# Activation, Regulation and Transcription of the Human and Murine Globin Loci

Mariken de Krom

MMII

# Activation, regulation and transcription of the human and murine globin loci

*Activatie, regulatie en transcriptie van de humane en muizen globine loci*

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
Rector Magnificus

Prof. Dr. Ir. J.H. van Bemmel

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op
Woensdag 20 november om 13.45 uur

Door

Mariken de Krom

Geboren te Schiedam

## Promotiecommissie

| | |
|---|---|
| Promotor: | Prof.dr. F.G. Grosveld |
| Overige leden: | Prof.dr. R. Kanaar |
| | Dr.ir. D.N. Meijer |
| | Dr. J.N.J. Philipsen |

Dit proefschrift is tot stand gekomen binnen de vakgroep Celbiologie en Genetica van de faculteit der Geneeskunde en Gezondheidswetenschappen van de Erasmus Universiteit te Rotterdam. De vakgroep maakt deel uit van het Medisch Genetisch Centrum Zuidwest Nederland. Het onderzoek werd financieel gesteund door NWO (ZonMw).

“Some things need to be believed to be seen”

-*Guy Kawasaki*

# Contents

## Scope of thesis

Haemoglobin is the carrier of oxygen in the bloodstream. It is a tetrameric protein composed of two α-globin chains and two β-globin chains. These chains are encoded by the α- and β-globin gene loci. The globin loci are located on separate chromosomes and are composed of several developmentally regulated genes and regulatory elements located both upstream of the gene loci and around the individual genes. The α-cluster is composed of αMRE5'ζ-$\alpha_1$-$\alpha_2$-3' and the β-cluster consists of LCR 5'-ε-$^{G}$γ-$^{A}$γ-δ-β-3'.

Lesions in the globin clusters give rise to blood disorders known as haemoglobinopathies, such as α- and β- thalassemias, Hereditary Persistence of Foetal Haemoglobin (HPFH) and sickle cell anaemia. The study of the molecular basis of these disorders led to the identification of many regulatory elements of the globin loci, including the locus control region (LCR). The β-globin locus is one of the best-studied multi gene loci and has been extensively used as a model system for the study of gene regulation.

The aims of this thesis are to study the mechanism of transcriptional activation of the globin loci, the role of the hypersensitive sites in LCR function and the further characterisation of putative γ-globin regulatory elements.

Chapter 1, the introduction, will give a broad outline of chromatin structure and transcription, as well as detailed background on haemoglobin and the regulation of globin gene loci. Chapter 2 concentrates on fundamental aspects of the transcriptional activation process of the mouse globin gene loci, with the aim of to serve as a model system for understanding transcriptional activation in general. Chapter 3 describes the development of novel methodology that employs homologous recombination in *E. coli* in order to manipulate the entire 180 kb human β-globin locus, for example, through the introduction of specific mutations and/or deletions. Chapter 4 describes the analysis of transgenic mice with targeted deletions in the human β-globin locus which remove two elements, located 3' to the $^{A}$γ-globin gene, and which have been implicated in the developmental regulation of the γ-globin gene. Chapter 5 describes an application of this technique in deleting two hypersensitive sites from the human β-globin locus LCR, with the aim of to study LCR function in chromatin organisation and gene activation. Chapter 6 is the general discussion.

# *CHAPTER 1*
# *INTRODUCTION*

## Introduction

The first part of the general introduction will be addressing common features of the regulation of eukaryotic transcription, with the aim of giving a broad outline of factors involved in the transcriptional process. The second part of the introduction will be concerned with erythropoiesis, haemoglobin and the structure and regulation of the mouse and human globin genes.

## Transcription

The process of transcriptional regulation is complex and involves many different players. DNA sequence itself, in the form of *cis*-regulatory elements of genes, chromatin and its higher order structures, histone modification and the diverse chromatin modification complexes, all play specific and important roles in eukaryotic gene regulation. In the following paragraphs these features will be addressed.

## Chromatin

Eukaryotic chromosomes contain 2m of DNA when stretched and need to be packaged to fit into the nucleus of the cell (with a diameter of 10μm). The packaging of the DNA to form a compact structure is achieved with the aid of specific proteins that are complexed with the DNA to form chromatin (Fig.1). All DNA processes in the nucleus, e.g. replication and transcription, take place in the context of chromatin. The picture that is emerging is that chromatin does *not* present an obstacle to these processes (as was previously thought) but, instead, plays a leading role in their regulation. In this part of my thesis I will consider the role of chromatin in transcriptional regulation. The general structure of chromatin will be addressed first.

### *Chromatin organisation*

The primary proteins involved in chromatin structure are the histones. There are five types of histones that fall into two categories. Firstly, there are the core histones H2A, H2B, H3 and H4 and, secondly, the linker histone H1. The core histones are evolutionary conserved both in size and amino acid sequence (de Lange *et al.*, 1969). They are small proteins, rich in arginine and lysine residues (van Holde, 1989). They contain a globular domain, which is important for histone-histone and histone-DNA interactions, as well as a N-terminal tail domain, with the exception of histone H2A that also has a tail at the C-terminus (Bohm and Crane-Robinson, 1984). The linker histones are less well conserved, are rich in lysine residues and contain a globular domain, which is flanked on both sides by a tail.

The core histones interact with DNA and with each other to form a histone octamer (Fig. 1D). The core histone interacts with the DNA via its globular domain. DNA is wrapped around the histone octamer. The octamer is composed of a tetramer formed by H3 and H4, which, after binding to DNA, is bound by two dimers of H2A-H2B resulting in the final histone octamer (Eickbush and Moudrianakis, 1978; Hayes *et al.*,1990, 1991 and Arents *et al.*,1991). The octamer forms the basis of the nucleosome. The nucleosome is the fundamental repeating unit of chromatin and consists of the histone octamer and approximately 146 bp of DNA wrapped around the octamer in 83 bp superhelical loops (Fig. 1D) (Kornberg, 1974). The DNA helix is thus packaged into nucleosomal cores giving a characteristic “beads on a string” appearance of the chromatin fiber when visualised using electron microscopy (Olins and Olins, 1974). The first level of organisation of DNA into nucleosomal cores forms the 10nm chromatin fiber (Fig. 1B).

The nucleosome model has been analysed in detail by solving the nucleosomal structure at a resolution of 2.8 Angstrom (Luger *et al.*, 1997) and 1.9 Angstrom (Davey *et al.*, 2002), showing a twist of the DNA molecule on the surface of the nucleosome. The “twisting” of the DNA around the nucleosomes results in the formation of minor and major grooves along the backbone of the helix. The minor grooves form the channels through which the N-terminal tails of histone H2B and H3 pass. The

tails of histones H2A and H4 pass through the gyres of the DNA superhelix. In these ways the tails not only interact with the DNA but may also interact with neighbouring nucleosomes. The tails are targets for modifications such as acetylation and phosphorylation due to their localisation on the surface of the nucleosome. Such modifications can lead to changes in the binding affinity of nucleosomes for DNA, for example, as a result of the neutralisation of the positively charged tails by acetylation which leads to a decrease in affinity of the nucleosome for DNA, thus resulting in alterations in local chromatin states.

Individual nucleosomes are spaced by approximately 80 bp of linker DNA. Micrococcal nuclease digestions followed by sucrose gradient and mobility shift assays of the smallest (mononucleosomal) fractions, showed that part of the linker DNA is bound by histone H1 (Varshavsky *et al.*, 1976). This histone is required for the maintenance of the stability of the core histones in the histone octamer and the stability of the higher order chromatin structures (Dasso *et al.*, 1994; Wolffe, 1998 and Carruthers and Hansen, 2000). The tails of histone H1 bind to the DNA within the nucleosome and with the linker DNA (Fig. 1D), with the C-terminal tail folding into α-helices that associate with the major groove of the linker DNA (Clark *et al.*, 1988). The backbone of the DNA is neutralised by the histone H1 tails and the 10nm fiber can thus fold into a higher order structure, compacting the DNA into what is known as the 30nm fiber (Fig. 1C) (Clark and Kimura, 1990). Although linker histones are important for the neutralisation of the backbone of DNA, it is the tails of the core histones that are essential for the folding of chromatin into the 30nm fiber. Using defined chromatin model systems with core histones lacking their tails, it was demonstrated that the tail-less chromatin fiber could not fold into 30nm fibers, even with linker histones present (Carruthers and Hansen, 2000). Thus tails of the core histones and the nucleosomes play an important role in internucleosomal contacts within the 30nm fiber and are critical for the self-assembly of the condensed fibers in higher order structures (Fletcher and Hansen, 1996 and Anderson and Widom, 2000).

The 30nm fiber itself is organised into loops estimated to be between 30-100 kb in size. These sizes were estimated using electron microscopy and by distance measurements using fluorescent *in situ* hybridisation (Yokata *et al.*, 1995). These loops have been proposed to play an important role in the regulation of gene expression, e.g. in the interaction between enhancers and promoters over long distances. Two nucleoprotein structures, which have been identified using different nuclear extraction methods have been proposed to play a role in this loop formation: the matrix associated regions (MARs) and the scaffold attachment regions (SARs) (Mirkovitch *et al.*, 1984 and Cockerill and Garrard, 1986). MARs and SARs do not have consensus sequences but contain AT-rich stretches recognised by linker histone H1 and topoisomerase II. This recognition is thought to lead to the anchoring of the 30nm chromatin fiber to the chromosome scaffold or nuclear matrix. Proteins already bound to the MARs and SARs then secure the chromatin fiber into structural loops (Laemmli *et al.*, 1992), thus achieving an even further packaging of the DNA. Other proteins binding to MAR and SAR elements, like SATB1, GATA3 and p300, have been implicated to play a role in gene regulation by introducing chromatin changes through binding to MARs and SARs (Kieffer *et al.*, 2002 and Martens *et al.*, 2002). SATB1 and GATA3 have both been shown to bind MARs in the CD8 gene, and via this binding have been suggested to play a role as epigenetic regulators of CD8 expression (Kieffer *et al.*, 2002).

### *Euchromatin and heterochromatin*

Chromatin has been cytologicaly divided into heterochromatin and euchromatin. (Heitz, 1928). Heterochromatin is densely stained (by carmin acetic acid) throughout the cell cycle indicative of a constitutively condensed chromatin structure. Euchromatin, on the other hand, is more lightly stained and decondenses as the cell progresses from the metaphase after mitosis to the interphase of the cell cycle.

Heterochromatic genomic domains consist predominantly of repetitive DNA, including satellite sequences, and is mostly found in pericentromeric and telomeric regions. Heterochromatin contains generally few genes whereas euchromatin is gene-rich. Other characteristics that distinguish heterochromatin from euchromatin include: (i) higher order chromatin structure, which shows a much more regular nucleosomal organisation for heterochromatin (Wallrath and Elgin, 1995 and Sun *et al.*, 2001); (ii) differences in histone modifications, with heterochromatin being rich in

methylated/hypoacytelated histones whereas euchromatin is enriched in acetylated/non-methylated histones (Strahl and Allis, 2000 and Rea *et al.*, 2000); (iii) differences in replication timing, with heterochromatin replicating late in S-phase versus early for euchromatin. Euchromatin is thus generally viewed as the chromatin compartment that supports transcription whereas heterochromatin represses transcription.

Two key observations have coupled formation of heterochromatic structures to silencing of genes that are normally found in euchromatin. The first is X-inactivation in dosage compensation, in which the inactive X-chromosome showed the same cytological staining and molecular characteristics, such as hypoacetylated histones, as heterochromatin, indicative of heterochromatinisation as cause of silencing (Schoenherr and Tilghman, 2000).

The second observation was first described in *Drosophila* in which a chromosomal translocation of the *white* gene, normally residing in euchromatin, placed it close to pericentromeric heterochromatin. This gave rise to a variegated phenotype of white and red patches in the *Drosophila* eye (Muller, 1930). Variegated *white* expression appeared to be due to the variable silencing of the gene in different cells of the eye tissue. Thus in some cells the *white* gene expresses giving rise to red colour, whereas in other cells the gene is inactivated giving rise to *white* colour. The variability in expression of the *white* gene in the eye is thought to reflect differences in the extent of heterochromatinisation at the site of the chromosomal translocation between different eye cells. It is thought that heterochromatin is laid down in a window of opportunity early in development and spreads from so-called nucleation centres along the chromatin fiber in the centromeres, with the extent of spreading varying from one cell to another. However, once heterochromatin has been laid down it becomes fixed thus "freezing" the differences in heterochromatin spreading between cells, which are then clonally inherited in subsequent cell generations. Therefore, in a cell where the *white* gene has been heterochromatinised it will be inactive, whereas in another cell in the same tissue where the gene has not been embedded in heterochromatin, it will be active. Variegated expression patterns also appear to be stable through subsequent cell divisions and are thus clonally inherited. This effect on gene expression is referred to as position effect variegation (PEV) and is defined as the heritable silencing through multiple cell divisions resulting from translocation or integration of a gene in a position close to heterochromatin.

Mutations affecting PEV have been extensively studied and have resulted in the identification of more than 30 genetic modifiers of PEV (Jenuwein and Allis, 2001). These can be divided into two groups, with antagonizing effects. The first group consists of the Su(Var) group of proteins, which when mutated suppress variegation. The second group is the E(Var) group, which when mutated enhance the variegation of expression of the integrated gene. It can therefore be inferred from these phenotypes that the Su(Var) proteins are normally involved in promoting heterochromatinisation, and hence silencing, whereas the E(Var) proteins normally counteract heterochromatinisation and hence promote expression (Wallrath, 1998 and Eissenberg *et al.*, 1990).

One of the best studied proteins of the first group of modifiers is heterochromatin protein 1 (HP1), which is encoded by Su(var) 2-5 (Eissenberg *et al.*, 1990 and Eissenberg *et al.*, 1992). Mutations in HP1 increase expression of the *white* gene, whereas an additional copy of HP1 reduces *white* expression, thus enhancing variegation (Cryderman *et al.*, 1998). Recent studies aimed at understanding the basis for HP1-mediated repression have shown that histone modifications play a role in placing heterochromatic marks. First it was shown that histone deacetylase inhibitors result in disruption of heterochromatin and HP1 binding (Taddei *et al.*, 2001). Secondly, a link between methylation of H3 and heterochromatin formation was identified. The Su(var)39h genes encode histone methyltransferases and have been shown to have high specificity for lys-9 of histone H3 (Rea *et al.*, 2000). A connection between a histone deacetylase and Su(var)39h has recently been shown in *Drosophila* thus providing a mechanism to convert an acetylated histone into a methylated one (Czermin *et al.*, 2001). The link between Su(var)39h and HP1 was found in primary mouse fibroblasts from double null Su(var)39h mice. These studies, using immunostaining to detect HP1 distribution, indicated that the methylation of H3 lys-9 by Su(var)39h is important for HP1 localisation and heterochromatinisation (Bannister *et al.*, 2001 and Lachner *et al.*, 2001). E(Var) proteins on the other hand interact with proteins that remodel chromatin to allow activation of transcription, like the SWI/SNF and Brahma complexes (Tsukinama and Wu, 1997 and Kal *et al.*, 2000).

Besides the Su(var) and E(var) proteins other factors influence the variegation of gene expression. The integration of β-globin and CD2 transgenes containing a full locus control region in

pericentromeric regions resulted in normal levels of expression of the transgenes (Festenstein *et al.*, 1996 and Milot *et al.*, 1996). Deletion of part of the LCR, however, led to PEV, which suggest that the loss of hypersensitive sites results in a decrease in accessibility of the regulatory elements of the transgene to the transcriptional machinery. These studies led to the formulation of a probability model for LCR function in which sufficient binding sites for positively acting factors would completely overcome the packaging of DNA into heterochromatin (Festenstein and Kioussis, 2000).

Also transcription factors have an influence on PEV. Studies in which the doses of transcription factors like EKLF and Sp1 in variegating mouse lines was either overexpressed or reduced, showed a higher or lower expression of the variegating transgene, respectively, compared to endogenous genes, which do not show a PEV (Lundgren *et al.*, 2000 and McMorrow *et al.*, 2000). All the results presented above indicate that the decision to silence or activate a gene in a heterochromatic environment is the result of a balance between positive and negative factors.

### *Nuclear organisation and gene expression*

The organisation of euchromatin and heterochromatin in discrete nuclear domains also plays an important role in gene expression. Within the nucleus euchromatin is mainly found in the interior of the nucleus, whereas heterochromatin is found near the nuclear periphery and around the nucleolus. The different chromatin domains delineate different subnuclear compartments, which exert different effects on the transcriptional regulation of genes. For example, a group of proteins, shown to accumulate in heterochromatic domains in the nucleus, are the Sir-proteins, which play a role in gene silencing in yeast (Gasser, 2001).

An example of this is demonstrated when genes become integrated at telomeres and, as a result, become repressed through association with silencer proteins (Sir 2p, 3p and 4p), which bind to histones H3 and H4 and deacetylate the amino-terminal tails of these histones (Grunstein, 1998). The sites of Sir-mediated binding have been localised by double DNA *in situ* hybridisations and immunofluorescence and shown to be clustered in specific foci within the nucleus around the nuclear periphery. Subsequently it has been shown that the nuclear organisation of the telomeres is of critical importance for the repression via Sir proteins. Disruption of the telomeric organisation proved to result in delocalisation of the Sir-proteins and a loss of telomeric silencing (Gotta *et al.*, 1996 and Galy *et al.*, 2000). The HM-locus of *Saccharomyces cerevisiae* has also been shown to be silenced by the interaction of Sir-proteins with silencers present in the HM loci. Studies on HM silencing in *Saccharomyces cerevisiae* show that the localisation of a locus to the nuclear periphery helps to establish transcriptionally silent heterochromatin domains. It was also shown that the HM-locus could be silenced, even with a defective silencer, by artificially anchoring the locus to the nuclear periphery (Andrulis *et al.*, 1998).

Studies in yeast have shown that repressor proteins are present in the nucleus in limiting amounts and are recruited to specific heterochromatic compartments (Gasser, 2001). In addition, the spatial distribution of repetitive DNA forms subcompartments in the nucleus that favour the packaging of chromatin in a repressive state (Gasser, 2001). In higher eukaryotes a nuclear compartmentalisation similar to that observed in yeast has also been shown. Genes that are not transcribed are recruited to heterochromatic compartments and when activated “move” to the periphery of those compartments to be transcribed (Brown *et al.*, 1997 and 1999 and Francastel *et al.*, 1999). Using immunofluorescence *in situ* hybridisations, the Ikaros transcription factor was shown to be localised in constitutively heterochromatic centromeric foci in interphase nuclei. Several genes which are regulated by Ikaros were tested for their nuclear localisation in expressing and non-expressing cells. It was shown that inactive genes associate with Ikaros-heterochromatin foci, whereas the active genes are not associated with the heterochromatin foci (Brown *et al.*, 1997).

Furthermore, not the transcriptional activation of a gene but the hyperacetylation of the histones in the promoter region of the gene that “signals” the re-localisation of the gene within the nucleus. This has been shown using a β-globin gene without a LCR. When integrated in a heterochromatic domain after transfection in MEL cells, the hyperacetylation of the β-promoter resulted in the relocalisation of the gene to the periphery of the heterochromatic compartment that it was previously located in (Schubeler *et al.*, 2000). Also studies with a λ5 transgene inserted in the mouse γ satellite repeat showed using *in situ* hybridisation techniques that while a repressed gene was localised in a

heterochromatic compartment, the binding of specific transcription factors relocated the transgene to the periphery of the compartment without activating it. Transcriptional activation of the gene required a strong transactivator, which could presumably overcome the repressing activity of the repressor proteins (Lundgren *et al.*, 2000). These observations indicate a specific sequence of events for gene activation which is related to the gene's subnuclear localisation, histone modification and the action of specific transcription factors.

## Chromatin and transcription

Transcription takes place in a chromatin context. A wealth of recent research evidence has revealed the many ways in which the cell not only deals with chromatin in transcription utilises it as a significant regulatory factor (Narlikar, 2002).

### *Nucleosome positioning*

Nucleosomes assembled on DNA generally render the DNA inaccessible to the transcriptional machinery *in vitro* (Svaren and Horz, 1996). However, DNA that is being transcribed or replicated remains nucleosomal, thus indicating that nucleosomes do not form an impossible obstacle for these processes (Studitsky *et al.*, 1995 and Felsenfeld 1996).

The orientation of a nucleosome on DNA can be defined by its translational and rotational positioning. The rotational position defines which of the DNA sequences wrapped around the nucleosome are facing outwards, towards the surrounding solution and may thus be accessible to transcription factors. The translational position of the nucleosome determines the precise site where the DNA and histone contacts begin and end.

The positioning of nucleosomal arrays on promoters and its function in repression or activation of genes has been studied most extensively in yeast, for example, using the PHO5 gene. Promoters of genes of mouse and human origin have also been studied to gain better insight into the role of positioned nucleosomes in the accessibility of transcription factors to their recognition sites regulating the induction of transcription.

The yeast PHO5 gene encodes a secreted acid phosphatase and is activated when phosphate is limiting in the cell. The PHO5 gene otherwise remains silent by the positioning of four nucleosomes over its promoter (Svaren and Horz, 1997). The activation of PHO5 requires two factors, PHO4 and PHO2, and the repositioning of the nucleosomes since at an active PHO5-promoter no nucleosomes are detected (as measured by an increase in deoxyribnuclease I (DNase I) accessibility). PHO4 binds to the sequences UASp1 and UASp2 present in the linker DNA between two precisely positioned nucleosomes. PHO4 binding leads to local chromatin remodelling allowing additional transcription factor binding and PHO5 activation (Haswell and O'Shea, 1999). The precise mechanism of the nucleosome remodelling by PHO2 and PHO4 remains unclear, however, recent evidence using yeast nuclear extracts has shown that PHO4 and PHO2 do not need nucleosome remodelling factors, but do need an ATP-dependent activity and core histone acetylation using acetyl CoA (Terrell *et al.*, 2002 and Haswell and O'Shea 1999).

The mammary tumour virus (MMTV) -promoter has also been studied in understanding how regulatory factors can recognise their cognate sequences when these are embedded in nucleosomal arrays. The MMTV-promoter is induced by glucocorticoids via the glucocorticoid receptor (GR) and NF1. Two GR-bindingsites sites are exposed on the surface of a nucleosome by the rotational positioning of the nucleosomes in the MMTV-promoter. This positioning makes it possible for a GR-dimer to bind to its sites on the nucleosome. Binding of the hormone to a GR-bindingsites site induces a change in the translational positioning of the nucleosomes at the MMTV-promoter, such that the binding sites for NF1 become accessible and, together with the glucocorticoid receptor, NF1 can activate the MMTV-promoter (Eisenfeld *et al.*, 1997 and Belikov *et al.*, 2001).

Furthermore, nucleosome rearrangements can be induced by transcription factors. This was originally observed for HNF3 and the foetal serum albumin enhancer. It was found that the nucleosomal organisation of the liver-specific enhancer was random in all tissues except liver where it is active. In liver, the chromatin structure of the enhancer consisted of an ordered array of three precisely positioned nucleosomes. Using *in vitro* chromatin assembly assays it was shown that for this

nucleosomal organisation the binding of proteins related to HNF3 to the enhancer was required. This led to the suggestion that certain transcription factors can induce nucleosomal rearrangements (McPherson *et al.*, 1993). Other transcription factors like Gal4, TFIIIA, Sp1, TBP, GATA4, Fos and Jun have also been shown to play a role in nucleosomal rearrangements. They bind to their recognition sites on the nucleosomal DNA and do so in competition with histones (Cirillo and Zaret, 1999, Blomquist *et al.*, 1999 and Ng *et al.*, 1997). Their binding changes the nucleosomal structure and subsequent binding of other transcription factors can take place. Studies with the HNF3 and GATA4 transcription factors, that bind to the albumin gene enhancer, provide good examples. To test how these two can bind to their recognition sites in "repressed" chromatin, *in vitro* experiments were done using nucleosomal arrays containing albumin enhancer sequences which were compacted with linker histones. HNF3 and GATA4 were able to bind to their recognition sites and open the nucleosomal arrays without the presence of ATP-dependent remodelling factors, whereas other transcription factors like NF1 and C/EBP could not. The opening of chromatin by HNF3 is mediated by a high affinity DNA binding site and the C-terminal domain of the protein which binds histone H3 and histone H4 and thus facilitates nucleosome rearrangements (Cirillo *et al.*, 2002).

