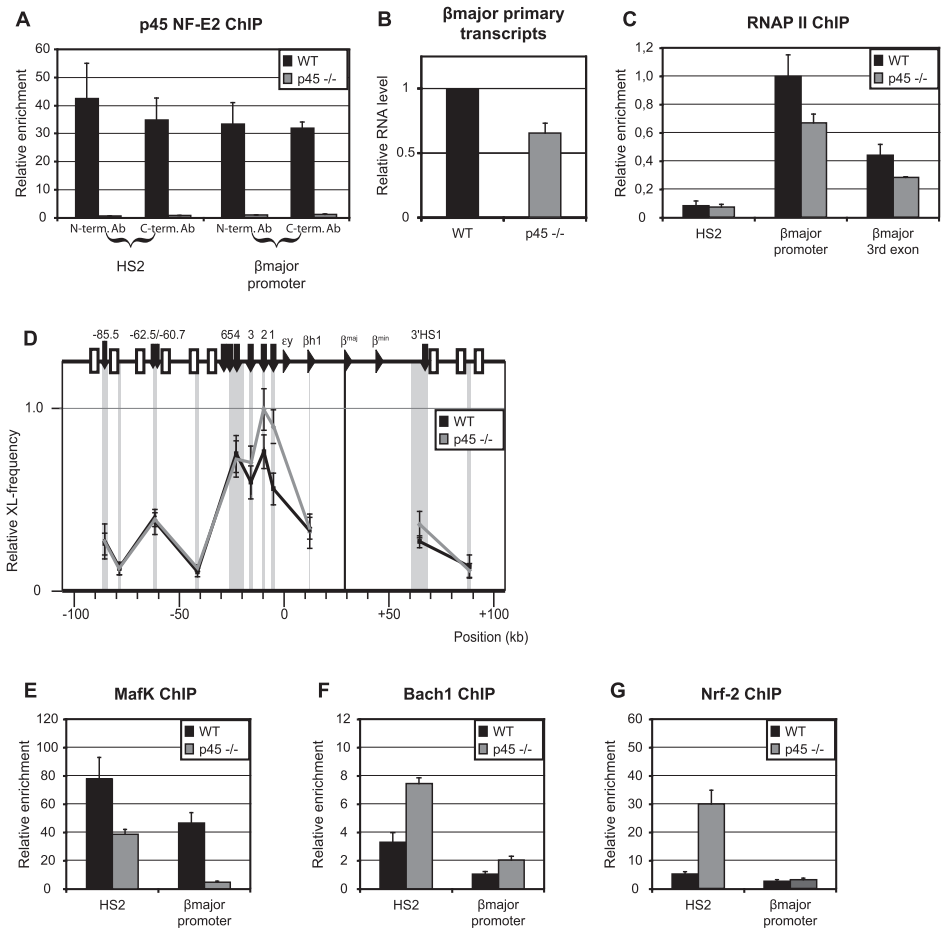

Figure 3.5

p45 NF-E2 is dispensable for β-globin ACH formation.

A. Chromatin immunoprecipitation (ChIP) using antibodies against either the N-terminal or C-terminal part of p45 NF-E2. Plotted are relative enrichments measured for sites in the β-globin locus in E14.5 liver of p45 NF-E2 knockout embryos (p45 -/-) and wild-type (WT) littermates.

B. Relative β-globin primary transcript levels (as determined by quantitative RT-PCR) in E14.5 liver of p45 NF-E2 knockout (p45 -/-) embryos and wild-type (WT) littermates (set to 1).

C. Chromatin immunoprecipitation (ChIP) using an antibody against RNAP II. Plotted are relative enrichments measured for sites in the β-globin locus in E14.5 liver of p45 NF-E2 knockout mice (p45 -/-) and wild-type (WT) littermates. Highest value set to 1.

D. Quantitative 3C analysis on E14.5 liver of p45 NF-E2 knockout embryos (p45 -/-) and wild-type (WT) littermates. Shown are cross-linking frequencies between a β_{major} restriction fragment and several other fragments across the β-globin locus. Data were normalized against cross-linking frequencies measured for two XPB restriction fragments and highest value was set to 1. Size and position of restriction fragments analyzed are indicated by grey vertical shades; black shading represents the fixed fragment β_{major}. Highest value set to 1.

E,**F**,**G**. Chromatin immunoprecipitation (ChIP) using an antibody against MafK (**E**), Bach1 (**F**) and Nrf-2 (**G**). Plotted are relative enrichments measured for sites in the β-globin locus in E14.5 liver of p45 NF-E2 knockout embryos (p45 -/-) and wild-type (WT) littermates. In all panels the standard error of the mean (SEM) is indicated.

the locus. We found that binding of MafK to the β_{major} promoter was almost completely abolished in p45 NF-E2 knockout mice. In contrast, MafK-binding to HS2 was still considerable, albeit 2-fold reduced (Figure 3.5E). This raised the possibility that MafK associated with other proteins at HS2. Indeed, binding of Bach1 increased two-fold and some Bach1 was also found associated with the β_{major} promoter in the absence of p45 NF-E2 (Figure 3.5F). A more dramatic increase in recruitment however was observed when we analysed Nrf-2-binding. Binding of this factor to HS2 increased 6-fold in p45 knockout mice compared to wildtype (Figure 3.5G). No increase in binding efficiency of Nrf-2 was observed at the β_{major} gene promoter in p45 knockout liver cells, consistent with the almost complete absence of MafK at this site. Thus, other MafK dimerization partners, most prominently Nrf-2, appear to physically replace p45 NF-E2 at HS2, while these complexes are almost completely absent at the β_{major} promoter in p45 NF-E2 knockout mice. Collectively, these data show that the transcription factor p45 NF-E2, unlike GATA-1 and EKLF, is not required for the formation of the β-globin ACH.

Discussion

We explored the use of I/11 cells as a model system to study differentiation-dependent β-globin gene activation. I/11 cells are erythroid progenitor cells that upon induction by physiologically relevant stimuli faithfully recapitulate the terminal erythroid differentiation program to mature into enucleated erythrocytes (Dolznig et al., 2001; von Lindern et al., 2001). We show by several criteria that differentiated I/11 cells do express the β-like globin genes at the very high levels observed *in vivo*, while expression prior to induction is ~20-fold lower and comparable to that observed in mice lacking the LCR. This suggests that upon differentiation of I/11 cells, β-globin gene expression proceeds from an LCR-independent to an LCR-dependent mode. In agreement with this idea is the observation that I/11 cell induction coincides with the establishment of frequent LCR-gene contacts. We show that ACH formation and the induction of high β-globin transcription rates is accompanied by an increased ratio of positive (AcH3) versus negative (2MeH3K9/K27) histone modifications at the *cis*-regulatory sites of the β-globin locus. Importantly, we also show increased recruitment of RNAP II to the promoter of the active β_{major} gene. Previously, ChIP experiments have shown that deletion of the LCR caused only a twofold reduction of RNAP II at the β_{major} promoter and a more dramatic reduction (5-fold) at the third exon of β_{major}, leading to the conclusion that the LCR primarily acts by enhancing the transition from initiation to elongation (Sawado et al., 2003). However, we would argue that this is difficult to conclude from ChIP experiments, since this assay cannot provide a measure for transcription re-initiation. ChIP involves the fixation of cells, which usually takes 5-10 minutes. RNAP II reloading onto the β_{major} promoter occurs multiple times during this timeframe, but ChIP fails to appreciate this frequency. The fact that we find increased RNAP II recruitment to the promoter once LCR-gene contacts are established reopens the possibility that the LCR also functions to efficiently recruit

RNAP II to the active β-globin genes. NF-E2, GATA-1 and EKLF are the three lineage-restricted transcription factors most prominently associated with β-globin gene regulation. Previous work has demonstrated that the latter two factors are required for the stabilization of long-range DNA contacts between the β-globin LCR and the active adult β-globin-like genes (Drissen et al., 2004; Vakoc et al., 2005). Here, we demonstrate that GATA-1 and EKLF already bind to cognate sites in the β-globin locus at a stage of erythroid differentiation that precedes ACH formation. This shows that binding per se of these factors to the locus is not sufficient for ACH formation. We find that the efficiency of their recruitment is higher in differentiated cells that do form an ACH, which raises the possibility that binding levels are important for GATA-1 and EKLF to mediate ACH formation. An alternative and not mutually exclusive possibility would be that at later stages of erythroid differentiation the proteins contain different modifications and/or are recruited as parts of different protein complexes in order to mediate β-globin ACH formation (Rodriguez et al., 2005). The role of NF-E2 in the spatial conformation of the β-globin locus has not been investigated before. The factor was reported to be absent from HS2 and the β_{major} gene in non-induced MEL cells (Brand et al., 2004), but was found present at the human β-globin locus in primary multipotent haematopoietic progenitor cells (Bottardi et al., 2006). We also find that p45 NF-E2 is already (abundantly) present at the LCR and the β_{major} promoter in I/11 progenitor cells that do not show frequent LCR-gene contacts and still express the β_{major} gene at basal levels. Thus, binding of p45 NF-E2 per se is not sufficient to confer high expression to the β-like globin genes. We show that in p45 NF-E2 knockout mice long-range contacts in the β-globin locus are formed normally and hence we conclude that p45 NF-E2 is dispensable for β-globin ACH formation.

p45 NF-E2 is a member of the cap 'n collar (CNC) subfamily of basic-leucine zipper (BZIP) transcription factors, to which also Nrf-1, Nrf-2, Nrf-3, Bach1 and Bach2 belong. They can form heterodimers with the small Maf proteins that contain a basic DNA-binding domain but lack a transactivation domain (Motohashi et al., 2002). The Bach factors lack a canonical transactivation domain and act as repressors in reporter assays (Yoshida et al., 1999), but the Nrf factors are all potential candidates to functionally compensate for the loss of p45 NF-E2. However, Nrf-1, -2 and -3 knockout mice each show normal β-globin gene expression (Chan et al., 1998; Derjuga et al., 2004; Farmer et al., 1997; Kuroha et al., 1998; Martin et al., 1998). Moreover, while the early embryonic lethality of Nrf-1 null mice compromises the combination of p45 NF-E2 and Nrf-1 deficiencies, mice lacking both p45 NF-E2 and Nrf-2, or p45 NF-E2 and Nrf-3, did not show an erythroid phenotype beyond that seen with loss of p45 NF-E2 alone (Derjuga et al., 2004; Kuroha et al., 1998; Martin et al., 1998). Based on this it was suggested that Nrf-2 and Nrf-3 do not complement p45 NF-E2 function. Here we demonstrate however that Nrf-2 binding to HS2 strongly increases when p45 is absent. This was not the case at the β_{major} promoter, where MafK binding was also lost in the absence of p45 NF-E2. This shows that different heterodimers have different affinities for DNA target sites. Whether the physical replacement of p45 NF-E2 by Nrf-2

at HS2 also has functional consequences for β-globin gene regulation remains an open question. In this respect it should be mentioned that deletion of HS2 from the endogenous mouse β-globin locus, which removed the most prominent p45 NF-E2-binding sites in the locus, caused a drop in β-globin gene expression similar to that seen in p45 NF-E2 knockout mice (Fiering et al., 1995). This may suggest that the absence of a dramatic effect on β-globin gene expression is not the consequence of related factors compensating functionally for the loss of p45 NF-E2. However, compound knockout mice or knockdown cells lacking all NF-E2 related factors may need to be studied to unambiguously address this issue.

Experimental procedures

Culturing l/11 cells and primary fetal liver cells

In order to generate primary fetal liver cells, livers were isolated from B6 E12.5 embryos, resuspended by repeated pipetting and applied through a 40 μm cell strainer (BD Falcon). Next, primary E12.5 liver cells were cultured as described previously for l/11 cells (Dolznig et al., 2001; von Lindern et al., 2001). In short, cells were cultured in serum free stem cell medium (StemPro-34 SFM, Gibco BRL) containing 0.5 U/ml erythropoietin (a kind gift from Ortho Biotech), 100 ng/ml SCF and 1μM dexamethasone. Proliferating cells were expanded and kept at a density between 1.5 x 10^6 and 4 x 10^6 cells/ml. To induce differentiation, proliferating cells were washed twice with Hanks' Balanced Salt Solution and seeded at 2-3 x 10^6 cells/ml in differentiation medium, containing 5 U/ml erythropoietin and 1 mg/ml transferrin (Sigma-Aldrich). While differentiating, cells were kept at densities of 2-6 x 10^6 cells/ml. Differentiation status was monitored by measuring cellular size distribution using an electronic cell counter (CASY-1, Schärfe-System). Differentiated cells were harvested when size distribution ranged between 6-7 μm, which typically took 24-32 hours for primary fetal liver cells and 40-48 hours for l/11 cells.

Analysis of gene expression

Total RNA was isolated from approximately 1 x 10^6 cells using TRIzol reagent (Invitrogen) according to the manufacturers instructions. 1.5 μg of RNA was treated for 1 hour with DNAseI (Invitrogen) to remove genomic DNA contamination. cDNA synthesis was performed using Superscript II RNase H-Reverse transcriptase (Invitrogen) according to the manufacturers instructions using either 500 ng oligo $(dT)_{12\text{-}18}$ or 200 ng random hexamers as primers. Quantification of transcripts was performed on Opticon II real-time PCR machines (MJ Research) using Platinum Taq DNA polymerase (Invitrogen) and SYBR Green (Sigma-Aldrich). The following PCR program was used: 2 min 94 °C, 45 cycles of 30 sec 94 °C, 30 sec 55 °C, 30 sec 60 °C and 15 sec 75 °C (during which measurements were taken), followed by 10 minutes chain extension. Expression levels were normalized to levels measured for control genes (hypoxanthine guanine phosphoribosyl transferase (*HPRT*), ribonuclease/angiogenin inhibitor 1 (*Rnh1*) and *18S*).

Chromatin Immunoprecipitation (ChIP)

ChIP analysis was performed as described in the Upstate protocol (http://www.upstate.com), except that cells were cross-linked using 2% formaldehyde for 5 minutes at room temperature. Real-time quantification

of precipitated DNA sequences (with typical fragment sizes between 300 to 600 bps) was performed on Opticon II PCR machines (MJ Research) using Platinum Taq DNA polymerase (Invitrogen) and SYBR Green (Sigma-Aldrich) using the following cycling conditions: 2 min 94 °C, 45 cycles of 30 sec 94 °C, 1 min 55 °C, 15 sec 72 °C and 15 sec 75 °C (during which measurements were taken). Enrichment was calculated relative to a control gene and all values were normalized to input measurements. The following antibodies were used: anti-RNAP II (N-20; sc-899), anti-GATA-1 (N-6; sc-265), anti-p45 NF-E2 C-term. (C-19; sc-291), anti-p45 NF-E2 N-term. (H-230; sc-22827), anti-MafK (C-16; sc-477) and anti-Nrf-2 (H-300; sc-13032), all obtained from Santa Cruz Biotechnology, anti-acetyl histone H3 (#06-599) obtained from Upstate, anti-di-methyl Histone H3 K9/K27 (ab7312) and anti-histone H3 (ab1791) obtained from Abcam. Anti-EKLF (5-V) was kindly provided by J. Philipsen, anti-Bach1 antiserum (A1-5) was kindly provided by K. Igarashi (Sun et al., 2004). Data were normalized to input and enrichment was measured over control gene (*Necdin* or *Amylase*, giving identical results).

3C analysis

3C analysis using HindIII as restriction enzyme was essentially performed as described (Splinter et al., 2004). Real-time quantification of ligation products was performed on Opticon II PCR machines (MJ Research) using Platinum Taq (Invitrogen) and double-dye oligonucleotides (5'FAM and 3'TAMRA) as probes. The following PCR program was used: 2 min 94 °C, 45 cycles of 15 sec 94 °C and 90 sec 60 °C. Data were normalized to interaction frequencies measured in the XPB locus to account for differences in quality and quantity of DNA templates.

Acknowledgements

We thank dr. R. Shivdasani for providing p45 NF-E2 knockout mice, dr. K. Igarashi for providing the anti-Bach1 antiserum, dr. J. Philipsen for providing the EKLF antibody and critical reading of the manuscript and M. Vendeloo for assistance with experiments. This work was financially supported by grants from the Dutch Scientific Organisation (NWO) to F.G. (912-03-009) and to W.L. (912-04-082). The authors declare no competing financial interests.

4

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

Based on: Genes & Development 20, 2349-2354 (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 neighboring 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 *in vitro* (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 acts 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 (Figure 4.1A), which contains 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 (Bulger et al., 2003; Farrell et al., 2002). 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) were found to 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 l/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 analysed β-globin DNA contacts in cells lacking the CTCF protein. Analysis was focused on embryonic day 12.5 (E12.5) erythroid progenitor cells because they can be expanded *ex vivo* (Dolznig et al., 2001) and lack stable LCR-gene contacts, and 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 β-globin locus in these cells (Figure 4.1F), while the protein was absent from HS5 and 3'HS1 in brain cells not showing these loops (data not shown). Since CTCF-null mice die early during embryogenesis, a conditional knockout mouse model was generated by inserting two lox sites upstream of and downstream from 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 (Figure 4.1B). 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 with wild-type (Figure 4.1C/D/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 (Figure 4.1F). To investigate β-globin locus conformation in these cells by 3C technology, we used a novel Taqman probe-based quantitative PCR (Q-PCR) strategy to accurately quantify 3C ligation efficiencies (Figure 4.1G). 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 (Figure 4.1H). 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 (Figure 4.1H and 4.1I). 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

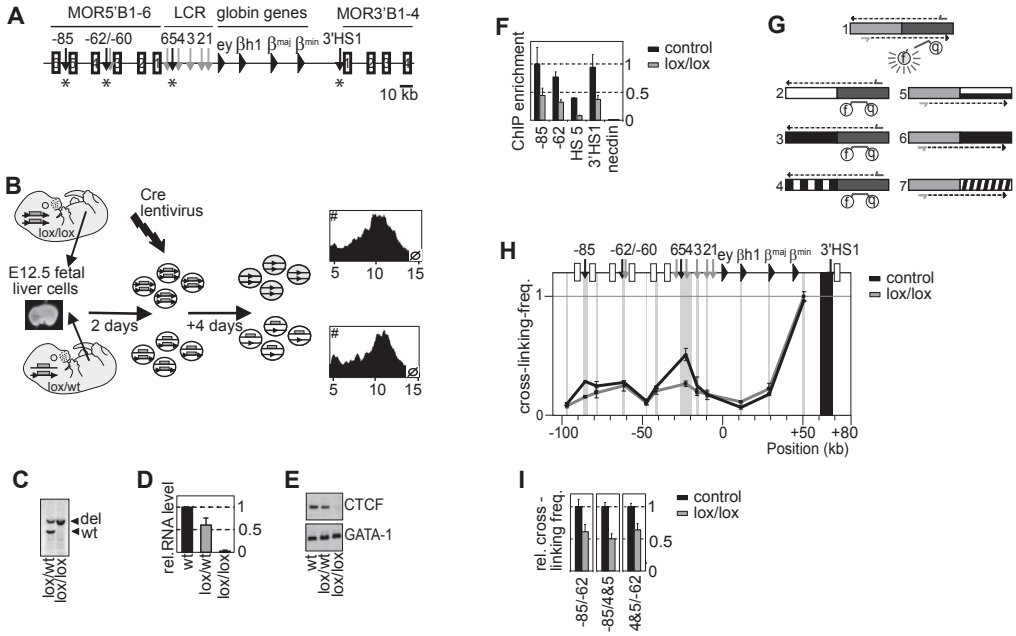

Figure 4.1

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 (black arrows and asterisk) are indicated.

B. Strategy to delete CTCF. Plots show a similar cellular size distribution for homozygous and heterozygous conditional knock-out cells.

C. Southern blot analysis showing complete deletion of CTCF conditional knock-out alleles.

D. CTCF mRNA levels (as determined by Q-RT-PCR) in untreated wild-type (level set to 1) and Cre-recombined lox/wt and lox/lox cells. Standard deviation is indicated.

E. Western blot analysis of CTCF protein and GATA-1 protein (control, stripped and re-hybridized blot).

F. CTCF ChIP analysis. Lox/lox (grey) vs. control cells (black).

G. Q-PCR analysis of ligation frequencies obtained by 3C. The approach entails a primer-probe combination that is specific for a particular restriction fragment (dark grey), with the probe hybridizing to the opposite strand as compared to the PCR primer. A second PCR primer hybridizes to the fragment (light grey) 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).

H. 3C analysis demonstrating reduced interaction frequencies between 3'HS1, HS4/5 and HS-85 in lox/lox cells (grey), as compared with control cells (black).

I. 3C analysis, demonstrating reduced interaction frequencies between the other CTCF-binding sites in the β-globin locus in lox/lox cells (grey) vs. control cells (black).

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 (data not shown). Moreover, we did not find activation of any of the mouse olfactory genes immediately surrounding the β-globin locus (MOR5B1-3 and MOR3B1-4) (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.

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. Bandshift assays confirmed that these alterations completely abolished CTCF binding *in vitro* (data not shown). Targeting was performed in embryonic stem (ES) cells that were established from a cross between the two inbred strains 129 and C57BL/6 (B6) and was directed to the B6 allele. Two additional, non-conserved nucleotides were changed 70 base pairs (bp) downstream from the core CTCF-binding site to allow allele-specific analysis of CTCF binding to 3'HS1 by ChIP. Moreover, an extra HindIII restriction site was introduced ~850 bp downstream from 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.

Definitive erythroid progenitors were generated from the ES cells *in vitro* (Carotta et al., 2004) 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) (Figure 4.2A). We validated the cells as a model system for erythroid differentiation (Carotta et al., 2004) and 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 (Figure 4.2B). Thus, the change of 4 nucleotides in the core CTCF-binding site also prevented binding of CTCF to 3'HS1 *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 from 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

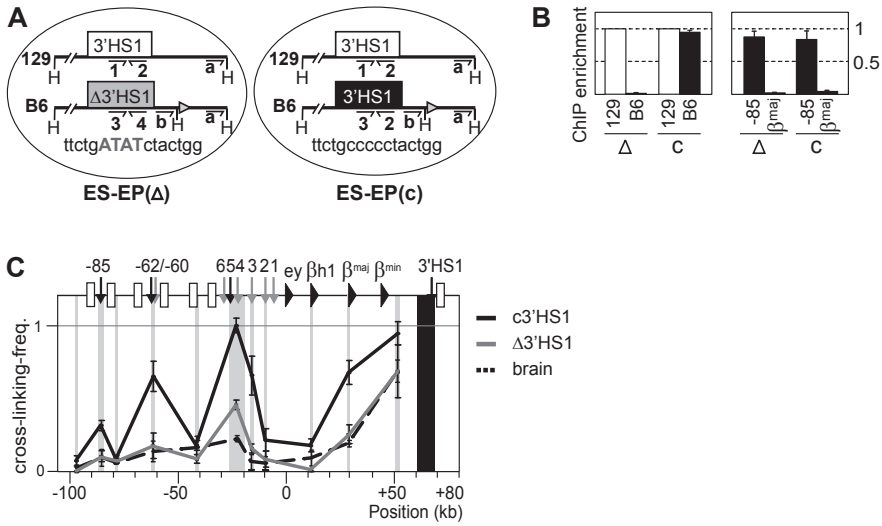

Figure 4.2

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(Δ3'HS1) harbors four targeted nucleotide changes in the core CTCF-binding site of 3'HS1 on the B6 allele (grey). ES-EP(c) contains wild-type 3'HS1 on the B6 allele (black). The nontargeted, intact 129 allele is in white. For ChIP, each 3'HS1 CTCF-binding site can be analysed with a unique primer pair (#1-4). An extra HindIII site targeted downstream from 3'HS1 allows exclusive 3C analysis of B6 allele (with 3C-primer 'b').

B. ChIP on undifferentiated ES-EP cell lines, with antibody against CTCF. Left: 3'HS1 alleles in the two ES-EP lines (Δ and c). Right: positive (HS-85) and negative (β_{major}) controls.

C. 3C analysis with primer 'b' (see above). Note that interaction frequencies with mutated 3'HS1 (grey) are reduced compared with wild-type 3'HS1 (black). Black dashed line: 3'HS1 interactions in fetal brain, analysed with primer 'a' and plotted for comparison.

HS-85 (Figure 4.2C). 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 nonexpressing fetal brain cells (Figure 4. 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.

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 (MORs) 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 MORs, or 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, it was previously 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

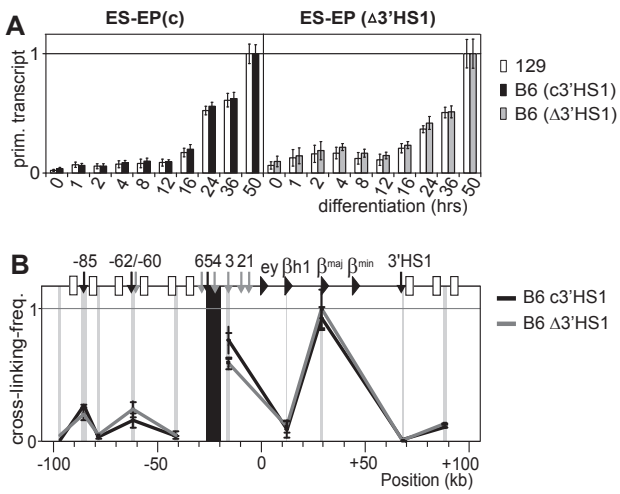

Figure 4.3

Targeted nucleotide changes in 3'HS1 do not affect β-globin gene expression.

A. Ongoing β_{major} transcription as measured by Q-RT-PCR, using 129- and B6-specific primers against intron 2 of the β_{major} gene. 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. Note that primer 'a' was used near 3'HS1, which on the B6 allele analyses a small (0.5 kb) fragment downstream from (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 shows that analysis is restricted to the B6 allele.

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 the LCR and genes, which may facilitate their productive interaction later during differentiation. If true, the 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. 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 allele and B6 allele, both in ES-EP(c) and ES-EP(Δ3'HS1) cells (Figure 4.3A). 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 εy and β_{h1}, which were off in both cell lines before and after differentiation (data not shown).

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 Taqman probe for a HindIII fragment encompassing HS4 and HS5 that signals exclusively from the B6 allele. 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 (Figure 4.3B). This demonstrated 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 (~15 kb) region of highly repetitive DNA is present ~3 kb 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 dimethyl H3K9 and dimethyl H3K27, we found that these repressive marks were abundantly present on both sides of 3'HS1, but not inside 3'HS1 (Figure 4.4A). 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 (Figure 4.4C). 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 nontargeted 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

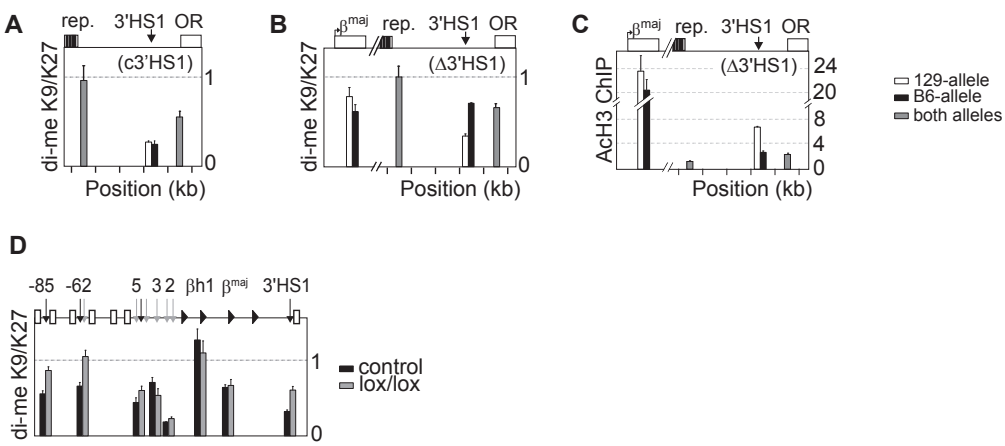

Figure 4.4

Histone modifications in the absence of CTCF binding.