Not all transcriptional activators are able to compete with the histones in the nucleosomal array. This could be because of inaccessibility of the DNA targets site for the transcription factor through nucleosomal positioning or because of very low affinity of the transcriptional activator for nucleosomal DNA. To facilitate the binding of these transcription factors two distinct types of enzymatic activities have been described: enzymatic histone tail modifications and ATP-dependent chromatin remodelling.

### *Histone tail modifications*

The tails of the core histones in nucleosomes are rich in lysine and arginine residues, which can be targets for posttranslational modifications. Unmodified histone tails possess a positive charge, which results in a closer interaction with the negatively charged nucleosomal DNA. Thus, histone tail modifications that change the positive charge are likely to influence interactions with the DNA.

To date, several histone tail modifications have been described (Strahl and Allis, 2000 and Turner, 2000). Regulated acetylation and deacetylation of specific lysine residues of histone H3 and histone H4 are correlated with gene activation and silencing, respectively. Phosphorylation of histone H3 ser-10 has been suggested to be important for transcription activation and chromosome condensation during mitosis (Cheung *et al.*, 2000). The methylation of arginine residues of arg-3, -17, -26 of histone H3 and arg-3 of histone H4 have been shown to play a role in gene activation (Chen *et al.*, 1999 and Wang *et al.*, 2001), whereas methylation of lys-9 of histone 3 plays a role in gene silencing through heterochromatinisation (Rea *et al.*, 2000). Another histone tail modification is the ubiquitination of lys-123 at the C-terminal tail of histone H2B. Mutation of the ubiquitination site of histone H2B causes defects in meiosis and mitosis in yeast (Robzyk *et al.*, 2000). Histone H1 has also been shown to be ubiquitinated by *Drosophila* TAFII250 (Pham and Sauer, 2000). TAFII250 is recruited to promoters, and loss of TAFII250 in the fly embryo has been shown to result in a reduced expression of the Dorsal activator (Pham and Sauer, 2000). This leads to the suggestion that ubiquitination of histone H1 can regulate chromosomal gene activity in a promoter specific manner.

The experimental evidence so far, suggests that histone tail modifications form part of an enzymatic cascade, which leads to specific changes in chromatin structure resulting in the repression or activation of transcription. The end result on transcription or other nuclear functions dictated by particular combinations of histone tail modifications has been called the histone code (review Jenuwein and Allis, 2001).

Histone tail acetylation has been a major focus of investigation. Acetylation of the four core histones, at specific lysine residues, occurs in all animals and plants studied to date (Csordas, 1990). The first evidence that acetylation of histones correlates with transcriptional activity came from studies in yeast (Grunstein *et al.*, 1992) and immunofluorescence studies using specific α-Acetyl-histone antibodies (Jeppesen and Turner, 1993). Transcriptionally active domains have thus been correlated with general histone hyperacetylation (Hebbes *et al.*, 1994 and Lee *et al.*, 1993), whereas inactive domains appear hypoacetylated (Turner *et al.*, 1992). Acetylation occurs on specific lysine residues. For heterochromatin in yeast and *Drosophila* it has been shown that only H4-lys-12 is

acetylated, whereas in euchromatin of yeast, *Drosophila* and humans different combinations of acetylated lysines of histone H3 and histone H4 have been found (Turner *et al.*, 1992; Clarke *et al.*, 1992; O'Neil and Turner, 1995; Bannister *et al.*, 2000 and Rojas *et al.*, 1999). Transcriptionally active chromatin has been shown to be acetylated at lys-14 of histone H3 and at lys-8 and -12 of histone H4 (Cheung *et al.*, 2000; Lo *et al.*, 2000 and Wang *et al.*, 2001). Newly synthesised histones appear to be acetylated at different residues than the histones in active chromatin. In newly synthesised histones, histone H4 is acetylated at lys-5 and -12 and at histone H2A lys-5 (Grant and Berger, 1999).

The effect of acetylation of the N-terminal histone tails is the neutralisation of the positive charged on the lysine residues, resulting in a decrease of histone tail affinity for negatively charged nucleosomal DNA. This has an effect on nucleosome conformation and inter-nucleosomal interactions (Hamiche *et al.*, 1999; Langst *et al.*, 1999; Whitehouse *et al.*, 1999; Clapier *et al.*, 2000 and Oliva, *et al.*, 1990), resulting in a "loosening up" of the nuclesosomal structure thus making the DNA more accessible for the transcription machinery (Lee *et al.*, 1993; Vetesse-Dadey *et al.*, 1996 and Sewack *et al.*, 2001).

An important tool in determining the histone tail acetylation status of specific genes of interest and relating it to transcriptional activation, has been the use of specific antibodies against acetylated histones H3 and H4 in the immunoprecipitation of formaldehyde cross-linked chromatin (review Orlando, 2000). Due to the fact that formaldehyde cross-links are reversible, specific DNA sequences can be recovered and tested for enrichment in acetylated histones (Saitoh and Wada, 2000, review by Forsberg and Bresnick, 2001). An example is a study in which the acetylation patterns of histones H3 and H4 of the human β-globin locus were analysed in mouse erythroleukemic (MEL) cells. The acetylation status of the complete locus was compared to that of the locus after a deletion removing HS2-5 of the LCR and another deletion removing HS2-5 as well as an additional 27 kb of upstream sequences. The full locus and the first deletion showed similar basal levels of acetylation throughout the locus, whereas the second deletion line showed a pattern of hypo-acetylation. Furthermore, the full locus, which is transcriptionally active, showed peaks of histone H3 acetylation at the LCR and the active promoters, whereas in both of the deleted loci, which are transcriptionally inactive, no such peaks were observed (Schubeler *et al.*, 2000).

The enzymes responsible for the acetylation of the histone tails are the histone acetyltransferases (HATs). There are two classes of HATs: the nuclear HATs involved in transcriptional regulation and the cytoplasmic HATs involved in the acetylation of newly synthesised histones. The first nuclear HAT was identified in 1996 by Brownwell *et al*. and corresponds to yeast GCN5. GCN5 had already been identified as transcriptional co-activator (Georgakopoulos and Thireos, 1992) and was later shown, using yeast mutants, to be important for gene activation and acetylation of H3-lys-14 and H4-lys-8 and -16 (Kuo *et al.*, 1996 and 1998).

Following identification of GCN5 as a histone acetyltransferase, more proteins were recognised with HAT activity. Many of them are part of multi-protein complexes recruited to promoters, by interaction with DNA-bound activator proteins (Utley *et al.*, 1998) thus playing a direct role in activation of transcription (Larschan and Winston, 2001, Bhaumik and Green, 2001). Examples include PCAF, a human transcriptional co-activator similar to GCN5, p300/CBP, which are global co-activators and TFII250, which is a TBP-associated factor and plays a direct role in transcription initiation (Sterner and Berger 2000).

Besides histone tails, some transcription factors, including p53, EKLF, TFIIEβ, HMG1, GATA1 and ACTR are also substrates for HATs, potentially influencing their roles in the transcriptional process (Gu and Roeder, 1997; Boyes *et al.*, 1998; Chen *et al.*, 1999; Marizo *et al.*, 2000 and Zhang and Bieker, 1998). In the cases of p53, EKLF and GATA1, the acetylation site(s) map next to the DNA-binding domains. Acetylation may thus affect the DNA binding properties of these factors, for example, it may have a stimulatory effect (Gu and Roeder, 1997; Boyes *et al.*, 1998; Zhang and Bieker, 1998 and Martinez-Balbas *et al.*, 2000). On the other hand, acetylation of HMG1 results in disruption of DNA binding, because the acetylated lysines fall directly within the DNA-binding domain (Ugrinova *et al.*, 2001). Therefore, the common view that acetylation is positively affecting transcription does not always seem to hold true when it comes to the acetylation of transcription factors.

Other proteins like α-tubulin (Sterner *et al.*, 1979 and L'Herault and Rosenbaum, 1985) and the importin-α family of nuclear importer factors (Bannister *et al.*, 2000) are also target for acetylases.

These findings suggest that acetyltransferases have a wide range of proteins as substrates: DNA-binding proteins (histones and transcription factors), non-nuclear proteins (α-tubulin) and proteins that shuttle between the nucleus and the cytoplasm. This, in turn, indicates that acetylation has diverse consequences: chromatin and nucleosome remodelling, DNA-binding (see transcription factors), or protein-protein interactions, for example in the generation of a recognition site for bromodomain binding via the acetylation of histones (Dhalluin *et al.*, 1999). Finally, acetylation also seems to influence protein stability, as a correlation has been described between the acetylation of α-tubulin and the stability of microtubules (Takemura *et al.*, 1992).

The effect of acetyltransferases can be antagonised by the histone deacetylases or HDACs, first isolated by Taunton *et al.* (1996). The antagonistic effect of HDACs results in repression of transcription. There are three classes of HDACs that have been described so far (Taunton *et al.*, 1996; Verdel and Khochbin, 1999 and Imai *et al.*, 2000). The first class is related to the yeast protein RPD3, which was the first HDAC to be described (Taunton *et al.*, 1996). The second class of HDAC are related to yeast HDA1 (Rundlett *et al.*, 1996), later human and mouse homologues were also identified (Verdel and Kochbin, 1999; Fischle *et al.*, 1999 and Grozinger *et al.*, 1999). The third class takes its name from the SIR2 protein first identified in yeast, but which also includes recently identified mammalian homologs (Imai *et al.*, 2000). The existence of three different classes indicate that deacetylases, like acetyltransferases, have diverse activities.

Another class of enzymes that has gained a lot of attention in the unravelling of histone modifications and their role in chromatin remodelling and transcriptional regulation, are the histone methyltransferases. The modification of histones by methylation was described for the first time in 1964 by Murray, though it has gained considerable importance only in the last three years. Methylation takes place at lys-4, -9, and -27 of histone H3 and lys-20 of histone H4. The first methyltransferase to be discovered was Suv39 which methylates lys-9 of histone H3 (Rea *et al.*, 2000).

The enzymatic activity of histone methyltransferases resides in their SET domain (Kouzarides, 2002). At the moment, four groups of proteins have been identified possessing a distinct SET domain, the Suv39 group and the SET1, SET2 and RIZ groups. The proteins in the first group specifically methylate lys-9 of histone H3 (O'Carrol *et al.*, 2000; Tachibana *et al.*, 2001 and Yang *et al.*, 2002), whereas proteins belonging to the SET1 group have been shown to methylate lys-4 of histone H3. Proteins belonging to the SET2 group methylate histone H3, although the specific residues have not yet been identified. Finally, for the RIZ group no methylase activity has yet been demonstrated (Kouzarides, 2002).

Methylation of histones can correlate with either a repressed state or an active state of transcription. Methylation of lys-9 on histone H3 leads to the recruitment of HP1 thus resulting in heterochromatinisation and silencing. However, before methylation of lys-9 on histone H3 can occur, this has to be deacetylated. Deacetylation of lys-9 on histone H3 is a result of the recruitment of deacetylases by repressor proteins to the site where histone methylation has to take place. An example of such a repressor protein is the retinoblastoma co-repressor protein (RB). Studies in RB null cells showed that the *cyclin E*-promoter is undermethylated at lys-9 on histone H3 and no association of HP1 with the *cyclin E*-promoter occurs (Nielsen *et al.*, 2001). This indicates that RB activity plays a role in the methylation of lys-9 on histone H3 and the subsequent recruitment of the HP1 repressor (Brehm and Kouzarides, 1999).

On the other hand, methylation of lys-4 of histone H3 is correlated with an active state of transcription. Studies using the mating type loci in fission yeast have shown that inactive chromatin is enriched in lys-9 methylation and devoid of lys-4 methylation of histone H3. The reverse was found in transcriptionally active chromatin (Litt *et al.*, 2001 and Noma *et al.*, 2001).

Besides lysine methylation, the arginine residues arg-3, -17 and -26 of histone H3 and arg-3 of histone H4 are also targets for methyltransferases. To date, there are five arginine methyltransferases known. Only recently, using chromatin immunoprecipitation assays, has it become clear that arginine methylation can be correlated with an active state of transcription, much like histone acetylation (Ma *et al.*, 2001 and Bauer *et al.*, 2002).

An example of the histone code and the interplay between histone modifications is that of the modification status of lys-9 of histone H3. When acetylated by HATs, lys-9 on histone H3 codes for an active status of transcription, however, when it becomes deacetylated by repressor proteins such as

RB, lys-9 on histone H3 becomes the target for methyltransferases. The resulting methylation of lys-9 on histone H3 now codes for an inactive transcription state (Noma *et al.*, 2001).

Another example is that of acetylation of lys-14 of histone H3, which correlates with a transcriptional active state. This acetylation is preceded by and depends on the phosphorylation of ser-10 of histone H3 (Cheung *et al.*, 2000 and Lo *et al.*, 2000). Another combination of histone modifications that marks a transcriptionally active state is the methylation of arg-3 on histone H4, which precedes the acetylation of lys-8 and lys-12 (Wang *et al.*, 2001).

This implies that existing histone modifications can "recruit" new modifications, leading to the recruitment of proteins, or protein complexes, that alter the chromatin structure. This alternating results in either activation or repression of transcription (Strahl and Allis, 2000).

### *ATP-dependent chromatin remodelling*

In order to facilitate transcription, nucleosomal arrays in the chromatin fiber often have to be remodelled or disrupted in such a way that transcription factors can bind to their recognition sites. Besides the histone tail modifications, ATP-dependent chromatin remodelling complexes have also been implicated in this process.

Two genetic screens for altered gene expression in *Saccharomyces cerevisiae*, one showing the importance of Snf genes for SUC2 expression and the second showing the necessity of Swi genes for HO expression, led to the discovery of the first chromatin remodelling complex SWI/SNF (Winston and Carlson, 1992). The complex has a molecular weight of 2 MDa and contains 11 proteins, of which the Swi2/Snf2 protein is responsible for nucleosome disruption. Complexes homologous to SWI/SNF include BRG1 and hBRM in human and the Brahma complex in *Drosophilla* (Khavari *et al.*, 1993 and Muchardt and Yaniv, 1993).

The disruption of the nucleosomal array by SWI/SNF is an ATP-dependent process. The SWI2/SNF2 subunit contains an ATP-ase and a helicase domain, both of which are necessary for chromatin remodelling (Cote *et al.*, 1994 and Kwon *et al.*, 1994).

Genome-wide expression studies in yeast, have shown that the action of SWI/SNF is limited to specific promoters rather than to chromosomal domains. (Sudarsanam *et al.*, 2000). The question remains as to how SWI/SNF binds to target genes since there are no known DNA binding activities in the complex . Two models have been proposed for the binding of SWI/SNF to its targets. The first model suggests that SWI/SNF is recruited to promoters by RNA Pol II (Wilson *et al.*, 1996 and Cho *et al.*, 1998) and the second model argues that SWI/SNF is recruited to promoters by transcriptional activators and thus before RNA Pol II is present (Neely *et al.*, 1999 and Natarajan *et al.*, 1999). Several *in vitro* and *in vivo* studies in yeast provide evidence for the occurrence of both models (Cho *et al.*, 1998; Wilson *et al.*, 1996 and Gregory *et al.*, 1999). Support for the first model is provided by studies showing co-immunoprecipitation of components of the SWI/SNF complex with the RNA Pol II holoenzyme. Using antibodies against SRB, which has been shown to be tightly associated with the C-terminal repeat domain of RNA Pol II in the holoenzyme, and SWI/SNF components, in both cases showed co-immunoprecipitation of SRB with SWI/SNF components. Subsequent purification of the RNA Pol II holoenzyme showed that the SWI/SNF proteins co-elute with known components of the holoenzyme. These data indicate that SWI/SNF, together with SRB, is bound to the RNA-pol II holoenzyme. This binding facilitates the recruitment of SWI/SNF to the gene promoters by RNA-Pol II and the subsequent remodelling of the chromatin template (Wilson *et al.*, 1996).

An example of the second model is provided by studies on the HO gene in yeast. Chromatin immunoprecipitation studies at the HO-promoter have shown that SWI/SNF is recruited to the promoter via an interaction with the swi5 transcription activator (Cosma *et al.*, 1999). These data are consistent with *in vitro* studies, which have shown that SWI/SNF, when purified or in whole cell extracts, directly interacts with swi5 (Neely *et al.*, 1999). After recruitment by swi5 of the SWI/SNF to the HO-promoter, this can remodel the nucleosomal array thus allowing the binding of other factors involved in the transcription of the HO gene.

The two models mentioned also imply different roles for SWI/SNF remodelling at different promoters in transcriptional activation. When a transcriptional activator, like swi5, can bind strongly to DNA, SWI/SNF may play a role in further remodelling the nucleosomal array to allow other proteins to bind such that transcriptional activation can take place. However, when a transcriptional

activator has only a low binding activity for DNA, as is the case for activators of the Gal4 gene (Burns and Peterson, 1997), SWI/SNF activity is necessary to remodel the nucleosomal array before the first activator can bind, which will then be followed by the binding of other proteins of the transcriptional complex (Neely *et al.*, 1999; Natarajan *et al.*, 1999 and Wallberg *et al.*, 2000). Interaction between SWI/SNF and activators could thus vary at different promoters or SWI/SNF could function in multiple ways at a single promoter depending on the activator present.

Interactions of human SWI/SNF homologs with tissue specific transcriptional activators have also been reported. The hSWI/SNF-containing E-RC1 complex purified from erythroid cells directly interacts with the zinc finger domain of EKLF and is functionally important for the efficient transcription of the β-globin gene in *in vitro* assays (Armstrong *et al.*, 1998; Kadam *et al.*, 2000 and Lee *et al.*, 1999).

Besides the "classical" role in gene activation, SWI/SNF also displays repressor activities (Sudarnasam *et al.*, 2000 and Holstege *et al.*, 1998). Possible mechanisms by which this repressors activity could take place have been reported. It has been shown that SWI/SNF can remodel nucleosomes between two states, "inactive" and "active" with equal ability (Schnitzler *et al.*, 1998), which could indicate that the SWI/SNF can also repress genes by creating an inactive nucleosomal state. SWI/SNF could bind repressors, instead of transcriptional activators, and finally the combination of SWI/SNF with histone modifications could also lead to repression (Strahl and Allis, 2000). The Mi-2 complex (now called NuRD for nucleosome remodelling and deacetylation) demonstrates that last point. This complex is composed of the Mi-2 protein, which is a member of the Snf2 superfamily of ATP-ases and the RPD3 protein, which is a HDAC. More complexes containing an ATP-ase and HDAC subunit have been identified and are thought to repress transcription through the combined properties of ATP-dependent remodelling and HDAC activity (Tong *et al.*, 1998 and Zhang *et al.*, 1998).

Additional ATP-dependent chromatin remodelling complexes besides SWI/SNF have also been identified and can be classified into four families depending on the ATPase activities they contain: SWI/SNF, ISWI, Mi-2 and Ino-80 (Cairns 1998; Alfas and Kingston, 2000; Shen *et al.*, 2000; and Vignal *et al.*, 2000).

The ISWI family has also been studied extensively. Family members of ISWI all contain the same ATP-dependent chromatin remodelling subunit ISWI (Corona *et al.*, 1999). Three ISWI-containing members have been isolated from *Drosophila*: NURF, ACF and CHRAC (Ito *et al.*, 1997; Varga-Weisz *et al.*, 1997 and Tsukiyama and Wu, 1995). NURF contains 4 subunits and has a molecular mass of 500 kDa (Tsukiyama *et al.*, 1995 and Xiao *et al.*, 2001). Like SWI/SNF, NURF is involved in allowing the binding of specific activators to DNA, for example, GAGA to the hsp70-promoter (Tsukiyama *et al.*, 1994) or of activator proteins to the Gal4-E4-promoter (Mizuguchi *et al.*, 1997 and 2001). NURF binds stably to nucleosomes in an ATP-dependent manner and facilitates nucleosome repositioning to allow factor binding. Unlike SWI/SNF, however, ISWI ATPase activity is simulated by nucleosomal DNA, whereas SWI/SNF also acts on free DNA. ACF contains 2 subunits and can order an evenly spaced array of randomly assembled nucleosomes on DNA and can further mobilize nucleosomes to facilitate interactions with DNA-binding proteins (Ito *et al.*, 1997 and 1999). Related to ACF is CHRAC, which contains 4 subunits with the acf1 subunit also present in the ACF complex (Ebenharter *et al.*, 2001). CHRAC also appears to function as a nucleosome-spacing factor thus enhancing the accessibility of chromatin (Langst and Becker, 2001).

To facilitate transcription, co-operation between the chromatin remodelling complexes and complexes affecting histone modifications must take place. For the HO gene in yeast, it has been shown that both SWI/SNF and the HAT complex SAGA are recruited one after the other to the promoter. Both complexes are necessary for transcriptional activation (Bhoite *et al.*, 2001).

In conclusion, the process of chromatin remodelling in transcriptional activation encompasses three collaborating events: first the ATP-dependent chromatin remodelling complexes are recruited to the target of activation or repression, where they alter nucleosomal structure, secondly the histone modifying enzymes like HATs, HDACs and methyltransferases are recruited, which further modify the chromatin structure, and enable the third step, the binding of a third group of proteins, which alter the non-histone part of the chromatin thus affecting transcriptional activity.

*Polycomb and Trithorax*

Two other groups of proteins that regulate transcription at the chromatin level are the Polycomb group (PcG) and the Trithorax group (TrxG). In general the PcG proteins are repressors that maintain the off-state of a gene whereas the TrxG proteins are activators that maintain transcription (Simon and Tamkun, 2002). These complexes appear to be involved in the maintenance and not in the establishment of active or repressive states of chromatin and transcription (Simon, 1995 and Kennison, 1995).

The function of PcG was first suggested by mutations in *Drosophila*. These mutants showed posterior transformations caused by the derepression of homeotic genes (Simon, 1995). Many of the TrxG members were identified in genetic screens as suppressors of the phenotypes of PcG mutations (Kennison and Tamkun, 1988).

It is now clear that PcG and TrxG are not involved in the establishment of the precise patterns of expression of the homeotic loci, which are set up by (short-lived) specific transcription factors earlier in *Drosophila* development (Bienz and Muller, 1995). After the "decay" of these proteins, the repressed or activated states of homeotic genes are maintained by the PcG and TrxG proteins (Pirrotta, 1998).

Mammalian homologs of PcG and TrX-G have also been identified (Jacobs and van Lohuizen, 2002). In addition to their function in the control of homeotic gene expression, PcG and Trx-G complexes are also involved in the control of several other processes including cell proliferation, haematopoiesis, neuronal development, sex determination and cell cycle regulation (Jacobs and van Lohuizen, 2002 and Lund and van Lohuizen, 2002). Oncogenic and tumour suppressive activities have been implicated for PcG and Trx-G proteins. An example of this is the mammalian TrX-G MLL homologue, which has been shown to be involved in myeloid and lymphoid leukemias (Rubnitz *et al*., 1996 and Corral *et al.*, 1996). The mammalian PcG BMIl gene has also been associated with lymphogenesis and was originally identified in a screen for oncogenes (van Lohuizen *et al.*, 1991 and Haupt *et al.*, 1991). The BMI1 gene has been shown to interact with distinct heterochromatin domains in tumour cell lines. The Ink4 tumour suppressor locus has been shown to be a critical target of the transcriptional repressor BMI1. In BMI1 knockout mice an increase in apoptosis is observed, which can be rescued by deleting the Ink4 locus (Jacobs *et al.*, 1999). Recent experiments, in which lung tumours were compared with healthy lung samples, indeed showed an inverse correlation between the expression of the Ink4 locus and the BMI1 gene. Ink4 expression levels were high in normal tissues and low in the tumour tissues, with BMI1 expression levels showing the opposite patterns. These data support the suggestion for a role of BMI1 in carcinogenesis (Vonlanthen *et al.*, 2001).