A. ChIP enrichment for di-meH3K9/K27 in undifferentiated control ES-EP(c) cells on the 129 allele (white), B6 allele (black), or on both alleles (grey). Note that values in **A-C** were normalized to input and therefore grey bars also represent enrichment per allele.

B,**C**. ChIP enrichment for di-meH3K9/K27 (**B**) and acetylated histone H3 (**C**) in undifferentiated ES-EP(Δ3'HS1) cells. Black bars: B6 allele, white bars: 129 allele. Note increased methylation (**B**) and decreased acetylation (**C**) only at mutated 3'HS1 on B6-allele

D. ChIP enrichment for di-meH3K9/K27 at the β-globin locus in control (black) and conditional CTCF knockout (lox/lox) (grey) E12.5 erythroid progenitors cells. Note that reduced levels of CTCF cause an increase in di-meH3K9/K27 specifically at the CTCF-binding sites HS-85, HS-62, HS5 and 3'HS1.

accompanied by an increase of di-meH3K9/K27 and concomitant decrease of AcH3 at 3'HS1 (Figure 4.4B and 4.4C). We found no indication for spreading of the methyl mark into the locus, either locally (compare levels of enrichment in Figure 4.4A and Figure 4.4B) or at the β_{major} gene, which locates more inside the locus (Figure 4.4B, 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 (Figure 4.3A).

We considered the possibility that the spreading of di-meH3K9/K27 into the locus requires the 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 (Figure 4.1C/D/E). 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 (Figure 4.4D). 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 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 (de Laat and Grosveld, 2003; Dillon and Sabbattini, 2000).

CTCF organizes 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, have shown previously 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 fibers 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 into 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.

Experimental procedures

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 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^{0}C in a SW28 rotor. Virus stocks were stored at –80^{0}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. 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^{0}C). 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 quantitive 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^{0}C, 45 cycles of 30 sec 94^{0}C, 1 min 62^{0}C, 15 sec 72^{0}C and 15 sec 75^{0}C (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 available upon request.

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^{0}C for 2 min and 44 cycles of 15 s at 94^{0}C and 90 s at 60^{0}C. Probe and primer sequences are available upon request.

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^{0}C for 2 min, 44 cycles of 30 s at 94^{0}C, 60 s at 55^{0}C, 15 s at 72^{0}C and 15 s at 75^{0}C (during which measurements are taken). Enrichment was calculated relative to *Necdin* and values were normalized to input measurements. Primer sequences are available upon request. Antibodies used: Anti-acetyl-Histone H3 (#06-599, Upstate); anti-di-methyl Histone H3 K9/K27 (ab7312, Abcam).

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). Site-directed mutagenesis was performed on an internal 683 bp NdeI-NdeI fragment (coordinates: 67033-67716), using QuikChange® Multi Site-Directed Mutagenesis Kit (Stratagene). Oligonucleotide used to change CTCF-binding site: CGGAAATCAGCGGAACACTTCTGATA TCTACTGGTATGCAACAGG. 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. 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 linearized 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 analysed 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.

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 eryhrocytes was performed as described (Carotta et al., 2004) (Epo was a kind gift from Ortho-Biotech), 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.

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 Organisation 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.

The nuclear organisation of peri-centromeric and peri-telomeric regions

Work in progress

Summary

Unravelling the fundamental principles and mechanisms that underlie genomic organisation is one of the major challenges in modern cell biology. Here we use recently developed 4C technology to study the nuclear organisation of both peri-centromeric and peri-telomeric chromosomal regions in murine B-lymphocytes. Analysis was focused on unique sequences near the ends of three chromosomes (1, 11 and 18), with each of these regions having distinct transcriptional activities. We show that there is a strong correlation between the transcriptional status of peri-centromeric and peri-telomeric regions and their interacting loci in the interphase nucleus, providing additional evidence that active and inactive chromatin domains are separated in the nucleus. Moreover, interchromosomal interactions with peri-centromeric regions were strongly biased towards other peri-centromeric regions. Interestingly, peri-telomeric regions also showed a preference to interact in *trans* with other peri-telomeric regions. We argue that 4C data provide evidence for self-organising principles underlying chromatin folding, where the preferred nuclear environment of a genomic locus is determined not only by its own functional properties but also by the properties of neighbouring DNA segments and, by extrapolation, of the entire chromosome. Our data show that both the transcriptional status of a genomic region and its proximity to defined repetitive sequences on the linear chromosome template strongly influence the positioning of a locus in the nucleus.

Introduction

The eukaryotic nucleus is characterized by the existence of distinct structural and functional units that show high levels of organisation. The folding and packaging of the genetic material into chromosomes is a typical example of this structural organisation. Identifying the mechanisms and rules that determine how chromosomes are spatially organised and how this three-dimensional organisation influences transcriptional regulation is currently one of the major challenges in cell biology.

Genomic organisation in the cells nucleus is non-random and two general types of interphase organisation have been described. One is the polarized Rabl configuration seen in e.g. flies and plants in which centromeres and telomeres reside at opposite poles of the nucleus as a consequence of anaphase chromosome segregation (Dong and Jiang, 1998; Hochstrasser et al., 1986). In mammalian cells, this polarized configuration degrades after telophase and a second type of domain organisation arises in which individual chromosomes occupy relatively defined locations in the nuclear space known as chromosome territories (Cremer and Cremer, 2001). There are conflicting ideas about which properties determine the localisation of chromosome territories in mammalian cells. Several types of radial distribution have been reported, either based on chromosome gene density (Boyle et al., 2001; Croft et al., 1999) or chromosome size (Sun et al., 2000), but

results differ between cell types and developmental stage (Bolzer et al., 2005; Parada et al., 2004). The recent finding that human lymphocytes show extensive intermingling of chromosome territories adds another level of complexity (Branco and Pombo, 2006). Thus, there are no simple parameters dictating chromosomal territorial arrangement and consequently the precise physiological relevance of chromosome and gene positioning for transcriptional regulation is still largely unclear. It is known however, that transcription can occur throughout the entire nucleus as nascent RNAs can be seen deep within chromosome territories (Abranches et al., 1998; Verschure et al., 1999) and sites of transcriptional activity are distributed throughout the entire nucleus (Cmarko et al., 1999; Jackson et al., 1993; Kimura et al., 2002; Wansink et al., 1993). In agreement, condensed chromatin domains are physically accessible for large macromolecules indicating that transcriptional silencing is not due to physical inaccessibility of gene loci for transcription factors (Verschure et al., 2003). However, some gene clusters and individual genes migrate upon change in their transcriptional status during differentiation and development. In general, this relocalisation is measured using Fluorescent In Situ Hybridization (FISH) technology against large nuclear landmarks such as chromosome territories, the nuclear periphery and centromeres. For example, the nuclear periphery is assumed to be an inactive compartment and repositioning of activated genes towards the nuclear interior has been documented for many loci in different cell types (Hewitt et al., 2004; Kosak et al., 2002; Williams et al., 2006; Zink et al., 2004), although some genes were detected at the nuclear periphery irrespective of their transcriptional status (Hewitt et al., 2004).

Interesting observations were made in several species and cell types, where a correlation was found between gene silencing and intranuclear positioning towards centromeres. In *Drosophila*, the introduction of a heterochromatic block into the *brown* locus strongly reduced expression of the gene and results in more frequent association of the locus with centromeric heterochromatin (Csink and Henikoff, 1996; Dernburg et al., 1996). Evidence that this silencing is a more widespread phenomenon and also applies for normal, unmodified genes came from studies in B- and T-lymphocytes where the transcriptionally inactive status of lymphoid genes correlates with their intranuclear proximity to centromeric heterochromatin (Brown et al., 1999; Brown et al., 1997). Subsequent reports established the important role of centromeric heterochromatin in gene silencing. For example, the silencing of *IL4* and *INFγ* during thymocyte development correlates with their proximity to centromeric heterochromatin (Grogan et al., 2001); the process of allelic exclusion of immunoglobulin heavy and light-chain genes appears to be regulated by location of excluded alleles in pericentromeric heterochromatic compartments (Skok et al., 2001); lineage choice in double positive thymocytes is predicted by the spatial repositioning of *CD4* and *CD8* coreceptor alleles to centromeric domains (Merkenschlager et al., 2004) and in erythroid cells a functional enhancer was shown to antagonize the silencing of a linked transgene by preventing its localization to centromeric heterochromatin (Francastel et al., 1999). These observations indicate that nuclear positioning of genes relative to centromeric

heterochromatin might regulate the proper execution of gene expression programmes. However, centromeric repositioning is not a prerequisite for the repression of every gene. This is illustrated by the subnuclear localization of the silenced human α- and β-globin loci in cycling lymphocytes, where α-globin was not associated with centromeres, whereas the β-globin locus together with other inactive genes was located near peri-centric heterochromatin (Brown et al., 2001).

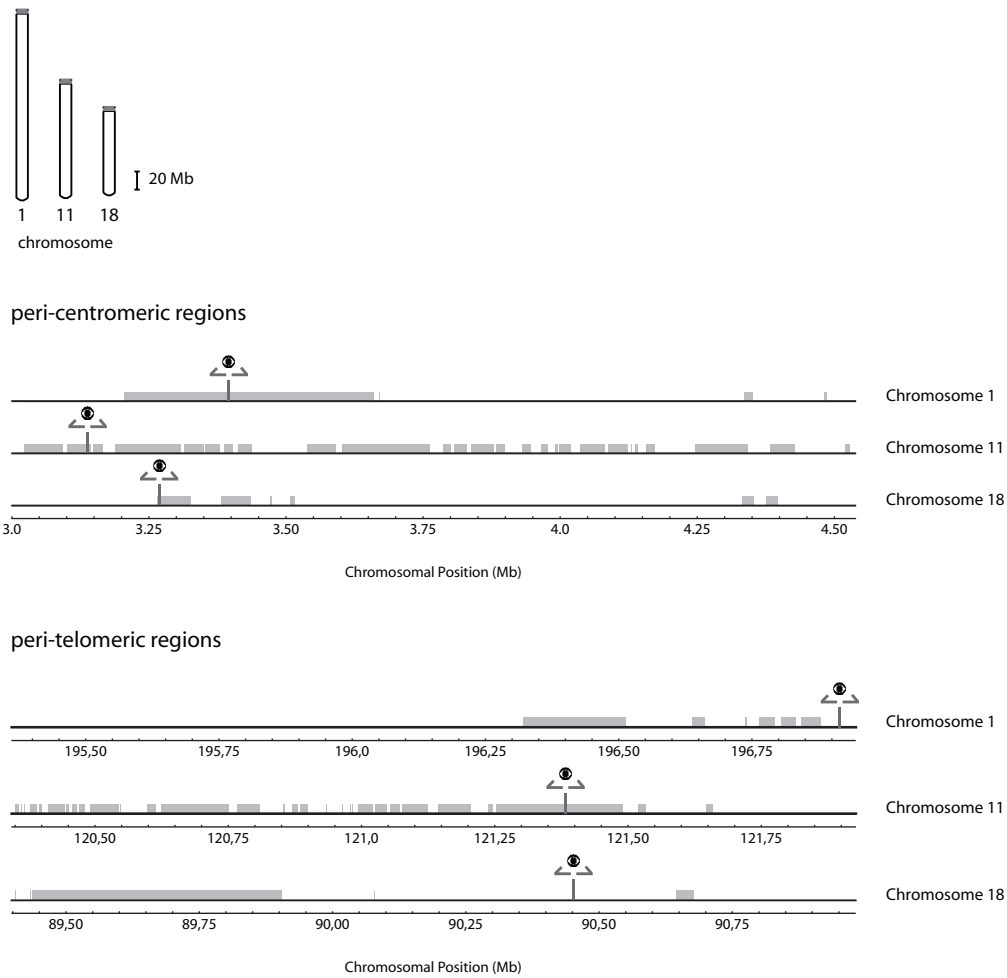

Figure 5.1

Overview of selected centromeric and telomeric regions for 4C analysis.

Schematic overview of the peri-centromeric and peri-telomeric regions of chromosome 1, 11 and 18. Total chromosomal size is 197 Mb for chromosome 1, 122 Mb for chromosome 11 and 91 Mb for chromosome 18. Depicted are the first sequenced 1.5 Mb from the centromeric region and the last sequenced 1.5 Mb from the telomeric region of chromosome 1, 11 and 18. Annotated genes are indicated by gray rectangles. The position of every first possible unique primer set (bait sequence) used for 4C analysis is indicated by 👁. Chromosomal positions were based on NCBI build m36.

Centromeric heterochromatin consists of repetitive DNA satellite sequences which have the tendency to cluster in interphase nuclei forming so-called chromocenters (Jin et al., 2000; Martou and De Boni, 2000; Weierich et al., 2003). This clustering occurs in every cell type but becomes more prominent during later stages of cellular differentiation (Alcobia et al., 2003; Beil et al., 2002). Another type of repetitive satellite sequences known as telomeres are located at the ends of the chromosomal arms. Whether positioning near heterochromatic telomeric domains also influences transcription in mammalian cells is still unclear (Baur et al., 2001; Quina and Parreira, 2005; Tham and Zakian, 2002). It is also not known whether mammalian telomeres, like their yeast counterparts (Funabiki et al., 1993; Maillet et al., 1996), cluster in the interphase nucleus, although some evidence exist that they do occasionally get together (Scherthan et al., 1996; Weierich et al., 2003).

Studies concerning gene positioning and nuclear organisation have mainly been carried out using FISH technology, which enables simultaneous visualization of selected parts of the genome. For a better understanding of both centromeric and telomeric nuclear organisation we used 4C technology, which enables to systematically screen the entire genome in an unbiased manner for DNA loci that contact a given locus in the nuclear space (Simonis et al., 2006). We focused our analysis on three peri-centromeric and three peri-telomeric regions in murine B-lymphocytes. We show that the nuclear environment of the selected regions is dependent on their transcriptional status. Transcriptionally active regions predominantly contacted other active regions both in *cis* and in *trans*, while inactive regions preferentially interacted with regions lacking active genes. Additionally, interchromosomal interactions of peri-centromeric regions were strongly biased towards other peri-centromeric regions. Interestingly, peri-telomeric regions also showed a preference to interact in *trans* with other peri-telomeric regions. Thus, transcriptional activity and proximity to defined repetitive sequences on the linear chromosome template both strongly influence the preferred nuclear environment of a locus.

Results

Selection of centromeric and telomeric regions for 4C analysis

To study the nuclear environment of centromeres and telomeres we applied 4C technology to the centromeric and telomeric chromosomal regions of chromosome 1, 11 and 18. Not only do these three chromosomes differ in size (ranging from 91 Mb for chromosome 18 to over 197 Mb for chromosome 1), they also show considerable differences in gene density at their chromosomal ends (Figure 5.1). For all selected chromosomes, we designed the first possible unique 4C primer set from the centromeric and telomeric sequence ends. Using this criterion, all 4C primers used were designed within 0.5 Mb from the annotated chromosomal ends, with the closest primerset being 142 kb and the most distal primerset being 408 kb away from the annotated sequence ends (Figure 5.1). Gene positioning and nuclear organisation have been

studied extensively in both B- and T-lymphocytes (Brown et al., 1999; Brown et al., 1997; Grogan et al., 2001; Merkenschlager et al., 2004; Skok et al., 2001). Therefore, we used a homogeneous population of primary B-lymphocytes for our 4C analysis to study the nuclear environment of peri-centromeric and peri-telomeric regions.

4C technology involves the selective amplification of cross-linked DNA fragments ligated to a restriction fragment of choice. All DNA fragments captured by this so-called 'bait'-sequence in a pool of cells are amplified via inverse PCR using two 4C primers that amplify the circularized ligation products. This material is subsequently labelled and hybridised to tailored microarrays containing 400.000 individual probes that each analyse a different DNA interaction across the entire mouse genome. We previously showed that clusters of positive hybridisation signals on the chromosome template reveal interacting regions. In order to identify these interacting regions, a running mean algorithm with a window size of approximately 60 kb was applied to the measured dataset. We then used the running mean distribution of a randomly shuffled dataset to set a threshold value, allowing a false discovery rate of 5%. Importantly, only genomic clusters that meet this threshold value in two independent replicate experiments are defined as interacting with the 4C bait sequence (Simonis et al., 2006).

Chromosomal ends interact with distinct regions of different transcriptional activity

First we analysed the distribution of intra-chromosomal long-range interactions of the centromeric and telomeric region of chromosome 1. Replicate experiments from both the centromere and telomere showed highly reproducible interaction patterns in *cis* (Figure 5.2A). Interactions were detected along the entire length of chromosome 1 and, as seen before, the strongest signals can be found within a 2-4 Mb region around the chromosomal position of the centromeric and telomeric bait sequences (Simonis et al., 2006). We identified 32 regions interacting with the peri-centromeric region and 45 regions that interacted with the peri-telomeric region of chromosome 1 (Figure 5.2A).

At first glance, the two ends of the chromosome appeared to contact a mutually exclusive set of chromosomal regions. This becomes even more evident if we zoom in on selected 20 Mb chromosomal regions (Figure 5.2B). A comparison with the distribution of genes along chromosome 1 suggests that the peri-telomeric region preferentially interacts with genic regions while the peri-centromeric region predominantly contacts non-genic regions (Figure 5.2B). To analyse whether these interactions were dictated by the transcriptional status of genomic regions, gene expression was analysed in CD19+ B-lymphocytes using Affymetrix expression arrays. Interacting regions identified by 4C were subsequently characterized with respect to gene density and activity. This revealed that the peri-centromeric region primarily contacts gene-poor, transcriptionally inactive regions whereas the peri-telomeric region interacts with gene-rich, transcriptionally active regions (Figure 5.2B). Thus, the great majority (87%) of

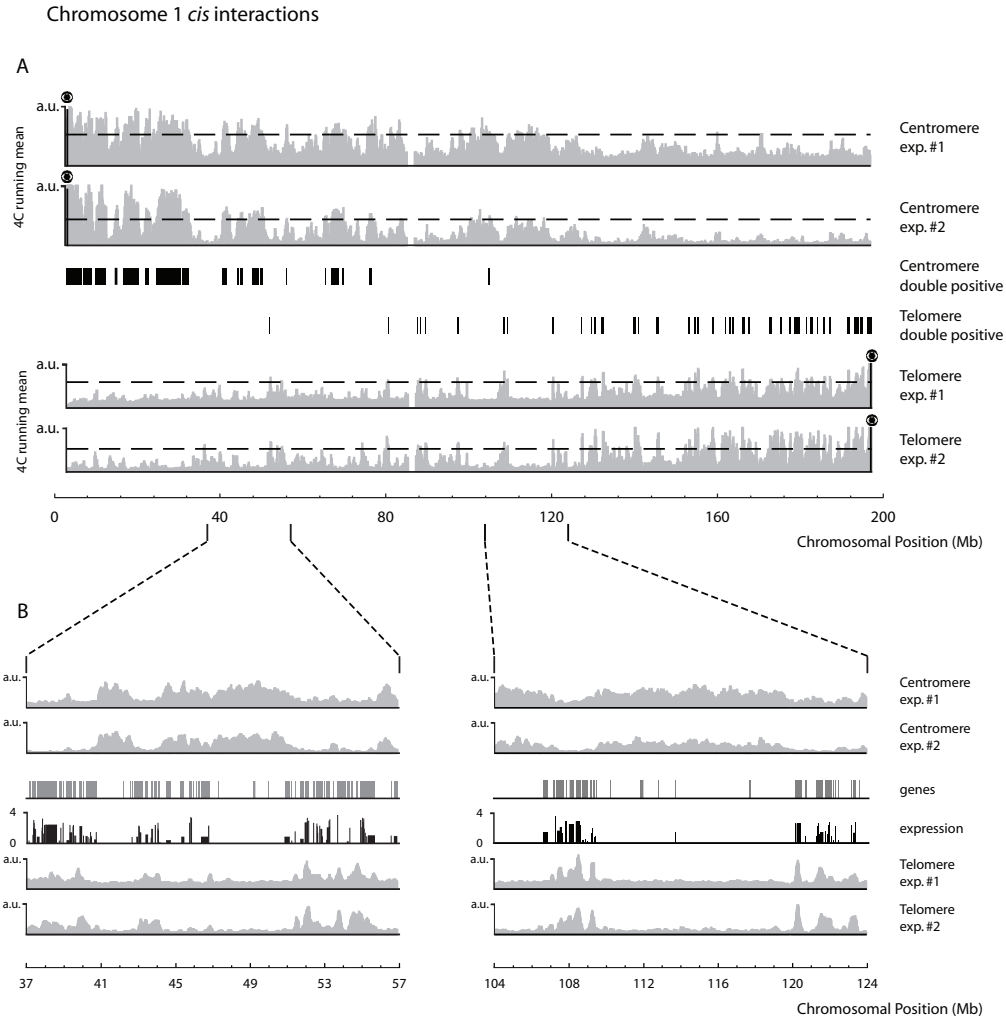

Figure 5.2

Centromeric and telomeric long-range intrachromosomal interactions for chromosome 1.

A. Overview of interactions in *cis* from both the peri-centromeric and peri-telomeric regions of chromosome 1. Shown are processed 4C data for two independent experimental samples (exp. #1 and exp. #2) analysing intrachromosomal interacting regions for both the centromeric and telomeric region of chromosome 1. Chromosomal position of used 4C primer sets is indicated by 👁. Chromosomal regions that meet the threshold value (indicated by a dashed line) in both duplicates were termed double positive and are indicated by black boxes.

B. Detailed overview of two randomly selected 20 Mb regions of chromosome 1 shown in **A**. Compared are duplicate processed 4C data sets for both the centromeric and telomeric regions of chromosome 1, the location of genes (middle, genes) and micro-array expression analysis in CD19-positive B-lymphocytes (log scale, middle, expression) plotted along a 20 Mb chromosomal axis. Data shows that the peri-centromeric region of chromosome 1 preferentially interacts with transcriptionally inactive regions while the peri-telomeric region of chromosome 1 associates with areas containing actively transcribed genes. Chromosomal positions were based on NCBI build m36.

centromere 1 interacting loci did not show any detectable gene activity, whereas almost all interacting loci of telomere 1 (96%) contained one or more actively transcribed genes (Figure 5.3A).

Subsequent analysis of other chromosomes shows that the difference in transcriptional activity between centromere- and telomere-associated regions is not a general phenomenon. Telomere 18 contacts primarily regions without any gene activity (89%), whereas centromere 18 is mainly associated with transcriptionally active loci (80%). The sub-centromeric and sub-telomeric regions of chromosome 11 and telomere 11 mainly interact with transcriptionally active loci (90% and 87%, respectively) (Figure 5.3A). In some of these cases though, the number of identified *cis*-interacting regions is relatively low (for example, with our standard stringent conditions we identify only 10 and 8 interacting regions for centromere 18 and telomere 11, respectively). However, even when we lower the threshold by removing the saturated probes within the first 5 Mb around the bait sequence, which results in an increase of identified interacting regions to 15 and 19, respectively, the nature of the interacting regions remains the same (data not shown). Thus, as both peri-centromeric and peri-telomeric regions can interact with transcriptionally active and inactive regions, the data indicate that centromeric or telomeric localization per se is not predictive for the transcriptional status of associated regions *in cis*.

We recently demonstrated that the active β-globin locus in fetal liver interacts with transcribed loci whereas the inactive β-globin locus in fetal brain contacts other silent loci. In contrast, the housekeeping gene *Rad23a*, which is constitutively expressed, is contacting other actively transcribed loci in both tissues (Simonis et al., 2006). We reasoned that the transcriptional status of peri-centromeric and peri-telomeric regions may also influence the nature of their nuclear environments. Indeed, the transcriptionally active peri-centromeric and -telomeric bait sequences (centromere 11 and 18, telomere 1 and 11) interacted primarily with transcriptionally active loci (Figure 5.3B). On the other hand, the peri-centromeric region of chromosome 1 and the peri-telomeric region of chromosome 18 are transcriptionally silent and preferentially interacted with regions without gene activity (Figure 5.3A and 5.3B).

In addition to interactions with regions on the same chromosome we identified associated regions in *trans* for all analysed sub-centromeric and sub-telomeric regions (Figure 5.4A). Similar to what we had observed for intrachromosomal contacts, interchromosomal contacts were made with regions having the same transcriptional status. Only for chromosome 18 this corrrelation was somewhat less obvious (see Discussion). Thus, our data provide additional evidence that active and inactive chromatin domains separate in the interphase nucleus (Simonis et al., 2006).

Peri-centromeric and peri-telomeric regions preferentially interact with correspondingly positioned regions on other chromosomes

We subsequently looked at the distribution of interchromosomal associations with the selected centromeric and telomeric regions. Centromeric regions interact with 34 other regions in

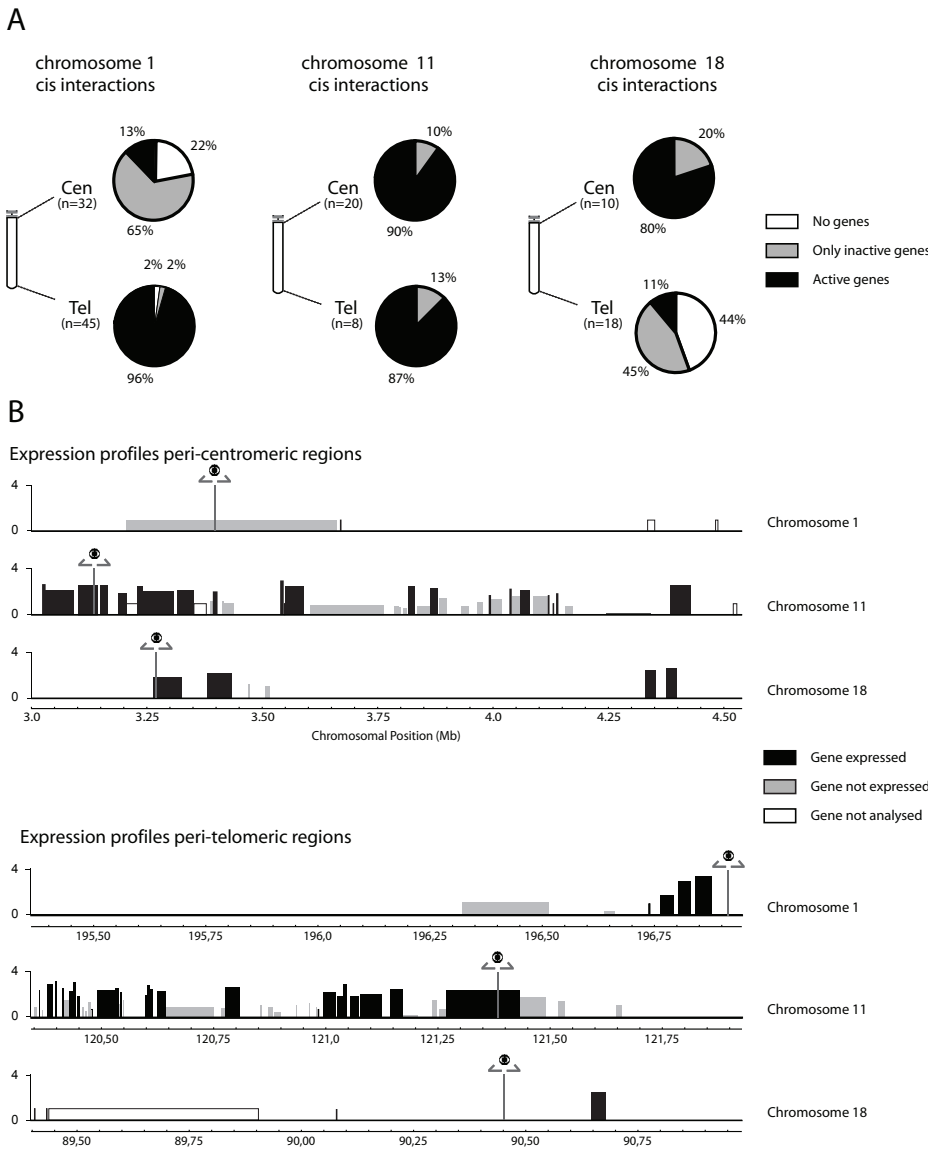

Figure 5.3

Characterization of *cis*-interacting regions and expression profiles of selected centromeric and telomeric regions.

A. Identification and transcriptional characterization of *cis*-interacting regions for all analysed peri-centromeric and peri-telomeric regions. Indicated are the total number of intramolecular interactions detected for each individual chromosomal end (n=x) together with the characterization of the identified regions in terms of gene content and gene activity.

B. Overview of the expression profile of the selected peri-centromeric and peri-telomeric regions. Data shown are based on micro-array expression analysis in CD19-positive B-lymphocytes (log scale), indicating expressed genes (value > log 1.698), non-expressed genes (value < log 1.698) and genes not present on the expression array. Chromosomal positions were based on NCBI build m36.

total, which are divided over 16 different chromosomes (Figure 5.4B), providing additional evidence for a random distribution of chromosome territories in cell populations. However, the peri-centromeric regions showed a clear preference to interact with regions located at similar positions on other chromosomes. Thus, 7 out of the 34 trans-interactions identified are positioned within 1.5 Mb of another centromere. Interestingly, also the sub-telomeric regions showed a strong preference to interact with other sub-telomeric regions in trans: 10 out of 40 *trans*-interactions with the three peri-telomeric regions are positioned within 1.5 Mb of another telomeric end.

These results demonstrate that both peri-centromeric and peri-telomeric regions have the tendency to associate with regions correspondingly positioned on other chromosomes. Clearly, this must be the consequence of clustering of the repetitive DNA sequences capping the chromosomes. Centromeric clustering is well established, but our data indicate that telomeric clustering also occurs and may be more common in mammalian cells than previously anticipated. Based on this data we conclude that both the transcriptional status of a genomic region and its proximity to repetitive sequences on the linear chromosome template influence the positioning of a locus in the nucleus.

Discussion and future experiments

We used 4C technology to study the nuclear organisation of three peri-centromeric and peri-telomeric regions in murine B-lymphocytes. Using 4C primer sets designed as close to both the centromeric and telomeric chromosomal ends as possible we were able to systematically screen the entire genome in an unbiased manner for DNA loci that contacted the selected peri-centromeric and peri-telomeric regions in nuclear space. We show that the nuclear environment of peri-centromeric and peri-telomeric regions is dependent on the transcriptional status of the selected regions as transcriptionally active regions predominantly contacted other active regions both in *cis* and in *trans*, while inactive regions preferentially interacted with regions lacking active genes. Together with data previously obtained in other tissues for the β-globin locus and the *Rad23a* gene (Simonis et al., 2006), these results indicate that there is a strong correlation between the transcriptional status of a region and the transcriptional status of its interacting loci in the interphase nucleus. As a consequence, active and inactive chromatin domains separate in the nucleus. Importantly, these principles not only determine the interaction patterns of individual loci but eventually underlie the folding and long-range DNA interactions of whole chromosomes.

Data obtained so far by 4C analysis are in strong agreement with the concept of self-organisation as a model of how genomes are organised (Misteli, 2001; Misteli, 2007). This model states 'that the sum of all functional properties of a chromosome (e.g. the frequency and linear distribution of its active and inactive regions) determines its positioning' (Misteli,

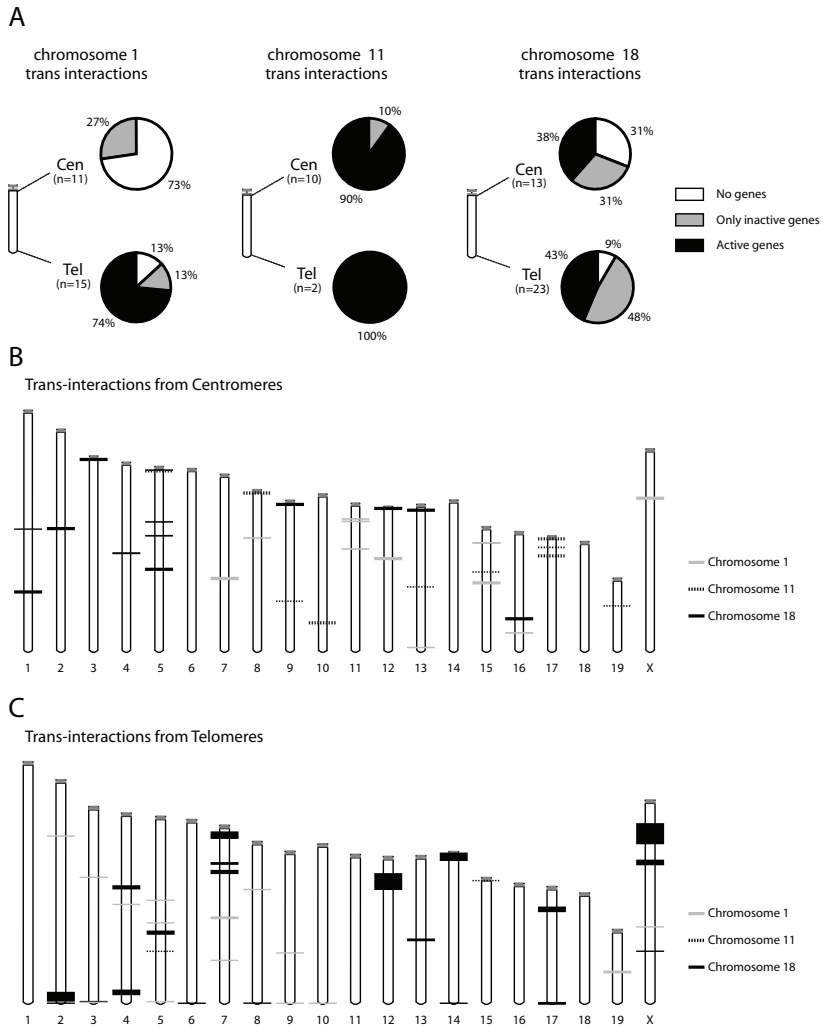

Figure 5.4

Characterisation and distribution of *trans*-interacting regions of selected centromeric and telomeric regions.
A. Identification and transcriptional characterization of *trans*-interacting regions for all analysed peri-centromeric and peri-telomeric regions. Indicated are the total number of intermolecular interactions detected for each individual chromosomal end (n=x) together with the characterization of the identified regions in terms of gene content and gene activity.

B. Distribution of identified interchromosomal interactions for the centromeric chromosomal ends of chromosome 1, 11 and 18. Interacting regions are drawn to scale and are on average 200-300 kb. Total number of *trans*-interactions is 34 (Cen 1=11, Cen 11=10, Cen 18=13) and 7 regions are positioned within 1.5 Mb of the corresponding centromere (Cen 1=0, Cen 11=3, Cen 18=4).

C. Distribution of identified interchromosomal interactions for the telomeric chromosomal ends of chromosome 1, 11 and 18. Interacting regions are drawn to scale and are on average 200-300 kb, except telomere 18 that interacts with regions on average 1.6 Mb in size. Total number of *trans*-interactions is 40 (Tel 1=15, Tel 11=2, Tel 18=23) and 10 regions are positioned within 1.5 Mb of the corresponding centromere (Tel 1=4, Tel 11=0, Tel=6).

Note that in **B** and **C** adjacent, non-overlapping interactions may appear as one region due to the scale of the figure.

2007). 4C data show that transcriptionally active regions (e.g. centromere and telomere 11) associate with other active regions, whereas gene-poor and transcriptionally inactive regions (e.g. centromere 1) preferentially interact with transcriptionally silent loci. This demonstrates that each chromosomal region, active or inactive, has its own preferred nuclear environment. This automatically implies that the nuclear positioning of a locus also depends on the properties of neighbouring segments and, by extrapolation, of the entire chromosome, as predicted by models of self-organisation. Gene-dense regions that are highly transcribed, such as the peri-centromeric region of chromosome 11 and the region around *Rad23a,* and gene-poor regions that are completely inactive, such as the β-globin locus in non-expressing tissue and the sub-centromeric region of chromosome 1 most clearly show an interaction pattern strictly biased towards regions of corresponding activity. Not surprisingly, regions that can not so easily be classified as active or inactive do not show such well defined nuclear environments. This is illustrated by the peri-centromeric region of chromosome 18, which is situated at the 5' end of a gene that is expressed just above threshold levels in a relatively gene-poor, transcriptionally silent chromosomal region. In *cis*, we find it to interact with transcriptionally active regions (80%), while in *trans* this region interacts predominantly with loci without gene activity (62%).

Chromatin movement is limited during most stages of the cell cycle, as illustrated by the relatively stable positioning of gene loci and chromosomes during most of G1, S and G2 (Abney et al., 1997; Chubb et al., 2002; Gerlich and Ellenberg, 2003). Early in G1 however, chromatin is more mobile (Thomson et al., 2004). It is therefore tempting to speculate that the nuclear environment of a chromosomal region (and consequently an entire chromosome) will mainly be determined in a short timeframe following mitosis when regions decondense and are not yet physically constrained to interact with other DNA segments. Most likely, random collision will establish contacts between sites that will be stabilised depending on relative affinities. Contacts formed will affect the freedom of movement of other chromosomal segments on the same chromatin fiber. This process eventually results in the nuclear positioning of loci and overall chromosomal conformation based on the order of established contacts and relative affinities between sites. The stochastic nature of this process inevitably leads to cell-to-cell differences in chromatin architecture, which is probably reflected by the many intramolecular and intermolecular interactions that we identify by 4C technology for each peri-centromeric and peri-telomeric region analysed. Consequently, it is impossible to predict the exact nuclear position of a given locus in a given cell.

Some chromatin properties may be dominant over others in determining chromatin folding and repetitive DNA sequences presumably play an important role in genomic organisation. A typical example of interacting repetitive sequences is the nucleolus, which is formed as a result of interchromosomal clustering of ribosomal gene arrays. Similarly, centromeres, which consist of repetitive DNA satellite sequences, also show a tendency to cluster in interphase nuclei of different cell types (Jin et al., 2000; Martou and De Boni, 2000; Weierich et al., 2003). Here we

showed preferential associations between peri-centromeric, but also between peri-telomeric regions, indicating interactions between repetitive chromosomal end sequences. Moreover, a 4C probe analysing the telomeric region of chromosome 18 was designed within a repetitive, gene-poor region and showed preferential association with relatively large-sized regions in *trans* (Figure 5.4B). These regions often spanned large chromosomal gaps devoid of unique probe sequences, indicating the repetitive nature of these regions. We therefore think these results probably reflect another example of clustering of repetitive sequences and we argue that the spatial separation between repetitive and non-repetitive sequences is an important parameter determining interphase nuclear organisation. Future experiments are needed to confirm these interactions and to determine which type of repetitive sequences are present in these associated regions. Spatial repositioning of many genes relative to centromeric heterochromatin correlates with gene silencing and chromocenters may play a role in stable gene repression (Brown et al., 1999; Brown et al., 1997; Grogan et al., 2001; Merkenschlager et al., 2004; Skok et al., 2001). As 4C data consistently show that active and inactive regions tend to separate in the nucleus, it is also possible that centromeric repositioning is not a prerequisite for stable gene silencing, but rather a consequence of this phase separation. Our analysis shows that both active and inactive peri-centromeric regions can form clusters with other peri-centromeric regions. It would be interesting to see how the active peri-centromeric clusters are positioned relative to the actual chromocenters, which can be visualized with DNA satellite repeat probes, as their presence within chromocenters would challenge the idea that these are nuclear areas exclusively repressing gene activity. It is also possible that silenced genes reposition to certain clusters of centromeres that are surrounded by transcriptionally inactive peri-centromeric regions. Indeed, the 4C data suggest that certain centromeres, and telomeres, have a preference to interact with a certain subset of other centromeres or telomeres. FISH studies using chromosome specific peri-centromeric probes should reveal whether this is the case and also whether such an organisation is conserved between different cell types.

Taken together, we demonstrate that there is a strong correlation between the transcriptional status of an analysed region and that of its interacting regions in the interphase nuclei, indicating that active and inactive chromatin domains separate in the interphase nucleus. Moreover, we show that linear chromosome proximity to repetitive DNA regions is an important parameter determining nuclear positioning. The results presented here are in strong agreement with the concept of self-organisation underlying genome organisation. The sum of all chromosomal functional properties results in preferential, yet probabilistic, associations that determine the folding and positioning of chromosomal regions. Using 4C technology we are now capable of measuring the interactions of genomic regions and chromosomes. Careful mapping of these associations should reveal patterns of positioning and eventually lead to the identification of mechanisms and rules that underlie nuclear organisation.

Experimental procedures

Isolation of murine B-lymphocytes

B-lymphocytes were isolated from spleens of 8-10 weeks old female B6 mice, resuspended by repeated pipetting and applied through a 40 µm cell strainer (BD Falcon) using ice-cold MACS-buffer (PBS pH 7.2, 0.5% BSA and 2 mM EDTA). Next, single-cell suspensions from individual spleens were magnetically labelled with CD19 mouse microbeads (order nr. 130-052-201, Miltenyi Biotec) according to the manufacturers instructions followed by isolation of CD19+ B-lymphocytes using an autoMACS (Miltenyi Biotec) magnetic cell separator. Typically, approximately 30 x 10^6 CD19+ B-lymphocytes were isolated from one spleen with a purity of 90-95%, as judged by FACS analysis using a CD19 FITC-conjugated antibody (order nr. 130-091-328, Miltenyi Biotec). Finally, all murine B-lymphocytes were pooled and treated for expression analysis or 4C analysis.

Expression analysis

Total RNA was isolated from approximately 1 x 10^6 CD19+ B-lymphocytes using TRIzol reagent (Invitrogen) according to the manufacturers instructions. Two independent microarrays were performed according to Affymetrix protocol (mouse 430_2 arrays) and for each probe set the measurements of two microarrays were averaged. Additionally, when multiple probe sets represented the same annotated gene according to National Center for Biotechnology (NCBI) build m36 they were averaged. Genes with an average Affymetrix expression value >50 were called expressed. The 4C probe set analysing the pericentromeric region of chromosome 1 is situated within the X Kell blood group precursor related family member 4 (*Xkr4*) gene (ENSMUSG00000051591) which is not present on the Affymetrix expression array. Therefore, 1.5 µg of RNA was treated for 1 hour with DNAseI (Invitrogen) to remove genomic DNA contamination. cDNA synthesis was performed using Superscript IIRNase H-Reverse transcriptase (Invitrogen) according to the manufacturers instructions using 500 ng oligo (dT)12-18 as primers. Quantification of transcripts was performed on Opticon II realtime PCR machines (MJ Research) using Platinum Taq DNA polymerase (Invitrogen) and SYBR Green (Sigma-Aldrich). The following PCR program was used: 2 min 94 °C, 45 cycles of 30 sec 94 °C, 30 sec 55 °C, 30 sec 60 °C and 15 sec 75 °C (during which measurements were taken), followed by 10 minutes chain extension. Expression levels of *Xkr4* were compared to levels measured for the following control genes present on the Affymetrix expression array: hypoxanthine guanine phosphoribosyl transferase (*HPRT*), erythroid kruppel-like factor (*EKLF*) and Dan-domain family member 5 (*Dand5*).

4C sample preparation

Preparation of experimental 4C samples from CD19-positive B-lymphocytes was essentially performed as described before (Simonis et al., 2006). For all samples, *DpnII* was used as a secondary, frequently cutting restriction enzyme. The resulting circular ligation products were linearized by restriction of the bait sequence between the primary and secondary restriction enzyme recognition site, using the following tertiary restriction enzymes: *PstI* (for centromere 1, telomere 1 and telomere 18), *BspMI* (for centromere 1), *SspI* (for centromere 18) and *SphI* (for telomere 11). Digested products were purified using a QIAquick nucleotide removal (250) column (Qiagen). PCR reactions were performed as described before (Simonis et al., 2006) using the following first possible unique 4C primer sets from the centromeric and telomeric sequence ends: centromere 1 #1 5'-TTAGTTACTGGGCTTTGTTTAGC-3' / #2 5'TGTGCCCTATATCAACGTGG-3'; centromere 11 #1 5'-GGCATTGGTTATTGGTCGTAG-3' / #2 5'-AACATGAAGACCGCTGTAATG-3';

centromere 18 #1 5'GTCTTTCAAAGTTTCAAGCTCC-3' / #2 5'- CTGTCATGAATAGTGTCCTATGC-3' ; telomere 1 #1 5' GTACTCAAAATTCATAAAGTAGTTCC-3' / #2 5'-TTAGACCTGAGGCAACGTAC-3'; telomere 11 #1 5'-CTGAGCAGTCTTTAGTGGTAG-3' / #2 5'-CCCAATGGTTGATTTTCATGG-3'; telomere 18 #1 5'-CTGTAGTCATATCTACTCTCGTG-3' / #2 5'-GATGAGCTAGCAGTCTTACATC-3'. Purified 4C template was labelled and hybridized to arrays according to standard chromatin immunoprecipitation (ChIP)-chip protocols (Nimblegen Systems).

4C arrays and data analysis

4C array design and data analysis were essentially performed as described before (Simonis et al., 2006), except the arrays we used for this study (Nimblegen Systems) covered all 21 mouse chromosomes completely. Each array was hybridized with two independently processed experimental samples labelled with alternate dye orientations. The ratio of sample-to-genomic DNA 4C signal was calculated for each individual probe. For the analysis of *cis*-interactions, a running mean algorithm was performed separately for each experimental sample with a window size of 29 probes (on average 60 kb) and compared with a running mean performed across randomised data. Consequently, all measurements were judged relative to the amplitude and background noise of that specific array. The threshold level was determined using a top-down approach to establish the minimal value for which the false discovery rate (FDR) < 0.05. Next, duplicate experiments were compared and windows that met the threshold in both duplicates were considered positive. When comparing randomised data, no windows were above threshold in both duplicates. Positive windows directly adjacent on the chromosome template were joined (no gaps allowed), creating positive areas. In defining interacting regions *in trans*, we took a similar approach. However, here we applied a running median algorithm, using a window size of 29 probes on 4C-sample/genomic DNA signal ratios. The threshold was set at a FDR of 0. Thus, a region was called interacting when, in both duplicates, the median signal was higher than any signal found in the respective randomized data sets.

Acknowledgments

We thank R. Hendriks and P.F. van Loo for help with experiments.

6

Discussion

Discussion

Regulated expression of protein coding genes throughout development is of fundamental importance for the survival of an organism. In order to control this process, the eukaryotic genome has evolved several regulatory DNA elements that ensure proper spatio-temporal transcription of genes. It is important to understand how these *cis*-regulatory elements communicate over large distances to control gene expression. Using the β-globin locus as a model system, it was shown previously that in erythroid cells regulatory DNA elements of the LCR physically contact downstream genes via a looping mechanism to regulate their expression (Tolhuis et al., 2002). In addition to spatial clustering of the active genes and regulatory elements in the LCR, 3C technology also revealed erythroid specific long-range DNA interactions with distal hypersensitive regions while the intervening chromatin containing the inactive β-globin genes loops out. This spatial configuration of transcriptional regulatory elements in the β-globin locus of erythroid cells is referred to as the Active Chromatin Hub (ACH) (Tolhuis et al., 2002). In erythroid progenitor cells that express β-globin genes at basal levels, a pre-structure consisting of interactions between the outer hypersensitive sites and the 5' side of the LCR is present and result in a spatial conformation called the Chromatin Hub (CH) (Palstra et al., 2003). Upon induction of erythroid differentiation, a fully functional ACH is formed when these sites cluster with the rest of the LCR and the actively transcribed gene. It is thought that during erythroid differentiation these *cis* regulatory DNA elements create a nuclear compartment dedicated to RNA polymerase II transcription of the β-globin genes (Palstra et al., 2003). Presumably, the spatial clustering of regulatory elements in an ACH results in a high local concentration of DNA-binding sites for cognate transcription factors and their interacting partners, which consequently accumulate at the site. Efficiency of transcription is proportional to the concentration of transcription factors involved (Droge and Muller-Hill, 2001) and local accumulation of *trans*-acting factors may indeed be obtained by spatial clustering of cognate binding sites. Therefore contacts formed in the context of the ACH may be essential for the high transcription rate of the β-globin genes. Although experimental data concerning higher order chromatin structure of a locus is limited, there are a number of observations and models that may explain how ACH long-range interactions are established and maintained.

ACH formation and maintenance

Formation of the ACH most likely depends on the affinity between *cis* regulatory elements in the β-globin locus and the transcription factors that bind them. It has been proposed that transcription factors bind their cognate binding sites via a mass-action mechanism (Grosveld, 1999). In such a system, a transcription factor occasionally binds its cognate binding site when this is (temporarily) available. Next, the factor dissociates or, within a small window of opportunity, a second binding site becomes available (either directly on the first factor associated or indirectly via the action of the first bound factor) that can be bound by another transcription factor. This process would

repeat itself and induces a chain reaction that eventually results in the formation of individual transcription factor complexes. The significance of the β-globin LCR in this mechanism is that it contains a large number of these transcription factor binding sites. A stable interaction between regulatory sites eventually leading to higher order chromatin structure (such as an ACH) can only be established if both sites are in close spatial proximity and if affinity exists between transcription factors bound at these sites. This suggests that influencing the *trans*-acting environment will influence ACH formation and maintenance. Indeed, using 3C technology it was previously shown that EKLF and GATA-1 play a pivotal role in the establishment and/or maintenance of the β-globin ACH (Drissen et al., 2004; Vakoc et al., 2005). Data on erythroid progenitor cells in Chapter 3 of this thesis demonstrates that these factors already bind to cognate sites in the β-globin locus at a stage of erythroid differentiation that precedes ACH formation (Kooren et al., 2007). This indicates that either the concentration of EKLF and GATA-1 is important to mediate ACH formation or they first need to be modified in order to execute their structural role. Alternatively, they mediate the recruitment of other proteins important for ACH formation during later stages of erythroid differentiation. In favour of the latter, EKLF and GATA-1 bind DNA directly but are not known to homo-multimerise and therefore it is unlikely that these factors are directly responsible for bringing distant DNA elements together. They do interact however with large protein complexes such as CBP/p300 and Mediator which are known to interact simultaneously with factors bound at enhancers and promoters and could in this way stabilise promoter-LCR contact (Chan and La Thangue, 2001; Kuras et al., 2003). Therefore, promoter-LCR contacts in the context of the β-globin ACH must be thought of as a multi-component entity containing many different proteins and DNA elements and multiple interactions between the LCR and the β-globin promoter are required to stabilise the spatial conformation of the locus. This is illustrated by the deletion of both HS3 and the β-globin promoter which results in a strong destabilisation of the ACH, while the ACH is maintained when each of these sites are deleted individually (Patrinos et al., 2004). Evidently, some of the protein components are crucial, like e.g. EKLF or GATA-1 and removal of these factors destabilises the structure in such a way that promoter-LCR contacts can not be firmly established or maintained. Not all transcription factors that bind specifically to regulatory elements residing in the β-globin locus are required for establishing interactions between these *cis*-regulatory DNA elements. This is illustrated by the role of the erythroid transcription factor p45 NF-E2 in ACH formation described in Chapter 3 of this thesis. Despite the fact that this factor binds to elements in both the β-globin LCR and the promoters of the genes (Forsberg et al., 2000; Johnson et al., 2002; Kang et al., 2002) and is essential for β-globin gene transcription in MEL cells (Johnson et al., 2001; Kotkow and Orkin, 1995; Lu et al., 1994), we show that p45 NF-E2 is not required for ACH formation (Kooren et al., 2007). We find that redundant factors bind to the β-globin locus and physically replace p45 NF-E2 in its absence, which might explain both the unaffected structural organisation of the locus and the mild erythroid phenotype in mice lacking this transcription factor. Whether this physical

replacement has any functional consequences remains unclear and additional experiments using e.g. cells lacking all p45 NF-E2 related factors need to be undertaken to address this question properly. If functional redundancy does exist, this is however not complete as these knock-out mice display levels of β-globin gene expression down to 60-70% of wildtype (Kooren et al., 2007), similar to levels found after complete deletion of HS2 from the endogenous mouse β-globin locus (Fiering et al., 1995).

In addition to individual regulatory elements and their bound transcription factors it has been suggested that gene loci and even whole chromosomes are structurally organised by the process of RNA polymerase II (RNAPII) transcription. The presence of transcribing RNAP II at enhancers and promoters was suggested to produce loops by targeting these sites to the same 'transcription factory '(Chakalova et al., 2005; Cook, 2002; Cook, 2003; Marenduzzo et al., 2007). Long-range DNA interactions of transcribed genes in these transcription factories could be established by a molecular crowding mechanism of bound RNAP II (Marenduzzo et al., 2006). Alternatively, regulatory elements like the β-globin LCR could relocate to preformed transcription factories and reel in the chromatin fiber until an activated promoter is encountered, thereby generating chromatin loops (Chakalova et al., 2005; Muller et al., 2007; West and Fraser, 2005). The prediction that emanates from this model is that inhibition of transcription influences DNA folding and eliminates looping (Faro-Trindade and Cook, 2006; Marenduzzo et al., 2007). However, a recent study performed in our group combines 3C and 4C technology with high-resolution cryo-FISH shows that upon inhibition of transcription by α-amanitin there is no large-scale change in both intra- and interchromosomal long-range DNA interactions of the β-globin locus and the *Rad23a* locus (Palstra et. al., submitted). Furthermore, it was found that after transcriptional inhibition the erythroid transcription factors EKLF, GATA-1 and p45 NF-E2 remain bound to their cognate binding sites in the β-globin locus. This demonstrates that in the absence of ongoing RNAP II transcription long-range DNA interactions, including the β-globin ACH, are maintained and therefore argues strongly against models that suggest that the genome is organised around clusters of active RNAP II. Altogether, these data supports the notion that a minimum amount of transcription factor bound *cis* regulatory elements is required to establish and maintain the β-globin ACH. This is in agreement with the idea that the LCR functions as a holocomplex in which cooperativeness between hypersensitive sites is required for stabilising LCR-gene contacts and for the overall spatial conformation of the β-globin locus.