There are specific elements through which PcG and TrxG confer their action, the so-called Polycomb and Trithorax response elements (PRE and TRE). These elements can vary in size from 100 bp to a few kb and different PcG or TrxG complexes can bind and confer their action via discrete response elements within the same regulator region of the gene (Tillib *et al.*, 1999).

It appears that TRE and PRE elements are located close to each other, sometimes separated by 30-40 bp (Brock and van Lohuizen, 2001). This suggests an intermingling of these elements in the activation and repression of genes. Evidence for this came from studies in *Drosophila* in which it has been shown that maintenance of repression through the *Fab-7* and *Scr* PREs is affected by both PcG and TrxG mutations (Hagstrom *et al.*, 1997; Gindhart and Kaufman, 1995 and Cavalli and Paro, 1995). It was therefore suggested to rename PREs and TREs as maintenance elements (ME) (Cavalli and Paro, 1995).

How repression or activation through MEs is regulated is not clear. It is not a question of competition because both PcG and TrxG can bind to the same ME regardless of the activity of the target locus. This was shown in studies in which the repression of a gene by PcG proteins was reversed by the co-expression of Trx-G proteins (Cavalli and Paro, 1995 and Zink and Paro, 1995). This shows that although a target locus is suppressed, Trx-G proteins can still bind to their recognition site and reverse the activity of the target locus. It could be a case of other proteins recruiting the PcG or TrxG to the right ME, or rather a histone modification that marks the ME such that only one of the two protein groups can bind (Ekwall *et al.*, 1997).

Insight in the molecular basis of PcG-mediated repression and TrxG-mediated activation has been obtained through the isolation and characterization of protein complexes. Presently there are two

complexes that contain PcG proteins: a 1-2M Da complex, PRC1 (Shao *et al.*, 1999 and Saurin *et al.*, 2001), and a 600 kDa complex, ESC/E(Z) (Tie *et al.*, 2001 and Chang *et al.*, 2001). In addition, four complexes containing TrxG proteins have been identified: a 2 MDa BRM complex, a 2 MDa ASH1complex, a 0.5 MDa ASH2 complex, and a 1 MDa TRX complex (TAC1) (Papoulas *et al.*, 1998 and Petruk *et al.*, 2001).

These complexes appear to have different activities and may thus play different roles in gene repression and activation. PRC1 inhibits the remodelling of nucleosomal arrays by SWI/SNF (Breiling *et al.*, 1999 and Francis *et al.*, 2001) and interacts with TBP-associated factors, as shown by co-purifications (Saurin *et al.*, 2001). The functional consequence of this interaction is, however, not clear. The second PcG complex ESC-E(Z) co-immunoprecipitates and co-fractionates with HDAC1 and HDAC2 in extracts from human cells (van der Vlag and Otte, 1999). Purification of the complex from *Drosophila* extracts followed by mass spectrometric analysis and co-immunoprecipitations confirm this finding (Tie *et al.*, 2001). This indicates that HDAC function in PcG repression is linked via the ESC-E(Z) complex. For the trithorax complexes also different activities have been described. The BRM complex has been suggested to play a role in ATP-dependent chromatin remodelling, since *in vitro* studies have shown that the BRM complex catalyses ATP-dependent alterations in nucleosomal organisation (Kal *et al.*, 2000) and TAC1 has been shown to have a role in histone modification (Petruk *et al.*, 2001). The characterisation of the TAC1 complex has shown that TAC1 contains a member of the CBP/p300 HAT family. When this subunit of TAC1 is mutated there is a decrease in transcription of HOX genes and reporter genes (Petruk *et al.*, 2001).

How do these complexes together play a role in the maintenance of "on" and "off" states of chromatin? A model for the multistep action of TrxG and PcG complexes has been proposed (Simon and Tamkun, 2002). In the Trx-G pathway, TAC1 first acetylates histone tails, which may lead to the recruitment of the BRM remodelling complex, thus remodelling the nucleosomal array and rendering the DNA more accessible to the transcriptional machinery. In the PcG pathway, the ESC-E(Z) complex, which acts earlier in development, first deacetylates nucleosomes, thus creating a histone code that attracts subsequent binding of PRC1. When PCR1 is bound, it can maintain the silenced state by counteracting remodelling complexes (Simon and Tamkun, 2002).

### *DNA methylation*

DNA methylation represents another mechanism for global gene repression. When a promoter and/or other regulatory sequences of a gene become methylated, gene expression is repressed. The C-5 position of cytosine in 5'-CpG-3' dinucleotides is the target for DNA methylases. CpGs from both DNA strands are symmetrically methylated to result in the final silencing of a gene (Antequera and Bird, 1993 and Ohki *et al.*, 2001). 60-90% of all the CpG sequences present in mammalian genomic DNA are methylated.

CpGs that are not methylated are often clustered in so-called CpG-islands, usually found in the regulatory elements (Antequera and Bird, 1993). It has been previously shown that, at least in some cases, DNA methylation can prevent the binding of transcription factors to their DNA binding sites, as is the case for CREB, for example (Iguchi-Ariga and Schaffner, 1989). However, other transcription factors, like Sp1, can still bind to their methylated recognition site, indicating that another mechanism is also playing a role in the inhibition of the binding of transcription factors by methylation. *In vitro* studies have shown that specific proteins (MeCP1 and MeCP2) can repress transcription by binding to methylated DNA, thus preventing the binding of transcription factors (Boyes and Bird, 1991 and 1992). *In vivo* both mechanisms may play a role in transcriptional regulation.

In addition, a relationship between DNA methylation and chromatin structure was hinted at when early on it was shown that artificially methylated DNA can result in a distinctive chromatin structure when integrated into the genome (Keshet *et al.*, 1986). Further studies implicated a connection between histone deacetylation and DNA methylation when it was shown that a specific inhibitor for HDACs, trichostatin A (TSA), can substitute for the DNA-demethylating agent 5-aza-2'-deoxycytidine, in that both could restore transcription from a repressed methylated template (Kass *et al.*, 1997; Chen and Pikaard, 1997 and Eden *et al.*, 1998). Another link between DNA methylation and histone deacetylation was the finding that MeCP2, a protein that specifically binds to methylated DNA via its methyl CpG-binding domain (Lewis *et al.*, 1992; Nan *et al.*, 1993 and Nan *et al.*, 1996), co-

purifies with a component of the mSin3A/HDAC complex (Nan *et al*., 1998 and Jones *et al.*, 1998). Co-immunoprecipitations showed that mSin3A is the preferred partner of MePC2, suggesting that MeCP2 recruits HDAC to the chromatin template of methylated DNA (Nan *et al*., 1998). This interaction can provide a mechanistic link between DNA-methylation and chromatin structure and transcriptional repression.

Recently, another intriguing connection between DNA-methylation and histone modifications has been made. It was shown in *Neurospora* that the dim2 gene, which encodes a protein with a C-terminal domain homologous to known histone methyltransferases, is responsible for all cytosine methylation (Kouzminova and Selker, 2001). This observation may provide a direct evolutionary link between DNA methylation and histone methylation in transcriptional repression.

There are two classes of DNA methylases: the de novo methylases, Dnmt3a and Dnmt3b, and maintenance methylase, Dnmt1. The de novo methylases add a methyl group to unmethylated CpG-pairs on both strands of DNA without high sequence specificity (Okano *et al.*, 1998 and Lyko *et al.*, 1999). Maintenance methylation by Dnmt1 provides the methylation of hemi-methylated CpG-pairs on the newly replicated DNA strand (Lyko *et al.*, 1999 and Bestor *et al.*, 2000).

Both de novo methylases and maintenance methylases play a role during mammalian embryogenesis. During the embryo implantation stage, genome-wide methylation patterns are lost (with very few exceptions, e.g. imprinted regions). The new methylation patterns after this period of embryogenesis are established through the combined action of de novo methylases and subsequently by maintenance methylases (Okano *et al.*, 1999). Knockout mouse models for both types of methylases have shown that both Dnmt1 and Dnmt3b are necessary for embryonic development and Dnmt3a for postnatal development (Lie *et al.*, 1996 and Okano *et al.*, 1999).

Besides these enzymatic activities, Dnmt1 has been shown to interact with HDAC1 and HDAC2. It was shown that during late S-phase HDAC2 co-localises with Dnmt1 in heterochromatin (Rountree *et al.*, 2000). This interaction could indicate that Dnmt1 directly represses transcription together with HDAC. Dnmt3a and b have been implicated to play a role in differentiation and cell growth, because it methylates unmethylated CpG-pairs creating hemi-methylated pairs which can be methylated by the maintenance methylases. In addition, Dnmt3a and b have been implicated to play a role in altered methylation patterns in tumourigenesis (Nakao, 2001).

### *DNase sensitivity*

Chromatin associated with transcriptionally active states is more sensitive to the action of nucleases (Weintraub and Groudine, 1976 and Wood and Felsenfeld, 1982), for example, DNase I. DNase I nicks double stranded DNA in a non-sequence-specific manner, but with a distinct preference for active chromatin in contrast to inactive chromatin. This distinction has often been used as a molecular tool to study chromatin structure.

Two types of DNase I sensitivity have been recognised: general sensitivity and hypersensitivity. General sensitivity to DNaseI digestion is found in areas of gene domains that are transcriptional active or potentially active. Hypersensitivity to DNase I digestion is found in smaller areas of around 200-600 bp in size within the areas of general DNase I sensitivity. These sites are often located within the regulatory elements of genes like enhancers, promoters and locus control regions and are called hypersensitive sites. The first demonstration of a DNAse I hypersensitive site was shown by Wu *et al.* in 1979 in the *Drosophila* hsp 70 gene. Hypersensitive sites are thought to reflect areas with a less dense nucleosomal packaging which contain multiple sequences for promoter-specific DNA binding proteins (Emerson *et al.*, 1985). However, the exact structural basis for DNase I sensitivity is not clear. Chromatin features such as the absence of linker histone H1 and increased histone acetylation in DNase I sensitive areas have been suggested as an explanation (Smith *et al.*, 1984 and Hebbes *et al.*, 1994). DNase I hypersensitive sites can be classified as constitutive, inducible, tissue-specific and developmental stage specific (review Gross and Garrard, 1988).

An example of a gene locus containing well characterised DNase I hypersensitive sites is the human β-globin locus. Two kinds of hypersensitive sites were mapped within the locus. Minor hypersensitive sites are observed when higher concentrations of DNase I are used and have been found close to the promoters and local enhancers of the β-globin genes (Stalder *et al.*, 1980; Charnay *et al.*, 1984 and Forrester *et al.*, 1986). Major hypersensitive sites were mapped in the locus with lower concentrations

of DNase I. Of these, five map within the locus control region upstream of the genes and one maps to the 3' of the locus (Tuan *et al.*, 1985; Grosveld *et al.*, 1987 and Forrester *et al.*, 1987).

## Transcription initiation, elongation and termination

After chromatin has been made accessible by remodelling and histone modification, the transcriptional machinery can "land" on the DNA and the process of transcription can take place. Gene transcription is carried out by three classes of polymerases: RNA polymerase I, II and III. The three RNA polymerases have different gene specificities: RNA pol I is responsible for transcribing ribosomal RNA, RNA Pol II synthesises all the heterocellular genes and finally RNA Pol III is responsible for the transcription of tRNA, 5 S RNA and small nuclear RNA (Ogbourne and Antalis, 1998). The initiation of transcription is a highly regulated event and plays a major role in gene regulation. The transcription of the globin genes plays an important role in this thesis, therefore, I will focus on the role of RNA Pol II in the transcriptional process.

RNA Pol II cannot by itself initiate transcription at a specific promoter region and requires a set of transcription factors to do so. These factors have to assemble at the promoter of the gene before transcription can take place. This process can be influenced by regulatory signals within the cell and in this way transcription can be accelerated or repressed. The factors needed for transcriptional activation were identified as basal transcription factors (TF) which assemble into a complex at the promoter of the gene and subsequently recruit RNA Pol II to the promoter (Johnson and McKnight, 1989). Most of the TFs have two distinct domains, one for the binding to the specific regulatory sequences in the DNA and one that interacts with the transcription machinery and accelerates the rate of transcription initiation. The activator domains of TFs can be classified into three classes; acidic, glutamine-rich and proline-rich domains (Alberts *et al.*, 1994).

Some steps in the assembly of the transcriptional complex can be rate-limiting and the TFs play a role in overcoming this limiting step. For example the TFs with acidic domains have been shown to overcome the rate-limiting step of TFIIB entry into the complex (Lin and Green, 1991).

TFs themselves are also regulated. This can be through ligand binding, protein phosphorylation, addition of a secondary subunit or the release from a tight complex with a specific inhibitor (Berk, 1989 and Hunter and Karin, 1992).

For the specific and regulated initiation of transcription a large transcriptional complex of RNA Pol II together with TFs, has to be formed (Buratowski *et al.*, 1989 and Conaway and Conaway, 1993). The assembly of the transcriptional apparatus starts with the binding of TFIID to a TATA-sequence (Fig. 2). This sequence is located around 30 bp upstream of the transcription start site. TFIID itself is composed of several subunits: TBP and TBP-associated factors (TAFs) (Dynlacht *et al.*, 1991; Tanese *et al.*, 1991 and Zhou *et al.*, 1992). TBP is the factor responsible for binding to the TATA-box and the TAFs are needed in order to mediate transcription regulation by upstream activating factors. After TFIID, TFIIA is recruited to the complex. TFIIA is important for the stabilisation of TFIID on the promoter and to counteract inhibitory factors, which cause TFIID to dissociate (Zawel and Reinberg, 1993). The TFIID-TFIIA complex undergoes a conformational change to allow TFIIB to bind (Horikoshi *et al.*, 1988 and Chi and Carey, 1996). TFIIB then facilitates the binding of the non-phosphorylated form of the TFIIF-RNA-pol II sub-complex to the transcription complex (Maxon *et al.*, 1994). The complex thus assembled is thought to melt the DNA around the start site of transcription to form an open complex. To be able to continue into the next step of transcription, that of elongation, promoter clearance has to take place. For this step, TFIIH and TFIIE are thought to be responsible (Goodrich and Tjian, 1994). TFIIE is thought to recruit TFIIH to the initiation complex with TFIIH being responsible for the actual promoter clearance. TFIIH has many subunits of which one is a protein kinase. This kinase phosphorylates the C-terminal Domain (CTD) of RNA Pol II (Maxon *et al.*, 1994), resulting in the release of RNA Pol II from the initiation complex and transcription can proceed.

After these steps, the next action of transcription, elongation, can take place. There are two phases in transcription elongation. The early phase, which is characterised by a hypophosphorylated CTD of RNA Pol II (by DSIF and Factor 2) and a late elongation phase, which is characterised by a hyperphosphorylated CTD. The hypophosphorylated RNA Pol II has to be re-phosphorylated to

proceed with elongation. Several factors have been found to be involved and include DmsII, TFIIF, ELL and Elongin (for review see Shilatifard *et al.*, 1997). If re-phosphorylation of the CTD does not occur, elongation is blocked 500 bp downstream of the start site (Chodosh *et al.*, 1989). If CTD phophorylation does take place, the late elongation phase proceeds and nascent RNA (nRNA) is synthesised until it is stopped by a termination signal.

Termination of transcription by RNA polymerases, at least for RNA pol I and RNA Pol III, appears to be mediated by specific DNA sequences. RNA pol I recognises a specific18 bp recognition sequence downstream of the mature 3' end of the newly synthesised RNA (Proudfoot, 1991). RNA pPol III terminates at uridine 2, 3 or 4 in a stretch of four uridines which are surrounded by GC pairs at the 3'end of the gene (Platt, 1986). The specific termination signal for RNA Pol II, however, has still to be determined.

After termination of transcription the nascent RNA (nRNA) matures into mRNA. During maturation, both the 5'end and the 3'end of the nRNA are covalently modified. The 5'end is capped by the addition of a methylated G nucleotide. This cap plays an important role in the initiation of protein synthesis. The 3'end of the nRNA is modified by the addition of a poly-A tail. This tail is important for the export of mRNA from the nucleus and for the stability of the mRNA in the cytoplasm. It also serves as a recognition signal for the ribosomes. Furthermore the intron sequences are removed from the nRNA during the process called splicing, leaving just the mRNA (Alberts *et al.*, 1994).

## Transcriptional regulation

The regulation of transcriptional activation is achieved through the interactions of distinct elements such as enhancers, silencers and insulators. These elements are discussed below.

### *Enhancers and silencers*

Enhancers were originally identified in transient transfection assays as sequences that were able to activate transcription in an orientation and distance independent manner with respect to the promoter to which they are linked (Banerji *et al.*, 1981 and Moreau *et al.*, 1981).

Enhancers can be found located from within the promoter of a gene up to several kilobases downstream and/or upstream of a gene. The first enhancer to be identified using transient transfection assays was a 200 bp element from the SV40 virus (Banerji *et al.*, 1981). The sizes of enhancer elements vary between 50 bp to 1.5 kb and contain a collection of protein binding sites to which both tissue specific and ubiquitous transcription factors can bind. By multimerising transcription factor binding sites a functional enhancer can be created *in vitro* (Zenke *et al.*, 1986). Interactions with different proteins give rise to cell-type and developmental stage specificity in the action of enhancers (Dynan, 1989), the first tissue-specific enhancer transcribed is the B-cell specific enhancer located at the 3'-end of the rat IgH locus (Pettersson *et al.*, 1990).

Enhancers have to interact directly with the promoter of the gene, as they themselves cannot initiate transcription. Three models have been proposed for this interaction. The prevailing favoured model is that of DNA looping. The proteins bound to the enhancer interact with the proteins bound to the promoter and intervening DNA sequences are looped out (Ptashne, 1988 and Ptashne and Gann, 1997). Experiments in which it was shown that the transcription of a β-globin gene could be stimulated using a biotin-streptavidin bridge between the β-promoter and enhancer sequences which resided on different molecules (Muller *et al.*, 1989) and transvection studies in *Drosophila*, in which an enhancer is shown to activate a gene in trans (Henikoff,1997), have provided support for the looping model.

Two other models have been proposed. First, a tracking mechanism, in which the transcription factors use the enhancer site as a landing platform for the assembly of a nucleoprotein complex which then travels along the chromatin fiber to activate the first gene it encounters (Heredeen *et al.*, 1992). Second, the accessibility model, in which the enhancer provides a favourable chromatin environment for transcription of a gene, by antagonising repressive chromatin structures (Weintraub, 1988; Walters *et al.*, 1995 and 1996). The tracking and accessibility models are not able to explain the results of the trans-activation experiments described above and are thus less likely to explain the physical

interaction between an enhancer and promoter. The interaction models described also apply with some variations to the function of the LCR of the β-globin locus. These models are described in the section concerning the locus control region.

The presence of an enhancer identified in transfection assays as part of a transgene, is in most cases unable to direct normal expression levels of the transgene in mice or in stably transfected cells. Instead, expression levels of the transgene are often low and the tissue and developmental regulation are disturbed. This is the result of random integration of the transgene into the host chromosomes, such that integration of the same transgene at different positions often shows different expression patterns. Classical enhancers are not able to shield a transgene from these effects, resulting in the absence of normal expression levels. This led to the identification and formulation of a new functional element: the locus control region. Such an element was first identified in the human β-globin locus (Grosveld *et al.*, 1987)(for more information about the human β-globin locus control region see section on the locus control region). More LCR elements have been identified since then, like the LCR from the mouse β-globin locus and the LCR of the CD2 gene (Greaves *et al.*, 1989 and Festenstein and Kioussis, 2000).

A LCR is functionally defined as an element that can confer position independent, copy number dependent and tissue specific activation of a gene in transgenic mouse assays (Grosveld *et al.*, 1987). A LCR resembles an enhancer in that it contains a high concentration of binding sites for specific transcription factors and it can activate genes over long distances (Talbot *et al.*, 1989; Dillon and Grosveld, 1993 and Kioussis and Festenstein, 1997). At the same time, a LCR is distinguishable from a classical enhancer by its property to confer chromosomal position independent transcription in transgenic assays (Grosveld *et al.*, 1987 and Blom van Assendelft *et al.*, 1989). Whereas a LCR can counteract the negative influences of chromatin, the classical enhancer is affected by the structure of surrounding chromatin at the site of transgene integration (Wilson *et al.*, 1990). LCRs and enhancers are therefore defined by different functional assays. Transient assays in which DNA is not integrated in chromatin can be used to identify enhancers, whereas, LCR activity will not necessarily be observed in these assays (Hug *et al.*, 1992; Tuan *et al.*, 1989; Pruzina *et al.*, 1991; Philipsen *et al.*, 1990 and Fraser *et al.*, 1990). Some elements will merely show activity in one of the assays, such as HS3 and HS4 of the human β-globin LCR, which are only active in a chromatin context (Pruzina *et al.*, 1991; Philipsen *et al.*, 1990 and Fraser *et al.*, 1990), whereas other elements, like HS2 of the LCR, show activity in both assays (Tuan *et al.*, 1989; Sorrentino *et al.*, 1990 and Talbot and Grosveld, 1991). This indicates that there is functional overlap between a classical enhancer and the LCR.

Silencer elements resemble enhancers, in that they contain multiple factor binding elements and they act on a promoter in an orientation and position independent manner (Brand *et al.*, 1985). However, instead of activating a gene they suppress it. The first silencing elements were identified in the yeast mating type loci (Brand *et al.*, 1985), and subsequently in many more gene systems, like the human ε-globin gene (Cao *et al.*, 1989) and the human thyrotropin-β gene (Kim *et al.*, 1996). Often, enhancers and silencers work together to ensure that gene expression is tissue and developmental stage specific (Huang *et al.*, 1993 and Trepicchio *et al.*, 1993). To date, a number of silencers have been described, which all confer different actions on genes and their promoters (Ogbourne and Antalis, 1998).

Silencers are divided into two functional groups: classical silencers, as described by Brand, and negative regulatory elements (NREs). Classical silencers are thought to confer their action through the binding of repressor proteins and subsequent interaction with promoters, and/or by changing chromatin structure or via physical interference with the transcription initiation process by blocking TFIIH activity at the transcription start site (Kim *et al.*, 1996 and Liu *et al.*, 1996). However, no definite mechanism has yet been described.

NREs are position-dependent silencers and have been identified in promoters, introns and exons (Clark and Docherty, 1993). They passively repress transcription by binding repressor proteins and thereby physically inhibit the binding of transcription factors or other factors that play a role in transcriptional activation (Dong and Lim, 1996; Gumucio *et al.*, 1993 and Peters *et al.*, 1993). An example of NRE function can be seen in the deletion of an element of 70 bp in the third exon of the human α1-chimaerin gene. This deletion resulted in a 5-6 fold increase of promoter activity (Dong and Lim, 1996). Addition of the NRE element to a heterologous TK element showed an orientation independent but position dependent repression of promoter activity (Dong and Lim, 1996).

*Insulators*

The presence of elements like insulators was proposed when the idea of different functional domains in the genome was put forward. Inherent to this was the suggestion that elements must exist that would prevent enhancers from acting on the wrong domain (Eisenberg and Elgin, 1991).

Two predictions were made for the function of such elements. The first one predicts that when such an element is placed on both sides of a transgene it should protect, or insulate, this from chromosomal position effects. Secondly, when placed between an enhancer and a promoter, activation of the gene should be blocked by this element (Eisenberg and Elgin, 1991). Using these assays, Kellum and Schedl showed that the scs and scs' elements from the 87A7 *Drosophila* hsp70 heat shock locus could block the expression of a gene when placed between the enhancer and promoter, while other elements from the same locus could not. In the same studies they also showed that the scs and scs' elements themselves do not contain enhancer activities (Kellum and Schedl, 1992). In addition, using the mini *white* gene of *Drosophila* as a reporter, they showed by scoring for eye colour in transgenic flies that the scs and scs' elements could insulate a gene from position effects (Kellum and Schedl, 1991).