CTCF and the β-globin ACH

Apart from LCR-gene interactions, the mouse β-globin ACH is comprised of interactions between multiple CTCF binding sites around the locus that spatially cluster in the erythroid cell nucleus. All these interacting sites have been shown to bind the insulator protein CTCF and both 5'HS5 of the LCR and 3'HS1 downstream of the β-globin genes function as a CTCF dependent

enhancer-blocker (Bulger et al., 2003; Farrell et al., 2002). Chapter 4 of this thesis describes experiments in which the role of CTCF in long-range chromatin looping in the mouse β-globin locus is investigated. Using a conditional CTCF knockout approach and targeted disruption of the CTCF binding site in 3'HS1, it was shown that CTCF is required for long-range interactions between cognate CTCF binding sites (Splinter et al., 2006). According to the structural model of insulator function, the chromatin fibre is organised by the establishment of transcriptionally independent loop domains via clustering of insulator sequences. Indeed it was shown that the insulator protein CTCF organises higher order chromatin structure in the context of the β-globin ACH (Splinter et al., 2006). Moreover, this clustering of regulatory elements in CTCF-dependent chromatin loops is evolutionary conserved and can only be found in erythroid cells, implying a functional role in β-globin gene expression. The functional significance of these CTCF mediated loops is however not clear, as disruption of CTCF binding to 3'HS1 had no measurable effect on β-globin transcription (Splinter et al., 2006) and mouse models carrying a combination of deletions of the outer CTCF binding sites showed no aberrant β-globin gene expression (Bender et al., 2006).

The observation that CTCF binds to both sides of the mouse locus has led to the idea that these sequences might act as a barrier element flanking the locus protecting it from the encroachment of heterochromatin or, vice versa, preventing inappropriate activation of surrounding mouse olfactory receptor genes by the β-globin LCR (Burgess-Beusse et al., 2002; Farrell et al., 2002). This idea has mainly emanated from studies on the well-characterised chicken β-globin locus, where the flanking HSs mark an abrupt change in chromatin structure (Chung et al., 1993; Litt et al., 2001). Disruption of the mouse 3'HS1 CTCF binding site did however not result in inappropriate expression of olfactory receptor genes nor did it give rise to heterochromatinization into the β-globin locus. In fact, removal of CTCF resulted only locally in a loss of histone acetylation and a gain of histone methylation, demonstrating that 3'HS1 bound CTCF does not to insulate the murine β-globin locus. This lack of CTCF dependent barrier activity is not completely unexpected however, as boundary activity of the chicken 5'HS4 element is not dependent on CTCF but rather on the binding of another group of proteins collectively known as USF proteins (Bell and Felsenfeld, 1999; Bell et al., 1999; West et al., 2004).

Altogether, the functional relevance of CTCF mediated loops in the β-globin locus still remains unclear. If looped domains facilitate enhancer-promoter interactions, as predicted by the structural model of insulator function, it is to be expected that β-globin gene expression would benefit from these CTCF mediated loops, but no such effect is measured using current detection methods. To test whether positioning of a gene on a CTCF mediated loop together with its enhancer elements positively influences its expression we have initiated ES cell targeting experiments that introduced a 'marked' β-globin gene in the mouse β-globin locus just upstream of a wildtype 3'HS1. More than the endogenous β-globin genes that locate much further away from 3'HS1, this juxtaposed marked β-globin gene is expected to benefit from LCR-3'HS1

interactions. We also prepared a targeting construct that allows the insertion of this β_{marked} gene upstream of a mutated 3'HS1 element that lacks CTCF binding and consequently shows no interactions between 3'HS1 and other sites participating in the ACH. Comparing the relative expression levels of this β_{marked} gene between both situations should reveal whether CTCF mediated loops can indeed facilitate enhancer-mediated gene expression, as predicted from the structural model of enhancer function.

An alternative explanation for the lack of changes in β-globin expression after abrogation of the loop between 3'HS1 and the LCR is the presence of additional CTCF binding sites downstream of 3'HS1, capable of generating an alternative loop with an increased size. In this scenario, the β-globin genes still benefit fully from the presence of a chromatin loop between the LCR and a downstream CTCF site. Indeed, preliminary experiments using an adapted 4C protocol suggest the presence of additional downstream regulatory elements in the mouse β-globin locus and subsequent 3C analysis does not rule out interactions between these unidentified elements and the LCR (M. Simonis, personal communication). Whether these sites truly represent CTCF binding sites capable of participating in higher order chromatin conformation in the mouse β-globin locus needs to be investigated in more detail. In conclusion, CTCF clearly has a specialised function as an enhancer blocker for example in imprinted gene loci such as the Igf2/H19 locus, where it regulates the imprinting of both genes by partitioning alleles into an active or repressed chromatin loop domain (Bell and Felsenfeld, 2000; Hark et al., 2000; Kurukuti et al., 2006; Murrell et al., 2004). The functional relevance of CTCF binding to other sites in the genome, for example to sites flanking gene loci, remains to be investigated.

In addition to CTCF, a number of other proteins have been suggested to play role in the spatial organisation of chromatin. Among them is SATB1 (special AT-rich sequence binding protein 1), a protein expressed at high levels in thymocytes and found to be involved in the organisation of chromatin into distinct loops. SATB1 shows a 'cage-like' nuclear distribution surrounding heterochromatin and is, like CTCF, thought to direct chromatin folding by tethering specialised DNA sequences called matrix attachment regions (MARs) to the nuclear matrix (Cai et al., 2003; de Belle et al., 1998). Indeed, studies on SATB1 in the T_h2 locus using a combination of chromatin immunoprecipitation (ChIP) and 3C technology demonstrated that SATB1 is important for the formation of an active chromatin structure in this locus. In this structure, the transcriptionally active locus is folded into numerous small loops, all anchored to SATB1 at their base. This SATB1 mediated loop formation was shown to be important for transcriptional regulation of the interleukin genes, because SATB1 knockdown cells lacked these small loops and did no express these genes properly (Cai et al., 2006). Another study showed that SATB1 is necessary for an observed loop structure in the major histocompatibility (MCH) class I locus. Moreover, this spatial conformation requires a direct interaction between SATB1 and a protein constituent of the PML nuclear body. Silencing of either SATB1 or this PML protein dynamically alters chromatin architecture and affects the expression profile of a subset of MHC class I genes

(Kumar et al., 2007). These studies indicate that, like CTCF, SATB1 is a protein that governs transcription by orchestrating dynamic chromatin loop architecture. It has been suggested it may also mediate looping in the β-globin locus by bringing distal regulatory elements of the LCR in close proximity to the β-globin promoters (Ostermeier et al., 2003). SATB1 indeed binds to HS2 of the human LCR in K562 cells, but binding to the promoter regions of the β-globin genes in these cells is confined to the embryonic ε-promoter only (Wen et al., 2005), despite the presence of MARs capable of binding SATB1 in both the γ- and β-globin gene promoters (Cunningham et al., 1994; Jackson et al., 1995). Whether SATB1 is indeed a structural protein involved in the transcriptional regulation of all β-globin genes needs to be addressed in future experiments, in which e.g. 3C technology is combined with a model system that more faithfully recapitulates *in vivo* erythroid differentiation, such as the I/11 cell system.

Nuclear architecture uncovered by 4C

The recent development of 4C technology is expected to contribute importantly to a better understanding of nuclear architecture. This unbiased, high-throughput method provides researchers the opportunity to gain new insights into DNA folding by identifying interactions between any locus of interest and the entire genome. Although 4C has been introduced recently, the technology already contributed to a better understanding of nuclear organisation by applying it to a number of genomic regions. Chapter 5 of this thesis describes experiments in which the nuclear organisation of a number of peri-centromeric and peri-telomeric regions has been studied in more detail. In line with 4C data previously obtained for the β-globin locus and the *Rad23a* gene (Simonis et al., 2006), it was shown that there is a strong correlation between the transcriptional status of a genomic region and that of its interacting loci in the interphase nucleus. This differential preferred nuclear environment is not completely unexpected though, as classical cytological staining and light microscopy studies already revealed decades ago that inactive, condensed heterochromatin locates differently compared to active, less condensed euchromatin. Moreover, recent DNA FISH experiments visualised that gene-poor and gene-rich regions were distributed in an alternate fashion on the linear template of a 4.3 Mbp genomic region on mouse chromosome 14 (Shopland et al., 2006). If we combine these results with our recently obtained 4C data, a picture emerges in which active and inactive chromatin domains each have their own preferred nuclear environment and separate in the interphase nucleus. As each genomic region has its own preferred nuclear environment, it implies that it is impossible to predict the long-range DNA interactions of a given locus without knowing the properties of neighbouring segments and, by extrapolation, of the whole chromosome. This notion is in agreement with the concept of self-organising principles underlying chromosomal organisation. In such a self-organising system, a macromolecular complex determines its own structure based on the functional and dynamic interactions of its different components (Misteli, 2001a). By

modification of their subunits, self-organising structures can easily be changed, thereby ensuring structural stability without loss of flexibility. A role for self-organisation was suggested in the biogenesis of multiple cellular component e.g. the nucleolus, in which nucleolar proteins are continuously exchanged between the nucleolus and the nucleoplasm, generating an overall stable structure from dynamic components (Lewis and Tollervey, 2000; Misteli, 2001b). Additionally, self-organisation is thought to be structurally involved in the actin and microtubule cytoskeleton (Mitchison, 1992) and the Golgi complex (Misteli, 2001a). The spatial positioning of genomic regions and chromosomes as measured by 4C can similarly be explained by self-organising properties. The central idea is 'that the sum of all functional properties of a chromosome (i.e., the frequency and linear distribution of its active and inactive regions) determines its positioning' (Misteli, 2007). Indeed, chromosomal folding patterns determined by 4C indicate that both active and inactive chromatin separate in the interphase nucleus. Similarities in structural, biochemical and/or biophysical properties may exist between loci with comparable transcriptional activity, resulting in increased affinities between these chromosomal regions. This process eventually results in the nuclear positioning of loci and overall chromosomal conformation based on the order of established contacts between regions with high mutual affinity.

The transcriptional status of a chromosomal region is probably not the only parameter dictating chromosome folding and positioning. For example, repetitive DNA elements are abundantly present in the genome and examples exist of repetitive sequences coming together in the nuclear space. Two well known examples are the clustering of ribosomal gene arrays in nucleoli and the clustering of centromeric satellite sequence in structures known as chromocentres. 4C analysis of peri-centromeric and peri-telomeric regions presented in Chapter 5 has revealed additional direct evidence for the presence of clusters of repetitive sequences. These experiments revealed preferential associations between peri-centromeric, but also between peri-telomeric regions, indicating the existence of interactions between repetitive chromosomal end sequences. It is therefore reasonable to expect that the spatial separation between repetitive and non-repetitive regions is another important parameter determining interphase nuclear organisation. Of course, many repetitive sequences are transcriptionally inactive and gene-rich regions are often non-repetitive, so there is no strict separation possible between inactive versus active and repetitive versus non-repetitive DNA regions. Altogether, 4C analysis has revealed that the relative positioning of active and inactive regions as well as the distribution of repetitive DNA elements on the linear chromosomal template importantly determine the relative positioning and spatial conformation of chromosomes.

The ongoing development of 4C technology is expected to boost progress in our understanding of chromosomal folding patterns determining nuclear architecture and numerous exciting new research questions can be addressed by using this method. An important question is how DNA topology changes during the process of cellular differentiation and how these architectural changes relate with changes in gene expression. For example, one could use mouse ES cells,

differentiate them into erythroid progenitor (ES-EP) cells (Carotta et al., 2004) followed by maturation into mature red blood cells. Applying 4C on cells from these different developmental stages combined with expression analysis of multiple genes (e.g. constitutively expressed, early developmental and tissue-specific genes) should reveal a possible relation between chromosomal architecture and gene expression throughout differentiation. Another possible application of 4C technology is to adapt the method to such an extent that it will allow for the identification of transcription regulatory elements at a more local scale. Typically, DNA regulatory elements such as enhancers and silencers locate within 1 Mb from their target sequence. The current 4C method does not identify these nearby interacting elements due to the intrinsic limited dynamic range of microarrays. This results in complete saturation of hybridisation signals surrounding 2-5 Mb of the 4C bait sequence, which prevents quantitative analysis of nearby probes. A possible adaptation resolving this problem might be to decrease the number of PCR cycles needed to generate a 4C library, followed by hybridisation to regular microarrays. A promising alternative is the use of high throughput sequencing methods. Sequencing overcomes the limitations of the dynamic range and thereby allows for the identification of interactions on a more local scale. Moreover, it combines accurate quantification with the possibility of analysing the interactions of multiple 4C bait sequences in one run. Although currently still expensive and time-consuming, it is likely that technological improvements will make large scale sequencing the method of choice for analysing 4C data in the near future.

References

Abney, J.R., Cutler, B., Fillbach, M.L., Axelrod, D. and Scalettar, B.A. (1997) Chromatin dynamics in interphase nuclei and its implications for nuclear structure. *J Cell Biol*, 137, 1459-1468.

Abranches, R., Beven, A.F., Aragon-Alcaide, L. and Shaw, P.J. (1998) Transcription sites are not correlated with chromosome territories in wheat nuclei. *J Cell Biol*, 143, 5-12.

Albright, S.R. and Tjian, R. (2000) TAFs revisited: more data reveal new twists and confirm old ideas. *Gene*, 242, 1-13.

Alcobia, I., Quina, A.S., Neves, H., Clode, N. and Parreira, L. (2003) The spatial organization of centromeric heterochromatin during normal human lymphopoiesis: evidence for ontogenically determined spatial patterns. *Exp Cell Res*, 290, 358-369.

Allan, J., Mitchell, T., Harborne, N., Bohm, L. and Crane-Robinson, C. (1986) Roles of H1 domains in determining higher order chromatin structure and H1 location. *J Mol Biol*, 187, 591-601.

Amrolia, P.J., Ramamurthy, L., Saluja, D., Tanese, N., Jane, S.M. and Cunningham, J.M. (1997) The activation domain of the enhancer binding protein p45NF-E2 interacts with TAFII130 and mediates long-range activation of the alpha- and beta-globin gene loci in an erythroid cell line. *Proc Natl Acad Sci U S A*, 94, 10051-10056.

Andrews, N.C., Erdjument-Bromage, H., Davidson, M.B., Tempst, P. and Orkin, S.H. (1993a) Erythroid transcription factor NF-E2 is a haematopoietic-specific basic-leucine zipper protein. *Nature*, 362, 722-728.

Andrews, N.C., Kotkow, K.J., Ney, P.A., Erdjument-Bromage, H., Tempst, P. and Orkin, S.H. (1993b) The ubiquitous subunit of erythroid transcription factor NF-E2 is a small basic-leucine zipper protein related to the v-maf oncogene. *Proc Natl Acad Sci U S A*, 90, 11488-11492.

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

Bajic, V.B., Choudhary, V. and Hock, C.K. (2004) Content analysis of the core promoter region of human genes. *In Silico Biol*, 4, 109-125.

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

Baniahmad, A., Muller, M., Steiner, C. and Renkawitz, R. (1987) Activity of two different silencer elements of the chicken lysozyme gene can be compensated by enhancer elements. *Embo J*, 6, 2297-2303.

Bannister, A.J., Schneider, R. and Kouzarides, T. (2002) Histone methylation: dynamic or static? *Cell*, 109, 801-806.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C. and Kouzarides, T. (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature*, 410, 120-124.

Baur, J.A., Zou, Y., Shay, J.W. and Wright, W.E. (2001) Telomere position effect in human cells. *Science*, 292, 2075-2077.

Bean, T.L. and Ney, P.A. (1997) Multiple regions of p45 NF-E2 are required for beta-globin gene expression in erythroid cells. *Nucleic Acids Res*, 25, 2509-2515.

Behringer, R.R., Hammer, R.E., Brinster, R.L., Palmiter, R.D. and Townes, T.M. (1987) Two 3' sequences direct adult erythroid-specific expression of human beta-globin genes in transgenic mice. *Proc Natl Acad Sci U S A*, 84, 7056-7060.

Beil, M., Durschmied, D., Paschke, S., Schreiner, B., Nolte, U., Bruel, A. and Irinopoulou, T. (2002) Spatial distribution patterns of interphase centromeres during retinoic acid-induced differentiation of promyelocytic leukemia cells. *Cytometry*, 47, 217-225.

Bell, A.C. and Felsenfeld, G. (1999) Stopped at the border: boundaries and insulators. *Curr Opin Genet Dev*, 9, 191-198.

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.

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

Belmont, A.S., Braunfeld, M.B., Sedat, J.W. and Agard, D.A. (1989) Large-scale chromatin structural domains within mitotic and interphase chromosomes in vivo and in vitro. *Chromosoma*, 98, 129-143.

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

Bender, M.A., Byron, R., Ragoczy, T., Telling, A., Bulger, M. and Groudine, M. (2006) Flanking HS-62.5 and 3' HS1, and regions upstream of the LCR, are not required for beta-globin transcription. *Blood*, 108, 1395-1401.

Bender, M.A., Mehaffey, M.G., Telling, A., Hug, B., Ley, T.J., Groudine, M. and Fiering, S. (2000b) Independent formation of Dnasel hypersensitive sites in the murine beta-globin locus control region. *Blood*, 95, 3600-3604.

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

Bender, M.A., Roach, J.N., Halow, J., Close, J., Alami, R., Bouhassira, E.E., Groudine, M. and Fiering, S.N. (2001) Targeted deletion of 5'HS1 and 5'HS4 of the beta-globin locus control region reveals additive activity of the DNaseI hypersensitive sites. *Blood*, 98, 2022-2027.

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

Bi, X. and Broach, J.R. (2001) Chromosomal boundaries in S. cerevisiae. *Curr Opin Genet Dev*, 11, 199-204.

Bi, X., Yu, Q., Sandmeier, J.J. and Zou, Y. (2004) Formation of boundaries of transcriptionally silent chromatin by nucleosome-excluding structures. *Mol Cell Biol*, 24, 2118-2131.

Bickel, S. and Pirrotta, V. (1990) Self-association of the Drosophila zeste protein is responsible for transvection effects. *Embo J*, 9, 2959-2967.

Bird, A., Taggart, M., Frommer, M., Miller, O.J. and Macleod, D. (1985) A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*, 40, 91-99.

Bird, A.P. (1986) CpG-rich islands and the function of DNA methylation. *Nature*, 321, 209-213.

Blackwood, E.M. and Kadonaga, J.T. (1998) Going the distance: a current view of enhancer action. *Science*, 281, 60-63.

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

Blobel, G.A., Nakajima, T., Eckner, R., Montminy, M. and Orkin, S.H. (1998) CREB-binding protein cooperates with transcription factor GATA-1 and is required for erythroid differentiation. *Proc Natl Acad Sci U S A*, 95, 2061-2066.

Blom van Assendelft, G., Hanscombe, O., Grosveld, F. and Greaves, D.R. (1989) The beta-globin dominant control region activates homologous and heterologous promoters in a tissue-specific manner. *Cell*, 56, 969-977.
Bodine, D.M. and Ley, T.J. (1987) An enhancer element lies 3' to the human A gamma globin gene. *Embo J*, 6, 2997-3004.
Bolzer, A., Kreth, G., Solovei, I., Koehler, D., Saracoglu, K., Fauth, C., Muller, S., Eils, R., Cremer, C., Speicher, M.R. and Cremer, T. (2005) Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol*, 3, e157.
Bonifer, C. (2000) Developmental regulation of eukaryotic gene loci: which cis-regulatory information is required? *Trends Genet*, 16, 310-315.
Bottardi, S., Aumont, A., Grosveld, F. and Milot, E. (2003) Developmental stage-specific epigenetic control of human beta-globin gene expression is potentiated in hematopoietic progenitor cells prior to their transcriptional activation. *Blood*, 102, 3989-3997.
Bottardi, S., Ross, J., Pierre-Charles, N., Blank, V. and Milot, E. (2006) Lineage-specific activators affect beta-globin locus chromatin in multipotent hematopoietic progenitors. *Embo J*, 25, 3586-3595.
Boyer, L.A., Latek, R.R. and Peterson, C.L. (2004) The SANT domain: a unique histone-tail-binding module? *Nat Rev Mol Cell Biol*, 5, 158-163.
Boyle, S., Gilchrist, S., Bridger, J.M., Mahy, N.L., Ellis, J.A. and Bickmore, W.A. (2001) The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum Mol Genet*, 10, 211-219.
Branco, M.R. and Pombo, A. (2006) Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol*, 4, e138.
Brand, A.H., Breeden, L., Abraham, J., Sternglanz, R. and Nasmyth, K. (1985) Characterization of a "silencer" in yeast: a DNA sequence with properties opposite to those of a transcriptional enhancer. *Cell*, 41, 41-48.
Brand, M., Ranish, J.A., Kummer, N.T., Hamilton, J., Igarashi, K., Francastel, C., Chi, T.H., Crabtree, G.R., Aebersold, R. and Groudine, M. (2004) Dynamic changes in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics. *Nat Struct Mol Biol*, 11, 73-80.
Brasher, S.V., Smith, B.O., Fogh, R.H., Nietlispach, D., Thiru, A., Nielsen, P.R., Broadhurst, R.W., Ball, L.J., Murzina, N.V. and Laue, E.D. (2000) The structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimer. *Embo J*, 19, 1587-1597.
Brown, J.M., Leach, J., Reittie, J.E., Atzberger, A., Lee-Prudhoe, J., Wood, W.G., Higgs, D.R., Iborra, F.J. and Buckle, V.J. (2006) Coregulated human globin genes are frequently in spatial proximity when active. *J Cell Biol*, 172, 177-187.
Brown, K.E., Amoils, S., Horn, J.M., Buckle, V.J., Higgs, D.R., Merkenschlager, M. and Fisher, A.G. (2001) Expression of alpha- and beta-globin genes occurs within different nuclear domains in haemopoietic cells. *Nat Cell Biol*, 3, 602-606.
Brown, K.E., Baxter, J., Graf, D., Merkenschlager, M. and Fisher, A.G. (1999) Dynamic repositioning of genes in the nucleus of lymphocytes preparing for cell division. *Mol Cell*, 3, 207-217.
Brown, K.E., Guest, S.S., Smale, S.T., Hahm, K., Merkenschlager, M. and Fisher, A.G. (1997) Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell*, 91, 845-854.
Buck, M., Gallegos, M.T., Studholme, D.J., Guo, Y. and Gralla, J.D. (2000) The bacterial enhancer-dependent sigma(54) (sigma(N)) transcription factor. *J Bacteriol*, 182, 4129-4136.
Bulger, M., Bender, M.A., van Doorninck, J.H., Wertman, B., Farrell, C.M., Felsenfeld, G., Groudine, M. and Hardison, R. (2000) Comparative structural and functional analysis of the olfactory receptor genes flanking the human and mouse beta-globin gene clusters. *Proc Natl Acad Sci U S A*, 97, 14560-14565.
Bulger, M. and Groudine, M. (1999) Looping versus linking: toward a model for long-distance gene activation. *Genes Dev*, 13, 2465-2477.
Bulger, M., Schubeler, D., Bender, M.A., Hamilton, J., Farrell, C.M., Hardison, R.C. and Groudine, M. (2003) A complex chromatin landscape revealed by patterns of nuclease sensitivity and histone modification within the mouse beta-globin locus. *Mol Cell Biol*, 23, 5234-5244.
Bulger, M., van Doorninck, J.H., Saitoh, N., Telling, A., Farrell, C., Bender, M.A., Felsenfeld, G., Axel, R. and Groudine, M. (1999) Conservation of sequence and structure flanking the mouse and human beta-globin loci: the beta-globin genes are embedded within an array of odorant receptor genes. *Proc Natl Acad Sci U S A*, 96, 5129-5134.
Bultman, S.J., Gebuhr, T.C. and Magnuson, T. (2005) A Brg1 mutation that uncouples ATPase activity from chromatin remodeling reveals an essential role for SWI/SNF-related complexes in beta-globin expression and erythroid development. *Genes Dev*, 19, 2849-2861.
Buratowski, S. (1995) Mechanisms of gene activation. *Science*, 270, 1773-1774.
Burgess-Beusse, B., Farrell, C., Gaszner, M., Litt, M., Mutskov, V., Recillas-Targa, F., Simpson, M., West, A. and Felsenfeld, G. (2002) The insulation of genes from external enhancers and silencing chromatin. *Proc Natl Acad Sci U S A*, 99 Suppl 4, 16433-16437.
Burke, L.J., Zhang, R., Bartkuhn, M., Tiwari, V.K., Tavoosidana, G., Kurukuti, S., Weth, C., Leers, J., Galjart, N., Ohlsson, R. and Renkawitz, R. (2005) CTCF binding and higher order chromatin structure of the H19 locus are maintained in mitotic chromatin. *Embo J*, 24, 3291-3300.
Byrd, K. and Corces, V.G. (2003) Visualization of chromatin domains created by the gypsy insulator of Drosophila. *J Cell Biol*, 162, 565-574.
Cai, H.N. and Shen, P. (2001) Effects of cis arrangement of chromatin insulators on enhancer-blocking activity. *Science*, 291, 493-495.
Cai, S., Han, H.J. and Kohwi-Shigematsu, T. (2003) Tissue-specific nuclear architecture and gene expression regulated by SATB1. *Nat Genet*, 34, 42-51.
Cai, S., Lee, C.C. and Kohwi-Shigematsu, T. (2006) SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat Genet*, 38, 1278-1288.
Cao, S.X., Gutman, P.D., Dave, H.P. and Schechter, A.N. (1989) Identification of a transcriptional silencer in the 5'-flanking region of the human epsilon-globin gene. *Proc Natl Acad Sci U S A*, 86, 5306-5309.
Carotta, S., Pilat, S., Mairhofer, A., Schmidt, U., Dolznig, H., Steinlein, P. and Beug, H. (2004) Directed differentiation and mass cultivation of pure erythroid progenitors from mouse embryonic stem cells. *Blood*, 104, 1873-1880.
Carter, D., Chakalova, L., Osborne, C.S., Dai, Y.F. and Fraser, P. (2002) Long-range chromatin regulatory interactions in vivo. *Nat Genet*, 32, 623-626.
Caterina, J.J., Ciavatta, D.J., Donze, D., Behringer, R.R. and Townes, T.M. (1994) Multiple elements in human beta-globin locus control region 5' HS 2 are involved in enhancer activity and position-independent, transgene expression. *Nucleic Acids Res*, 22, 1006-