The assays led to the further description of additional insulators of which the best studied, in addition to the scs and scs' sequences (Kellum and Schedl, 1991), are the *gypsy* transposon (Corces and Geyer, 1993) and HS4 of the chicken β-globin locus (Chung *et al.*, 1993).

Enhancer blocking assays showed that the binding sites for *Drosophila* suppressor of Hairy-wing [su(hw)] protein in the *gypsy* retrotransposon also confer insulator properties (Corces and Geyer, 1993). Recent results suggest, that the gypsy insulator affects the enhancer-promoter interactions by affecting chromatin structure. It was shown that in the presence of the Su(Hw) protein and a second component, the modifier of mdg4 protein, accessibility of DNA for nucleases was increased in the promoter-proximal but not in the promoter-distal region (Chen and Corces, 2001).

The first vertebrate insulator described was the HS4 of the chicken β-globin locus (Chung *et al.*, 1993). This element was tested using the enhancer blocking assays in human erythroid cell-lines. It was shown that cHS4 can insulate a reporter gene containing the β-globin promoter from the effects of the LCR (Chung *et al.*, 1993). The minimal core required for insulation contains binding sites for the CTCF transcription factor. CTCF binding sites are necessary and sufficient for the insulating effects of HS4 in this assay (Bell *et al.*, 1999 and Bell *et al.*, 2001). The core of HS4 was also tested in the position effect insulation assay. The positioning of two cHS4 elements around a transgene showed that the cHS4 can protect a transgene from position effects, this has been shown both in *Drosophila* and transgenic mice (Chung *et al.*, 1993 and Pikaart *et al.*, 1998). However, for this insulation the CTCF sites were not necessary as was shown by mutation analysis. These observations suggest that there are two overlapping insulator activities within the HS4 of the β-globin locus.

Exactly how insulators work is still a matter of debate. Two models have been proposed to account for the basis of insulator function: the local interaction and the structural model (Zhan *et al.*, 2001 review). The first model suggests a local interaction between proteins bound to the insulator element and proteins bound to the enhancer. This blocks the interaction of proteins in the enhancer with the promoter and the gene is thus repressed. Variations on the principle put forward by this model are the decoy looping model (Fig.3A), in which the looping of the enhancer to the promoter is blocked by the formation of an additional loop (Geyer, 1997 and Gerasimova and Gorces, 1998), and the derailment of tracking model (Fig. 3B) (Dorsett, 1993), in which the spreading of activation signals from the enhancer are blocked. The two models can be distinguished by the directionality of the insulator and the bidirectionallity of the enhancer.

The structural model proposes that insulating function is coupled to a structural role in higher order nuclear organisation. This could be established by the formation of loops by insulator proteins and attachment to the nuclear scaffold at MARs and SARs (Cai and Chen, 2001 and Murayova *et al.*, 2001). This would constrain the chromatin thus hindering transcriptional activation.

The HS5 of the human β-globin locus has also been reported to contain insulator properties like cHS4 (Li and Stamatoyannopoulos, 1994). Studies in which the LCR is reversed, show a downregulation of the expression of all the globin genes, indicating that HS5 might shield the genes form the other hypersensitive sites (Tanimoto *et al.*, 1999). However, other studies show contrary results. For example, the deletion of HS5 of the mouse β-globin locus, which is highly homologous to the human HS5, has only a minimal effect on transcription, indicating that HS5 is not necessary to

protect the β-globin genes against surrounding chromatin (Bender *et al.*, 1998 and Farell *et al.*, 2000). Furthermore HS5 cannot shield the globin genes from the effect of the LCR in transgenis mice (Zafarana, thesis 2001). These observations indicate that HS5 by itself does not have (strong) insulator capacities. With the different studies contradicting each other it is not clear whether HS5 does or does not act as an insulator.

## Regulation of gene expression

The dogma on gene expression is, that transcription activation and gene expression are following a rate or analog model, also called deterministic or gradient model. This model implies that gene expression levels are regulated by the rate of transcription as response upon a stimulus in the environment of the cell. An other model for gene activation is a probability or stochastic model. This model states that the probability and frequency of gene activation determines the level of gene expression rather then the increase of gene expression per cell upon a stimulus.

Several *in vitro* studies using inducible reporter gene assays show evidence for a probability or stochastic gene expression model. They find bimodal expression patterns, either a cell does or does not express and upon induction just more cells start expressing instead of higher expression levels per cell (Femino *et al.*, 1998; Ross *et al.*, 1994 and Fiering *et al.*, 1990). An other feature of a probability or stochastic model, a normal distribution of expressing cells, is also noted in these studies.

If stochastic models are indeed true for general gene transcription, how are then the high and constant levels of gene expression established and/or maintained? The probability that a gene will express is based upon different probabilities, the intrinsic probability of the gene itself and the transcription probabilities of the gene's regulatory elements which are occupied by diverse transcription factors. These probabilities are multiplied during the process of transcriptional activation, leading to, if required, a high probability for gene expression. Constant expression levels of a gene are reached if the half-live of the mRNA or protein of a gene is longer then the time required for a second round of transcription. If the half-live is shorter then this period, the transcription of the gene will be observed in pulses (Hume, 2000).

Data on enhancer function too indicate a probability or stochastic nature of gene expression. Also on enhancer function the dogma is a rate model: the presence of an enhancer extends the rate of transcription and thereby increases levels of expression of a gene. However, this idea is the result of the early studies on enhancer function, which were done using bulk assays and not on single-cell level. This way no distinction could be made between a rate and a probability model (Fiering *et al.*, 2000). Single cell studies, using the β-gal gene with the SV-40 or β-globin HS2 enhancer, show now that it is not a higher level of gene expression per cell, but the number of cells expressing that is influenced by the enhancer (Walter *et al.*, 1995). Enhancers thus affect rather the on/off status of a cell then the rate of transcription of a gene. This indicates that enhancers act in a stochastic fashion to increase the probability that a gene will be transcribed (Fiering *et al.*, 2000).

If indeed a probability or stochastic model determines gene activation and expression it also indicates that any (inducible) gene will show a percentage of mono-allelic expressing cells, which has nothing to do with specific regulation of expression levels. Several studies describe and mention indeed mono-allelic expression or stochastic gene expression, however, these are all studies on genes involved in cellular responses, dosage compensation or lineage commitment (Goto and Monk, 1998; Nemazee, 2000; Chess *et al*., 1994; Held *et al*., 1999; Hollander *et al*., 1998; Bix and Locksley, 1998 and Riviere *et al*., 1998). Recently several studies have been published, which indeed show allelic expression patterns indicative of a stochastic nature for general gene activation and expression (Chapter 2; Elowitz *et al.*, 2002 and Obzudak *et al.*, 2002).

If all this leads to the acceptance of a probabilistic or stochastic nature for general gene expression, also the description of gene transcription has to be changed. "The production of mRNA occurs in *pulses*. The *mean frequency* of pulses is the major determinant of mRNA production and is determined by the *probability* of formation of a preinitiation complex" (Hume, 2000).

## Erythropoiesis

Erythropoiesis is a multi-step process in which early progenitor cells differentiate to become erythrocytes (Fig. 4). This process starts when a multipotent haematopoietic stem cell (HSC) undergoes multilineage commitment, followed by proliferation and maturation into the erythroid committed progenitor cell (Ogawa, 1993). This maturation takes place via a series of intermediate precursor cells, like burst and colony forming units (BFU-e and CFU-e). Besides the commitment to erythroid progenitors, cells originating from HSC, can also form all of the other cell lineages, such as lymphocytes, mast cells, megakaryocytes, macrophages and neutrophils (Metcalf, 1998).

The first haematopoietic cell, which can be identified, arises from the blood islands of the yolk sac, the primitive erythrocytes. These erythrocytes, however, cannot reconstitute the multiple haematopoietic lineages as definitive HCS can (Medvinsky *et al.*, 1993 and Rich, 1995). HCS are identified within the mouse embryo in the aortic/gonad/mesonephros (AGM) region, and are the key elements responsible for the maintenance of blood cell formation throughout life (Medvinsky and Dzierzak, 1996 and Sanchez *et al.*, 1996). A second origin of haematopoiesis in the embryo has been identified in the AGM. Cells from this region have been shown to reconstitute the multiple haematopoietic lineages in an irradiated adult recipient, and are definitive haematopoietic stem cells (Medvinsky and Dzierzak, 1996; Sanchez *et al.*, 1996 and Mukouyama *et al.*, 1998). It appears that there are two sites from which the haematopoietic cells in the embryo originate. First from the yolk sac producing large numbers of primitive erythrocytes, followed by the production of definitive haematopoietic stem cells in the AGM. After the embryonic phase the site of erythropoiesis shifts to the foetal liver and finally to the bone marrow, the sites of definitive erythropoiesis (Moore and Metcalf, 1970).

Primitive and definitive erythrocytes also differ morphologically, with the primitive erythrocyte nucleated and the definitive erythrocyte enucleated. Primitive and definitive erythrocytes also differ in the complement of globin genes that are activated, resulting in the formation of embryonic-, foetal- and adult-stage haemoglobin in man (Stamatoyannopoulos and Grosveld, 2001).

## Haemoglobin

Haemoglobin is synthesised in erythrocytes as a heterotetrameric protein and is responsible for the transport of oxygen. The heterotetramer consists of two α-like and two β-like globin chains, these polypeptides bind one haem group each. The α-like globin chains and β-like globin chains synthesised by each erythrocyte can differ depending on whether the cell is part of the primitive or definitive lineage and also depending on the species. This results in man in embryonic haemoglobin, type Gower I $\zeta_2\varepsilon_2$, type Gower II $\alpha_2\varepsilon_2$, type Portland I $\zeta_2\gamma_2$ type Portland II $\zeta_2\gamma_2$, foetal haemoglobin HbF, $\alpha_2\gamma_2$, and adult haemoglobin HbA and $HbA_2$, $\alpha_2\beta_2$ and $\alpha_2\delta_2$ (Bunn and Forget, 1986).

The three-dimensional protein structure of the globin protein family members has been resolved and consists of a four of α-helices, forming the haem-binding pocket, which is characteristic for all family members (Dickerson and Geis, 1983).

Although haemoglobin was first characterised in vertebrates it is highly conserved throughout evolution and present in invertebrates, plants, and also in several species of eubacteria, *Saccharomyces cerevisiae*, protist and protozoa (Hardison, 1996). This shows the importance of globin throughout evolution.

## Haemoglobinopathies

Several hereditary blood disorders, the haemoglobinopathies, have been described as a result of mutations or deletions in the α- and/or β-globin gene loci. The analysis of the molecular basis for these disorders has been very important in the understanding of globin gene regulation.

*Sickle cell disease*

Sickle cell disease results from a single T→A base change causing the substitution of one amino acid from valine to glutaminic acid in the amino terminus of the β-globin chain, resulting in Sickle haemoglobin HbS. This change causes the β-globin chain to polymerise when it is deoxygenated and to form aggregates in the cell (for review see Stamatoyanopoulos and Grosveld, 2001). As a result of this, the shape of red blood cells changes from a round shape to that of a sickle. In acute cases, sickling causes vaso-occlusion and severe anaemia. Occurrence of sickle cell anaemia is correlated with areas where malaria is endemic. People heterozygous for HbS are relative resistant to malaria infection caused by *Plasmodium falciparum* (Allison, 1957). In these areas carriers have a selective advantage in survival explaining the high frequency within the population.

*Thalassemias*

Thalassemias are the most common single gene disorders in the world. The disease is characterised by abnormal globin gene expression, which results in the reduction of the α- or β-globin chains, giving rise to either α- or β-thalassemias. The reduced production of one of the globin chains results in the accumulation of the intact globin chains in the erythroid cells. The intact chains precipitate in the erythroid precursor cells and form inclusion bodies. The inclusion bodies cause membrane damage and premature destruction of the erythroid cells, both mature and precursor erythroid cells, resulting in ineffective erythropoiesis and anaemia (Orkin, 1986 and Thein, 1993).

Thalassemias are defined both clinically and genetically. Using the clinical definition, thalassemias are divided into major forms, which are severe and transfusion dependent and minor forms, which are asymptomatic and often resemble the carrier state or trait.

The genetic classification describes the globin chains affected and the amount of chain production. The most common thalassemias genetically defined are the α-thalassemias, with the absence ($\alpha^0$-thalassemia) or reduction ($\alpha^+$-thalassemias) of the α-globin chains. β-thalassemias are similarly classified as $\beta^0$-thalassemia and $\beta^+$-thalassemias. The δβ-thalassemias are characterised by the production of HbF in adult life and are genetically classified by the amount of γ chains that are produced. $(\delta\beta)^0$ and $(^A\gamma\delta\beta)^0$-thalassemias, are similar to HPFH condition which is discussed below. Most of the α-thalassemias are caused by deletions of the locus, while the majority of the β-thalassemias are caused by non-deletion mutations.

Although uncommon, deletion types of the β-thalassemias have played an important role in gaining insight on the expression mechanisms of the globin genes and the discovery of important regulatory elements (Collins and Weissman, 1984 and Stamatoyannopoulos and Nienhuis, 1996). Two deletion types, the Dutch and the Hispanic thalassemias have played an important role. The Dutch deletion has a size of 100 kb and removes the whole β-globin locus leaving just the β gene and its promoter intact (van der Ploeg *et al.*, 1980; Kiousis *et al.*, 1983 and Taramelli *et al.*, 1986). The Hispanic deletion removes 30 kb upstream of the β-globin locus leaving the genes and their proximal regulatory elements intact (Driscoll *et al.*, 1989). In both cases there is no globin expression, although the genes are intact. It is now clear that the region upstream of the genes, deleted in both the Dutch and the Hispanic thalassemias, contains an important regulatory element, identified as the locus control region, which is necessary for the normal expression of the globin genes (Grosveld *et al.*, 1987).

Non-deletion β-thalassemias are caused by a variety of mutations (Huisman and Carver, 1998). The majority of the non-deletion β-thalassemias are caused by either insertion or deletion of a nucleotide or a nonsense mutation, all leading to a frame shift and/or a preliminary stop codon. Furthermore, mutations which affect the start codon or the correct splicing of the β-globin RNA have also been described. Also the non-deletion β-thalassemias have played a role in the characterisation of regulatory elements. For example mutations in the CACCC box of the β-globin promoter result in down regulation of the β-globin gene expression (Kulozik *et al.*, 1991). This element is the binding site for the erythroid-specific transcription factor EKLF (Wijgerde *et al.*, 1996).

*HPFH*

Hereditary Persistence of Foetal Haemoglobin (HPFH) is not considered a disease but rather a condition, because individuals with HPFH are clinically normal. HPFH disorders are heterogeneous and show an elevated level of HbF in adult life. Molecular analysis of HPFH conditions has shown that in some cases the β-globin cluster is intact. These are called the non-deletion HPFH conditions. In some cases, the 3'end of the locus, encompassing the δ and the β gene is deleted, thus giving rise to deletion HPFH. This last group resembles δβ-thalassemias which are also associated with elevated HbF. Distinction between the two disorders is made on a number of features: δβ-thalassemia patients show between 5-15% HbF in adult life and hypochromic and microcytic red cells, whereas HPFH heterozygotes show between 15-30% HbF and normal red cells.

Several hypotheses have been put forward to explain the differences in the HbF levels observed between HPFH and δβ-thalassemias. They fall into three categories: deletion of regulatory sequences within the genecluster; powerful enhancers downstream of the locus brought closer to the γ genes due to the large deletions; and competition between the γ- and β-genes for the activation by the LCR.

The first hypothesis was based on the comparison of deletions causing either δβ-thalalassemia or HPFH. Since the report of these initial studies, many more deletions causing either βδ-thalassemia or HPFH have been described. These include deletions causing δβ-thalassemia which have a 5' breakpoint upstream of all previously reported HPFH 5' breakpoints, thus deleting the same or even more of the possible regulatory sequences within the region (Wood, 1993). These data make the first hypothesis less likely. Transgenic studies, in which several sequences of the region have been deleted (Peterson *et al.*, 1995; Zhang *et al.*, 1997 and Calzolari *et al.*, 1999), also suggest that there are no essential silencer regions elements outside the γ-genes. Still it could be that some intergenic sequences act as positive or negative regulators and that the deletion of these could determine the observed phenotypes.

The second hypothesis is based on a HPFH deletion in which the β 3'enhancer is brought 25 kb closer to the γ genes (Feingold and Forget, 1989) and also on the HPFH1 and HPFH2 deletions in which a region normally located 120 kb from the γ genes and containing enhancer properties is brought within 10 kb of the γ genes (Feingold and Forget, 1989). This latter element has been shown in transgenic mouse models to give rise to elevated HbF (Arcasoy *et al.*, 1997).

The last hypothesis is based on the observation that elevated HbF is observed only when both the δ and the β genes are deleted. Furthermore in non-deletion HPFH there is evidence for competition between the γ and the β genes in the adult. However, this competition for the LCR cannot explain the differences observed between δβ-thalassemias and HPFH in HbF levels, it can only explain why higher levels of γ expression are observed in the adult. Another argument against this hypothesis is that in transgenic mice containing a LCR coupled to the γ genes, no γ expression is observed in the adult transgenic mice (Dillon and Grosveld, 1991).

In conclusion neither of the three hypotheses can explain all of the conditions and the differences observed between HPFH and the δβ-thalassemias in increased HbF. These hypotheses are not mutually exclusive and it is therefore likely that a combination of the mechanisms and different contributions of each mechanism between the various deletions account for the observed HbF increases and differences between δβ-thalassemias and HPFH deletions.

Non-deletion HPFH result from point mutations in the promoters of the γ genes (mutations in the $^{A}\gamma$ -promoter give rise to $^{A}\gamma$ HPFH and mutations in the $^{G}\gamma$-promoter give rise to $^{G}\gamma$ HPFH). These HPFHs are characterised by the δ-and β-globin chain synthesis in *cis* with the HPFH determinant. Transgenic mouse studies in which the $^{A}\gamma$-promoter is mutated, at -117 causing Greek non-deletion HPFH, indeed showed elevated $^{A}\gamma$ expression in the adult mice, in contrast to control mice which carried the same locus but without the mutation (Peterson *et al.*, 1995). An other study, in which the same mutation was introduced in mice, also showed persistence of γ-globin expression in the adult and a concomitant decrease in β-globin expression (Berry *et al.*, 1992).

The HPFH phenotypes are of clinical importance since they modulate the effect of both the β-thalassemia and HbS mutations. Patients heterozygous for both HPFH and either β-thalassemia or HbS show a much milder condition (Fessas and Stamatoyannopoulos, 1964 and Stamatoyannopoulos *et al.*,

1975). It is thus of great interest to elucidate the mechanisms for γ gene silencing and how these are apparently reversed in HPFH phenotypes, in order to develop therapeutic alternatives for the treatment of haemoglobinopathies.

## Structure of globin gene loci

It is thought that the α-and β-globin geneclusters originated from one ancestral globin gene and have been duplicated and diverged into two geneclusters during evolution. In the beginning these geneclusters were linked, as suggested by the fact that in primitive vertebrates, such as the zebrafish *Danio rerio* (Chan *et al.*, 1997) and the frog *Xenopus* (Hosbach *et al.*, 1983), the α-and β-globin genes are still physically linked. In other vertebrates the split of the α- and β-globin genes onto separate clusters is thought to have taken place about 300 million years ago, before mammals and birds diverged during evolution. This has been deduced from data showing that in both mammals and birds the globin genes are situated on separate chromosomes (Deisseroth *et al.*, 1976 and Hughes *et al,*. 1979).

In mammals, the α- and β-globin geneclusters are located on different chromosomes and are independently regulated. In humans, the β-globin genecluster resides on chromosome 11p (Deisseroth *et al.*, 1978) and the α-globin genecluster on chromosome 16p (Deisseroth *et al.*, 1977). In mouse the β-globin genecluster is located on chromosome 7 (Jahn *et al*., 1980), and the α-globin genecluster on chromosome 11 (Fig. 5).

The mammalian α- and β-globin geneclusters differ in a number of respects: they are embedded in different chromatin environments, with the α-cluster residing in a constitutively 'open', CG-rich chromatin domain, containing mostly housekeeping genes (Craddoch *et al.*, 1995). The β-cluster is found in a 'closed', AT-rich chromatin domain and is flanked by olfactory receptor genes. The different chromatin environments are also reflected on the replication timing of the two loci, such that the α-cluster is replicated early in S-phase in most cell types (Epner *et al.*, 1981 and Furst *et al.*, 1981), whereas the β-cluster is late replicating except in eryhroid cells (Kitsberg *et al.*, 1993 and Aladjem *et al.*, 1998). Furthermore the α-cluster does show differential methylation (Bird *et al.*, 1987), whereas the β-cluster shows methylation in non-erythroid cells (van der Ploeg and Flavell, 1980). Other differences between the clusters are found in specific DNA sequences, for example, the α-cluster contains Alu-repeats, whereas the β-cluster contains LINE-elements and Alu-repeats; the α-contains CpG-islands (Pondel *et al.*, 1995), which are not found in the β-genecluster (Collins and Weisman, 1984). Finally the α-cluster does not have MARs, in contrast to the β-cluster (Fischel-Ghodsian *et al.*,1987 and Jarman and Higgs, 1988).

The human β-globin cluster is composed of five developmentally regulated genes, arranged in the order in which they are expressed during development: 5'ε-$^{A}$γ-$^{G}$γ-δ-β 3'. The locus is regulated by an element located upstream of the genes called the locus control region (LCR) (Grosveld *et al.*, 1987). This element comprises five tissue-specific DNase I hypersensitive sites. The LCR will be discussed in more detail below.

The ε gene is expressed in the blood islands of the yolk sac and is detectable between the third and the tenth week of gestation in man. At about five weeks of gestation the site of haematopoiesis changes to the foetal liver. At the same time the switch from ε to γ expression starts to take place and is complete by around ten weeks of gestation (Huehns *et al.*, 1964). Until the twentieth week of gestation the foetal liver remains the main site of erythropoiesis and it gradually switches to the spleen and bone marrow, until at around birth the bone marrow becomes the main site of erythropoiesis (Bloom and Bartelmez, 1940; Knoll and Pingel, 1949 and Wintrobe and Shumacker, 1935). During the same period γ expression decreases and around birth switches to expression of the adult β and δ genes, with δ-globin making only a minor (3%) contribution to adult globin chain synthesis (Fig. 6).

The mouse β-globin locus contains two embryonic globins, the εy and βH1 genes, which are expressed in the embryonic yolk sac. The two adult globin genes, $\beta_{min}$ and $\beta_{maj}$, are expressed during the foetal liver and adult bone marrow stages with the $\beta_{maj}$ being the dominant gene. In mice, therefore, there are no distinct foetal genes and the switch from embryonic to adult globin gene expression takes place around day 11.5 of gestation (Farace *et al.*, 1984 and Chada *et al.*, 1986). The

mouse β-globin locus is also regulated by a LCR, which contains six DNase I hypersensitive sites (Moon and Ley, 1990).

The major α-globin locus regulatory element comprises only one DNase I hypersensitive site, located 40 kb, in the case of the human locus, and 26 kb, in the case of the mouse locus, upstream of the ζ gene (Higgs *et al.*, 1990). The adult α gene is expressed during all developmental stages and is only in the embryonic stage accompanied by expression of the embryonic-specific ζ-globin gene (Rohrbaugh and Hardison, 1983; Leder *et al.*, 1985). The ζ-gene follows the expression of the ε gene. Whereas the α gene is expressed throughout development, it is from week six of gestation that the adult expression levels are reached. Both human and mouse α-globin geneclusters express the same genes and have very similar expression profiles.

Pseudo genes with structural homology to the globin genes are also present in the globin gene loci and appear to be the result of gene duplication events. The duplication is followed by mutation and inactivation of the duplicated gene, which is followed by subsequent mutations due to loss of selective pressure. In the β-globin cluster there is one pseudo gene, ψβ while three pseudo genes are present in the α-globin genecluster: ψζ and $\psi\alpha_1$ $\psi\alpha_2$ (Forget, 2001).

## Structure and regulation of the globin genes

Human globin genes are relatively small and contain three exons and two introns. The exons code for 141 and 146 amino acid peptides for the α- and the β-globin chains, respectively. Intron 2 (IVS-2) appears to be important for polyadenylation and for the release of the transcript from the template, such that the transport from the nucleus to the cytoplasm can take place (Collis *et al.*, 1990 and Antoniou *et al.*, 1998).

Each gene is flanked by promoters, enhancers and silencers, important for the correct tissue-specific and developmental stage-specific expression. These elements are thought to interact with the LCR, in the case of the β-globin genes and with the αMRE, in the case of the α-globin genes, in order to activate transcription. The fact that the hypersensitive sites of the LCR and αMRE as well as the proximal regulatory elements of the individual genes contain some of the same transcription factor binding sequences, suggest an interaction between the different regulatory elements. The globin gene promoters, while sharing several characteristics, also have unique, distinguishing features. The α- and β-globin gene promoters contain a TATA box and a CCAAT box (Efstratiadis *et al.*, 1980), but differ otherwise (Fig. 7A). Because the main part of this thesis will concentrate on the human β-globin genecluster I will only briefly discuss the promoters and regulation of the individual α-globin genes and concentrate more in detail on the human β-globin genes.

All β-globin gene promoters show with minor variations the same recognition sequences for transcription factors. The sequences identified are ATAA, CCAAT and CACCC located respectively at -30, -70-80 and -80-140 nucleotides from the cap-site (Meyers *et al.*, 1986). A fourth sequence which is also found in all promoters is the (A/T) GATA (A/G) motif (Martin *et al.*, 1989 and Tsai *et al.*, 1989).

Looking at the genes individually, differences in both regulatory sequences and regulation of the genes can be observed. The individual genes and their regulatory elements are discussed below.

### *The ε-globin gene*

The region containing the ε gene and its regulatory sequences spans approximately 3.7 kb of DNA, with 2 kb of upstream sequences containing regulatory elements. In this region the TATA, CACCC and CCAAT boxes and GATA1 binding sites have been identified. These sites play a role both in the activation and the developmental stage-specific silencing of the ε gene, as has been shown in a number of transfection experiments using cultured cells (Gong and Dean, 1993; Gong *et al.*, 1991 and Walters and Martin, 1992).

Transcription factors Sp1 (Yu *et al.*, 1991), FKLF1 (Asano and Stamatoyannopoulos, 1999) and FKLF2 (Asano and Stamatoyannopoulos, 2000) have been shown to bind to the CACCC box in the ε-promoter, although Sp1 binding does not appear to have a significant influence on ε-globin expression

(Yu *et al.*, 1991). Of the two FKLFs, FKLF1 is the predominant factor activating ε expression, whereas FKLF 2 activates ε expression to a lesser degree in stable and transient transfection assays. However, the *in vivo* role for both FKLF1 and 2 remains to be determined (Asano and Stamatoyannopoulos, 1999 and Asano and Stamatoyannopoulos, 2000).

Using transgenic mice, several sequences have been indicated to play a role in the developmental stage-specific silencing of the ε gene. A silencer element from –304 to –179 5' of the ε gene was identified containing binding sites for GATA and YY1 factors (Cao *et al.*, 1989). Binding of GATA1 at –208 together with binding of YY1 at position –269 appear to be involved in ε repression (Raich *et al.*, 1995). Deletion of this silencer element in the context of a YAC containing the complete human β-globin locus, did not lead to high levels of ε expression in the foetal stage as would be expected, but instead to a decline of ε expression in the yolk sac (Liu *et al.*, 1997). Together with the data from Raich *et al*., this suggests that the 5'silencer of ε has a dual role both in suppressing ε in the foetal stage and in maintaining ε expression in the embryonic stage. Furthermore, is has been shown that the binding of NF-E3, which is immunologically related to COUP-TF, to the DR1 element located near the CCAAT box, results in the repression of ε expression (Filipe *et al.*, 1999). The replacement of the DR1 element by a high-affinity CACC-binding site for EKLF and a 4 bp substitution in the ε-globin CAAT sequence, also disrupting a DR element, led to the identification of a protein complex which mediates the suppression of ε-globin transcription during definitive erythropoeisis, called DRED (direct repeat erythroid-definitive) (Tanimoto *et al.*, 2000). Recently, two core components of DRED, nuclear orphan receptors TR2 and TR4 have been shown to bind to the DR1 element present in the ε-globin promoter and to repress ε-globin expression in definitive erythroid cell (Tanabe *et al.*, 2002).

The ε gene also needs the LCR for expression since an ε-globin gene alone, or a human β-globin locus with deletions in the LCR, result in undetectable ε expression in transgenic mice (Raich *et al.*, 1990; Shih *et al.*, 1990; Navas *et al.*, 1998 and Chapter 5 of this thesis). Additionally, ε-globin is said to be autonomously regulated during development, in that it does not require the presence of other globin genes in order to be silenced in the foetal and adult stages in transgenic mice (Shih *et al.*, 1990).

### *The γ-globin genes*

The regulation of the γ-globin genes has been studied intensively because even slightly elevated HbF in the adult, as seen in HPFH conditions, can alleviate the effects of β-thalassemia and sickle cell disease. Some mutations resulting in HPFH map within transcription factor binding sites in γ regulatory sequences, giving rise to the creation of new protein binding sites or the destruction of existing ones.

The promoters of the two γ genes are identical and contain the conserved CCAAT and CACCC boxes, a TATA box, DRE 1 sites and an OCT1 binding site (ATTTGCAT) flanked by two GATA1 binding sites. Between the CCAAT and TATA boxes a so-called stage selector element (SSE) has been reported (Jane *et al.*, 1992 and Jane *et al.*, 1993).

Different transcription factors bind to these sequences and play a role in the regulation of the γ genes. Transcription factors suggested to act as activators for γ expression are SSP, binding to the SSE (Jane *et al.*, 1992), CP1 binding to the CCAAT box (Skalnik *et al.*, 1991) and FKLF2 binding to the CACCC box (Asano *et al.*, 2000). The binding of SSP to SSE has been proposed to provide the γ gene with a competitive advantage over the β gene in the foetal stage. Transgenic studies in which the SSE has been mutated show a down-regulation of γ expression only when the gene is in competition with the β gene (Jane *et al.*, 1992). CP1 is ubiquitously expressed and interacts with both CCAAT boxes, however, there is no *in vivo* evidence to support CP1's role as a positive regulator of γ gene expression. Mutations in the CACCC box of the γ-promoter have been shown in transgenic mice to result in severe decrease of expression, indicating that the binding of transcription factors to this site plays an important role in γ-globin gene regulation (Stamatoyannopoulos *et al.*, 1993 and Duan *et al.*, 2001). Of these factors, Sp1, Sp3 do not appear to play a role *in vivo* (Marin *et al.*, 1997 ) and for BKLF/TEF2 there is no evidence at all for an effect on γ-expression (Crossley *et al.*, 1996). FKLF1 and FKLF2 bind to the γ-CACCC box *in vitro* and have been implicated as transcriptional activators, however, their role *in vivo* remains to be determined (Asano *et al.*, 1999 and Asano *et al.*, 2000).

Other transcription factors binding to the γ-promoter, such as CDP1 (Skalnik *et al.*, 1991), NF-E3, DRED and GATA1 (Gumucio *et al.*, 1988; Mantovani *et al.*, 1988 and Berry *et al.*, 1992), have been suggested to act as transcriptional repressors, although no *in vivo* experimental evidence, e.g. from gene knockouts, exists to support these assertions. CDP, at least *in vitro*, acts as a transcriptional repressor and competes with CP1 for binding at the CAAT-boxes (Skalnik *et al.*, 1991). A G →A mutation in the distal CAAT-box showed decreased binding of NF-E3 and GATA1 and led to the association of these factors with γ-globin gene repression. However, other mutations which decrease NF-E3 and GATA1 binding to the CAAT-box did not result in an increase of γ expression in transgenic mice (Ronchi *et al.*, 1996). Thus, the exact role of these factors in the γ-globin gene repression is still under debate. Mutations in the DR 1 binding site of DRED resulted in an elevation of γ expression in the adult, at least *in vitro*, indicative of a suppressor role of DRED for γ expression in the adult (Tanabe *et al.*, 2002).

Other binding sites for transcription factors have been found in the region upstream of the γ-promoter, having been associated with mutations that result in HPFH. Two of these regions which have been associated with elevated γ expression, are the –175 and –200 regions (Surrey *et al.*, 1988; Stoming *et al.*, 1989 and Jane *et al.*, 1995).

Regulatory sequences besides the promoter include an enhancer element reported at 750 bp downstream of the γ genes (Bodine and Ley, 1987) and sequences –382 to –370 5' to the $^{A}$γ-promoter, which have been shown to contain an adult specific silencer in transgenic studies using constructs containing the γ region with different truncations of the 5'region of the $^{A}$γ-promoter coupled to a μLCR (Stamatoyannopoulos *et al.*, 1993). The 3' γ gene enhancer however does not significantly affect γ gene transcription when deleted in transgenic mice carrying a YAC construct containing the β-globin locus (Puruker *et al.*, 1990 and Liu *et al.*, 1998). Instead this element has been proposed to play a role in the protection against chromosomal position effects and in the stabilisation of the interaction of the LCR with the γ-promoters (Li and Stamatoyannopoulos, 1994 and Stamatoynnopoulos *et al.*, 1997). Like the ε gene, the γ gene was also shown to be autonomously regulated. When a γ gene is linked to a LCR, it is silenced or expressed at very low levels in adult transgenic mice, indicative of an autonomous control of the γ gene (Dillon and Grosveld, 1991).

### *The δ-globin gene*

The next genes to be developmentally activated are the adult δ- and β-globin genes. The δ gene is very similar to the β gene in its 5' region but is distinct in its 3' region (Spritz *et al.*, 1980 and Martin *et al.*, 1983). The similarity of the 5' region between the β and the δ genes would suggest a similar mode of regulation. This, however, is not the case and is due to the deletion of the CACCC box and mutations found in the CCAAT box. The former is the major reason for the low δ expression levels, because this deletion results in the loss of an EKLF binding site (Donze *et al.*, 1996 and Tang *et al.*, 1997). Addition of the β-globin CACCC box to the δ gene, results in a 10-fold upregulation of δ expression in transfection assays (Donze *et al.*, 1996).

### *The β-globin gene*

β-globin is the dominantly expressed gene in the adult stage. CCAAT and CACCC boxes have also been identified in the β-promoter. In contrast to the γ genes, however, there are two CACC boxes and one CCAAT box in the β-promoter. Both boxes bind proteins involved in the activation of the β gene. CP1, GATA1 and NF-E6 bind to the CCAAT box (Antoniou and Grosveld, 1990; Antoniou *et al.*, 1988; de Boer *et al.*, 1988; Berry *et al.*, 1992 and Wall *et al.*, 1996). Of these, CP1 is thought to be a positive regulator, at least *in vitro*. GATA1 binds weakly and may not be of functional importance (Li *et al.*, 1998). NF-E6 seems to have a role *in vivo* since overexpression in transgenic mice of a dominant negative NF-E6 mutant leads to a shift in the ratio of γ to β expression, resulting in a higher expression of γ (Zafarana *et al.*, 2000).

The CACCC boxes in the β-promoter bind several factors *in vitro* (Hartzog and Myers, 1993), however, *in vivo* studies have shown that EKLF is the functional protein binding at this site (Miller and Bieker, 1993 and Feng *et al.*, 1994). EKLF has been shown to be the major regulator of β gene

expression. In transgenic mice heterozygous for the EKLF gene knockout, lower β-globin levels were observed (Nuez *et al.*, 1995). The complete gene knockout of EKLF proved to be lethal due to severe anaemia immediately after the γ- to β-globin expression switch (Perkins *et al.*, 1995 and Nuez *et al.*, 1995).

There are a number of reasons for the profound effect of EKLF on β gene expression but not on ε and γ gene expression. The β-globin CACC box has a much higher affinity for binding EKLF than those of ε and γ. In addition, the CACC box in the γ-promoter is flanked by a CCTTG repeat which has been shown to be a repressor for the recruitment of EKLF (Donze *et al.*, 1995 and Lee *et al.*, 2000).

Besides the promoter, two other regulatory elements have been described. An enhancer element located downstream of the poly-A signal containing four GATA1 sites has been shown to stimulate the activity of a linked promoter in transfection studies (Antoniou *et al.*, 1988). Furthermore, the deletion of this enhancer resulted in a decrease of β expression in transgenic mice (Liu *et al.*, 1997). Another enhancer element described in cell transfection and transgenic mouse studies is located near the junction of intron 2 and exon 3 (Antoniou *et al.*, 1988 and Behringer *et al.*, 1987). Its *in vivo* role in the context of the whole locus has not been determined yet.

In contrast to the other genes silencer elements have not been identified in the neighbourhood of the β gene.

### *The ζ-globin gene*

The ζ-globin promoter contains a TATA-box, a CAAT-box, a CACC-box and DR-repeats. Transcription factors involved in the regulation of the ζ gene include CP2, which binds to the CAAT-box (Lim *et al.*, 1992), Sp1-like proteins, which bind in the CACC-box region (Watt *et al.*, 1990 and Yu *et al.*, 1990) and GATA1, which has a strong binding site overlapping the Sp1-binding site, the two proteins binding in a competitive manner. There is also a GATA1 site in the upstream part of the promoter. Both GATA1 sites are necessary for interactions with αMRE, as has been shown in transient transfection studies (Zhang *et al.*, 1993).

Besides the promoter, another positive regulatory element containing a GATA1 site flanked at the 3' by a CACC-box, has also been described (Sabath *et al.*, 1995).

Like the ε gene the developmental regulation of the ζ gene is autonomous: all elements required for the silencing of the ζ gene can be found in the sequences flanking it (Sabath *et al.*, 1993; Albitar *et al.*, 1991 and Pondel *et al.*, 1992).

### *The α-globin gene*

The promoter of the α-globin gene differs from those of all other globin genes. There is no CACC-box, but there is a GC-rich area which forms part of a methylation-free island extending into the gene (Flint *et al.*, 1997 and Shewchunk and Hardison, 1997). It was shown in transfection studies that the α gene could be expressed in non-erythroid cells without additional enhancer elements, probably due to the presence of the methylation-free island (Humphries *et al.*, 1982). For the expression of α-globin in transgenic mice, the α gene requires the presence of the αMRE or the β-globin LCR (Hanscombe *et al.*, 1991 and Higgs *et al.*, 1990).

Transcription factors that bind to the α-promoter overlap those that bind to the β-globin-like promoters, like GATA1 and CP1, but there are also proteins that only bind to the α-promoter, like the inverted repeat protein (Lim *et al.*, 1992; Kim *et al.*, 1990; Lim *et al.*, 1993 and Swendeman *et al.*, 1994).

## The Locus Control Region

The existence of important regulatory sequences upstream of the β-globin genes became clear from the molecular analysis of large deletions in the β-globin locus that give rise to thalassemias (van der Ploeg *et al.*, 1980; Driscoll *et al.*, 1989 and Kulozik *et al.*, 1991). In particular, the Dutch and Hispanic thalassemias are characterised by the deletion of 100 kb and 40 kb of sequence, respectively, upstream

of the gene locus. These deletions left part, or all, of the β-globin gene locus intact but the genes were transcriptionally inactive. Although the intact genes are still able to express, as was shown by cloning and expressing them in transfection studies (Kioussis *et al.*, 1983; Ryan *et al.*, 1989). In the deleted locus the genes are silent and embedded in an inactive chromatin structure (Kioussis *et al.*, 1983; Forrester *et al.*, 1990 and Schubeler *et al.*, 2000).

DNase I hypersensitivity studies in the region usptream of the β-globin gene locus revealed the presence of five tissue-specific hypersensitive sites located between 6 kb and 25 kb 5' of the ε gene (Forrester *et al.*, 1986, Tuan *et al.*, 1985). These sites have been termed 5'hypersensitive site 1-5 (5'HS1- 5'HS5). The importance of these hypersensitive sites for globin expression was demonstrated in transgenic mouse studies, where this region was coupled to a β-globin gene. Expression levels proportional to transgene copy number and comparable (per copy) to endogenous mouse globin levels were observed (Grosveld *et al.*, 1987).

Transgenic mouse studies in which the LCR was coupled to a β-globin gene showed that the LCR can drive tissue-specific expression of the transgene independently of its (random) site of chromosomal integration, thus conferring copy number dependent levels of expression (Grosveld *et al.*, 1987 and Blom van Assendelft *et al.*, 1989). These properties form the defining characteristics of LCRs and suggest that one of the fundamental aspects of LCR function is the organisation of a chromatin domain that will support transcriptional activation (Festenstein and Kioussis, 1997 and 2000; Fraser and Grosveld, 1998; Grosveld, 1999).

Although the LCR was suggested by both transgenic mouse assays and the analysis of deletions in thalassemias as having chromatin activating capacities, recent studies in mice in which the LCR was deleted from the endogenous mouse β-globin locus have led to a discussion about this function of the LCR (Reik *et al.*, 1998; Epner *et al.*, 1998 and Bender *et al.*, 2000). In mice carrying a deletion of the endogenous mouse β-globin LCR, it was shown that the locus maintained DNase I hypersensitivity, however gene expression levels had dropped to just a small fraction of the wild type expression levels (Epner *et al.*, 1998; Reik *et al.*, 1998 and Bender *et al.*, 2000). From this, it was suggested that the LCR is not involved in the chromatin opening of the gene domain. One possible reason for these observations could be that there are differences between the mouse and the human LCRs (Higgs, 1998 and Grosveld, 1999). However, a satisfying conclusion reconciling the differences between the human genetic data on human β-globin LCR function and the LCR deletions in mice, has not yet been suggested.

### *The hypersensitive sites*

The five hypersensitive sites of the LCR can be sub-divided into the erythroid-specific, HS1-4 (Forrester *et al.*, 1986 and Tuan *et al.*, 1985), and the constitutive HS5 (Tuan *et al.*, 1985 and Dhar *et al.*, 1990) although additional studies indicate that HS5 is present in most haematopoietic cells (Li *et al.*, 1999 and Zafarana *et al.*, 1995). Construction of a micro-LCR (μLCR) containing small regions holding each hypersensitive site showed that these retain the functional activity of the LCR (Talbot *et al.*, 1989). Mapping of each individual hypersensitive site revealed core sequences of around 250-500 bp (Philipsen *et al.*, 1990; Talbot and Grosveld, 1991; Pruzina *et al.*, 1991 and Lowrey *et al.*, 1992).

All the hypersensitive sites contain binding sites for the erythroid specific factors NF-E2 and GATA1, as well as GT-sequences to which factors like EKLF and Sp1 can bind (Talbot *et al.*, 1990; Strauss and Orkin, 1992 and Ikuta *et al.*, 1996). Hypersensitive sites 2, 3 and 4, contain these binding sites, however, in different combinations (Fig. 7B) (Talbot *et al.*, 1990 and 1991; Pruzina *et al.*, 1991; Stamatoyannopoulos *et al.*, 1995; Walters *et al.*, 1991; Strauss and Orkin, 1992 and Ikuta and Kan, 1991).

HS2 contains two NF-E2, two GATA1, one Sp1 and two Tal1/USF binding sites (Ney *et al.*, 1990; Talbot and Grosveld, 1991 and Lui *et al.*, 1992). Mutagenesis of the GATA1 sites in a synthetic HS2 fragment shows reduced activity of the HS2 when coupled to a β-globin gene and transfected into MEL cells (Ellis *et al.*, 1993). The significance of the NF-E2 sites was shown both in transient transfection assays using MEL cells and in transgenic mouse studies demonstrating that they are necessary for full LCR and HS2 activity (Caterina *et al.*, 1991; Moi and Kan, 1990 and Liu *et al.*, 1992).

HS3 contains one NF-E2 site, a triple tandem repeat of GATA1 sites and GT-sequences (Philipsen *et al.*, 1990 and Strauss and Orkin, 1992). In transgenic mice it has been demonstrated that the GATA1 sites are required for LCR activity (Philipsen *et al.*, 1993). The GT sequence in HS3 is bound by EKLF *in vivo* (Gillemans *et al.*, 1998). The lack of binding of EKLF at the GT-sequence of HS3 results in changes in chromatin structure, as detected by DNase I sensitivity studies (Gillemans *et al.*, 1998 and Wijgerde et al., 1996).

Finally, HS4 has an AP1/NF-E2 site followed by a Sp1 site and two GATA1 binding sites (Pruzina *et al.*, 1991 and Lowrey *et al.*, 1992). The GATA1 sites in HS4 are inverted and are required for hypersensitive site formation (Lowrey *et al.*, 1992 and Stamatoyannopoulos *et al.*, 1995). Other factors that have been reported to interact with the LCR include USF and YY1. The functional relevance of their binding is not clear.

### *The role of the hypersensitive sites in the LCR*

The role of each hypersensitive site within the LCR has been investigated in transgenic mouse studies and cell transfection assays. Transgenic mice in which a single hypersensitive site was coupled to a β-globin gene, showed different levels of expression. HS3 showed approximately 70% of full LCR activity, HS2 and HS4 30% and HS1 less then 10%, while HS1 did not show any activity (Fraser *et al.*, 1990). Similar results were obtained using stable transfections in MEL cells (Collis *et al*., 1990).

In transient transfection assays, only HS2 showed classical enhancer activity (Ney *et al.*, 1990 and Tuan *et al.*, 1989). This enhancer activity is dependent on the tandem repeat of NF-E2 sites, which are only found in HS2 (Ney *et al.*, 1990 and Pruzina *et al.*, 1991). The role of HS3 became apparent in single copy transgenic studies, since only HS3 was able to drive β-globin gene expression. The other hypersensitive sites tested needed to integrate as multiple copies to activate globin expression (Ellis *et al.*, 1996 and Ellis *et al.*, 1993). From these studies it was suggested that HS3 contains chromatin opening activities.

The role of the hypersensitive sites has also been studied in transgenic mice carrying the complete human globin locus bearing a deletion of each one of the hypersensitive sites. The results of these studies showed a loss of chromatin opening activity by the LCR, resulting in position effects and lower expression levels of the transgenes (Milot *et al.*, 1996; Peterson *et al.*, 1996; Bungert *et al.*, 1995; Bungert *et al.*, 1999 and Chapter 5 of this thesis). Integration of the transgene in pericentromeric regions resulted in two types of position effects. Some lines showed a classical position effect, PEV, in which a sub-population of the cells does not express the transgene. Other lines showed a new kind of position effect, in which the transgene is expressed in all cells but for a shorter period of time during the cell cycle. This type of position effect was called cell timing position effect. A detailed description and discussion of the various HS-deletion studies in transgenic mice, is given in Chapter 5.

The observations on the hypersensitive sites led to the suggestion that the LCR functions as one unit called a holocomplex, which interacts with the β-globin genes in a developmental order and with only one gene at the time and that there is a developmental stage specificity in the interaction of the hypersensitive sites with the globin genes (Fraser *et al.*, 1993 and Dillon *et al.*, 1996).

### *αMRE, the locus control element of the α-globin locus*

In contrast to the β-globin locus, the α-globin genecluster does not have an equivalent of a locus control region but one hypersensitive site located 40 kb in the human locus and 26 kb in the mouse locus upstream of the ζ gene (Higgs *et al.*, 1990). The critical region in αMRE has been localised in a 350 bp core element and contains binding sites for NF-E2, GATA1 and Sp1 (Jarman *et al.*, 1991). Deletion of this site in a MEL cell line containing a human chromosome 16, resulted in the down-regulation of the α genes (Bernet *et al.*, 1995). When the element was coupled to the α-globin genes in transgenic mice correct tissue specific and developmentally regulated expression was observed, although the α gene was silenced in the adult (Higgs *et al.*, 1990). In the absence of αMRE, α-globin transgenes do not normally express in transgenic mice. These results indicate, that this HS site is very important for the expression of the α-globin genes.