1011.
Chada, K., Magram, J. and Costantini, F. (1986) An embryonic pattern of expression of a human fetal globin gene in transgenic mice. *Nature*, 319, 685-689.
Chakalova, L., Debrand, E., Mitchell, J.A., Osborne, C.S. and Fraser, P. (2005) Replication and transcription: shaping the landscape of the genome. *Nat Rev Genet*, 6, 669-677.
Chambeyron, S. and Bickmore, W.A. (2004) Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev*, 18, 1119-1130.
Chambeyron, S., Da Silva, N.R., Lawson, K.A. and Bickmore, W.A. (2005) Nuclear re-organisation of the Hoxb complex during mouse embryonic development. *Development*, 132, 2215-2223.
Chan, H.M. and La Thangue, N.B. (2001) p300/CBP proteins: HATs for transcriptional bridges and scaffolds. *J Cell Sci*, 114, 2363-2373.
Chan, J.Y., Kwong, M., Lu, R., Chang, J., Wang, B., Yen, T.S. and Kan, Y.W. (1998) Targeted disruption of the ubiquitous CNC-bZIP transcription factor, Nrf-1, results in anemia and embryonic lethality in mice. *Embo J*, 17, 1779-1787.
Chao, W., Huynh, K.D., Spencer, R.J., Davidow, L.S. and Lee, J.T. (2002) CTCF, a candidate trans-acting factor for X-inactivation choice. *Science*, 295, 345-347.
Chen, C.J., Deng, Z., Kim, A.Y., Blobel, G.A. and Lieberman, P.M. (2001) Stimulation of CREB binding protein nucleosomal histone acetyltransferase activity by a class of transcriptional activators. *Mol Cell Biol*, 21, 476-487.
Cheng, X., Reginato, M.J., Andrews, N.C. and Lazar, M.A. (1997) The transcriptional integrator CREB-binding protein mediates positive cross talk between nuclear hormone receptors and the hematopoietic bZip protein p45/NF-E2. *Mol Cell Biol*, 17, 1407-1416.
Chi, T.H., Wan, M., Zhao, K., Taniuchi, I., Chen, L., Littman, D.R. and Crabtree, G.R. (2002) Reciprocal regulation of CD4/CD8 expression by SWI/SNF-like BAF complexes. *Nature*, 418, 195-199.
Cho, E.J., Takagi, T., Moore, C.R. and Buratowski, S. (1997) mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev*, 11, 3319-3326.
Choi, O.R. and Engel, J.D. (1988) Developmental regulation of beta-globin gene switching. *Cell*, 55, 17-26.
Chubb, J.R., Boyle, S., Perry, P. and Bickmore, W.A. (2002) Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr Biol*, 12, 439-445.
Chung, J.H., Whiteley, M. and Felsenfeld, G. (1993) A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. *Cell*, 74, 505-514.
Cmarko, D., Verschure, P.J., Martin, T.E., Dahmus, M.E., Krause, S., Fu, X.D., van Driel, R. and Fakan, S. (1999) Ultrastructural analysis of transcription and splicing in the cell nucleus after bromo-UTP microinjection. *Mol Biol Cell*, 10, 211-223.
Collins, F.S., Metherall, J.E., Yamakawa, M., Pan, J., Weissman, S.M. and Forget, B.G. (1985) A point mutation in the A gamma-globin gene promoter in Greek hereditary persistence of fetal haemoglobin. *Nature*, 313, 325-326.
Constancia, M., Dean, W., Lopes, S., Moore, T., Kelsey, G. and Reik, W. (2000) Deletion of a silencer element in Igf2 results in loss of imprinting independent of H19. *Nat Genet*, 26, 203-206.
Cook, P.R. (2002) Predicting three-dimensional genome structure from transcriptional activity. *Nat Genet*, 32, 347-352.
Cook, P.R. (2003) Nongenic transcription, gene regulation and action at a distance. *J Cell Sci*, 116, 4483-4491.
Corces, V.G. and Geyer, P.K. (1991) Interactions of retrotransposons with the host genome: the case of the gypsy element of Drosophila. *Trends Genet*, 7, 86-90.
Cosma, M.P (2002) Ordered recruitment: gene-specific mechanism of transcription activation. *Mol Cell*, 10, 227-236.
Cosma, M.P., Panizza, S. and Nasmyth, K. (2001) Cdk1 triggers association of RNA polymerase to cell cycle promoters only after recruitment of the mediator by SBF. *Mol Cell*, 7, 1213-1220.
Cote, J., Quinn, J., Workman, J.L. and Peterson, C.L. (1994) Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science*, 265, 53-60.
Cowieson, N.P., Partridge, J.F., Allshire, R.C. and McLaughlin, P.J. (2000) Dimerisation of a chromo shadow domain and distinctions from the chromodomain as revealed by structural analysis. *Curr Biol*, 10, 517-525.
Cremer, T. and Cremer, C. (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet*, 2, 292-301.
Croft, J.A., Bridger, J.M., Boyle, S., Perry, P., Teague, P. and Bickmore, W.A. (1999) Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol*, 145, 1119-1131.
Crossley, M., Merika, M. and Orkin, S.H. (1995) Self-association of the erythroid transcription factor GATA-1 mediated by its zinc finger domains. *Mol Cell Biol*, 15, 2448-2456.
Crossley, M., Tsang, A.P., Bieker, J.J. and Orkin, S.H. (1994) Regulation of the erythroid Kruppel-like factor (EKLF) gene promoter by the erythroid transcription factor GATA-1. *J Biol Chem*, 269, 15440-15444.
Csink, A.K. and Henikoff, S. (1996) Genetic modification of heterochromatic association and nuclear organization in Drosophila. *Nature*, 381, 529-531.
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) The regulatory element 3' to the A gamma-globin gene binds to the nuclear matrix and interacts with special A-T-rich binding protein 1 (SATB1), an SAR/MAR-associating region DNA binding protein. *Blood*, 84, 1298-1308.
de Belle, I., Cai, S. and Kohwi-Shigematsu, T. (1998) The genomic sequences bound to special AT-rich sequence-binding protein 1 (SATB1) in vivo in Jurkat T cells are tightly associated with the nuclear matrix at the bases of the chromatin loops. *J Cell Biol*, 141, 335-348.
de Laat, W. (2007) Long-range DNA contacts: romance in the nucleus? *Curr Opin Cell Biol*, 19, 317-320.
de Laat, W. and Grosveld, F. (2003) Spatial organization of gene expression: the active chromatin hub. *Chromosome Res*, 11, 447-459.
de Villiers, J., Olson, L., Banerji, J. and Schaffner, W. (1983) Analysis of the transcriptional enhancer effect. *Cold Spring Harb Symp Quant Biol*, 47 Pt 2, 911-919.
Defossez, P.A. and Gilson, E. (2002) The vertebrate protein CTCF functions as an insulator in Saccharomyces cerevisiae. *Nucleic Acids Res*, 30, 5136-5141.
Dekker, J. (2006) The three 'C' s of chromosome conformation capture: controls, controls, controls. *Nat Methods*, 3, 17-21.
Dekker, J., Rippe, K., Dekker, M. and Kleckner, N. (2002) Capturing chromosome conformation. *Science*, 295, 1306-1311.
Delgado, S., Gomez, M., Bird, A. and Antequera, F. (1998) Initiation of DNA replication at CpG islands in mammalian chromosomes. *Embo*

J, 17, 2426-2435.
Demers, C., Chaturvedi, C.P., Ranish, J.A., Juban, G., Lai, P., Morle, F., Aebersold, R., Dilworth, F.J., Groudine, M. and Brand, M. (2007) Activator-mediated recruitment of the MLL2 methyltransferase complex to the beta-globin locus. *Mol Cell*, 27, 573-584.
Derjuga, A., Gourley, T.S., Holm, T.M., Heng, H.H., Shivdasani, R.A., Ahmed, R., Andrews, N.C. and Blank, V. (2004) Complexity of CNC transcription factors as revealed by gene targeting of the Nrf3 locus. *Mol Cell Biol*, 24, 3286-3294.
Dernburg, A.F., Broman, K.W., Fung, J.C., Marshall, W.F., Philips, J., Agard, D.A. and Sedat, J.W. (1996) Perturbation of nuclear architecture by long-distance chromosome interactions. *Cell*, 85, 745-759.
Dillon, N. and Festenstein, R. (2002) Unravelling heterochromatin: competition between positive and negative factors regulates accessibility. *Trends Genet*, 18, 252-258.
Dillon, N. and Grosveld, F. (1991) Human gamma-globin genes silenced independently of other genes in the beta-globin locus. *Nature*, 350, 252-254.
Dillon, N. and Sabbattini, P. (2000) Functional gene expression domains: defining the functional unit of eukaryotic gene regulation. *Bioessays*, 22, 657-665.
Dillon, N., Trimborn, T., Strouboulis, J., Fraser, P. and Grosveld, F. (1997) The effect of distance on long-range chromatin interactions. *Mol Cell*, 1, 131-139.
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.
Dolznig, H., Kolbus, A., Leberbauer, C., Schmidt, U., Deiner, E.M., Mullner, E.W. and Beug, H. (2005) Expansion and differentiation of immature mouse and human hematopoietic progenitors. *Methods Mol Med*, 105, 323-344.
Dong, F. and Jiang, J. (1998) Non-Rabl patterns of centromere and telomere distribution in the interphase nuclei of plant cells. *Chromosome Res*, 6, 551-558.
Dorsett, D. (1993) Distance-independent inactivation of an enhancer by the suppressor of Hairy-wing DNA-binding protein of Drosophila. *Genetics*, 134, 1135-1144.
Dorsett, D. (1999) Distant liaisons: long-range enhancer-promoter interactions in Drosophila. *Curr Opin Genet Dev*, 9, 505-514.
Dostie, J. and Dekker, J. (2007) Mapping networks of physical interactions between genomic elements using 5C technology. *Nat Protoc*, 2, 988-1002.
Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C., Green, R.D. and Dekker, J. (2006) Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res*, 16, 1299-1309.
Driscoll, M.C., Dobkin, C.S. and Alter, B.P. (1989) Gamma delta beta-thalassemia due to a de novo mutation deleting the 5' beta-globin gene activation-region hypersensitive sites. *Proc Natl Acad Sci U S A*, 86, 7470-7474.
Drissen, R., Palstra, R.J., Gillemans, N., Splinter, E., Grosveld, F., Philipsen, S. and de Laat, W. (2004) The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev*, 18, 2485-2490.
Drissen, R., von Lindern, M., Kolbus, A., Driegen, S., Steinlein, P., Beug, H., Grosveld, F. and Philipsen, S. (2005) The erythroid phenotype of EKLF-null mice: defects in hemoglobin metabolism and membrane stability. *Mol Cell Biol*, 25, 5205-5214.
Droge, P. and Muller-Hill, B. (2001) High local protein concentrations at promoters: strategies in prokaryotic and eukaryotic cells. *Bioessays*, 23, 179-183.
Dunaway, M. and Droge, P. (1989) Transactivation of the Xenopus rRNA gene promoter by its enhancer. *Nature*, 341, 657-659.
Dunn, K.L., Zhao, H. and Davie, J.R. (2003) The insulator binding protein CTCF associates with the nuclear matrix. *Exp Cell Res*, 288, 218-223.
Eden, S., Constancia, M., Hashimshony, T., Dean, W., Goldstein, B., Johnson, A.C., Keshet, I., Reik, W. and Cedar, H. (2001) An upstream repressor element plays a role in Igf2 imprinting. *Embo J*, 20, 3518-3525.
Eissenberg, J.C. and Elgin, S.C. (1991) Boundary functions in the control of gene expression. *Trends Genet*, 7, 335-340.
Ellis, J., Talbot, D., Dillon, N. and Grosveld, F. (1993) Synthetic human beta-globin 5'HS2 constructs function as locus control regions only in multicopy transgene concatamers. *Embo J*, 12, 127-134.
Ellis, J., Tan-Un, K.C., Harper, A., Michalovich, D., Yannoutsos, N., Philipsen, S. and Grosveld, F. (1996) A dominant chromatin-opening activity in 5' hypersensitive site 3 of the human beta-globin locus control region. *Embo J*, 15, 562-568.
Enver, T., Raich, N., Ebens, A.J., Papayannopoulou, T., Costantini, F. and Stamatoyannopoulos, G. (1990) Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. *Nature*, 344, 309-313.
Epner, E., Reik, A., Cimbora, D., Telling, A., Bender, M.A., Fiering, S., Enver, T., Martin, D.I., Kennedy, M., Keller, G. and Groudine, M. (1998) The beta-globin LCR is not necessary for an open chromatin structure or developmentally regulated transcription of the native mouse beta-globin locus. *Mol Cell*, 2, 447-455.
Erickson, R.L., Hemati, N., Ross, S.E. and MacDougald, O.A. (2001) p300 coactivates the adipogenic transcription factor CCAAT/enhancer-binding protein alpha. *J Biol Chem*, 276, 16348-16355.
Evans, T., Reitman, M. and Felsenfeld, G. (1988) An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes. *Proc Natl Acad Sci U S A*, 85, 5976-5980.
Fang, X., Xiang, P., Yin, W., Stamatoyannopoulos, G. and Li, Q. (2007) Cooperativeness of the higher chromatin structure of the beta-globin locus revealed by the deletion mutations of DNase I hypersensitive site 3 of the LCR. *J Mol Biol*, 365, 31-37.
Farmer, S.C., Sun, C.W., Winnier, G.E., Hogan, B.L. and Townes, T.M. (1997) The bZIP transcription factor LCR-F1 is essential for mesoderm formation in mouse development. *Genes Dev*, 11, 786-798.
Faro-Trindade, I. and Cook, P.R. (2006) Transcription factories: structures conserved during differentiation and evolution. *Biochem Soc Trans*, 34, 1133-1137.
Farrell, C.M., Grinberg, A., Huang, S.P., Chen, D., Pichel, J.G., Westphal, H. and Felsenfeld, G. (2000) A large upstream region is not necessary for gene expression or hypersensitive site formation at the mouse beta -globin locus. *Proc Natl Acad Sci U S A*, 97, 14554-14559.
Farrell, C.M., West, A.G. and Felsenfeld, G. (2002) Conserved CTCF insulator elements flank the mouse and human beta-globin loci. *Mol Cell Biol*, 22, 3820-3831.
Faustino, P., Lavinha, J., Marini, M.G. and Moi, P. (1996) beta-Thalassemia mutation at -90C-->T impairs the interaction of the proximal CACCC box with both erythroid and nonerythroid factors. *Blood*, 88, 3248-3249.
Feng, W.C., Southwood, C.M. and Bieker, J.J. (1994) Analyses of beta-thalassemia mutant DNA interactions with erythroid Kruppel-like

factor (EKLF), an erythroid cell-specific transcription factor. *J Biol Chem*, 269, 1493-1500.
Festenstein, R., Tolaini, M., Corbella, P., Mamalaki, C., Parrington, J., Fox, M., Miliou, A., Jones, M. and Kioussis, D. (1996) Locus control region function and heterochromatin-induced position effect variegation. *Science*, 271, 1123-1125.
Fiering, S., Epner, E., Robinson, K., Zhuang, Y., Telling, A., Hu, M., Martin, D.I., Enver, T., Ley, T.J. and Groudine, M. (1995) Targeted deletion of 5'HS2 of the murine beta-globin LCR reveals that it is not essential for proper regulation of the beta-globin locus. *Genes Dev*, 9, 2203-2213.
Filippova, G.N., Cheng, M.K., Moore, J.M., Truong, J.P., Hu, Y.J., Nguyen, D.K., Tsuchiya, K.D. and Disteche, C.M. (2005) Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development. *Dev Cell*, 8, 31-42.
Filippova, G.N., Thienes, C.P., Penn, B.H., Cho, D.H., Hu, Y.J., Moore, J.M., Klesert, T.R., Lobanenkov, V.V. and Tapscott, S.J. (2001) CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. *Nat Genet*, 28, 335-343.
Flanagan, J.F., Mi, L.Z., Chruszcz, M., Cymborowski, M., Clines, K.L., Kim, Y., Minor, W., Rastinejad, F. and Khorasanizadeh, S. (2005) Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature*, 438, 1181-1185.
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.
Forrester, W.C., Epner, E., Driscoll, M.C., Enver, T., Brice, M., Papayannopoulou, T. and Groudine, M. (1990) A deletion of the human beta-globin locus activation region causes a major alteration in chromatin structure and replication across the entire beta-globin locus. *Genes Dev*, 4, 1637-1649.
Forrester, W.C., Thompson, C., Elder, J.T. and Groudine, M. (1986) A developmentally stable chromatin structure in the human beta-globin gene cluster. *Proc Natl Acad Sci U S A*, 83, 1359-1363.
Forsberg, E.C., Downs, K.M. and Bresnick, E.H. (2000) Direct interaction of NF-E2 with hypersensitive site 2 of the beta-globin locus control region in living cells. *Blood*, 96, 334-339.
Forsberg, E.C., Johnson, K., Zaboikina, T.N., Mosser, E.A. and Bresnick, E.H. (1999) Requirement of an E1A-sensitive coactivator for long-range transactivation by the beta-globin locus control region. *J Biol Chem*, 274, 26850-26859.
Francastel, C., Magis, W. and Groudine, M. (2001) Nuclear relocation of a transactivator subunit precedes target gene activation. *Proc Natl Acad Sci U S A*, 98, 12120-12125.
Francastel, C., Walters, M.C., Groudine, M. and Martin, D.I. (1999) A functional enhancer suppresses silencing of a transgene and prevents its localization close to centrometric heterochromatin. *Cell*, 99, 259-269.
Fraser, P., Hurst, J., Collis, P. and Grosveld, F. (1990) DNaseI hypersensitive sites 1, 2 and 3 of the human beta-globin dominant control region direct position-independent expression. *Nucleic Acids Res*, 18, 3503-3508.
Fraser, P., Pruzina, S., Antoniou, M. and Grosveld, F. (1993) Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression on the globin genes. *Genes Dev*, 7, 106-113.
Fujiwara, Y., Browne, C.P., Cunniff, K., Goff, S.C. and Orkin, S.H. (1996) Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc Natl Acad Sci U S A*, 93, 12355-12358.
Funabiki, H., Hagan, I., Uzawa, S. and Yanagida, M. (1993) Cell cycle-dependent specific positioning and clustering of centromeres and telomeres in fission yeast. *J Cell Biol*, 121, 961-976.
Fuss, S.H., Omura, M. and Mombaerts, P. (2007) Local and cis effects of the H element on expression of odorant receptor genes in mouse. *Cell*, 130, 373-384.
Gaszner, M. and Felsenfeld, G. (2006) Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet*., 7, 703-713.
Gaszner, M., Vazquez, J. and Schedl, P. (1999) The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer-promoter interaction. *Genes Dev*, 13, 2098-2107.
Gause, M., Morcillo, P. and Dorsett, D. (2001) Insulation of enhancer-promoter communication by a gypsy transposon insert in the Drosophila cut gene: cooperation between suppressor of hairy-wing and modifier of mdg4 proteins. *Mol Cell Biol*, 21, 4807-4817.
Gerasimova, T.I., Byrd, K. and Corces, V.G. (2000) A chromatin insulator determines the nuclear localization of DNA. *Mol Cell*, 6, 1025-1035.
Gerber, M. and Shilatifard, A. (2003) Transcriptional elongation by RNA polymerase II and histone methylation. *J Biol Chem*, 278, 26303-26306.
Gerlich, D. and Ellenberg, J. (2003) Dynamics of chromosome positioning during the cell cycle. *Curr Opin Cell Biol*, 15, 664-671.
Gershenzon, N.I. and Ioshikhes, I.P. (2005) Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. *Bioinformatics*, 21, 1295-1300.
Geyer, P.K. (1997) The role of insulator elements in defining domains of gene expression. *Curr Opin Genet Dev*, 7, 242-248.
Geyer, P.K. and Corces, V.G. (1992) DNA position-specific repression of transcription by a Drosophila zinc finger protein. *Genes Dev*, 6, 1865-1873.
Ghirlando, R., Litt, M.D., Prioleau, M.N., Recillas-Targa, F. and Felsenfeld, G. (2004) Physical properties of a genomic condensed chromatin fragment. *J Mol Biol*, 336, 597-605.
Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N.P. and Bickmore, W.A. (2004) Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell*, 118, 555-566.
Gillemans, N., McMorrow, T., Tewari, R., Wai, A.W., Burgtorf, C., Drabek, D., Ventress, N., Langeveld, A., Higgs, D., Tan-Un, K., Grosveld, F. and Philipsen, S. (2003) Functional and comparative analysis of globin loci in pufferfish and humans. *Blood*, 101, 2842-2849.
Gillemans, N., Tewari, R., Lindeboom, F., Rottier, R., de Wit, T., Wijgerde, M., Grosveld, F. and Philipsen, S. (1998) Altered DNA-binding specificity mutants of EKLF and Sp1 show that EKLF is an activator of the beta-globin locus control region in vivo. *Genes Dev*, 12, 2863-2873.
Giresi, P.G., Kim, J., McDaniell, R.M., Iyer, V.R. and Lieb, J.D. (2007) FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res*, 17, 877-885.
Glover, D.M., Leibowitz, M.H., McLean, D.A. and Parry, H. (1995) Mutations in aurora prevent centrosome separation leading to the formation of monopolar spindles. *Cell*, 81, 95-105.
Gombert, W.M., Farris, S.D., Rubio, E.D., Morey-Rosler, K.M., Schubach, W.H. and Krumm, A. (2003) The c-myc insulator element and matrix attachment regions define the c-myc chromosomal domain. *Mol Cell Biol*, 23, 9338-9348.

Goodwin, A.J., McInerney, J.M., Glander, M.A., Pomerantz, O. and Lowrey, C.H. (2001) In vivo formation of a human beta-globin locus control region core element requires binding sites for multiple factors including GATA-1, NF-E2, erythroid Kruppel-like factor, and Sp1. *J Biol Chem*, 276, 26883-26892.
Greil, F., van der Kraan, I., Delrow, J., Smothers, J.F., de Wit, E., Bussemaker, H.J., van Driel, R., Henikoff, S. and van Steensel, B. (2003) Distinct HP1 and Su(var)3-9 complexes bind to sets of developmentally coexpressed genes depending on chromosomal location. *Genes Dev*, 17, 2825-2838.
Gresh, L., Bourachot, B., Reimann, A., Guigas, B., Fiette, L., Garbay, S., Muchardt, C., Hue, L., Pontoglio, M., Yaniv, M. and Klochendler-Yeivin, A. (2005) The SWI/SNF chromatin-remodeling complex subunit SNF5 is essential for hepatocyte differentiation. *Embo J*, 24, 3313-3324.
Grewal, S.I. and Moazed, D. (2003) Heterochromatin and epigenetic control of gene expression. *Science*, 301, 798-802.
Gribnau, J., de Boer, E., Trimborn, T., Wijgerde, M., Milot, E., Grosveld, F. and Fraser, P. (1998) Chromatin interaction mechanism of transcriptional control in vivo. *Embo J*, 17, 6020-6027.
Grogan, J.L., Mohrs, M., Harmon, B., Lacy, D.A., Sedat, J.W. and Locksley, R.M. (2001) Early transcription and silencing of cytokine genes underlie polarization of T helper cell subsets. *Immunity*, 14, 205-215.
Grosveld, F. (1999) Activation by locus control regions? *Curr Opin Genet Dev*, 9, 152-157.
Grosveld, F., van Assendelft, G.B., Greaves, D.R. and Kollias, G. (1987) Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell*, 51, 975-985.
Groudine, M., Eisenman, R., Gelinas, R. and Weintraub, H. (1983) Developmental aspects of chromatin structure and gene expression. *Prog Clin Biol Res*, 134, 159-182.
Guarente, L. (1988) UASs and enhancers: common mechanism of transcriptional activation in yeast and mammals. *Cell*, 52, 303-305.
Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R. and Young, R.A. (2007) A chromatin landmark and transcription initiation at most promoters in human cells. *Cell*, 130, 77-88.
Gutierrez, L., Nikolic, T., van Dijk, T.B., Hammad, H., Vos, N., Willart, M., Grosveld, F., Philipsen, S. and Lambrecht, B.N. (2007) Gata1 regulates dendritic-cell development and survival. *Blood*, 110, 1933-1941.
Guy, L.G., Kothary, R., DeRepentigny, Y., Delvoye, N., Ellis, J. and Wall, L. (1996) The beta-globin locus control region enhances transcription of but does not confer position-independent expression onto the lacZ gene in transgenic mice. *Embo J*, 15, 3713-3721.
Hagege, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., de Laat, W. and Forne, T. (2007) Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc*, 2, 1722-1733.
Hagstrom, K., Muller, M. and Schedl, P. (1996) Fab-7 functions as a chromatin domain boundary to ensure proper segment specification by the Drosophila bithorax complex. *Genes Dev*, 10, 3202-3215.
Hall, I.M., Shankaranarayana, G.D., Noma, K., Ayoub, N., Cohen, A. and Grewal, S.I. (2002) Establishment and maintenance of a heterochromatin domain. *Science*, 297, 2232-2237.
Hampsey, M. and Reinberg, D. (2003) Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation. *Cell*, 113, 429-432.
Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N. and Grosveld, F. (1991) Importance of globin gene order for correct developmental expression. *Genes Dev*, 5, 1387-1394.
Hansen, J.C. (2002) Conformational dynamics of the chromatin fiber in solution: determinants, mechanisms, and functions. *Annu Rev Biophys Biomol Struct*, 31, 361-392.
Hardison, R. (1998) Hemoglobins from bacteria to man: evolution of different patterns of gene expression. *J Exp Biol*, 201, 1099-1117.
Hardison, R., Slightom, J.L., Gumucio, D.L., Goodman, M., Stojanovic, N. and Miller, W. (1997) Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. *Gene*, 205, 73-94.
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.
Hart, C.M., Zhao, K. and Laemmli, U.K. (1997) The scs' boundary element: characterization of boundary element-associated factors. *Mol Cell Biol*, 17, 999-1009.
Hassan, A.H., Prochasson, P., Neely, K.E., Galasinski, S.C., Chandy, M., Carrozza, M.J. and Workman, J.L. (2002) Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell*, 111, 369-379.
Hatzis, P. and Talianidis, I. (2002) Dynamics of enhancer-promoter communication during differentiation-induced gene activation. *Mol Cell*, 10, 1467-1477.
Heard, E. and Disteche, C.M. (2006) Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev*, 20, 1848-1867.
Heitz E. (1928) Das Heterochromatin der Moose. *Jahrb. Wiss. Bot.*, 69, 762-818.
Herendeen, D.R., Kassavetis, G.A. and Geiduschek, E.P. (1992) A transcriptional enhancer whose function imposes a requirement that proteins track along DNA. *Science*, 256, 1298-1303.
Hewitt, S.L., High, F.A., Reiner, S.L., Fisher, A.G. and Merkenschlager, M. (2004) Nuclear repositioning marks the selective exclusion of lineage-inappropriate transcription factor loci during T helper cell differentiation. *Eur J Immunol*, 34, 3604-3613.
Hochstrasser, M., Mathog, D., Gruenbaum, Y., Saumweber, H. and Sedat, J.W. (1986) Spatial organization of chromosomes in the salivary gland nuclei of Drosophila melanogaster. *J Cell Biol*, 102, 112-123.
Hogan, G.J., Lee, C.K. and Lieb, J.D. (2006) Cell cycle-specified fluctuation of nucleosome occupancy at gene promoters. *PLoS Genet*, 2, e158.
Horike, S., Cai, S., Miyano, M., Cheng, J.F. and Kohwi-Shigematsu, T. (2005) Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat Genet*, 37, 31-40.
Horowitz-Scherer, R.A. and Woodcock, C.L. (2006) Organization of interphase chromatin. *Chromosoma*, 115, 1-14.
Huang, Y., Fang, J., Bedford, M.T., Zhang, Y. and Xu, R.M. (2006) Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science*, 312, 748-751.
Hug, B.A., Wesselschmidt, R.L., Fiering, S., Bender, M.A., Epner, E., Groudine, M. and Ley, T.J. (1996) Analysis of mice containing a targeted deletion of beta-globin locus control region 5' hypersensitive site 3. *Mol Cell Biol*, 16, 2906-2912.
Igarashi, K., Kataoka, K., Itoh, K., Hayashi, N., Nishizawa, M. and Yamamoto, M. (1994) Regulation of transcription by dimerization of erythroid factor NF-E2 p45 with small Maf proteins. *Nature*, 367, 568-572.
Iizuka, M. and Smith, M.M. (2003) Functional consequences of histone modifications. *Curr Opin Genet Dev*, 13, 154-160.