## Haemoglobin switching

The developmental expression patterns of globin genes is characterised by the switching of expression of one globin gene to another. Developmental switching occurs twice in the human β-globin locus, from expression of ε-globin to γ and from γ to β. In the mouse there is only developmental switch in the expression of the β-globin genes from the embryonic εy/βH1 to the adult $\beta_{min}$ and $\beta_{maj}$ genes.

Transgenic mouse studies have been used to elucidate the mechanism of switching. Integration of the human γ or β genes without the LCR in transgenic mice, resulted in very low levels of globin expression, but expression was tissue- and developmental-stage specific (Kollias *et al.*, 1986; Trudel *et al.*, 1987; Trudel and Costantini, 1987 and Townes *et al.*, 1985). These observations indicate that the elements responsible for developmental-stage specificity lie within the regions flanking the genes.

Transgenic studies with constructs containing the whole β-globin locus, based on both ligated cosmid and YAC constructs (Stouboulis *et al.*, 1992; Gaensler *et al.*, 1993 and Peterson *et al.*, 1993), show levels of expression of the human globin genes similar to those of the endogenous mouse globin genes, as well as correct developmental switching, indicating that transgenic mice can be used to study the basis of human β-globin switching in the context of the full human β-globin locus (Fig. 8).

There is, however, one difference between the switching in humans and the switching of human β-globin genes in transgenic mice that has to be taken into account. The switch from γ to β takes place around birth in humans, whereas in transgenic mice this is accelerated so that γ to β switching takes place around day 12.5/14.5dpc in the foetal liver.

Switching is not a progressive process in which γ expression is followed by β expression. It is more of a dynamic process in which the γ- and β-globin genes are alternately transcribed during development. This was shown in transgenic mouse studies in which the transcription sites of the γ- and β-globin genes were visualised *in vivo* using primary transcript *in situ* hybridisation techniques. These studies showed that some cells had both γ and β transcripts on the same allele. There were also cells which showed β mRNA in the cytoplasm and a γ signal in the nucleus (Wijgerde *et al.*, 1995 and Gribnau *et al.*, 1998). These results indicate that the LCR flip-flops between the γ and β genes during a period of overlap in the expression of these genes and that the switching from γ to β takes place gradually.

Studies on the basis of haemoglobin switching led to the proposal of a dual mechanism by which switching is regulated: autonomous gene silencing and gene competition.

### *Autonomous globin gene silencing*

Regulation of different human globin genes has been studied in several transgenic mouse models. Transgenic mice containing just the ε gene with 5' and 3' flanking sequences do not express at any developmental stage. Addition of the LCR to the ε gene, however, resulted in detectable expression only in the embryonic stage (Raich *et al.*, 1990 and Shih *et al.*, 1990). This observation suggests that the ε gene is autonomously silenced during development and that it does not require the presence of additional globin genes as would have been predicted in a competitive model. Deletion in transgenic mice of a putative silencer element located upstream of the ε gene, resulted in the continued expression of the ε gene into definitive erythroid cells albeit at low levels (Raich *et al.*, 1992 and Cao *et al.*, 1989).

The γ gene also appears to be regulated by autonomous silencing. Transgenic mice with a γ gene coupled to the LCR show expression of the gene in the foetal stage but no expression in the adult stage (Dillon and Grosveld, 1991; Enver *et al.*, 1989; Enver *et al.*, 1990 and Behringer *et al.*, 1990). Mutations in the promoter sequences of the γ genes, which give rise to a HPFH phenotype, suggest that the promoter of the γ genes play a role in the process of autonomous silencing ( Berry *et al.*, 1992 and Ronchi *et al.*, 1996).

Studies with the β gene coupled to the LCR, show immediate activation of the transgene at the embryonic stage which persists all the way into the adult stage (Enver *et al.*, 1990 and Behringer *et al.*,

1990). This suggests that the β-globin gene when linked by itself to the LCR is not appropriately regulated during development in transgenic mice.

*Gene competition*

The result that the β-globin gene on its own is not developmentally regulated, led to the idea that the correct β expression is regulated through gene competition. Support for the competition model in human β-globin gene regulation came from experiments in which transgenic mice with the LCR coupled to the β gene (Enver *et al.*, 1990) and transgenic mice with a γ gene preceding the β gene were compared (Hanscombe *et al.*, 1991). These studies showed that the presence of the γ gene restored the normal developmental expression patterns of the β-globin gene thus restricting its expression in the foetal liver and adult blood stages.

The functional role of this competition between the genes is thought to be related to keeping balanced β-like globin chain production. That competition is indeed important for balanced haemoglobin production is shown by studies with non-deletion HPFH subjects. In these subjects, γ expression levels are increased with β expression levels decreased to an extent equivalent to the increase in γ expression, indicative of regulation by competition (Giglioni *et al.*, 1984).

Transgenic mouse studies have indicated that several factors play a role in gene competition: gene order (Hanscombe *et al.*, 1991 and Tanimoto *et al.*, 1999), distance between a gene and the LCR (Dillon *et al.*, 1997) and dosage of transcription factors present (Wijgerde *et al.*, 1996).

In transgenic mouse studies where the positions of the β- and the γ-globin genes were changed with respect to the LCR, correct timing of β gene expression depended on a place of the β gene further away from the LCR than the competing γ genes. Placing the γ genes away from the LCR resulted in premature silencing due to competition from the more LCR-proximal β gene (Hanscombe *et al.*, 1991). These results suggest that the difference in relative distance from the LCR plays an important role in gene competition for the LCR. The importance of gene order has also been shown in transgenic mice in which either the gene order or the LCR had been reversed (Tanimoto *et al.*, 1999). Reversal of the order of genes rendered ε-globin the 3'-most and β the 5'-most genes with respect to the LCR. The expression profile of the human globin genes in these transgenic mice was completely changed, with the ε gene no longer expressing, whereas expression of the β gene was found throughout development. One conclusion from these studies was that it is necessary for the embryonic genes to be proximal to the LCR for their transcriptional activation.

The importance of proximity to the LCR was demonstrated in a study in which an additional "marked" β-globin gene was inserted at two different positions in the β-globin locus. The first position is in place of the ε-globin gene proximal to the LCR and the second position is just upstream of the δ gene, i.e. distally to the LCR. The effects of placing the $\beta^{marked}$ gene proximally or distally to the LCR, on the expression of the β-globin gene in its native position and on the developmental regulation of all globin genes in the locus were assessed in transgenic mice (Dillon *et al.*, 1997). The expression levels of the $\beta^{marked}$ at the position just in front of δ compared to the β gene, showed that $\beta^{marked}$ is expressed at 75% and β at 25% of the total expression level of human β compared to mouse β. Both transcriptional interference or competition for the LCR could be causing this result. If the former were the case, placing the $\beta^{marked}$ at the position of the ε gene should decrease the effect of $\beta^{marked}$ on the β gene. The analysis of the expression levels of $\beta^{marked}$ at the ε position versus β, showed that β gene expression is severely reduced to approximately 10%, whereas the $\beta^{marked}$ was expressed at 90% of the total level of human β expression. In addition, expression of the γ-globin genes at the embryonic stage was similar to the levels of $\beta^{marked}$ expression. In the early foetal liver stage, however, there is only expression of $\beta^{marked}$ with no detectable expression for γ-globin. This indicated that normally β is repressed during embryonic and foetal stages owing to its distal position to the LCR. This indicates that indeed the distance from the LCR plays an important role in the correct developmental expression of the human globin genes (Dillon *et al.*, 1997).

Finally the role of trans-acting factors in gene competition has been shown in transgenic studies using compound EKLF knockout/human β-locus transgenic mice (Wijgerde *et al.*, 1996). Mice heterozygous for the knockout allele of EKLF (hum$\beta^{+/+}$/EKLF$^{+/-}$) show

a decrease in β transcription and a reciprocal increase of γ transcription. These data indicate that EKLF plays a role in the γ versus β competition and suggest that EKLF is potentially important in stabilising the interaction of the LCR with the β-promoter, thus giving it a competitive advantage (Wijgerde *et al.*, 1996).

In conclusion, autonomous gene silencing and gene competition account for the developmental regulation of globin gene switching, with silencing appearing to be more important for the embryonic to foetal switch and gene competition for the foetal to adult switch.

### *Models of gene regulation by the LCR*

Three models haven been proposed to explain the basis of activation by the LCR: the accessibility model (Martin *et al.*, 1996), the scanning/tracking model (Tuan *et al.*, 1992 and Kong *et al.*, 1997) and the looping model (Fig. 9) (Stamatoyannopoulos *et al.*, 1991; Epner *et al.*, 1992; Dillon *et al.*, 1993; Grosveld *et al.*, 1993 and Hanscombe *et al.*, 1991).

The first model envisions that the LCR's function is to open up the chromatin structure over the entire globin gene domain thus rendering it accessible to transcription factor binding at the regulatory elements of the individual genes. The developmental expression of the genes is then the result of stage-specific binding of transcription factors and transcriptional interference. According to this model, the genes behave independently to each other and there is no competition between them for interaction with the LCR. Since several studies have clearly indicated the existence of gene competition between the globin genes (Enver *et al.*, 1990; Hanscombe *et al.*, 1989 and Giglioni *et al.*, 1984), this model is unlikely to account for the basis of LCR function.

The second model of scanning/tracking suggests that the LCR binds a transcriptional activator complex, which starts scanning along the DNA fiber of the locus activating the first promoter poised for transcription that it encounters. This model can account for the results obtained in the studies on gene order and distance of genes from the LCR (Dillon *et al.*, 1997 and Tanimoto *et al.*, 1999), however, it is difficult to explain the alternating expression of the γ and β genes in the same cell observed in foetal liver cells (Wijgerde *et al.*, 1995) and the order/distance parameter which plays a role in competition.

The third model is that of looping. This model envisions direct interactions of the LCR as a holocomplex and genes in the locus via the "looping out" of intervening DNA sequences. The presence of transcription factors at the gene promoter will secure the binding of the LCR and activation can take place. The LCR could activate one gene and then loop directly to the next gene, thus explaining γ and β transcription at the same time. The presence of the appropriate transcription factors and the strength of binding of the LCR to the promoter will determine the duration of time that the LCR will be present at a promoter and thus the expression levels of a gene. The looping model can also account for the results obtained in the gene order and distance experiments. In this model a gene closer to the LCR would interact more frequently with the LCR than a gene that is more distal, resulting in the higher expression of the more proximal gene. When the proximal gene is placed closer to the distal gene, then the frequency of interaction with the LCR would become less and the advantage of the proximal gene is reduced, resulting in smaller differences in expression between the proximal and distal genes. This is what was observed in the $\beta^{marked}$ experiments. This model can explain all the results thus far obtained for the expression patterns of the genes and by the time that I will defend my thesis there will be direct proof for this model.

## Transcription factors and globin expression

Transcription factors play important roles in the regulation of globin genes. In the following paragraphs the most important factors will be described and their actions on the globin genes summarized.

### *GATA1*

GATA1 was the first member to be identified of a family consisting of six proteins, all recognising the consensus GATA motif (Orkin, 1992). The GATA motif is found in almost all regulatory elements

of the globin genes. At first it was thought that GATA1 was erythroid specific, however, it is also present in other haematopoietic lineages such as mast cells, megakaryocytes and eosinophils (Martin *et al.*, 1990 and Crotta *et al.*, 1990) and in the Sertoli cells of testes (Ito *et al.*, 1993).

GATA proteins are characterised by two zinc-fingers which interact with the major groove of the DNA helix (Omichinski *et al.*, 1993). The C-terminal finger is required for the binding of the GATA-motif, whereas the N-terminal finger is important for the stabilisation of this binding (Tsai *et al.*, 1989 and Trainor *et al.*, 1990) and for the interaction with other factors like, for example, friend of GATA (FOG) (Tsang *et al.*, 1997 and Tsang *et al.*, 1998).

GATA1 is thought to carry out several functions. GATA1 overexpression can dominantly affect lineage selection in cell lines (Kulessa *et al.*, 1995 and Visvader *et al.*, 1992). For instance the introduction of GATA1 in a myeloid cell line resulted in the induction of megakaryocytic differentiation (Visvader *et al.*, 1992). Studies indicate that GATA1 plays a role in the regulation of a cascade of downstream pathways in cellular differentiation. GATA1 has also been shown to play an important role in the balance between erythroid cell proliferation and survival (Weiss *et al.*, 1994). Furthermore, it has been shown that upon induction of GATA1 overexpressing MEL cells, cyclin A-dependent kinase activity was decreased much less in the GATA1 overexpressing than in control cells. In the same study it was also shown that GATA1 binds to the retinoblastoma protein. The data together led to the conclusion GATA1 regulates differentiation by affecting the cell-cycle apparatus (Whyatt *et al.*, 1997).

GATA1 knockout embryos die at embryonic day 10 or 11 from severe anaemia. This is caused by the production of erythroid precursors arrested at the pre-erythroblast stage, which then undergo apoptosis (Pevny *et al.*, 1991; Fujiwara *et al.*, 1996 and Weiss *et al.*, 1994). Overexpression of GATA1 showed the opposite effect of stimulation of proliferation of pro-erythroblast cells resulting in inhibition of differentiation (Whyatt *et al.*, 1997). Finally, GATA1 plays a role as transcriptional activator, which correlates with the presence of GATA1 binding sites in the promoters of the globin genes and the core regions of the hypersensitive sites of the LCR. In conventional reporter assays in heterologous cells, GATA1 has been shown to act as a transcriptional activator (Martin and Orkin, 1990). Furthermore GATA1 has been reported to have effects on the expression of the globin genes (see the paragraph on gene regulation). Studies in which GATA1 knockout cells were tested for rescue of differentiation by different forms of GATA1, showed that its transcriptional activation function can be dissociated from its survival and differentiation function (Weiss *et al.*, 1997).

These studies also led to the suggestion that GATA1 probably needs a transcriptional co-activator. In agreement with this suggestion GATA1 has been shown to interact with several other transcription factors via its zinc finger domain. Examples of these factors are EKLF, Sp1 (Merika and Orkin, 1995), p300/CBP (Blobel *et al.*, 1998) and FOG (Tsang *et al.*, 1997 and Tsang *et al.*, 1998). The precise function of these interactions remains to be elucidated. The interactions of GATA1 with different transcription factors suggest that it has a primary role in the formation of a haematopoietic transcription factor complex at specific sites in the globin locus, potentially controlling expression at different developmental time-points (Orkin, 2000). Furthermore, the interaction with p300/CBP, might be a way via which histone acetyltransferases are brought to specific DNA sites (Blobel *et al.*, 1998), resulting in the acetylation of histones and enhancement of transcription of the globin genes (Boyes *et al.*, 1998).

*NF-E2*

NF-E2 was the second erythroid-specific factor to be identified (Mignotte *et al.*, 1989). It was initially found to bind AP-1 sites in the promoter of the human porphobillinogen deaminase (PBGD) gene. Subsequent studies showed that the AP-1 sites present in HS2 enhanced expression of reporter constructs in transfected cells (Ney *et al.*, 1990). The same activation by HS2 was also observed in transgenic mice (Talbot *et al.*, 1990 and Caterina *et al.*, 1994). NF-E2 binds as a heterodimer and consists of a haematopoietic subunit called p45 NF-E2 and a more widely expressed subunit called p18 NF-E2 or MafK (Andrews *et al.*, 1993a and 1993b). p45 NF-E2 contains the transcriptional activation domain. Both subunits are family members of the basic leucine zipper family. *In vitro* studies using MEL cells support the idea that NF-E2 plays a role in globin expression. MEL cells not expressing p45 NF-E2, cannot sustain high levels of globin expression. Reintroduction of the p45 NF-

E2 subunit restored expression of globin genes (Lu *et al.*, 1994). Furthermore, these studies also indicated that multiple p45 NF-E2 subunits are required for NF-E2-mediated activation (Bean and Ney, 1997). Using chromatin immunoprecipitations a recent study showed that the NF-E2 complex is recruited to both the LCR and the active globin promoters upon induction of MEL cells. This recruitment has been shown to correlate with a 100-fold increase in $\beta_{maj}$ globin expression. From these results it has been speculated that the recruitment of the NF-E2 complex to both the LCR and the active globin promoters may be a rate-limiting step in the globin gene expression (Sawado *et al.*, 2001).

Although shown to be important for globin expression *in vitro*, the *in vivo* role of NF-E2 is not clear. Knockout mice for p45 NF-E2 only show a subtle reduction in globin expression, however, they do suffer from the loss of production of circulating platelets (Shivadasani and Orkin, 1995 and Shivadasani *et al.*, 1995).

*EKLF*

CACC motifs are present in many gene promoters, including those of the β-globin genes, and are bound by a number of proteins, like Sp1 and Kruppel related proteins. EKLF has been identified by cDNA subtraction assays between lymphoid and erythroid transcripts (Miller and Bieker, 1993) and is highly erythroid specific (Southwood *et al.*, 1996).

EKLF contains three zinc fingers that bind specifically to the CCACACCCT sequence found in the β-promoter and HS3 of the LCR (Feng *et al.*, 1994 and Gillemans *et al.*, 1998). The transcriptional activity of EKLF seems to be downstream to that of GATA1, as indicated by the presence of GATA1 binding sites in the EKLF-promoter (Crossley *et al.*, 1994). Although present throughout development, with binding sites present in all globin gene promoters (except for the δ-promoter), EKLF only acts on the β gene. This has been shown in knockout mice for EKLF. These mice die from anaemia at the foetal stage due to a deficiency in β-globin synthesis. No effect on expression of the other globin genes is observed in these mice (Perkins *et al.*, 1995 and Nuez *et al.*, 1995; for effects of EKLF on the globin genes also see the paragraph on globin gene regulation). Overexpression of EKLF in transgenic mice, showed a reduction in platelets which suggests that EKLF could also play a role in the balance between megakaryocytic and erythroid lineages (Tewari *et al.*, 1998).

Furthermore, EKLF has been suggested to have an effect on the γ to β switch. In human β globin locus transgenic mice heterozygous for the knockout allele of EKLF ($hum\beta^{+/+}/EKLF^{+/}$), γ expression is increased with a concomitant reduction in β expression during the period of gene competition between γ and β genes (Wijgerde *et al.*, 1996 and Perkins *et al.*, 1996). Finally, EKLF has been shown to play a direct role in LCR function (Gillemans *et al.*, 1998). HS3, thought to play a role in the chromatin opening function of the LCR (Ellis *et al.*, 1996) contains binding sites for EKLF. DNase I hypersensitivity of HS3 is markedly reduced in EKLF knockout mice. In addition, EKLF has been shown to interact *in vitro* with the chromatin remodelling complex E-RC1 (Armstrong *et al.*, 1998). Combination of these data led to the suggestion that EKLF could play a role in LCR activation by binding to HS3 and recruiting E-RC1.

## Aim of the PhD project

A number of different elements play a role in the developmentally regulated expression of the human and mouse β- and α-globin geneclusters. The elements, which are involved in this process, have been discussed in the introduction above. During my PhD project I investigated the activation of the mouse α-and β-globin gene loci in their endogenous context. I was involved in the development of novel methodology that allowed the manipulation of the human β-globin locus in the context of a 185 kb PAC by homologous recombination in *E. coli*. I further applied this methodology in deleting separately HS2 and HS3 from the human β-globin LCR and assaying the effects these deletions had on the regulation of the locus in transgenic mice. Finally, I was involved in a project examining the role in the regulation of γ gene expression of putative regulatory sequences located downstream of the $^{A}\gamma$-globin gene in the context of the human β-globin locus in transgenic mice.

The aim of the first project was to investigate the basic mechanisms of transcriptional activation of the mouse globin gene loci. We made use of *in situ* hybridisation techniques to detect nuclear and cytoplasmic patterns of globin gene expression in 14.5 dpc mouse foetal liver cells. We were able to provide strong evidence that globin gene activation takes place in a stochastic, all-or-nothing manner which, once established, is clonally inherited in subsequent cell generations. The results of this project are described in chapter 2 of this thesis.

Chapter 3 describes work which extends our ability to modify the human β-globin in the context of a 185 kb PAC insert by homologous recombination. This method allows us to efficiently manipulate the locus while avoiding the limitations of the cosmid ligation approach.

The manipulation of the human β-globin locus by homologous recombination in *E. coli*, was applied in deleting two putative regulatory elements located downstream of the $^{A}\gamma$-globin gene and assaying the effects of this deletion on human globin gene regulation in transgenic mice. The two elements Enh and F have been associated with naturally occurring deletions that give rise to elevated γ-globin gene expression in the adult stage. The 5' breakpoints of these deletions map within the $^{A}\gamma$- and δ-globin intergenic region and it has been postulated that this region harbours *cis*-regulatory elements important for γ gene silencing in the adult stage. Consistent with this hypothesis, the Enh and F elements had previously shown to exhibit silence activity in transient transfection assays (Kosteas *et al.*, 1993and 1994). We tested whether this is indeed the case by deleting the two elements together in the context of the 185 kb human β-globin locus PAC. As described in Chapter 4, analysis of this deletion showed that Enh and F indeed act as locus-wide embryonic stage-specific transcriptional repressors, but are not involved in the regulation of γ switching in the foetal liver and adult stages.

In Chapter 5 we applied the method of homologous recombination in separately deleting HS2 and HS3 from the human β-globin LCR. This project extends on earlier work on the deletion of HS2 and HS3 in the context of a smaller (70 kb) human β-globin locus construct (Milot *et al.*, 1996). This work showed different chromosomal position effects in transgenic mice resulting from the deletion of HS2 and HS3. For example, HS2 deletion led to a classical effect of PEV, whereas deletion of HS3 resulted in a novel type of position effect called cell timing position effect (Milot *et al.*, 1996). These observations raised the prospect that the deletion of specific HS sites gave rise to different chromosomal position effects. These studies, however, were done in multiple copy mice with only few transgenic lines available. We extended this work by obtaining additional HS2- and HS3-deleted lines. Only single copy lines were analysed. The results of this study revealed no correlation between deletion of a specific hypersensitive site and a specific type of position effect, in contrast to the effects observed in the studies by Milot *et al.*, 1996. The results of the study and discussion on the differences between this study and earlier studies on transgenic mice carrying the human or mouse β-globin locus with deletions in the LCR are described in Chapter 5.

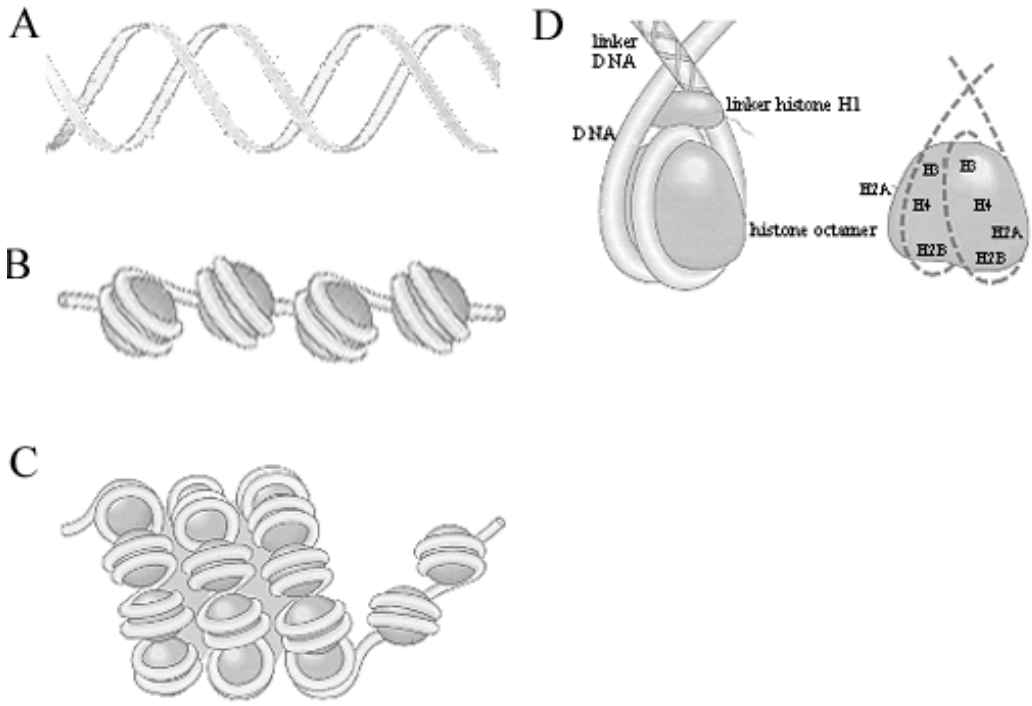



***Figure 1:*** *Chromatin structure*—A: DNA helix, B: "beads on a string", DNA wrapped around histone octamers resulting in a 10nm fiber, C: compaction of the chromatin fiber into the 30nm fiber, D: left, a schematic representation of the organisation of a nucleosome; right, the organisation of the core histones in the histone octamer. (adapted from Wolfe; Molecular and cellular biology, 1993).

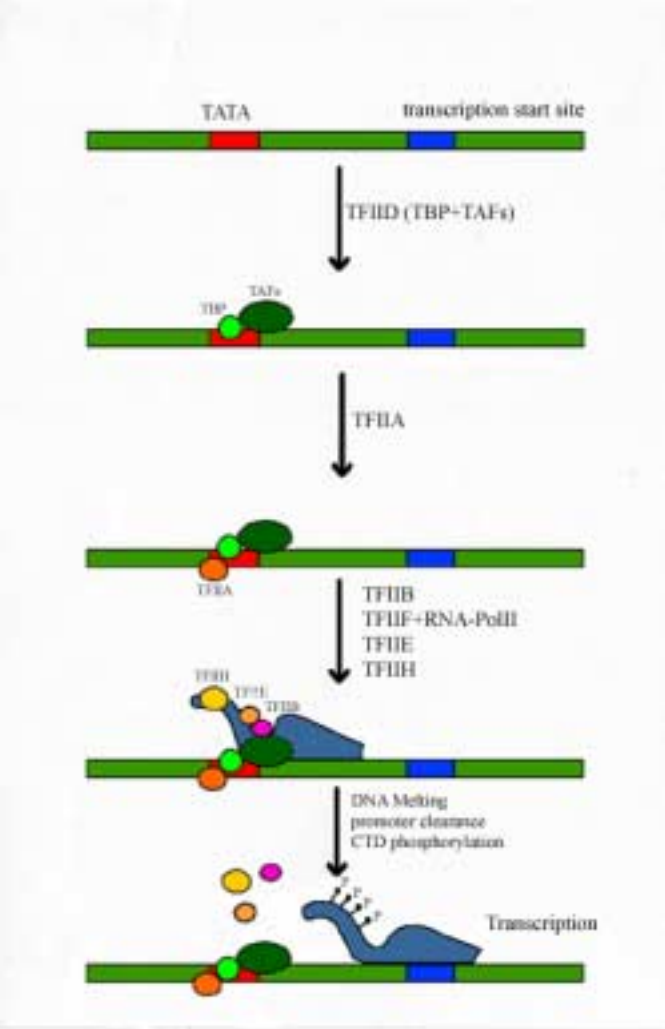



***Figure 2****: Transcription initiation*—The start of transcription initiation, binding of the TFIID complex, composed of TATA box binding protein (TBP) and TBP-associated factors, to the TATA box, followed by the recruitment of TFIIA, which stabilises the TFIID-DNA interaction. This complex undergoes a conformational change allowing TFIIB to bind and the recruitment of the TFIIF-RNA-Pol II complex, which is followed by the binding of the two last components of the complex, TFIIE and TFIIH. The final step of initiation includes, DNA melting; promoter clearance, for which the TFIIH and TFIIIE are responsible; and the CTD phosphorylation of the RNA-Pol II, done by TFIIH. The phosphorylated RNA-Pol II is released from the complex and transcription can take place. (adapted from Ogbourne and Antalis, 1998).

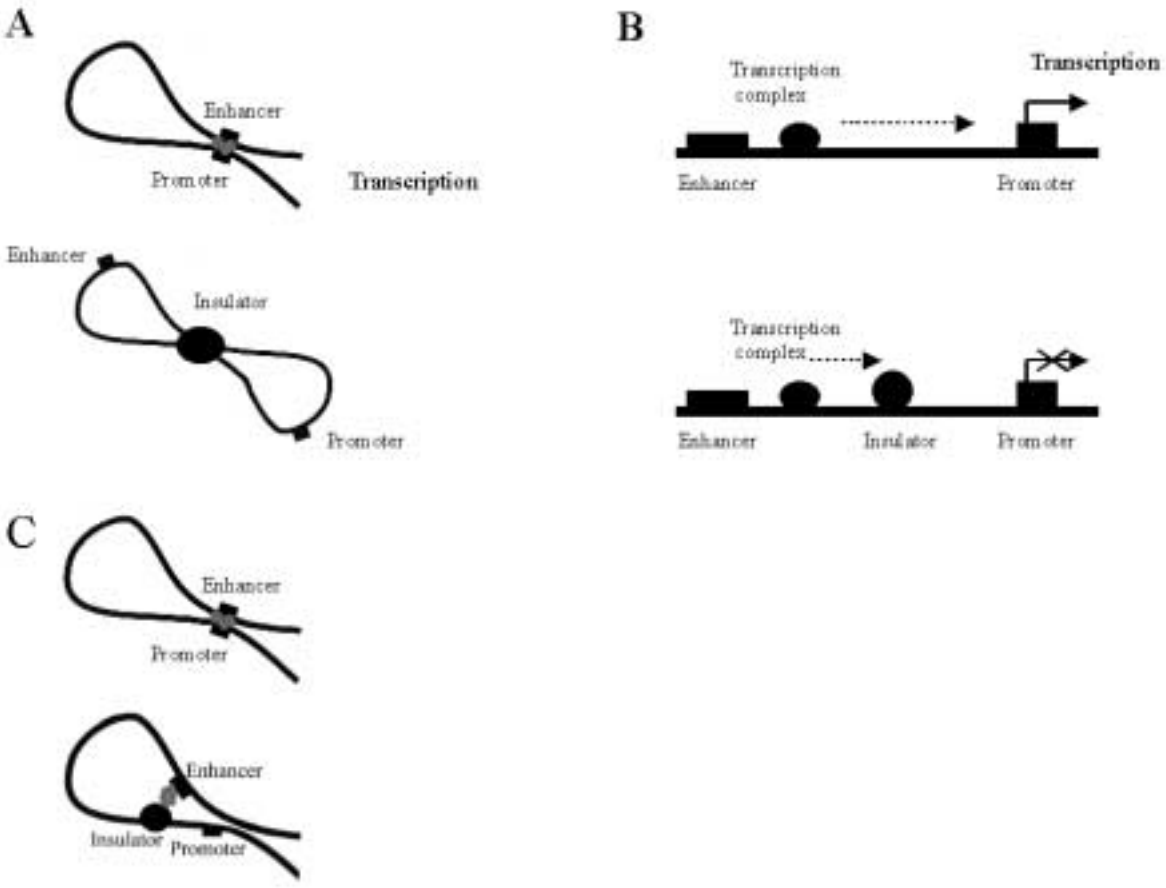



***Figure 3:*** *Local interaction models for insulator function*—A: the decoy looping model, the interaction between an enhancer and a promoter, as in the top panel, is prevented by the formation of an additional loop by insulator binding, placing enhancer and promoter in two separate loops, bottom panel. B: the derailment or tracking model, in which the spreading of activation signals, as in top panel, is blocked by the binding of an insulator, bottom panel. C: straight competition.

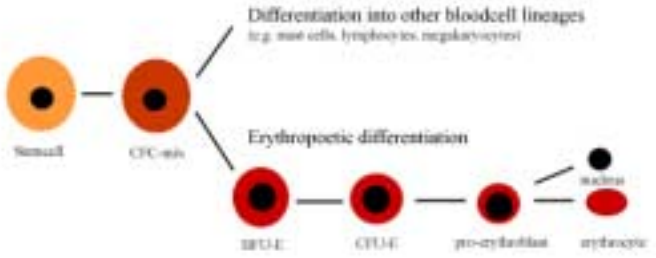



***Figure 4:*** *Erythroid differentiation pathway*—The erythrocytes arise from a pluripotent stem cell which divides and produces more pluripotent stem cells or commited progenitor cells (CFC-mix). Cells from the CFC-mix start to differentiate into the different bloodcell lineages under the influence of specific growth factors. During erythropoetic differentiation the nucleus becomes smaller because of chromatin condensation and in the final step of differentiation enucleation takes place, resulting in the erythrocyte. Accumulation of haemoglobin in the red blood cells takes place during the terminal phase of differentiation.
BFU-E: burst forming unit erythroid; CFU-E: colony forming unit erythroid; CFC-mix: colony forming cells.

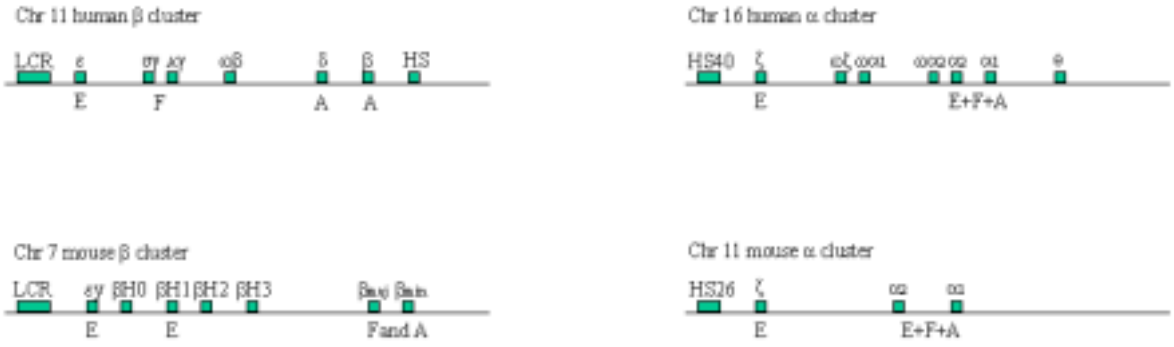



***Figure 5:*** *Schematic representation of the α- and the β-globin loci of human and mouse*—E: gene expressed during embryonic phase; F: gene expressed during foetal phase; A: gene expressed in the adult.

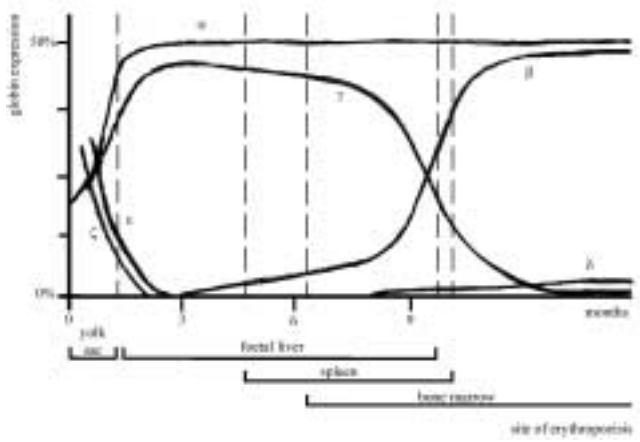



***Figure 6:*** *Expression patterns of the human α-and β-globin genes*—The site of erythropoeisis during development is depicted below the expression profiles of the genes.

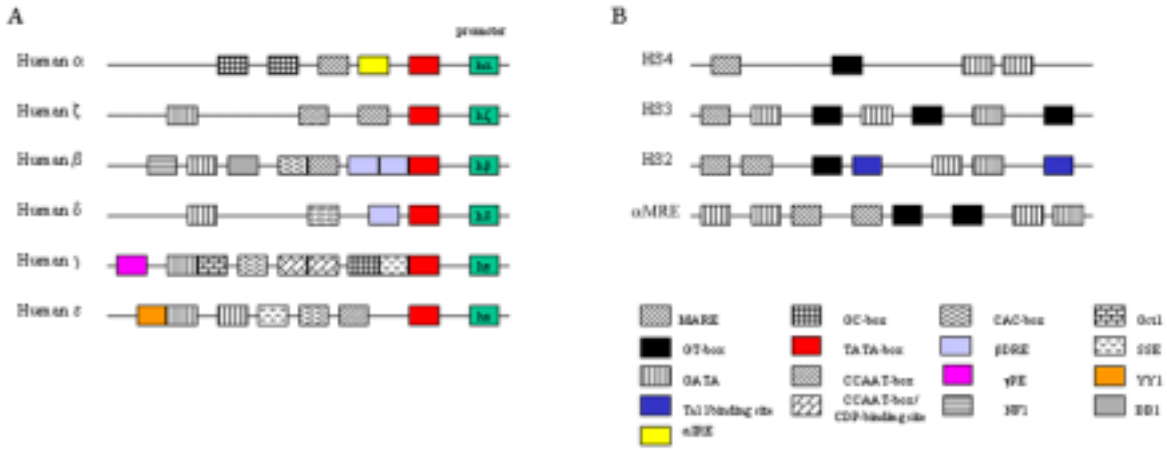



***Figure 7:*** *Schematic representation of the regulatory elements of the human α- and β-globin loci*—A: representation of the transcription factor binding sites in the regions upstream of the individual globin gene promoters. B: representation of the transcription factor binding sites present in hypersensitive sites 2, 3 and 4 of the human β-globin LCR and the αMRE of the human α-globin locus.

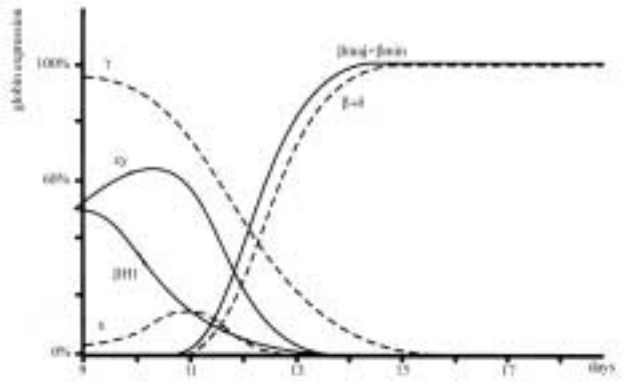


***Figure 8:*** *Comparison of the expression patterns of the endogenous mouse β-globin genes and the expression patterns of the human β-globin genes in transgenic mice*—Dotted lines represent the human genes and the black lines the mouse genes.

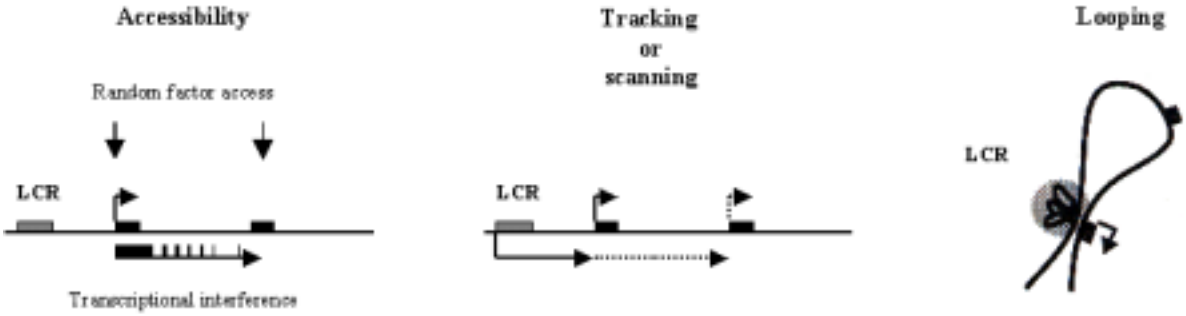



***Figure 9:*** *Long-range activation models proposed for the activation of the human β-globin genes by the LCR*—The arrows on the black boxes indicate transcriptional active genes. The arrow beneath the genes in the accessibility model indicates, that the transcriptional interference caused by an upstream gene decreases with distance.

# *CHAPTER 2*
# *STOCHASTIC PATTERNS IN GLOBIN GENE EXPRESSION ARE ESTABLISHED PRIOR TO TRANSCRIPTIONAL ACTIVATION AND ARE CLONALLY INHERITED*

Mariken de Krom, Mariette van de Corput, Marieke von Lindern, Frank Grosveld and John Strouboulis

*Molecular Cell* Vol **9**, 1319-1326 (2002)

## Summary

We have undertaken a detailed characterization of mouse globin gene expression patterns in the nucleus and cytoplasm of single erythroid cells. We demonstrate an imbalance of α- versus β-globin expression in a significant proportion of cells both in nuclear transcription patterns and cytoplasmic mRNA levels. Clonal cell analysis showed these expression patterns to be clonally inherited, while analysis of a multicopy transgenic locus showed an all-or-none effect in the activation of all the genes in one locus. These data provide strong evidence for a stochastic basis of globin gene activation resulting in heritable all-or-none expression patterns.

## Introduction

A stochastic basis for transcriptional activation can be viewed as a sequence of events that combines a random component, such as transcription factor binding, with a selective step, such as cell commitment, so that only certain outcomes of the random event persist (Bateson, 1979). In this model, cis regulatory elements, such as enhancers, or an increase in an inducing signal, such as a transcription factor, increase the probability that a gene will be transcriptionally activated (Ko, 1990).

One implication of the probabilistic nature of a stochastic model is that the individual alleles of a gene will be regulated independently. Support for this comes from random X chromosome inactivation in dosage compensation (reviewed in Goto and Monk, 1998), from allelic exclusion of immunoglobulin genes in immune responses (reviewed in Nemazee, 2000), and from random monoallelic expression of genes involved in lineage commitment, such as olfactory receptors or T and NK cell receptors (Chess *et al*., 1994; Held *et al*., 1999), and in cellular responses to external stimuli, such as the cytokine genes IL-2 and IL-4 in activated T cells (Hollander *et al*., 1998; Bix and Locksley, 1998; Riviere *et al*., 1998). In all these cases, stochastic choices in allelic expression patterns of a limited number of genes give rise to a diverse repertoire of cells, each with a highly restricted specificity. Upon stimulation/induction, the cell is committed and clonally expanded to provide a specific response (Bix and Locksley, 1998; Chess, 1998; Shulman and Wu, 1999; Held *et al*., 1999). The advantage of a stochastic mechanism is that it creates diversity by allowing choice at low cost. While this predicts that individual cells will vary, the net result for a cell population will be a stable outcome since large numbers of cells, each with the same probabilities for a particular event, are involved (Michaelson, 1993).

The nuclear expression patterns of α- and β-globin genes appear consistent with the predictions of a stochastic model (Wijgerde *et al*., 1995, 1996; Gribnau *et al*., 1998; Trimborn *et al*., 1999). Total mouse globin mRNA levels reflected the number of actively transcribing genes in nuclei, suggesting that, when transcribing, genes are fully on. Moreover, different combinations of mouse globin expression patterns were observed, suggesting that each globin locus is independently regulated.

Here, we examine the apparent stochastic basis for globin gene activation by calculating the proportion of genetically identical mouse erythroid cells that fail to activate one or both of the α- and β-globin alleles. This results in an imbalance in α- versus β-globin transcription patterns which is reflected in α/β mRNA ratios in the cytoplasm. We also provide evidence that globin allelic transcription patterns are clonally inherited and that decisions for transcription occur in an all-or-none fashion at a step prior to transcriptional activation.

## Results

### *Independent Activation of Globin Gene Alleles*

*In situ* hybridisation was used to detect globin nuclear primary transcripts and/or cytoplasmic mRNA in 14.5 days post coitus (dpc) mouse foetal liver cells, which express the adult α-globin and $\beta_{maj}$- and $\beta_{min}$-globin genes (Trimborn *et al*., 1999). The efficiency of hybridisation in these assays was

shown to be ~99% (data not shown). We then measured α- and β-globin primary transcription foci using intron-specific probes. A range of allelic expression combinations was observed with the majority of cells expressing two α and two β alleles (Figures 1A–1B). Active alleles per cell were counted and grouped as shown in the table in Figure 1. From this, two important points emerge. First, while >73% of cells (Figure 1C, column 4) exhibit balanced transcription of 2α- and 2β-globin alleles, the remainder of the cells (~17% of the total, Figure 1C) fail to transcribe at least one globin allele. Second, α- and β-globin transcription is clearly unequal, as more cells express predominantly or exclusively α-globin, compared to cells expressing mostly β-globin (18% versus 5%; Figure 1C, columns 1 + 2 + 3 versus columns 6 + 7 + 8). This is contrary to predictions if one assumed an equal activation potential for α- and β-globin alleles. The number of cells expressing only β alleles is not statistically significant, as it is below the 1% limit of hybridization efficiency.

We next calculated the probability of activation for each α (pα)- and β (pβ)-globin gene. For α-globin, pα can be calculated from the 2α-expressing cells ($p\alpha^2 = 87.2\%$ [sum of columns 2–4, Figure 1C]; therefore pα is 93.4%). Similarly, pβ is calculated to be 88.4% (sum of columns 4, 6, and 7, Figure 1C). This would result in an expected 20.5% or 1.3% of cells expressing a single or no β gene, respectively. In fact, the observed number of single or no β allele-expressing cells is 12.4% and 9.3%, respectively (columns 3, 5, and 8, and 1 and 2, respectively, Figure 1C). This discrepancy is caused by the high incidence of α-only-expressing cells. The trivial explanation for the higher number of α-only-expressing cells (contamination with embryonic cells expressing α-globin) was excluded by *in situ* hybridisation of 14.5dpc foetal liver cells using probes specific for mouse α-globin and the mouse εy-embryonic-globin gene (data not shown).

### *α-Globin Is Activated Earlier Than β- in Erythroid Differentiation*

We can think of three plausible explanations for the high number of α-only-expressing cells. First, mouse α-globin genes could, on average, be transcribed earlier than β genes in erythroid cell maturation. Second, the α genes could be expressed at a specific cell cycle stage when the β genes are not active (e.g., in G1 phase). Third, the α-only cells could result from discontinuous transcription, as has been previously described for the α- and β-globin genes (Wijgerde *et al*., 1995; Gribnau *et al*., 1998), with the β genes transcribing less frequently than the α genes. This would result in cells with α on in the nucleus, but with the β gene(s) temporarily off.

To address the first possibility, we size fractionated mouse foetal liver cells into large immature erythroid cells and small, more mature erythroid cells. *In situ* hybridisation with α- and β-globin exon probes, which detect cytoplasmic mRNA and nuclear primary transcripts, showed that the large immature erythroid cells have one or two α-globin primary transcripts and little or no cytoplasmic α mRNA. No β primary transcripts or mRNA were present (Figure 2A). Thus, these cells have just started transcribing the α-globin genes ahead of the β genes and have not yet built up detectable mRNA in the cytoplasm. The vast majority of the more mature cells contain both α- and β-globin primary transcripts and mRNA (Figure 2B). Re-counting shows a similar distribution of primary transcripts as in Figure 1C, with the exception that the α-only cells or cells expressing more α than β have now decreased (Figure 1D). We conclude that the α-globin locus becomes transcriptionally active before the β-globin locus.

The observation that cells with α-only signals in the nucleus had little or no α signal and no β signal in the cytoplasm argues against the possibilities that the α genes are activated at a stage of the cell cycle when β is off or that the α genes are transcribed more frequently than β. If these possibilities were true, these cells would have both α and β cytoplasmic mRNA synthesized in previous cell cycles. Thus, the α-only cells are primarily due to the earlier activation of β-globin during erythroid cell differentiation.