Im, H., Grass, J.A., Johnson, K.D., Kim, S.I., Boyer, M.E., Imbalzano, A.N., Bieker, J.J. and Bresnick, E.H. (2005) Chromatin domain activation via GATA-1 utilization of a small subset of dispersed GATA motifs within a broad chromosomal region. *Proc Natl Acad Sci U S A*, 102, 17065-17070.

Im, H., Park, C., Feng, Q., Johnson, K.D., Kiekhaefer, C.M., Choi, K., Zhang, Y. and Bresnick, E.H. (2003) Dynamic regulation of histone H3 methylated at lysine 79 within a tissue-specific chromatin domain. *J Biol Chem*, 278, 18346-18352.

Ishii, K., Arib, G., Lin, C., Van Houwe, G. and Laemmli, U.K. (2002) Chromatin boundaries in budding yeast: the nuclear pore connection. *Cell*, 109, 551-562.

Ishii, K. and Laemmli, U.K. (2003) Structural and dynamic functions establish chromatin domains. *Mol Cell*, 11, 237-248.

Ito, E., Toki, T., Ishihara, H., Ohtani, H., Gu, L., Yokoyama, M., Engel, J.D. and Yamamoto, M. (1993) Erythroid transcription factor GATA-1 is abundantly transcribed in mouse testis. *Nature*, 362, 466-468.

Jackson, C.E., O'Neill, D. and Bank, A. (1995) Nuclear factor binding sites in human beta globin IVS2. *J Biol Chem*, 270, 28448-28456.

Jackson, D.A., Hassan, A.B., Errington, R.J. and Cook, P.R. (1993) Visualization of focal sites of transcription within human nuclei. *Embo J*, 12, 1059-1065.

Jackson, V. (1999) Formaldehyde cross-linking for studying nucleosomal dynamics. *Methods*, 17, 125-139.

Jacobs, S.A., Taverna, S.D., Zhang, Y., Briggs, S.D., Li, J., Eissenberg, J.C., Allis, C.D. and Khorasanizadeh, S. (2001) Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. *Embo J*, 20, 5232-5241.

Jenuwein, T. and Allis, C.D. (2001) Translating the histone code. *Science*, 293, 1074-1080.

Jin, Q.W., Fuchs, J. and Loidl, J. (2000) Centromere clustering is a major determinant of yeast interphase nuclear organization. *J Cell Sci*, 113 (Pt 11), 1903-1912.

Johnson, G.R. and Metcalf, D. (1977) Pure and mixed erythroid colony formation in vitro stimulated by spleen conditioned medium with no detectable erythropoietin. *Proc Natl Acad Sci U S A*, 74, 3879-3882.

Johnson, K.D., Christensen, H.M., Zhao, B. and Bresnick, E.H. (2001) Distinct mechanisms control RNA polymerase II recruitment to a tissue-specific locus control region and a downstream promoter. *Mol Cell*, 8, 465-471.

Johnson, K.D., Grass, J.A., Boyer, M.E., Kiekhaefer, C.M., Blobel, G.A., Weiss, M.J. and Bresnick, E.H. (2002) Cooperative activities of hematopoietic regulators recruit RNA polymerase II to a tissue-specific chromatin domain. *Proc Natl Acad Sci U S A*, 99, 11760-11765.

Jones, P.A. and Takai, D. (2001) The role of DNA methylation in mammalian epigenetics. *Science*, 293, 1068-1070.

Juven-Gershon, T., Hsu, J.Y. and Kadonaga, J.T. (2006) Perspectives on the RNA polymerase II core promoter. *Biochem Soc Trans*, 34, 1047-1050.

Kang, S.H., Vieira, K. and Bungert, J. (2002) Combining chromatin immunoprecipitation and DNA footprinting: a novel method to analyze protein-DNA interactions in vivo. *Nucleic Acids Res*, 30, e44.

Kataoka, K., Igarashi, K., Itoh, K., Fujiwara, K.T., Noda, M., Yamamoto, M. and Nishizawa, M. (1995) Small Maf proteins heterodimerize with Fos and may act as competitive repressors of the NF-E2 transcription factor. *Mol Cell Biol*, 15, 2180-2190.

Kataoka, K., Noda, M. and Nishizawa, M. (1994) Maf nuclear oncoprotein recognizes sequences related to an AP-1 site and forms heterodimers with both Fos and Jun. *Mol Cell Biol*, 14, 700-712.

Kellum, R. and Schedl, P. (1991) A position-effect assay for boundaries of higher order chromosomal domains. *Cell*, 64, 941-950.

Kellum, R. and Schedl, P. (1992) A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Mol Cell Biol*, 12, 2424-2431.

Kiekhaefer, C.M., Grass, J.A., Johnson, K.D., Boyer, M.E. and Bresnick, E.H. (2002) Hematopoietic-specific activators establish an overlapping pattern of histone acetylation and methylation within a mammalian chromatin domain. *Proc Natl Acad Sci U S A*, 99, 14309-14314.

Kim, J.K., Huh, S.O., Choi, H., Lee, K.S., Shin, D., Lee, C., Nam, J.S., Kim, H., Chung, H., Lee, H.W., Park, S.D. and Seong, R.H. (2001) Srg3, a mouse homolog of yeast SWI3, is essential for early embryogenesis and involved in brain development. *Mol Cell Biol*, 21, 7787-7795.

Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V. and Ren, B. (2007) Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell*, 128, 1231-1245.

Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D. and Ren, B. (2005) A high-resolution map of active promoters in the human genome. *Nature*, 436, 876-880.

Kim, Y.J., Bjorklund, S., Li, Y., Sayre, M.H. and Kornberg, R.D. (1994) A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell*, 77, 599-608.

Kimura, H., Sugaya, K. and Cook, P.R. (2002) The transcription cycle of RNA polymerase II in living cells. *J Cell Biol*, 159, 777-782.

Kina, T., Ikuta, K., Takayama, E., Wada, K., Majumdar, A.S., Weissman, I.L. and Katsura, Y. (2000) The monoclonal antibody TER-119 recognizes a molecule associated with glycophorin A and specifically marks the late stages of murine erythroid lineage. *Br J Haematol*, 109, 280-287.

Kioussis, D., Vanin, E., deLange, T., Flavell, R.A. and Grosveld, F.G. (1983) Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. *Nature*, 306, 662-666.

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

Ko, L.J. and Engel, J.D. (1993) DNA-binding specificities of the GATA transcription factor family. *Mol Cell Biol*, 13, 4011-4022.

Kobayashi, A., Ito, E., Toki, T., Kogame, K., Takahashi, S., Igarashi, K., Hayashi, N. and Yamamoto, M. (1999) Molecular cloning and functional characterization of a new Cap'n' collar family transcription factor Nrf3. *J Biol Chem*, 274, 6443-6452.

Kobor, M.S., Archambault, J., Lester, W., Holstege, F.C., Gileadi, O., Jansma, D.B., Jennings, E.G., Kouyoumdjian, F., Davidson, A.R., Young, R.A. and Greenblatt, J. (1999) An unusual eukaryotic protein phosphatase required for transcription by RNA polymerase II and CTD dephosphorylation in S. cerevisiae. *Mol Cell*, 4, 55-62.

Miele, A., Gheldof, N., Tabuchi, T.M., Dostie, J. & Dekker, J. Mapping chromatin interactions by chromosome conformation capture (3C). In *Current protocols in molecular biology* (eds. Ausubel, F.M. *et al.*), pp. 21.11.21–21.11-20. John Wiley & Sons, New York. 2006.

Kolesky, S.E., Ouhammouch, M. and Geiduschek, E.P. (2002) The mechanism of transcriptional activation by the topologically DNA-linked sliding clamp of bacteriophage T4. *J Mol Biol*, 321, 767-784.

Kollias, G., Hurst, J., deBoer, E. and Grosveld, F. (1987) The human beta-globin gene contains a downstream developmental specific

enhancer. *Nucleic Acids Res*, 15, 5739-5747.
Kollias, G., Wrighton, N., Hurst, J. and Grosveld, F. (1986) Regulated expression of human A gamma-, beta-, and hybrid gamma beta-globin genes in transgenic mice: manipulation of the developmental expression patterns. *Cell*, 46, 89-94.
Kong, S., Bohl, D., Li, C. and Tuan, D. (1997) Transcription of the HS2 enhancer toward a cis-linked gene is independent of the orientation, position, and distance of the enhancer relative to the gene. *Mol Cell Biol*, 17, 3955-3965.
Kooren, J., Palstra, R.J., Klous, P., Splinter, E., von Lindern, M., Grosveld, F. and de Laat, W. (2007) Beta-globin active chromatin Hub formation in differentiating erythroid cells and in p45 NF-E2 knock-out mice. *J Biol Chem*, 282, 16544-16552.
Kosak, S.T., Skok, J.A., Medina, K.L., Riblet, R., Le Beau, M.M., Fisher, A.G. and Singh, H. (2002) Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science*, 296, 158-162.
Kotkow, K.J. and Orkin, S.H. (1995) Dependence of globin gene expression in mouse erythroleukemia cells on the NF-E2 heterodimer. *Mol Cell Biol*, 15, 4640-4647.
Kowenz-Leutz, E. and Leutz, A. (1999) A C/EBP beta isoform recruits the SWI/SNF complex to activate myeloid genes. *Mol Cell*, 4, 735-743.
Kulozik, A.E., Bellan-Koch, A., Bail, S., Kohne, E. and Kleihauer, E. (1991) Thalassemia intermedia: moderate reduction of beta globin gene transcriptional activity by a novel mutation of the proximal CACCC promoter element. *Blood*, 77, 2054-2058.
Kumar, P.P., Bischof, O., Purbey, P.K., Notani, D., Urlaub, H., Dejean, A. and Galande, S. (2007) Functional interaction between PML and SATB1 regulates chromatin-loop architecture and transcription of the MHC class I locus. *Nat Cell Biol*, 9, 45-56.
Kuras, L., Borggrefe, T. and Kornberg, R.D. (2003) Association of the Mediator complex with enhancers of active genes. *Proc Natl Acad Sci U S A*, 100, 13887-13891.
Kuras, L. and Struhl, K. (1999) Binding of TBP to promoters in vivo is stimulated by activators and requires Pol II holoenzyme. *Nature*, 399, 609-613.
Kuroha, T., Takahashi, S., Komeno, T., Itoh, K., Nagasawa, T. and Yamamoto, M. (1998) Ablation of Nrf2 function does not increase the erythroid or megakaryocytic cell lineage dysfunction caused by p45 NF-E2 gene disruption. *J Biochem (Tokyo)*, 123, 376-379.
Kurukuti, S., Tiwari, V.K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., Lobanenkov, V., Reik, W. and Ohlsson, R. (2006) CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc Natl Acad Sci U S A*, 103, 10684-10689.
Kwon, H., Imbalzano, A.N., Khavari, P.A., Kingston, R.E. and Green, M.R. (1994) Nucleosome disruption and enhancement of activator binding by a human SW1/SNF complex. *Nature*, 370, 477-481.
Lachner, M. and Jenuwein, T. (2002) The many faces of histone lysine methylation. *Curr Opin Cell Biol*, 14, 286-298.
Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. and Jenuwein, T. (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature*, 410, 116-120.
Lee, C.H., Murphy, M.R., Lee, J.S. and Chung, J.H. (1999) Targeting a SWI/SNF-related chromatin remodeling complex to the beta-globin promoter in erythroid cells. *Proc Natl Acad Sci U S A*, 96, 12311-12315.
Lee, D.H. and Schleif, R.F. (1989) In vivo DNA loops in araCBAD: size limits and helical repeat. *Proc Natl Acad Sci U S A*, 86, 476-480.
Lemon, B. and Tjian, R. (2000) Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev*, 14, 2551-2569.
Leonard, M., Brice, M., Engel, J.D. and Papayannopoulou, T. (1993) Dynamics of GATA transcription factor expression during erythroid differentiation. *Blood*, 82, 1071-1079.
Lettice, L.A., Heaney, S.J., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E. and de Graaff, E. (2003) A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet*, 12, 1725-1735.
Lewis, J.D. and Tollervey, D. (2000) Like attracts like: getting RNA processing together in the nucleus. *Science*, 288, 1385-1389.
Li, G., Sudlow, G. and Belmont, A.S. (1998a) Interphase cell cycle dynamics of a late-replicating, heterochromatic homogeneously staining region: precise choreography of condensation/decondensation and nuclear positioning. *J Cell Biol*, 140, 975-989.
Li, J., Noguchi, C.T., Miller, W., Hardison, R. and Schechter, A.N. (1998b) Multiple regulatory elements in the 5'-flanking sequence of the human epsilon-globin gene. *J Biol Chem*, 273, 10202-10209.
Li, Q., Barkess, G. and Qian, H. (2006) Chromatin looping and the probability of transcription. *Trends Genet*, 22, 197-202.
Li, Q., Peterson, K.R., Fang, X. and Stamatoyannopoulos, G. (2002a) Locus control regions. *Blood*, 100, 3077-3086.
Li, Q. and Stamatoyannopoulos, G. (1994) Hypersensitive site 5 of the human beta locus control region functions as a chromatin insulator. *Blood*, 84, 1399-1401.
Li, Q., Zhang, M., Han, H., Rohde, A. and Stamatoyannopoulos, G. (2002b) Evidence that DNase I hypersensitive site 5 of the human beta-globin locus control region functions as a chromosomal insulator in transgenic mice. *Nucleic Acids Res*, 30, 2484-2491.
Li, X.Y., Virbasius, A., Zhu, X. and Green, M.R. (1999) Enhancement of TBP binding by activators and general transcription factors. *Nature*, 399, 605-609.
Lin, P.S., Dubois, M.F. and Dahmus, M.E. (2002) TFIIF-associating carboxyl-terminal domain phosphatase dephosphorylates phosphoserines 2 and 5 of RNA polymerase II. *J Biol Chem*, 277, 45949-45956.
Ling, J., Ainol, L., Zhang, L., Yu, X., Pi, W. and Tuan, D. (2004) HS2 Enhancer Function Is Blocked by a Transcriptional Terminator Inserted between the Enhancer and the Promoter. *J Biol Chem*, 279, 51704-51713.
Ling, J.Q., Li, T., Hu, J.F., Vu, T.H., Chen, H.L., Qiu, X.W., Cherry, A.M. and Hoffman, A.R. (2006) CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science*, 312, 269-272.
Ling, K.W. and Dzierzak, E. (2002) Ontogeny and genetics of the hemato/lymphopoietic system. *Curr Opin Immunol*, 14, 186-191.
Litt, M.D., Simpson, M., Recillas-Targa, F., Prioleau, M.N. and Felsenfeld, G. (2001) Transitions in histone acetylation reveal boundaries of three separately regulated neighboring loci. *Embo J*, 20, 2224-2235.
Liu, Q., Bungert, J. and Engel, J.D. (1997) Mutation of gene-proximal regulatory elements disrupts human epsilon-, gamma-, and beta-globin expression in yeast artificial chromosome transgenic mice. *Proc Natl Acad Sci U S A*, 94, 169-174.
Liu, Z. and Garrard, W.T. (2005) Long-Range Interactions between Three Transcriptional Enhancers, Active V{kappa} Gene Promoters, and a 3' Boundary Sequence Spanning 46 Kilobases. *Mol Cell Biol*, 25, 3220-3231.
Lomvardas, S., Barnea, G., Pisapia, D.J., Mendelsohn, M., Kirkland, J. and Axel, R. (2006) Interchromosomal interactions and olfactory receptor choice. *Cell*, 126, 403-413.
Lopes, S., Lewis, A., Hajkova, P., Dean, W., Oswald, J., Forne, T., Murrell, A., Constancia, M., Bartolomei, M., Walter, J. and Reik, W. (2003)

Epigenetic modifications in an imprinting cluster are controlled by a hierarchy of DMRs suggesting long-range chromatin interactions. *Hum Mol Genet*, 12, 295-305.
Lu, B.Y., Emtage, P.C., Duyf, B.J., Hilliker, A.J. and Eissenberg, J.C. (2000) Heterochromatin protein 1 is required for the normal expression of two heterochromatin genes in Drosophila. *Genetics*, 155, 699-708.
Lu, S.J., Rowan, S., Bani, M.R. and Ben-David, Y. (1994) Retroviral integration within the Fli-2 locus results in inactivation of the erythroid transcription factor NF-E2 in Friend erythroleukemias: evidence that NF-E2 is essential for globin expression. *Proc Natl Acad Sci U S A*, 91, 8398-8402.
Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. and Richmond, T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, 389, 251-260.
Magram, J., Chada, K. and Costantini, F. (1985) Developmental regulation of a cloned adult beta-globin gene in transgenic mice. *Nature*, 315, 338-340.
Mahmoudi, T., Katsani, K.R. and Verrijzer, C.P. (2002) GAGA can mediate enhancer function in trans by linking two separate DNA molecules. *Embo J*, 21, 1775-1781.
Maillet, L., Boscheron, C., Gotta, M., Marcand, S., Gilson, E. and Gasser, S.M. (1996) Evidence for silencing compartments within the yeast nucleus: a role for telomere proximity and Sir protein concentration in silencer-mediated repression. *Genes Dev*, 10, 1796-1811.
Maison, C. and Almouzni, G. (2004) HP1 and the dynamics of heterochromatin maintenance. *Nat Rev Mol Cell Biol*, 5, 296-304.
Maison, C., Bailly, D., Peters, A.H., Quivy, J.P., Roche, D., Taddei, A., Lachner, M., Jenuwein, T. and Almouzni, G. (2002) Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nat Genet*, 30, 329-334.
Marenduzzo, D., Faro-Trindade, I. and Cook, P.R. (2007) What are the molecular ties that maintain genomic loops? *Trends Genet*., 23, 126-133.
Marenduzzo, D., Micheletti, C. and Cook, P.R. (2006) Entropy-driven genome organization. *Biophys J*, 90, 3712-3721.
Martin, D.I. and Orkin, S.H. (1990) Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. *Genes Dev*, 4, 1886-1898.
Martin, D.I., Zon, L.I., Mutter, G. and Orkin, S.H. (1990) Expression of an erythroid transcription factor in megakaryocytic and mast cell lineages. *Nature*, 344, 444-447.
Martin, F., van Deursen, J.M., Shivdasani, R.A., Jackson, C.W., Troutman, A.G. and Ney, P.A. (1998) Erythroid maturation and globin gene expression in mice with combined deficiency of NF-E2 and nrf-2. *Blood*, 91, 3459-3466.
Martou, G. and De Boni, U. (2000) Nuclear topology of murine, cerebellar Purkinje neurons: changes as a function of development. *Exp Cell Res*, 256, 131-139.
Matthews, K.S. and Nichols, J.C. (1998) Lactose repressor protein: functional properties and structure. *Prog Nucleic Acid Res Mol Biol*, 58, 127-164.
McDevitt, M.A., Shivdasani, R.A., Fujiwara, Y., Yang, H. and Orkin, S.H. (1997) A "knockdown" mutation created by cis-element gene targeting reveals the dependence of erythroid cell maturation on the level of transcription factor GATA-1. *Proc Natl Acad Sci U S A*, 94, 6781-6785.
Melnikova, L., Juge, F., Gruzdeva, N., Mazur, A., Cavalli, G. and Georgiev, P. (2004) Interaction between the GAGA factor and Mod(mdg4) proteins promotes insulator bypass in Drosophila. *Proc Natl Acad Sci U S A*, 101, 14806-14811.
Merika, M. and Orkin, S.H. (1995) Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Kruppel family proteins Sp1 and EKLF. *Mol Cell Biol*, 15, 2437-2447.
Merkenschlager, M., Amoils, S., Roldan, E., Rahemtulla, A., O'Connor, E., Fisher, A.G. and Brown, K.E. (2004) Centromeric repositioning of coreceptor loci predicts their stable silencing and the CD4/CD8 lineage choice. *J Exp Med*, 200, 1437-1444.
Mignotte, V., Wall, L., deBoer, E., Grosveld, F. and Romeo, P.H. (1989) Two tissue-specific factors bind the erythroid promoter of the human porphobilinogen deaminase gene. *Nucleic Acids Res*, 17, 37-54.
Miller, I.J. and Bieker, J.J. (1993) A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. *Mol Cell Biol*, 13, 2776-2786.
Milot, E., Strouboulis, J., Trimborn, T., Wijgerde, M., de Boer, E., Langeveld, A., Tan-Un, K., Vergeer, W., Yannoutsos, N., Grosveld, F. and Fraser, P. (1996) Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell*, 87, 105-114.
Misteli, T. (2001a) The concept of self-organization in cellular architecture. *J Cell Biol*, 155, 181-185.
Misteli, T. (2001b) Protein dynamics: implications for nuclear architecture and gene expression. *Science*, 291, 843-847.
Misteli, T. (2007) Beyond the sequence: cellular organization of genome function. *Cell*, 128, 787-800.
Mitchison, T.J. (1992) Self-organization of polymer-motor systems in the cytoskeleton. *Philos Trans R Soc Lond B Biol Sci*, 336, 99-106.
Moi, P., Chan, K., Asunis, I., Cao, A. and Kan, Y.W. (1994) Isolation of NF-E2-related factor 2 (Nrf2), a NF-E2-like basic leucine zipper transcriptional activator that binds to the tandem NF-E2/AP1 repeat of the beta-globin locus control region. *Proc Natl Acad Sci U S A*, 91, 9926-9930.
Moon, A.M. and Ley, T.J. (1990) Conservation of the primary structure, organization, and function of the human and mouse beta-globin locus-activating regions. *Proc Natl Acad Sci U S A*, 87, 7693-7697.
Morcillo, P., Rosen, C., Baylies, M.K. and Dorsett, D. (1997) Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in Drosophila. *Genes Dev*, 11, 2729-2740.
Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P. and Chambon, P. (1981) The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res*, 9, 6047-6068.
Motohashi, H., O'Connor, T., Katsuoka, F., Engel, J.D. and Yamamoto, M. (2002) Integration and diversity of the regulatory network composed of Maf and CNC families of transcription factors. *Gene*, 294, 1-12.
Motohashi, H., Shavit, J.A., Igarashi, K., Yamamoto, M. and Engel, J.D. (1997) The world according to Maf. *Nucleic Acids Res*, 25, 2953-2959.
Mueller-Storm, H.P., Sogo, J.M. and Schaffner, W. (1989) An enhancer stimulates transcription in trans when attached to the promoter via a protein bridge. *Cell*, 58, 767-777.
Muller, F. and Tora, L. (2004) The multicoloured world of promoter recognition complexes. *Embo J*, 23, 2-8.
Muller, H.P., Matthias, P. and Schaffner, W. (1990) A transcriptional terminator between enhancer and promoter does not affect remote

transcriptional control. *Somat Cell Mol Genet*, 16, 351-360.
Muller, W.G., Rieder, D., Karpova, T.S., John, S., Trajanoski, Z. and McNally, J.G. (2007) Organization of chromatin and histone modifications at a transcription site. *J Cell Biol*, 177, 957-967.
Munkel, C., Eils, R., Dietzel, S., Zink, D., Mehring, C., Wedemann, G., Cremer, T. and Langowski, J. (1999) Compartmentalization of interphase chromosomes observed in simulation and experiment. *J Mol Biol*, 285, 1053-1065.
Muravyova, E., Golovnin, A., Gracheva, E., Parshikov, A., Belenkaya, T., Pirrotta, V. and Georgiev, P. (2001) Loss of insulator activity by paired Su(Hw) chromatin insulators. *Science*, 291, 495-498.
Murrell, A., Heeson, S., Bowden, L., Constancia, M., Dean, W., Kelsey, G. and Reik, W. (2001) An intragenic methylated region in the imprinted Igf2 gene augments transcription. *EMBO Rep*, 2, 1101-1106.
Murrell, A., Heeson, S. and Reik, W. (2004) Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. *Nat Genet*, 36, 889-893.
Naar, A.M., Taatjes, D.J., Zhai, W., Nogales, E. and Tjian, R. (2002) Human CRSP interacts with RNA polymerase II CTD and adopts a specific CTD-bound conformation. *Genes Dev*, 16, 1339-1344.
Narlikar, G.J., Fan, H.Y. and Kingston, R.E. (2002) Cooperation between complexes that regulate chromatin structure and transcription. *Cell*, 108, 475-487.
Navas, P.A., Peterson, K.R., Li, Q., Skarpidi, E., Rohde, A., Shaw, S.E., Clegg, C.H., Asano, H. and Stamatoyannopoulos, G. (1998) Developmental specificity of the interaction between the locus control region and embryonic or fetal globin genes in transgenic mice with an HS3 core deletion. *Mol Cell Biol*, 18, 4188-4196.
Ney, P.A., Andrews, N.C., Jane, S.M., Safer, B., Purucker, M.E., Weremowicz, S., Morton, C.C., Goff, S.C., Orkin, S.H. and Nienhuis, A.W. (1993) Purification of the human NF-E2 complex: cDNA cloning of the hematopoietic cell-specific subunit and evidence for an associated partner. *Mol Cell Biol*, 13, 5604-5612.
Ney, P.A., Sorrentino, B.P., Lowrey, C.H. and Nienhuis, A.W. (1990a) Inducibility of the HS II enhancer depends on binding of an erythroid specific nuclear protein. *Nucleic Acids Res*, 18, 6011-6017.
Ney, P.A., Sorrentino, B.P., McDonagh, K.T. and Nienhuis, A.W. (1990b) Tandem AP-1-binding sites within the human beta-globin dominant control region function as an inducible enhancer in erythroid cells. *Genes Dev*, 4, 993-1006.
Nielsen, A.L., Oulad-Abdelghani, M., Ortiz, J.A., Remboutsika, E., Chambon, P. and Losson, R. (2001) Heterochromatin formation in mammalian cells: interaction between histones and HP1 proteins. *Mol Cell*, 7, 729-739.
Nuez, B., Michalovich, D., Bygrave, A., Ploemacher, R. and Grosveld, F. (1995) Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature*, 375, 316-318.
Ogbourne, S. and Antalis, T.M. (1998) Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem J*, 331 (Pt 1), 1-14.
Okano, M., Bell, D.W., Haber, D.A. and Li, E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, 99, 247-257.
Oki, M., Valenzuela, L., Chiba, T., Ito, T. and Kamakaka, R.T. (2004) Barrier proteins remodel and modify chromatin to restrict silenced domains. *Mol Cell Biol*, 24, 1956-1967.
Orkin, S.H., Antonarakis, S.E. and Kazazian, H.H., Jr. (1984) Base substitution at position -88 in a beta-thalassemic globin gene. Further evidence for the role of distal promoter element ACACCC. *J Biol Chem*, 259, 8679-8681.
Orlando, V., Strutt, H. and Paro, R. (1997) Analysis of chromatin structure by in vivo formaldehyde cross-linking. *Methods*, 11, 205-214.
Orphanides, G., Lagrange, T. and Reinberg, D. (1996) The general transcription factors of RNA polymerase II. *Genes Dev*, 10, 2657-2683.
Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W. and Fraser, P. (2004) Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet*, 36, 1065-1071.
Ostermeier, G.C., Liu, Z., Martins, R.P., Bharadwaj, R.R., Ellis, J., Draghici, S. and Krawetz, S.A. (2003) Nuclear matrix association of the human beta-globin locus utilizing a novel approach to quantitative real-time PCR. *Nucleic Acids Res*, 31, 3257-3266.
O'Sullivan, J.M., Tan-Wong, S.M., Morillon, A., Lee, B., Coles, J., Mellor, J. and Proudfoot, N.J. (2004) Gene loops juxtapose promoters and terminators in yeast. *Nat Genet*, 36, 1014-1018.
Otero, G., Fellows, J., Li, Y., de Bizemont, T., Dirac, A.M., Gustafsson, C.M., Erdjument-Bromage, H., Tempst, P. and Svejstrup, J.Q. (1999) Elongator, a multisubunit component of a novel RNA polymerase II holoenzyme for transcriptional elongation. *Mol Cell*, 3, 109-118.
Oyake, T., Itoh, K., Motohashi, H., Hayashi, N., Hoshino, H., Nishizawa, M., Yamamoto, M. and Igarashi, K. (1996) Bach proteins belong to a novel family of BTB-basic leucine zipper transcription factors that interact with MafK and regulate transcription through the NF-E2 site. *Mol Cell Biol*, 16, 6083-6095.
Pabo, C.O. and Sauer, R.T. (1992) Transcription factors: structural families and principles of DNA recognition. *Annu Rev Biochem*, 61, 1053-1095.
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.
Pan, X., Ohneda, O., Ohneda, K., Lindeboom, F., Iwata, F., Shimizu, R., Nagano, M., Suwabe, N., Philipsen, S., Lim, K.C., Engel, J.D. and Yamamoto, M. (2005) Graded levels of GATA-1 expression modulate survival, proliferation, and differentiation of erythroid progenitors. *J Biol Chem*, 280, 22385-22394.
Pant, V., Mariano, P., Kanduri, C., Mattsson, A., Lobanenkov, V., Heuchel, R. and Ohlsson, R. (2003) The nucleotides responsible for the direct physical contact between the chromatin insulator protein CTCF and the H19 imprinting control region manifest parent of origin-specific long-distance insulation and methylation-free domains. *Genes Dev*, 17, 586-590.
Parada, L.A., McQueen, P.G. and Misteli, T. (2004) Tissue-specific spatial organization of genomes. *Genome Biol*, 5, R44.
Parseghian, M.H. and Hamkalo, B.A. (2001) A compendium of the histone H1 family of somatic subtypes: an elusive cast of characters and their characteristics. *Biochem Cell Biol*, 79, 289-304.
Patrinos, G.P., de Krom, M., de Boer, E., Langeveld, A., Imam, A.M., Strouboulis, J., de Laat, W. and Grosveld, F.G. (2004) Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. *Genes Dev*, 18, 1495-1509.
Pedersen, T.A., Kowenz-Leutz, E., Leutz, A. and Nerlov, C. (2001) Cooperation between C/EBPalpha TBP/TFIIB and SWI/SNF recruiting domains is required for adipocyte differentiation. *Genes Dev*, 15, 3208-3216.
Perkins, A.C., Gaensler, K.M. and Orkin, S.H. (1996) Silencing of human fetal globin expression is impaired in the absence of the adult

beta-globin gene activator protein EKLF. *Proc Natl Acad Sci U S A*, 93, 12267-12271.
Perkins, A.C., Sharpe, A.H. and Orkin, S.H. (1995) Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature*, 375, 318-322.
Perry, C. and Soreq, H. (2002) Transcriptional regulation of erythropoiesis. Fine tuning of combinatorial multi-domain elements. *Eur J Biochem*, 269, 3607-3618.
Peterson, C.L. (2002) HDAC's at work: everyone doing their part. *Mol Cell*, 9, 921-922.
Peterson, K.R. and Stamatoyannopoulos, G. (1993) Role of gene order in developmental control of human gamma- and beta-globin gene expression. *Mol Cell Biol*, 13, 4836-4843.
Pevny, L., Lin, C.S., D'Agati, V., Simon, M.C., Orkin, S.H. and Costantini, F. (1995) Development of hematopoietic cells lacking transcription factor GATA-1. *Development*, 121, 163-172.
Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.F., D'Agati, V., Orkin, S.H. and Costantini, F. (1991) Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature*, 349, 257-260.
Piacentini, L., Fanti, L., Berloco, M., Perrini, B. and Pimpinelli, S. (2003) Heterochromatin protein 1 (HP1) is associated with induced gene expression in Drosophila euchromatin. *J Cell Biol*, 161, 707-714.
Pikaart, M.J., Recillas-Targa, F. and Felsenfeld, G. (1998) Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. *Genes Dev*, 12, 2852-2862.
Prioleau, M.N., Nony, P., Simpson, M. and Felsenfeld, G. (1999) An insulator element and condensed chromatin region separate the chicken beta-globin locus from an independently regulated erythroid-specific folate receptor gene. *Embo J*, 18, 4035-4048.
Proudfoot, N.J., Furger, A. and Dye, M.J. (2002) Integrating mRNA processing with transcription. *Cell*, 108, 501-512.
Ptashne, M. (1986) Gene regulation by proteins acting nearby and at a distance. *Nature*, 322, 697-701.
Ptashne, M. and Gann, A. (1997) Transcriptional activation by recruitment. *Nature*, 386, 569-577.
Quina, A.S. and Parreira, L. (2005) Telomere-surrounding regions are transcription-permissive 3D nuclear compartments in human cells. *Exp Cell Res*, 307, 52-64.
Raich, N., Clegg, C.H., Grofti, J., Romeo, P.H. and Stamatoyannopoulos, G. (1995) GATA1 and YY1 are developmental repressors of the human epsilon-globin gene. *Embo J*, 14, 801-809.
Raich, N., Enver, T., Nakamoto, B., Josephson, B., Papayannopoulou, T. and Stamatoyannopoulos, G. (1990) Autonomous developmental control of human embryonic globin gene switching in transgenic mice. *Science*, 250, 1147-1149.
Raich, N., Papayannopoulou, T., Stamatoyannopoulos, G. and Enver, T. (1992) Demonstration of a human epsilon-globin gene silencer with studies in transgenic mice. *Blood*, 79, 861-864.
Raich, N. and Romeo, P.H. (1993) Erythroid regulatory elements. *Stem Cells*, 11, 95-104.
Recillas-Targa, F., Bell, A.C. and Felsenfeld, G. (1999) Positional enhancer-blocking activity of the chicken beta-globin insulator in transiently transfected cells. *Proc Natl Acad Sci U S A*, 96, 14354-14359.
Recillas-Targa, F., Pikaart, M.J., Burgess-Beusse, B., Bell, A.C., Litt, M.D., West, A.G., Gaszner, M. and Felsenfeld, G. (2002) Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc Natl Acad Sci U S A*, 99, 6883-6888.
Reik, A., Telling, A., Zitnik, G., Cimbora, D., Epner, E. and Groudine, M. (1998) The locus control region is necessary for gene expression in the human beta-globin locus but not the maintenance of an open chromatin structure in erythroid cells. *Mol Cell Biol*, 18, 5992-6000.
Reitman, M., Lee, E., Westphal, H. and Felsenfeld, G. (1993) An enhancer/locus control region is not sufficient to open chromatin. *Mol Cell Biol*, 13, 3990-3998.
Rekhtman, N., Radparvar, F., Evans, T. and Skoultchi, A.I. (1999) Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells. *Genes Dev*, 13, 1398-1411.
Richards, E.J. and Elgin, S.C. (2002) Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell*, 108, 489-500.
Ringrose, L., Chabanis, S., Angrand, P.O., Woodroofe, C. and Stewart, A.F. (1999) Quantitative comparison of DNA looping in vitro and in vivo: chromatin increases effective DNA flexibility at short distances. *Embo J*, 18, 6630-6641.
Rippe, K. (2001) Making contacts on a nucleic acid polymer. *Trends Biochem Sci*, 26, 733-740.
Rodriguez, P., Bonte, E., Krijgsveld, J., Kolodziej, K.E., Guyot, B., Heck, A.J., Vyas, P., de Boer, E., Grosveld, F. and Strouboulis, J. (2005) GATA-1 forms distinct activating and repressive complexes in erythroid cells. *Embo J*, 24, 2354-2366.
Romeo, P.H., Prandini, M.H., Joulin, V., Mignotte, V., Prenant, M., Vainchenker, W., Marguerie, G. and Uzan, G. (1990) Megakaryocytic and erythrocytic lineages share specific transcription factors. *Nature*, 344, 447-449.
Rozenblatt-Rosen, O., Rozovskaia, T., Burakov, D., Sedkov, Y., Tillib, S., Blechman, J., Nakamura, T., Croce, C.M., Mazo, A. and Canaani, E. (1998) The C-terminal SET domains of ALL-1 and TRITHORAX interact with the INI1 and SNR1 proteins, components of the SWI/SNF complex. *Proc Natl Acad Sci U S A*, 95, 4152-4157.
Sabatino, D.E., Cline, A.P., Gallagher, P.G., Garrett, L.J., Stamatoyannopoulos, G., Forget, B.G. and Bodine, D.M. (1998) Substitution of the human beta-spectrin promoter for the human agamma-globin promoter prevents silencing of a linked human beta-globin gene in transgenic mice. *Mol Cell Biol*, 18, 6634-6640.
Sachs, R.K., van den Engh, G., Trask, B., Yokota, H. and Hearst, J.E. (1995) A random-walk/giant-loop model for interphase chromosomes. *Proc Natl Acad Sci U S A*, 92, 2710-2714.
Saitoh, N., Bell, A.C., Recillas-Targa, F., West, A.G., Simpson, M., Pikaart, M. and Felsenfeld, G. (2000) Structural and functional conservation at the boundaries of the chicken beta-globin domain. *Embo J*, 19, 2315-2322.
Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y. and Hume, D.A. (2007) Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat Rev Genet*, 8, 424-436.
Santoso, B., Ortiz, B.D. and Winoto, A. (2000) Control of organ-specific demethylation by an element of the T-cell receptor-alpha locus control region. *J Biol Chem*, 275, 1952-1958.
Sawado, T., Halow, J., Bender, M.A. and Groudine, M. (2003) The beta -globin locus control region (LCR) functions primarily by enhancing the transition from transcription initiation to elongation. *Genes Dev*, 17, 1009-1018.
Sawado, T., Igarashi, K. and Groudine, M. (2001) Activation of beta-major globin gene transcription is associated with recruitment of NF-E2 to the beta-globin LCR and gene promoter. *Proc Natl Acad Sci U S A*, 98, 10226-10231.
Sayegh, C., Jhunjhunwala, S., Riblet, R. and Murre, C. (2005) Visualization of looping involving the immunoglobulin heavy-chain locus in

developing B cells. *Genes Dev*, 19, 322-327.
Scherthan, H., Weich, S., Schwegler, H., Heyting, C., Harle, M. and Cremer, T. (1996) Centromere and telomere movements during early meiotic prophase of mouse and man are associated with the onset of chromosome pairing. *J Cell Biol*, 134, 1109-1125.
Schoenherr, C.J., Levorse, J.M. and Tilghman, S.M. (2003) CTCF maintains differential methylation at the Igf2/H19 locus. *Nat Genet*, 33, 66-69.
Schotta, G., Ebert, A., Krauss, V., Fischer, A., Hoffmann, J., Rea, S., Jenuwein, T., Dorn, R. and Reuter, G. (2002) Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. *Embo J*, 21, 1121-1131.
Schubeler, D., Francastel, C., Cimbora, D.M., Reik, A., Martin, D.I. and Groudine, M. (2000) Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human beta-globin locus. *Genes Dev*, 14, 940-950.
Schubeler, D., Groudine, M. and Bender, M.A. (2001) The murine beta-globin locus control region regulates the rate of transcription but not the hyperacetylation of histones at the active genes. *Proc Natl Acad Sci U S A*, 98, 11432-11437.
Serizawa, S., Miyamichi, K., Nakatani, H., Suzuki, M., Saito, M., Yoshihara, Y. and Sakano, H. (2003) Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. *Science*, 302, 2088-2094.
Serizawa, S., Miyamichi, K. and Sakano, H. (2004) One neuron-one receptor rule in the mouse olfactory system. *Trends Genet*, 20, 648-653.
Shi, X., Hong, T., Walter, K.L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Pena, P., Lan, F., Kaadige, M.R., Lacoste, N., Cayrou, C., Davrazou, F., Saha, A., Cairns, B.R., Ayer, D.E., Kutateladze, T.G., Shi, Y., Cote, J., Chua, K.F. and Gozani, O. (2006) ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature*, 442, 96-99.
Shih, D.M., Wall, R.J. and Shapiro, S.G. (1993) A 5' control region of the human epsilon-globin gene is sufficient for embryonic specificity in transgenic mice. *J Biol Chem*, 268, 3066-3071.
Shivdasani, R.A. and Orkin, S.H. (1995) Erythropoiesis and globin gene expression in mice lacking the transcription factor NF-E2. *Proc Natl Acad Sci U S A*, 92, 8690-8694.
Shivdasani, R.A., Rosenblatt, M.F., Zucker-Franklin, D., Jackson, C.W., Hunt, P., Saris, C.J. and Orkin, S.H. (1995) Transcription factor NF-E2 is required for platelet formation independent of the actions of thrombopoietin/MGDF in megakaryocyte development. *Cell*, 81, 695-704.
Shopland, L.S., Lynch, C.R., Peterson, K.A., Thornton, K., Kepper, N., Hase, J., Stein, S., Vincent, S., Molloy, K.R., Kreth, G., Cremer, C., Bult, C.J. and O'Brien, T.P. (2006) Folding and organization of a contiguous chromosome region according to the gene distribution pattern in primary genomic sequence. *J Cell Biol*, 174, 27-38.
Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B. and de Laat, W. (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet*, 38, 1348-1354.
Sims, R.J., 3rd, Chen, C.F., Santos-Rosa, H., Kouzarides, T., Patel, S.S. and Reinberg, D. (2005) Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *J Biol Chem*, 280, 41789-41792.
Skok, J.A., Brown, K.E., Azuara, V., Caparros, M.L., Baxter, J., Takacs, K., Dillon, N., Gray, D., Perry, R.P., Merkenschlager, M. and Fisher, A.G. (2001) Nonequivalent nuclear location of immunoglobulin alleles in B lymphocytes. *Nat Immunol*, 2, 848-854.
Skok, J.A., Gisler, R., Novatchkova, M., Farmer, D., de Laat, W. and Busslinger, M. (2007) Reversible contraction by looping of the Tcra and Tcrb loci in rearranging thymocytes. *Nat Immunol*., 8, 378-387.
Smale, S.T. (2001) Core promoters: active contributors to combinatorial gene regulation. *Genes Dev*, 15, 2503-2508.
Solomon, M.J. and Varshavsky, A. (1985) Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. *Proc Natl Acad Sci U S A*, 82, 6470-6474.
Somma, M.P., Pisano, C. and Lavia, P. (1991) The housekeeping promoter from the mouse CpG island HTF9 contains multiple protein-binding elements that are functionally redundant. *Nucleic Acids Res*, 19, 2817-2824.
Southwood, C.M., Downs, K.M. and Bieker, J.J. (1996) Erythroid Kruppel-like factor exhibits an early and sequentially localized pattern of expression during mammalian erythroid ontogeny. *Dev Dyn*, 206, 248-259.
Spilianakis, C.G. and Flavell, R.A. (2004) Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat Immunol*, 5, 1017-1027.
Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R. and Flavell, R.A. (2005) Interchromosomal associations between alternatively expressed loci. *Nature*, 435, 637-645.
Splinter, E., Grosveld, F. and de Laat, W. (2004) 3C technology: analyzing the spatial organization of genomic loci in vivo. *Methods Enzymol*, 375, 493-507.
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.
Stamatoyannopoulos G. and Grosveld F. (2001) Hemoglobin switching. In *The molecular basis of blood diseases* (ed. G. Stamatoyannopoulos, P. Majerus, R. Perlmutter and H. Varmus), pp. 135-182. W.B. Saunders, Philidelphia.
Sterner, D.E. and Berger, S.L. (2000) Acetylation of histones and transcription-related factors. *Microbiol Mol Biol Rev*, 64, 435-459.
Strahl, B.D. and Allis, C.D. (2000) The language of covalent histone modifications. *Nature*, 403, 41-45.
Strouboulis, J., Dillon, N. and Grosveld, F. (1992) Developmental regulation of a complete 70-kb human beta-globin locus in transgenic mice. *Genes Dev*, 6, 1857-1864.
Sun, H.B., Shen, J. and Yokota, H. (2000) Size-dependent positioning of human chromosomes in interphase nuclei. *Biophys J*, 79, 184-190.
Sun, J., Brand, M., Zenke, Y., Tashiro, S., Groudine, M. and Igarashi, K. (2004) Heme regulates the dynamic exchange of Bach1 and NF-E2-related factors in the Maf transcription factor network. *Proc Natl Acad Sci U S A*, 101, 1461-1466.
Svejstrup, J.Q. (2002) Chromatin elongation factors. *Curr Opin Genet Dev*, 12, 156-161.
Svejstrup, J.Q. (2004) The RNA polymerase II transcription cycle: cycling through chromatin. *Biochim Biophys Acta*, 1677, 64-73.
Takahashi, N., Kawada, T., Yamamoto, T., Goto, T., Taimatsu, A., Aoki, N., Kawasaki, H., Taira, K., Yokoyama, K.K., Kamei, Y. and Fushiki, T. (2002) Overexpression and ribozyme-mediated targeting of transcriptional coactivators CREB-binding protein and p300 revealed their indispensable roles in adipocyte differentiation through the regulation of peroxisome proliferator-activated receptor gamma. *J Biol Chem*, 277, 16906-16912.
Takahashi, S., Onodera, K., Motohashi, H., Suwabe, N., Hayashi, N., Yanai, N., Nabesima, Y. and Yamamoto, M. (1997) Arrest in primitive erythroid cell development caused by promoter-specific disruption of the GATA-1 gene. *J Biol Chem*, 272, 12611-12615.

Talbot, D. and Grosveld, F. (1991) The 5'HS2 of the globin locus control region enhances transcription through the interaction of a multimeric complex binding at two functionally distinct NF-E2 binding sites. *Embo J*, 10, 1391-1398.

Talbot, D., Philipsen, S., Fraser, P. and Grosveld, F. (1990) Detailed analysis of the site 3 region of the human beta-globin dominant control region. *Embo J*, 9, 2169-2177.

Tamaru, H. and Selker, E.U. (2001) A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. *Nature*, 414, 277-283.

Tanabe, O., Katsuoka, F., Campbell, A.D., Song, W., Yamamoto, M., Tanimoto, K. and Engel, J.D. (2002) An embryonic/fetal beta-type globin gene repressor contains a nuclear receptor TR2/TR4 heterodimer. *Embo J*, 21, 3434-3442.

Tanimoto, K., Liu, Q., Bungert, J. and Engel, J.D. (1999) Effects of altered gene order or orientation of the locus control region on human beta-globin gene expression in mice. *Nature*, 398, 344-348.

Tanimoto, K., Liu, Q., Grosveld, F., Bungert, J. and Engel, J.D. (2000) Context-dependent EKLF responsiveness defines the developmental specificity of the human epsilon-globin gene in erythroid cells of YAC transgenic mice. *Genes Dev*, 14, 2778-2794.

Tewari, R., Gillemans, N., Harper, A., Wijgerde, M., Zafarana, G., Drabek, D., Grosveld, F. and Philipsen, S. (1996) The human beta-globin locus control region confers an early embryonic erythroid-specific expression pattern to a basic promoter driving the bacterial lacZ gene. *Development*, 122, 3991-3999.

Tewari, R., Gillemans, N., Wijgerde, M., Nuez, B., von Lindern, M., Grosveld, F. and Philipsen, S. (1998) Erythroid Kruppel-like factor (EKLF) is active in primitive and definitive erythroid cells and is required for the function of 5'HS3 of the beta-globin locus control region. *Embo J*, 17, 2334-2341.

Tham, W.H. and Zakian, V.A. (2002) Transcriptional silencing at Saccharomyces telomeres: implications for other organisms. *Oncogene*, 21, 512-521.

Thomson, I., Gilchrist, S., Bickmore, W.A. and Chubb, J.R. (2004) The radial positioning of chromatin is not inherited through mitosis but is established de novo in early G1. *Curr Biol*, 14, 166-172.

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

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

Townes, T.M., Lingrel, J.B., Chen, H.Y., Brinster, R.L. and Palmiter, R.D. (1985) Erythroid-specific expression of human beta-globin genes in transgenic mice. *Embo J*, 4, 1715-1723.

Tremethick, D.J. (2007) Higher-Order Structures of Chromatin: The Elusive 30 nm Fiber. *Cell*, 128, 651-654.

Trimborn, T., Gribnau, J., Grosveld, F. and Fraser, P. (1999) Mechanisms of developmental control of transcription in the murine alpha- and beta-globin loci. *Genes Dev*, 13, 112-124.

Tsai, S.F., Martin, D.I., Zon, L.I., D'Andrea, A.D., Wong, G.G. and Orkin, S.H. (1989) Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. *Nature*, 339, 446-451.

Tsang, A.P., Visvader, J.E., Turner, C.A., Fujiwara, Y., Yu, C., Weiss, M.J., Crossley, M. and Orkin, S.H. (1997) FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell*, 90, 109-119.

Tse, C., Sera, T., Wolffe, A.P. and Hansen, J.C. (1998) Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III. *Mol Cell Biol*, 18, 4629-4638.

Tuan, D., Kong, S. and Hu, K. (1992) Transcription of the hypersensitive site HS2 enhancer in erythroid cells. *Proc Natl Acad Sci U S A*, 89, 11219-11223.

Tuan, D., Solomon, W., Li, Q. and London, I.M. (1985) The "beta-like-globin" gene domain in human erythroid cells. *Proc Natl Acad Sci U S A*, 82, 6384-6388.

Tuan, D.Y., Solomon, W.B., London, I.M. and Lee, D.P. (1989) An erythroid-specific, developmental-stage-independent enhancer far upstream of the human "beta-like globin" genes. *Proc Natl Acad Sci U S A*, 86, 2554-2558.

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

Vakoc, C.R., Letting, D.L., Gheldof, N., Sawado, T., Bender, M.A., Groudine, M., Weiss, M.J., Dekker, J. and Blobel, G.A. (2005a) Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol Cell*, 17, 453-462.

Vakoc, C.R., Mandat, S.A., Olenchock, B.A. and Blobel, G.A. (2005b) Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. *Mol Cell*, 19, 381-391.

Verschure, P.J., van der Kraan, I., Manders, E.M., Hoogstraten, D., Houtsmuller, A.B. and van Driel, R. (2003) Condensed chromatin domains in the mammalian nucleus are accessible to large macromolecules. *EMBO Rep*, 4, 861-866.

Verschure, P.J., van Der Kraan, I., Manders, E.M. and van Driel, R. (1999) Spatial relationship between transcription sites and chromosome territories. *J Cell Biol*, 147, 13-24.

Vieira, K.F., Levings, P.P., Hill, M.A., Crusselle, V.J., Kang, S.H., Engel, J.D. and Bungert, J. (2004) Recruitment of transcription complexes to the beta -globin gene locus in vivo and in vitro. *J Biol Chem*., 279, 50350-50357.

von Lindern, M., Deiner, E.M., Dolznig, H., Parren-Van Amelsvoort, M., Hayman, M.J., Mullner, E.W. and Beug, H. (2001) Leukemic transformation of normal murine erythroid progenitors: v- and c-ErbB act through signaling pathways activated by the EpoR and c-Kit in stress erythropoiesis. *Oncogene*, 20, 3651-3664.

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.

Wansink, D.G., Schul, W., van der Kraan, I., van Steensel, B., van Driel, R. and de Jong, L. (1993) Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. *J Cell Biol*, 122, 283-293.

Wasylyk, B., Wasylyk, C., Augereau, P. and Chambon, P. (1983) The SV40 72 bp repeat preferentially potentiates transcription starting from proximal natural or substitute promoter elements. *Cell*, 32, 503-514.

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

Weierich, C., Brero, A., Stein, S., von Hase, J., Cremer, C., Cremer, T. and Solovei, I. (2003) Three-dimensional arrangements of centromeres and telomeres in nuclei of human and murine lymphocytes. *Chromosome Res*, 11, 485-502.

Weiler, K.S. and Wakimoto, B.T. (1995) Heterochromatin and gene expression in Drosophila. *Annu Rev Genet*, 29, 577-605.

Weintraub, H. and Groudine, M. (1976) Chromosomal subunits in active genes have an altered conformation. *Science*, 193, 848-856.
Weiss, M.J., Keller, G. and Orkin, S.H. (1994) Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. *Genes Dev*, 8, 1184-1197.
Weiss, M.J. and Orkin, S.H. (1995) Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis. *Proc Natl Acad Sci U S A*, 92, 9623-9627.
Weiss, M.J., Yu, C. and Orkin, S.H. (1997) Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. *Mol Cell Biol*, 17, 1642-1651.
Wen, J., Huang, S., Rogers, H., Dickinson, L.A., Kohwi-Shigematsu, T. and Noguchi, C.T. (2005) SATB1 family protein expressed during early erythroid differentiation modifies globin gene expression. *Blood*, 105, 3330-3339.
West, A.G. and Fraser, P. (2005) Remote control of gene transcription. *Hum Mol Genet*, 14 Suppl 1, R101-111.
West, A.G., Gaszner, M. and Felsenfeld, G. (2002) Insulators: many functions, many mechanisms. *Genes Dev*, 16, 271-288.
West, A.G., Huang, S., Gaszner, M., Litt, M.D. and Felsenfeld, G. (2004) Recruitment of Histone Modifications by USF Proteins at a Vertebrate Barrier Element. *Mol Cell*, 16, 453-463.
Whitelaw, E., Tsai, S.F., Hogben, P. and Orkin, S.H. (1990) Regulated expression of globin chains and the erythroid transcription factor GATA-1 during erythropoiesis in the developing mouse. *Mol Cell Biol*, 10, 6596-6606.
Whyatt, D.J., deBoer, E. and Grosveld, F. (1993) The two zinc finger-like domains of GATA-1 have different DNA binding specificities. *Embo J*, 12, 4993-5005.
Whyatt, D.J., Karis, A., Harkes, I.C., Verkerk, A., Gillemans, N., Elefanty, A.G., Vairo, G., Ploemacher, R., Grosveld, F. and Philipsen, S. (1997) The level of the tissue-specific factor GATA-1 affects the cell-cycle machinery. *Genes Funct*, 1, 11-24.
Wijgerde, M., Gribnau, J., Trimborn, T., Nuez, B., Philipsen, S., Grosveld, F. and Fraser, P. (1996) The role of EKLF in human beta-globin gene competition. *Genes Dev*, 10, 2894-2902.
Wijgerde, M., Grosveld, F. and Fraser, P. (1995) Transcription complex stability and chromatin dynamics in vivo. *Nature*, 377, 209-213.
Williams, C.J., Naito, T., Arco, P.G., Seavitt, J.R., Cashman, S.M., De Souza, B., Qi, X., Keables, P., Von Andrian, U.H. and Georgopoulos, K. (2004) The chromatin remodeler Mi-2beta is required for CD4 expression and T cell development. *Immunity*, 20, 719-733.
Williams, R.R., Azuara, V., Perry, P., Sauer, S., Dvorkina, M., Jorgensen, H., Roix, J., McQueen, P., Misteli, T., Merkenschlager, M. and Fisher, A.G. (2006) Neural induction promotes large-scale chromatin reorganisation of the Mash1 locus. *J Cell Sci*, 119, 132-140.
Wong, P.M., Chung, S.W., Chui, D.H. and Eaves, C.J. (1986) Properties of the earliest clonogenic hemopoietic precursors to appear in the developing murine yolk sac. *Proc Natl Acad Sci U S A*, 83, 3851-3854.
Wurtele, H. and Chartrand, P. (2006) Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology. *Chromosome Res*, 14, 477-495.
Xu, G.L., Bestor, T.H., Bourc'his, D., Hsieh, C.L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J. and Viegas-Pequignot, E. (1999) Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature*, 402, 187-191.
Yamada, T., Fischle, W., Sugiyama, T., Allis, C.D. and Grewal, S.I. (2005) The nucleation and maintenance of heterochromatin by a histone deacetylase in fission yeast. *Mol Cell*, 20, 173-185.
Yamamoto, M., Ko, L.J., Leonard, M.W., Beug, H., Orkin, S.H. and Engel, J.D. (1990) Activity and tissue-specific expression of the transcription factor NF-E1 multigene family. *Genes Dev*, 4, 1650-1662.
Yasuhara, J.C. and Wakimoto, B.T. (2006) Oxymoron no more: the expanding world of heterochromatic genes. *Trends Genet*, 22, 330-338.
Yoon, Y.S., Jeong, S., Rong, Q., Park, K.Y., Chung, J.H. and Pfeifer, K. (2007) Analysis of the H19ICR insulator. *Mol Cell Biol*, 27, 3499-3510.
Yoshida, C., Tokumasu, F., Hohmura, K.I., Bungert, J., Hayashi, N., Nagasawa, T., Engel, J.D., Yamamoto, M., Takeyasu, K. and Igarashi, K. (1999) Long range interaction of cis-DNA elements mediated by architectural transcription factor Bach1. *Genes Cells*, 4, 643-655.
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.
Yue, Z., Maldonado, E., Pillutla, R., Cho, H., Reinberg, D. and Shatkin, A.J. (1997) Mammalian capping enzyme complements mutant Saccharomyces cerevisiae lacking mRNA guanylyltransferase and selectively binds the elongating form of RNA polymerase II. *Proc Natl Acad Sci U S A*, 94, 12898-12903.
Yusufzai, T.M. and Felsenfeld, G. (2004) The 5'-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element. *Proc Natl Acad Sci U S A*, 101, 8620-8624.
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, W. and Bieker, J.J. (1998) Acetylation and modulation of erythroid Kruppel-like factor (EKLF) activity by interaction with histone acetyltransferases. *Proc Natl Acad Sci U S A*, 95, 9855-9860.
Zhang, W., Kadam, S., Emerson, B.M. and Bieker, J.J. (2001) Site-specific acetylation by p300 or CREB binding protein regulates erythroid Kruppel-like factor transcriptional activity via its interaction with the SWI-SNF complex. *Mol Cell Biol*, 21, 2413-2422.
Zhao, H. and Dean, A. (2004) An insulator blocks spreading of histone acetylation and interferes with RNA polymerase II transfer between an enhancer and gene. *Nucleic Acids Res*, 32, 4903-4919.
Zhao, Z., Tavoosidana, G., Sjolinder, M., Gondor, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K.S., Singh, U., Pant, V., Tiwari, V., Kurukuti, S. and Ohlsson, R. (2006) Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet*, 38, 1341-1347.
Zink, D., Amaral, M.D., Englmann, A., Lang, S., Clarke, L.A., Rudolph, C., Alt, F., Luther, K., Braz, C., Sadoni, N., Rosenecker, J. and Schindelhauer, D. (2004) Transcription-dependent spatial arrangements of CFTR and adjacent genes in human cell nuclei. *J Cell Biol*, 166, 815-825.
Zon, L.I., Yamaguchi, Y., Yee, K., Albee, E.A., Kimura, A., Bennett, J.C., Orkin, S.H. and Ackerman, S.J. (1993) Expression of mRNA for the GATA-binding proteins in human eosinophils and basophils: potential role in gene transcription. *Blood*, 81, 3234-3241.
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.

Summary

All genetic information required for the development and functioning of an organism is stored in billions of base pairs of deoxyribonucleic acid (or DNA). In eukaryotes, DNA is organised in large units called chromosomes that are located inside the cells nucleus. On these chromosomes reside functional units called genes, which encode for discrete hereditary characteristics, in most cases proteins. Multicellular organisms are composed of numerous different cell types, such as blood, skin, muscle and brain cells. This cellular diversity is extraordinary given the fact that all of these individual cell types harbour the same genetic material. The functional and morphological diversity among cells is dictated by the precise regulation of genes throughout development and different stages of cellular differentiation. Incorrect spatiotemporal activation and repression of genes can lead to cellular defects, which may eventually result in diseases such as cancer. Therefore, different mechanisms of control exist that regulate the transcription of genes in time and space, and understanding this transcriptional regulation in more detail is one of the key objectives in the field of molecular biology.

Different *cis*-regulatory DNA elements like promoters, enhancers and insulators are involved in the regulated expression of protein coding genes throughout development. An important question is how these distal *cis*-regulatory DNA elements communicate over distance to activate tissue- and developmental stage-specific gene expression.

In this thesis, the mouse β-globin locus is used as a model system to study transcriptional regulation in the context of the living nucleus. The genes of this locus all encode for the β-globin like protein that together with α-globin and heme forms a functional hemoglobin molecule, which is capable of binding oxygen and carbon dioxide in a reversible manner in red blood cells (erythrocytes). Besides the β-globin genes, the locus contains a number of well-characterised *cis*-regulatory DNA elements like enhancers, promoters and a locus control region (LCR). Previously, the structural organisation of the β-globin locus in its native context was studied by applying Chromosome Conformation Capture (3C) technology. These experiments showed that the locus adopts a red blood cell specific organisation in which *cis* regulatory elements of the LCR and the active β-globin genes spatially cluster. In addition to LCR-gene interactions these studies revealed erythroid specific long-range interactions with other sets of hypersensitive sites residing upstream and downstream of the β-globin locus. Collectively this spatial conformation of the β-globin locus was named the Active Chromatin Hub (ACH). However, little is known about the molecular events that accompany and underlie ACH formation in the β-globin locus during erythroid differentiation.

Chapter 3 of this thesis describes experiments in which recently established I/11 erythroid progenitor cells are used as a model system to study the molecular events involved in the establishment of contacts between the LCR and the β-globin genes. The results show that upon induction of I/11 cells, the β-globin like genes are transcribed at rates similar to those

observed *in vivo* and differentiation is accompanied by an increased ratio of positive versus negative chromatin modifications at the β_{major} promoter and the HSs of the LCR. In addition, the data show that binding of the erythroid-specific transcription factors GATA-1 and EKLF to the locus, while previously shown to be required, is not sufficient for ACH formation. Moreover, it was demonstrated that in p45 NF-E2 knockout mice long-range contacts in the β-globin locus are formed normally, indicating that the erythroid-specific transcription factor p45 NF-E2 is dispensable for ACH formation.

The role of the vertebrate insulator protein CTCF in long-range interactions in the β-globin locus is addressed in Chapter 4. CTCF binding sites flank the β-globin locus and have previously been shown to participate in erythroid specific spatial interactions in the context of the ACH. After conditional deletion of CTCF and targeted disruption of the 3'HS1 CTCF-binding site, long-range interactions in the β-globin locus were destabilised. However, β-globin gene transcription was not measurably affected by the loss of these interactions and disruption of CTCF binding only caused local loss of histone acetylation and gain of histone methylation. This data demonstrates that CTCF is directly involved in long-range chromatin looping in the β-globin locus and regulates the local balance between active and repressive chromatin marks.

3C technology enables researchers to study the structural organisation of individual loci at high resolution. Recently, a number of 3C based strategies have been developed that allow screening of the entire genome in an unbiased manner for DNA segments that physically interact with a DNA fragment of choice. Application of these new methods is expected to provide exciting new insights into the conformational structure of selected chromosomal regions in the genome. However, 3C and 4C based methods also have important limitations and preconditions that need to be recognised and addressed properly. Chapter 2 of this thesis therefore describes and evaluates the most commonly used 3C-based methods and addresses potentials and pitfalls of these technologies.

In Chapter 5 recently developed 4C technology was applied to study the nuclear organisation of both peri-centromeric and peri-telomeric chromosomal regions. The data show that there is a strong correlation between the transcriptional status of the selected peri-centromeric and peri-telomeric regions and their interacting loci in the interphase nucleus. Moreover, interchromosomal interactions with peri-centromeric regions were biased towards other peri-centromeric regions and also peri-telomeric regions showed a preference to interact *in trans* with other peri-telomeric regions. The results demonstrate that both the transcriptional status of a genomic region and its proximity to defined repetitive sequences on the linear chromosome template strongly influence the positioning of a locus in the nucleus.

Samenvatting

Alle erfelijke informatie die nodig is voor de ontwikkeling en het functioneren van een organisme ligt opgeslagen in het DNA, dat bij de mens bestaat uit ongeveer 3 miljard eenheden desoxyribonucleïnezuur. Bij veel meercellige organismen is dit DNA georganiseerd in grotere eenheden die chromosomen worden genoemd en zich bevinden in de kern van een cel. Deze chromosomen zijn opgebouwd uit aparte functionele eenheden die genen worden genoemd. Een gen bevat de informatie voor een specifieke erfelijke eigenschap; in de meeste gevallen een eiwit. Meercellige organismen, zoals de mens, zijn opgebouwd uit zeer veel verschillende soorten cellen. Zo bevinden zich in het lichaam bijvoorbeeld huidcellen, bloedcellen, spiercellen en hersenencellen, die allemaal hun eigen specifieke functie hebben. Deze diversiteit is opmerkelijk, omdat deze verschillende celtypen allemaal hetzelfde genetische materiaal bevatten. Het verschil in celtypen wordt echter niet bepaald door de genomische samenstelling van een cel, maar door de manier waarop deze genetische informatie wordt gebruikt. Iedere cel is namelijk in staat om heel nauwkeurig genen aan en uit te zetten en de functionele en uiterlijke verschillen tussen cellen zijn dan ook het gevolg van een precieze regulatie van genen tijdens de ontwikkeling van een cel. Deze nauwkeurige regulatie is zeer belangrijk; incorrecte activatie of onderdrukking van genen kan leiden tot allerlei celdefecten, die bijvoorbeeld kunnen resulteren in het ontstaan van kanker. De cel bezit veel verschillende mechanismen die zorgdragen voor correcte genregulatie en het begrijpen van deze zogenaamde transcriptionele regulatiesystemen is daarom een zeer belangrijk onderdeel van de moleculaire biologie.

Het DNA bevat verschillende regulerende elementen (bijvoorbeeld promoters, enhancers en insulators) die bij de transcriptie van genen betrokken zijn. Deze zogenaamde *cis*-regulerende elementen zijn echter vaak ver van elkaar verwijderd. Een belangrijke vraag is dan ook hoe deze elementen met elkaar communiceren over aanzienlijke afstanden en hoe deze communicatie uiteindelijk leidt tot een nauwkeurige regulatie van genactiviteit.

In dit proefschrift wordt gebruik gemaakt van het β-globine locus van de muis als een modelsysteem om genregulatie te onderzoeken in levende cellen. De genen die zich in dit locus bevinden bevatten allemaal de informatie voor de productie van het eiwit β-globine. β-Globine vormt samen met het eiwit α-globine en haem een functioneel eiwitcomplex genaamd hemoglobine, dat zorgt voor de binding van zuurstof en koolstofdioxide in rode bloedcellen (of erythrocyten). Naast de β-globine genen bevat het locus een aantal goed gekarakteriseerde *cis*-regulerende elementen zoals enhancers, promoters en een Locus Control Region (of LCR). Voorafgaand aan dit onderzoek is de ruimtelijke organisatie van het β-globine locus in rode bloedcellen onderzocht met behulp van een techniek genaamd Chromosome Conformation Capture (ook wel 3C genoemd). Deze experimenten hebben aangetoond dat het β-globine locus een rode bloedcel-specifieke organisatie heeft waarbij *cis*-regulerende elementen van de LCR en de actieve β-globine genen ruimtelijk bij elkaar komen. Naast deze interacties tussen

LCR en gen zijn in rode bloedcellen ook andere interacties aanwezig met een aantal nabijgelegen elementen. Gezamenlijk wordt deze ruimtelijke organisatie van het β-globine locus een Active Chromatin Hub (of ACH) genoemd. Er is echter nog maar weinig bekend omtrent de moleculaire mechanismen die ten grondslag liggen aan de vorming van een ACH gedurende de ontwikkeling van erythrocyten.

Hoofdstuk 3 van dit proefschrift beschrijft experimenten waarbij onderzoek is gedaan met zogenaamde l/11 cellen. Deze erythroide voorlopercellen zijn gebruikt als modelsysteem om de moleculaire mechanismen te onderzoeken die betrokken zijn bij de totstandkoming van interacties tussen de LCR en de β-globine genen. Activatie van l/11 cellen leidt tot β-globine expressieniveaus die vergelijkbaar zijn met de situatie *in vivo*. Daarnaast gaat deze erythroide differentiatie gepaard met meer positieve versus negatieve chromatine-modificaties op de promoter van het β_{major} gen en de LCR. Eerder onderzoek heeft aangetoond dat binding van de rode bloedcel-specifieke transcriptiefactoren GATA-1 en EKLF aan het locus noodzakelijk is voor ACH formatie. De experimenten in dit proefschrift laten echter zien dat uitsluitend binding van deze factoren aan het β-globine locus niet voldoende is voor de vorming van een ACH. Tenslotte is in Hoofdstuk 3 gebruik gemaakt van een muismodel waarbij de erythroide transcriptiefactor p45 NF-E2 is verwijderd. De resultaten laten zien dat in afwezigheid van p45 NF-E2 de contacten in het β-globine locus normaal gevormd kunnen worden, wat aangeeft dat p45 NF-E2 niet noodzakelijk is voor ACH formatie.

Onderzoek naar de rol van het insulator-eiwit CTCF bij interacties in het β-globine locus staat beschreven in Hoofdstuk 4 van dit proefschrift. Bindingsplaatsen voor CTCF flankeren het β-globine locus en in het verleden is aangetoond dat deze elementen deel uitmaken van de ACH. Dit gedeelte van het proefschrift laat zien dat door het eiwit CTCF uit cellen te verwijderen en een CTCF bindingsplaats dusdanig te veranderen dat het eiwit niet meer in staat is te binden, de interacties in het β-globine locus gedestabiliseerd worden. Echter, β-globine genexpressie wordt hierdoor niet meetbaar beïnvloed en het verlies van deze interacties resulteert uitsluitend in locaal verlies van histon-acetylatie en toename van histon-methylatie. De data laat zien dat CTCF direct betrokken is bij looping in het β-globine locus en dat het de locale balans tussen actieve en repressieve chromatine modificaties reguleert.

3C technologie stelt onderzoekers in staat om de ruimtelijke organisatie van een locus te onderzoeken met een hoge resolutie. Recentelijk zijn een aantal nieuwe, op 3C gebaseerde methoden ontwikkeld. De verwachting is dat toepassing hiervan zal leiden tot belangrijke nieuwe inzichten in de ruimtelijke organisatie van de geselecteerde chromosomale regio's. De op 3C en 4C gebaseerde methoden hebben echter ook belangrijke beperkingen en er moet worden voldaan aan bepaalde randvoorwaarden om deze technieken op een verantwoorde wijze toe te passen. Hoofdstuk 2 van dit proefschrift beschrijft en evalueert de momenteel meest gebruikte, op 3C gebaseerde methoden en behandelt de mogelijkheden en valkuilen van deze technologieën.

Tenslotte wordt in Hoofdstuk 5 de recent ontwikkelde 4C techniek toegepast om de nucleaire organisatie van zowel peri-centromerische en peri-telomerische chromosoomregio's te onderzoeken. De resultaten laten zien dat er een sterke correlatie is tussen de transcriptionele status van de geselecteerde peri-centromerische en peri-telomerische regio's en de gebieden waarmee ze een interactie aangaan in de kern. Daarnaast hebben de peri-centromerische gebieden een voorkeur om te interacteren met peri-centromerische regio's op andere chromosomen en laten de geselecteerde peri-telomerische regio's vooral interacties zien met peri-telomerische gebieden op andere chromosomen. Dit toont aan dat zowel de transcriptionele status van een regio op het genoom alsmede de nabijheid van repetitieve elementen een belangrijke invloed hebben op de positie van een locus in de kern.

Curriculum vitae

Naam:	Jurgen Alexander Kooren
Geboren:	24 februari 1979 te Rotterdam

1997-1997	VWO, Farel College, Ridderkerk.
1998-2001	HLO Biochemie, Hogeschool Brabant, Etten-Leur. Afstudeeronderzoek: Lead Discovery Unit, NV Organon, Oss. "Expression and characterisation of recombinant human glucocorticoid receptor for the detection of ligand-specific conformational changes." (dr. G. Zaman)
2001-2003	Doctoraalfase Biologie, Faculteit der Wiskunde en Natuurwetenschappen, Universiteit Leiden. Afstudeeronderzoek: Instituut voor Evolutionaire en Ecologische Wetenschappen, sectie Plantenecologie, Universiteit Leiden. "Analysis of dispersal patterns of *T. jacobaeae* by applying AFLP." (Prof.dr. E. vd Meijden)
2002-2003	Master Science Based Business, Faculteit der Wiskunde en Natuurwetenschappen, Universiteit Leiden. Afstudeeronderzoek: Merck, Sharp & Dohme, Human Health Europe, Haarlem. "Compliance and persistency in asthma treatments."
2003-2007	Promotieonderzoek, vakgroep Celbiologie, Erasmus MC Rotterdam. "β-Globin gene regulation and nuclear organisation." (Prof.dr. F. Grosveld en dr. W. de Laat)

List of publications

Simonis M*, **Kooren J*** and de Laat W.
'An evaluation of 3C-based methods capturing DNA interactions'.
Nature Methods 2007 Nov; 4(11): 895-901.

Kooren J, Palstra RJ, Klous P, Splinter E, Grosveld F and de Laat W.
'β-globin Active Chromatin Hub formation in differentiating erythroid cells and in p45 NF-E2 knockout mice'.
Journal Biol Chem. 2007 Jun; 282(22): 16544-16552.

Splinter E*, Heath H*, **Kooren J**, Palstra RJ, Klous P, Grosveld F, Galjart N, de Laat W.
'CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus'.
Genes and Development 2006 Sep; 20(17): 2349-54.

* These authors contributed equally

Dankwoord

Het boekje dat voor u ligt is het resultaat van 4 jaar onderzoek en enkele maanden intensief schrijfwerk. Uiteraard hebben veel mensen deel uitgemaakt van dit avontuur en ik wil hierbij dan ook iedereen bedanken die in welke vorm dan ook een steentje heeft bijgedragen. Toch wil ik nog een aantal mensen in het bijzonder bedanken voor hun steun en vertrouwen in de afgelopen jaren.

Frank, bedankt dat ik 4 jaar geleden de kans heb gekregen om deel uit te maken van jouw club. Het wetenschappelijke klimaat dat jij hebt gecreëerd en de daaruit voortvloeiende sfeer binnen de afdeling hebben er mede toe bijgedragen dat ik me de afgelopen jaren zeer op mijn plaats heb gevoeld. Trouwens, mocht je in de toekomst nog een schietschijf op doel nodig hebben om de eer van Celbiologie te verdedigen, je weet me te vinden.

Wouter, dank je wel dat ik de afgelopen 4 jaar je 'bonus-AIO' mocht zijn. Ondanks het feit dat onze groep de afgelopen jaren sterk is gegroeid heb je altijd tijd vrij kunnen maken om urgente zaken te bespreken. Je bezit de belangrijke eigenschap om hoofdzaken van bijzaken te onderscheiden (onmisbaar tijdens een promotietraject) en het is duidelijk dat ik zonder jouw hulp dit resultaat nooit had kunnen bereiken. Bedankt!

Vervolgens wil ik natuurlijk al mijn labgenoten van 702 bedanken. Ik denk dat ons lab laat zien dat goed wetenschap bedrijven uitstekend samen kan gaan met gezelligheid en collegialiteit. Door de jaren heen is de groep beetje bij beetje groter geworden, zonder dat dit een negatieve invloed heeft gehad op de manier waarop we met elkaar omgaan. Mocht er binnenkort nog een spelletjesavond worden georganiseerd ('wil er nog iemand bouwen?'), of gewoon iets gedronken worden in de kroeg ('bittergarnituurtje?'), je kunt me altijd bellen.

Robert-Jan, halverwege mijn tijd in Rotterdam ben je als AIO overgegaan naar het onzekere bestaan van een post-doc. Je bent iemand die wetenschap bedrijft omdat je er gewoon veel plezier aan beleeft en ik hoop dat het wetenschappelijke klimaat in Nederland ruimte laat voor mensen zoals jij. Ik hoefde maar met mijn mok te zwaaien en je begreep wat de bedoeling was (koffie-start) en onze wetenschappelijke discussies bij de koffieautomaat of gewoon slap ouwehoeren over van alles en nog wat (met als dieptepunt een nachtje doorzakken in een kelder in Lunteren) zal ik zeker gaan missen. Ik vond het een voorrecht om met je te mogen samenwerken en ik ben je dankbaar dat je mijn paranimf wilt zijn.

Erik, ook jij maakte al deel uit van het oorspronkelijke 3C-groepje toen ik in Rotterdam begon. Als ervaren analist was jij het eerste aanspreekpunt voor iedereen met vragen en jouw ervaring en advies waren voor mij onontbeerlijk ('voeg 1 µl CIP bij het digest'). Gelukkig ben je na een aantal jaren ook tot de conclusie gekomen dat je meer in je mars hebt en je bent nu begonnen aan het avontuur van een promotietraject (heb je trouwens je veiligheidscursus al gedaan?). Ik ben ervan overtuigd dat je over een aantal jaar ook je boekje zult afleveren en ik vind het erg fijn dat je mijn paranimf wilt zijn.

Daan, jij bent de volgende die het promotietraject zal gaan afronden en gezien je behaalde resultaten zal dat geen problemen opleveren. Mocht je nog een keer de behoefte hebben om te gaan squashen, mijn racket ligt klaar (wanneer ga je trouwens leren koffie drinken?).

Marieke, wij hebben het afgelopen jaar onze computerwerkplek gedeeld in een bijzonder kantoor/callcenter. Bedankt voor je geduld en goede tips met betrekking tot de 4C analyse en veel succes in het laatste jaar van jouw AIO-tijd. Petra, wees zuinig op mijn bench en zet de radio maar lekker voluit op de nieuwe CD van Racoon (wanneer komt die eigenlijk uit?). Marika, dank je wel dat je het ruim anderhalf jaar heb uitgehouden met mij als je directe stagebegeleider. Ik hoop dat ik je iets heb kunnen meegeven van het bijzondere gevoel van onderzoek doen en ik wens je veel succes in je toekomstige carrière, ongeacht of deze nu in of buiten het lab ligt.

Emilie, although we've only worked together for about one year I would like to thank you for your kindness en great interest in my personal situation during the last couple of months of my PhD. Hopefully, your efforts for our 4C project will result in a good publication and I wish you all the best with your scientific and personal life in France. Sanja, we started our PhD in Rotterdam at about the same time and I've enjoyed working together. Thank you for the excellent Serbian cookies you brought along after your holidays and 'mazzeltof' in your future life in Israel. Athina, thank you for sharing the 702 office with me and I wish you all the best in finishing your PhD.

Tevens wil ik graag de leden van mijn kleine commissie bedanken voor het lezen en verbeteren van mijn proefschrift; Dies, Sjaak en Niels, dank je wel. In het bijzonder gaat mijn dank uit naar Marieke, die niet alleen de moeite heeft genomen om mijn proefschrift door te lezen maar mij vooral in het begin van mijn promotietraject veel heeft geholpen bij het werken met erythroide cellen. Dank je wel voor je tijd, geduld en enthousiasme om me waar mogelijk te helpen. Marieke, Jasperine en Melle, dank je wel voor al jullie administratieve hulp en algemene ondersteuning in de afgelopen jaren. Ook wil ik Jopie en alle andere mensen van de spoelkeuken hartelijk bedanken. De mannen van de computerafdeling wil ik bedanken voor hun deskundige hulp en bereidheid om wanneer dan ook te helpen. Tom, bedankt dat je me hebt willen helpen bij het maken van de lay-out van dit proefschrift.

Verder wil ik graag nog een aantal mensen uit mijn persoonlijke omgeving extra bedanken. Henk, Erik, Les, Peet en Christiaan, dank jullie wel voor de interesse die jullie in mijn werk hebben getoond en voor jullie goede gezelschap. Onze vriendentripjes naar PSV (met als hoogtepunt natuurlijk ons bezoek aan Milaan en Liverpool) hebben mij altijd veel plezier gegeven en hebben samen met onze wekelijkse squash-sessies voor de broodnodige afleiding gezorgd.

Ook wil ik hierbij mijn broer Pascal, overige familie en schoonfamilie graag bedanken voor hun steun in de afgelopen jaren. In het bijzonder gaat mijn dank uit naar mijn beide ouders. Pa en ma, jullie hebben altijd achter mij gestaan en me onvoorwaardelijk gesteund bij al mijn keuzes.

Jullie hebben me de mogelijkheid gegeven om alles uit dit leven te halen wat erin zit en ik zal dit nooit vergeten.

Tot slot de meest belangrijke persoon in mijn leven. Irma, ik kan je nooit genoeg bedanken voor alle begrip, ondersteuning en liefde die je me altijd geeft. Ik hou van je.