We recalculated the activation probabilities for both α- and β-globin genes, using the mature erythroid cell fraction. We now found pα to be 94% ($p\alpha^2 = 88.4$, columns 2–4, Figure 1D). Conversely, pβ in fractionated cells is 96% ($p\beta^2 = 92.2$, columns 4, 6, and 7, Figure 1D). This predicts 7.7% of cells expressing a single β allele, which agrees with the 7.6% observed (columns 3, 5, and 8, Figure 1D). The non-β-expressing cells are predicted to be 0.16%, with 0.21% observed (columns 1,

Figure 1D). Hence, we conclude that after removal of the early erythroid cells, α and β expression patterns closely resemble a stochastic distribution, with a very similar probability for α- and β-globin activation.

### *Globin Primary Transcript Imbalance Is Maintained in the Cytoplasm*

The imbalance in α versus β allelic expression in the nucleus of erythroid cells is consistent with a stochastic basis in gene activation. However, it cannot be excluded that the distribution of nuclear transcription patterns observed may be due to discontinuous transcription where an α or a β gene may be temporarily off. If the distribution of nuclear primary transcripts is reflected in α/β mRNA ratios in the cytoplasm, it would confirm that the nuclear transcription patterns are stable and become fixed and would strongly argue for a stochastic activation mechanism.

We carried out *in situ* hybridisation on 14.5dpc foetal liver cells using exon-specific probes and quantitated the relative abundance of cytoplasmic α- and β-globin mRNA (Figures 2C–2F). The distribution obtained for α/β mRNA ratios is shown in Figure 1E. We found 26.7% of the cells having an α/β imbalance. Of those, 15.7% have α > β, with 5.2% of the cells having α/β > 2. The latter group very likely represents the cells that in the primary transcript analysis were expressing only α alleles or 2α1β alleles.

The overall distribution of α/β-globin mRNA ratios agrees with that obtained for the primary transcript analysis, except that the number of cells with α-only primary transcripts is higher than the number of cells with high α/β mRNA ratio in the cytoplasm, because this group also contains the early α-only-expressing cells which have undetectable levels of cytoplasmic α-globin mRNA. Thus, the distribution of globin mRNA in the cytoplasm reflects that of primary transcripts in the nucleus. We conclude that the primary transcript patterns observed in the nucleus cannot be explained by genes being temporarily off but instead demonstrate a stochastic basis for α- and β-globin gene activation.

### *Cellular Globin Expression Patterns Become Fixed during Differentiation*

The agreement of nuclear transcription patterns with mRNA levels suggests that early stochastic choices become fixed in differentiated erythroid cells. We therefore carried out methylcellulose colony assays on single immature erythroid cells, purified under conditions of cell differentiation that allow the visualization of globin transcription before cell enucleation. Individual colonies were harvested after around five divisions (32 cells) and assayed for globin primary transcript expression patterns. If choices in globin expression are not fixed and remain purely stochastic during differentiation, one would expect each colony to display the distribution of nuclear expression patterns observed in the bulk of foetal liver cells (Figure 1C). If, instead, stochastic decisions become fixed, a deviation from the expected stochastic patterns would be predicted (Figure 3A).

We find that the distribution of expression patterns of all the cells assayed in all the colonies is no different to that observed with bulk foetal liver cells. A stochastic model based on the calculated probabilities for α or β gene activation in late erythroid cells (94% and 96%, respectively) would predict the following: in 3.5% of the clones, 100% of the cells express 2α2β; in 21.2% of the clones, 90%of the cells express 2α2β (with the remaining having a different pattern); in 61.6% of the clones, 80% of the cells express 2α2β; in 11.9% of the clones, 70% of the cells express 2α2β, while only 0.6% of the colonies would have cell populations in which less than 60% of the cells express 2α2β. This would lead to a normal distribution (Figure 3B, black bars and curve). Instead, we find a totally different distribution (Figure 3B, gray bars). Many more clones than expected exclusively contain only 2α2β cells or have less than 60% of the cells per clone expressing 2α2β. Many fewer than expected clones contain 80% of cells expressing 2α2β (Figure 3B, gray bars). Importantly, all cells in a colony with patterns other than 2α2β had the same pattern, with two exceptions that contained two patterns different from 2α2β. This can be explained if two or three choices were made in these colonies and subsequently fixed (Figure 3A, right panel).

*All-or-None Activation of Human β-Globin Locus Transgenes*

In the experiments above, we measured transcriptional activation as the read-out for the whole process. We therefore asked whether the stochastic distribution we observe in globin expression patterns is determined prior to or at the transcriptional activation step by examining the activation of three copies of a 180 kb human β-globin locus PAC integrated in a head-to-tail fashion in transgenic mouse lines. If stochastic expression is determined at the transcriptional activation step, each β-locus in the three linked loci would transcribe independently, and, as a result, a substantial number of cells would express only one or two of the three β-loci.

To test this, we used two different three-copy transgenic mouse lines, one bearing a telomeric integration while the other line carried a nontelomeric, noncentromeric insertion. We assumed a similar probability of activation for human β-globin and mouse β-globin (pβ = 88% in nonfractionated cells), as they have equal expression levels per gene copy. If stochastic distribution is determined at the step of transcriptional activation, we would predict a 68% chance for all three loci to be active in the nucleus of one cell ($p\beta^3 = [0.88]^3 = 0.68$). The probability for all three loci to be active in, e.g., ten cells, is around 2% ($[0.68]^{10} = 0.02$).

Foetal liver cells from the three-copy transgenic lines were probed by *in situ* hybridisation with either a 12 kb LCR-derived probe (Figure 4A) or the entire 180 kb PACderived human β-globin locus (Figure 4E1). The LCR probe clearly shows the three separate LCR signals 180 kb apart. Hybridization with the PAC insert shows a brighter, more contiguous signal as more DNA is hybridized. We did a primary transcript hybridization to detect mouse α- and human β-locus transcription. Three types of cells were observed (Figure 4B): some had no α or β signals and are presumably nonerythroid or early erythroid cells (white arrow head, Figure 4B); 19% (telomeric line) and 17% (nontelomeric line) of the remaining cells had only α signals (Figure 4B, solid white arrow); while 81% and 83% of the remaining cells showed α and human β signals (Figure 4B, open white arrow). This fits well with the number predicted from the table in Figure 1, considering that these mice are heterozygous for the human β-globin locus. Deconvolution of confocal images showed the human primary transcripts appearing as a triple signal (Figure 4B, CCD image open arrow; Figure 4C, confocal image, two focal planes merged), while rotation of the nucleus showed the triple signals as short tracts of primary transcript (Figure 4D, confocal image, one focal plane).

We next analyzed 52 human β-expressing cells from the nontelomeric integrant by confocal microscopy for RNA (γ and β intron probes) and DNA (entire PAC). In 49 cells, primary transcripts were visualized as three discernible RNA "tracks" coincident with the human β locus DNA (Figures 4E1–4E3). We did not observe any cell with a single primary transcript track. The remaining 3 (out of 52) cells had coincident primary transcript and DNA signals which, under the confocal microscope, appeared as single large bright foci rather than tracks, indicating a less extended conformation of the DNA. In 9 out of 12 nonexpressing cells, we also observed three DNA signals, indicative of an extended conformation. Essentially identical results were obtained with the telomeric integrant line (n = 30 cells).

In summary, when human β-globin is transcribed, all three loci are active in an all-or-none fashion. We conclude that the stochastic distribution we observe for the human β-globin locus is determined at a step prior to actual transcriptional activation.

## Discussion

Our analysis revealed that a significant proportion of erythroid cells (23.2%) show an imbalance of 2α-versus 2β-globin gene expression, which is maintained in the levels of α- and β-globin cytoplasmic mRNA. In addition, patterns of expression were clonally inherited during erythroid cell differentiation, suggesting that when an allele is off, it remains off. The decision for expression appears to occur in an all-or-none fashion at a step prior to transcriptional activation. These observations have implications for globin gene regulation and the basis of gene activation in general.

### *A Stochastic Basis for Globin Gene Activation*

The α- and β-globin gene loci reside on different chromosomes and are regulated differently and independently. The α-globin locus, unlike the β-locus, lies in an early replicating, constitutively open chromatin domain and does not appear to require the presence in cis of a complete locus control region (LCR), an element required for the continued activation of the β-globin locus (review Higgs *et al*., 1998; Grosveld, 1999). This suggests that the α-globin locus may require fewer steps for activation from its silent state. In addition, there are differences in the transcription factor requirements of the two geneclusters.

The data on globin gene expression patterns presented here provide strong evidence for a stochastic model for transcriptional activation of both globin geneclusters. The higher probability for α-globin activation in the total cell population could result from a lower threshold for productive transcriptional activation, for example, by having more transcription factor binding sites. The constitutively open chromatin configuration of the α-locus could also place it in a nuclear subdomain (Brown *et al*., 2001) that may have a higher diffusion rate for transcription factors (Misteli, 2001). Consequently, α genes would, on average, be transcribed earlier than β genes during differentiation, while at later stages a much smaller proportion of α-only cells will remain. This was observed when we fractionated erythroid cells.

These observations could also be explained by a graded model, in which two alleles of a gene are equally regulated in a graded manner in response to an inducer. However, a graded model would need to invoke additional parameters to account for the lack of transcribing alleles in a significant proportion of the cells. Even so, this model would predict no globin mRNA imbalance in the cytoplasm, as opposed to what is observed in the nucleus. The agreement between primary transcript patterns and cytoplasmic mRNA levels, therefore, argues against the graded response model in this system.

### *Stochastic Expression Choices Are Fixed prior to Transcriptional Activation of a Locus*

We found that mouse globin allelic expression patterns are clonally inherited in differentiated cells. We also showed that, at least for a human multicopy β-globin transgenic locus, the decision to express occurs as an "all-or-none" choice prior to transcriptional activation, probably on a basis of mass action (Grosveld, 1999). Heritable on-or-off decisions are also observed in position effect variegation (PEV) and are thought to be regulated at the level of chromatin structure and subnuclear localization (Wakimoto, 1998; Milot *et al*., 1996). In PEV, mass action has been invoked to account for the stochastic, long-range spreading of heterochromatin in an all-or-none event, which may close down the expression of adjacent genes (Locke *et al*., 1988). Our results suggest that mass action in the β-globin locus works in the opposite direction, in that it opens up chromatin in the locus prior to transcriptional activation. Genetic evidence supports the β-locus LCR as playing a vital role in this all-or-none effect (Grosveld, 1999). To complete the parallel between PEV and globin locus activation, one would need to determine whether the chromatin of nonexpressing globin loci have failed to open, as has been observed in PEV (Wallrath and Elgin, 1995). However, this is technically unfeasible in the current *in vivo* system. Nevertheless, we suggest that the initial choice of activating or not a globin locus is made in an early window of opportunity and subsequently fixed by an as yet unknown parameter at the chromatin level.

### *Globin Gene Expression and Other Stochastic Activation Phenomena*

Stochastic choices in allelic expression have been described in X-inactivation, allelic exclusion, and monoallelic expression and have been proposed to play a role in giving rise to a diverse repertoire of cells, each with a highly restricted specificity to provide a specific response to a stimulus. Globin gene expression, however, does not fit the pattern set by the examples of stochastic choices above. Globin genes are transcribed at high levels in the erythroid lineage with the aim of synthesizing α- and β-globin protein chains in a 1:1 ratio. These cells are fully committed to haemoglobin accumulation and enucleation to form mature erythrocytes; hence they are not part of a cellular response pathway and play no role in cell lineage commitment decisions. It is in fact a disadvantage to the cell to not express

balanced α- and β-globin chains, as it shortens its lifetime. The α- and β-loci have therefore evolved a system whereby the probability of activation is sufficiently high to give a 1:1 ratio of α and β product in the total cell population at a loss of some non-1:1 cells that is sufficiently small so as not to result in a selective disadvantage.

It is tempting to suggest that genes in other tissues may not express for stochastic reasons, even if this were to occur less frequently than observed for the globin genes. For example, even if the chance for activating a gene were 99.9%, one in a million cells would express neither allele of that gene. Such phenomena would be unnoticed if small numbers of cells were assayed, but could play a role when large numbers of cells in a tissue are involved. For example, the failure to activate a tumour suppressor gene in the skin (even in one in a million cells) would increase the chance for transformation in a substantial number of cells. For that reason (among others), additional cellular mechanisms may have evolved to control the phenotypic consequences of such events.

## Material and Methods

### *In Situ Hybridisation*

Mouse 14.5dpc foetal liver cells were isolated, fixed, hybridised, and detected as previously described (Wijgerde *et al*., 1995; van de Corput and Grosveld, 2001). Probe sequences and hybridisation efficiencies have been previously reported (Gribnau *et al*., 1998; Trimborn *et al*., 1999). As control for colour and labelling differences, identical *in situ* hybridisations were performed reversing the fluorescent labels for the α- and β- globin intron and exon probes, yielding similar results. To compare levels of two mRNAs in single cells, a control two-colour mRNA *in situ* hybridisation was done using a single probe mixture with half of the probe mix labelled with digoxigenin (DIG) and the other half with dinitrophenol (DNP). Measuring red and green signal intensities in the cytoplasm revealed the background signal and the distribution intrinsic to the 1:1 ratio of red/ green signal for the same mRNA (Figures 2C and 2F) serving as the baseline distribution for equal amounts of two different mRNAs. Fluorescence was detected by epifluorescence/CCD or confocal microscopy.

DNA-RNA *in situ* hybridisation of the three-copy human β-globin locus was essentially as described by van Raamsdonk and Tilghman (2001). Human γ- and β-globin intron-specific biotinylated probes were used together with mouse α-globin DIG-labeled intron probes. DNA FISH probes included DIG-labelled human β-globin locus PAC insert and a 12 kb LCR probe fragment. Cells were analysed by epifluorescence microscopy or/and confocal laser scanning microscopy, imaging software provided by the Zeiss LSM 510, and deconvolution software (Scientific Volume Imaging, Huygens v. 2.0).

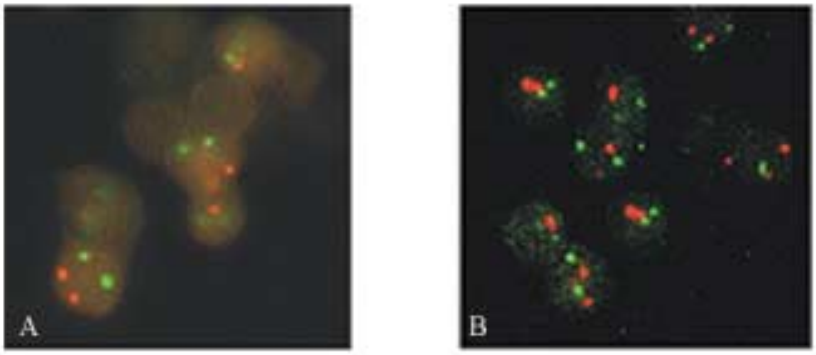
### *Percoll Density Gradient*

Mouse 14.5dpc foetal liver cell suspensions were layered on top of a 2 ml cushion of Percoll (SIGMA-Aldrich) with a density of 1.0735 g/ml and centrifuged at room temperature, 10 min at 2000 rpm (no break for last 300 rpm). The top and bottom fraction contained the immature (larger) and mature (smaller) erythroid cells, respectively. Fractionated cells were pelleted, resuspended in medium, and spotted onto slides for *in situ* hybridisation as above.

### *Single Cell Clonal Assay and Statistical Analysis*

Single immature erythroid cells from Percoll gradient fractions were obtained by limiting dilution and cultured in 50 µl conditioned medium with erythropoietin and stem cell factor, for 3 days. For the methylcellulose assay, 5,000–10,000 cells were mixed with 1ml of methylcellulose solution in cfu- or bfu-conditioned medium and grown for 3 days at 37°C in 10% $CO_2$. Single colonies were fixed for primary transcript *in situ* hybridisation on heavy teflon supercured coated slides (Nutacon), ten wells per slide, each well treated separately to prevent mixing. α- and β-globin allelic expression patterns in all cells in each well were counted. The predicted distribution was derived from the following formula: (n!/(n!-x!*x!))*(0.8n-x*0.2x), where n = 30 (total number of clones) and x is the number of clones

predicted with a specific allelic expression pattern. The probability for each allelic distribution observed per clone was calculated using the npar $\chi^2$ test of the SPSS9 package. The distribution of probabilities observed was statistically tested against the expected distribution using the one-sample Kolmogorov-Smirnov test. The observed distribution was statistically different at p level p = 0.00. In graphing the data, we grouped the percentage of clones with all cells expressing 2α2β genes (100%), then the percentage of clones with 90% of the cells expressing 2α2β with the other 10% of cells having another transcription pattern (almost always just one other pattern), followed by the percentage of clones with 80% 2α2β cells, followed by the percentage of clones with 70% 2α2β cells, and finally all the clones that had 60% or less 2α2β-expressing cells.




| | Expressing | 1α | 2α | 2α1β | 2α2β | 1α1β | 1α2β | 2β | 1β |
|---|---|---|---|---|---|---|---|---|---|
| | Alleles α<br>Alleles β | | | | | | | | |
| C | % cells with primary transcripts<br>sdv | 4.03<br>0.97 | 5.31<br>1.84 | 8.69<br>0.53 | 73.19<br>2.63 | 3.66<br>0.54 | 4.94<br>1.08 | 0.09<br>0.19 | 0.09<br>0.09 |
| D | % cells with primary transcripts (+sdv) after fractionation | 0.21<br>0.30 | 0.00<br>0.00 | 3.80<br>1.16 | 84.57<br>0.76 | 3.83<br>0.03 | 6.77<br>0.66 | 0.84<br>0.59 | 0.00<br>0.00 |
| E | % cells with different α/β mRNA ratios | Rα/β > ctrl = 15.7 (Rα/β > 2 = 5.2) | | | Rα/β = ctrl 73.3 | | Rα/β < ctrl = 11 (Rα/β < 2 = 4.2) | | |

***Figure 1:*** *Primary Transcript In Situ Hybridisations of 14.5dpc Foetal Liver Cells*—(A and B) Two representative hybridizations with DNP-labeled mouse α-globin intron probe (red) and DIG-labeled mouse β-globin intron probe (green). (A) One focal plane. (B) Stack of confocal images. Different allelic expression combinations are visible. Hybridization efficiency was measured using probes against the introns and the 3rd exon of the α- or β-globin genes. Less than 1% of the exon-positive cells failed to show preceding exon or intron signal, indicating ~99% hybridization efficiency (data not shown). (C–E) Summary of α- and β-globin expression patterns in the nucleus and cytoplasm of 14.5dpc foetal liver cells. (C) α- and β-globin primary transcript patterns counted and tabulated as a percentage of total cells (n = 1200). (D) Re-counting of α- and β-globin primary transcript signals in fractionated mature erythroid cell fraction (n = 500). (E) Ratios of α/β cytoplasmic mRNA levels per cell (n = 200) were plotted as histograms (Figures 2C–2E) after normalization of the control (α/α or β/β) to a ratio of 1.0. Abbreviations: sdv, standard deviation; Rα/β, α/β cytoplasmic mRNA ratio; ctrl, control (α/α or β/β).

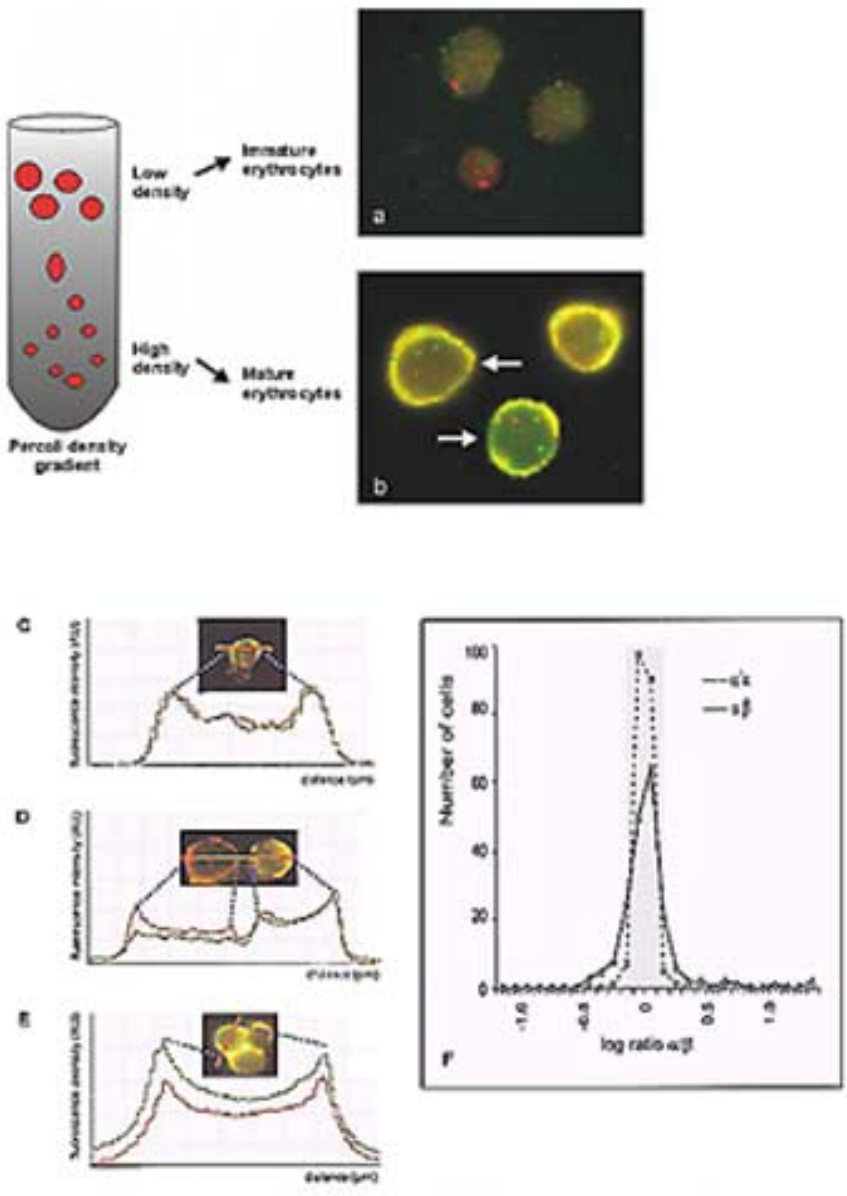



***Figure 2:*** *Visualization of Nuclear and Cytoplasmic RNA in Fractionated Erythroid Cells* —(A and B) Percoll density gradient size fractionation of erythroid cells. (A) Fraction enriched in large immature erythrocytes. Only α-globin nuclear staining is observed with no or very little cytoplasmic staining. No β-globin signal is detected. (B) Fraction enriched in small mature erythrocytes. Both α- and β-globin nuclear transcripts and cytoplasmic staining are observed (arrows). (C) Control mRNA *in situ* hybridisation using a mixture of one α-globin exon probe, half of which was labeled with DNP (red) and the other half with DIG (green). The two signals were overlaid and cytoplasmic intensities measured by drawing a line through single cells (n = 200) and plotted as two peaks in a histogram, one for the green signal (green line) and one for the red signal (red line). The average of two measurements was calculated. The ratio obtained served as control for background fluorescence and for the distribution intrinsic to the 1:1 ratio of red/green signal for the same mRNA. (D and E) Quantitation of signal intensities for α- (red) and β-globin (green) mRNAs. Three examples of cells with different α/β cytoplasmic ratios are shown. (F) Line representation of the ratios of α/β-globin mRNA in the cytoplasm as obtained from (C–E). The dotted line indicates the control distribution of α/α ratios, with the shaded area indicating cells with a 1:1 cytoplasmic mRNA ratios (β/β ratios are identical). 200 cells were scanned in randomly chosen fields. The ratios for signal intensities (α/α and α/β) were normalized to 1 using all ratios between 0.5 and 2.0. The cells were grouped with a ratio interval of a log of 0.1, and the number of cells falling in each group was plotted against the log of the ratios.

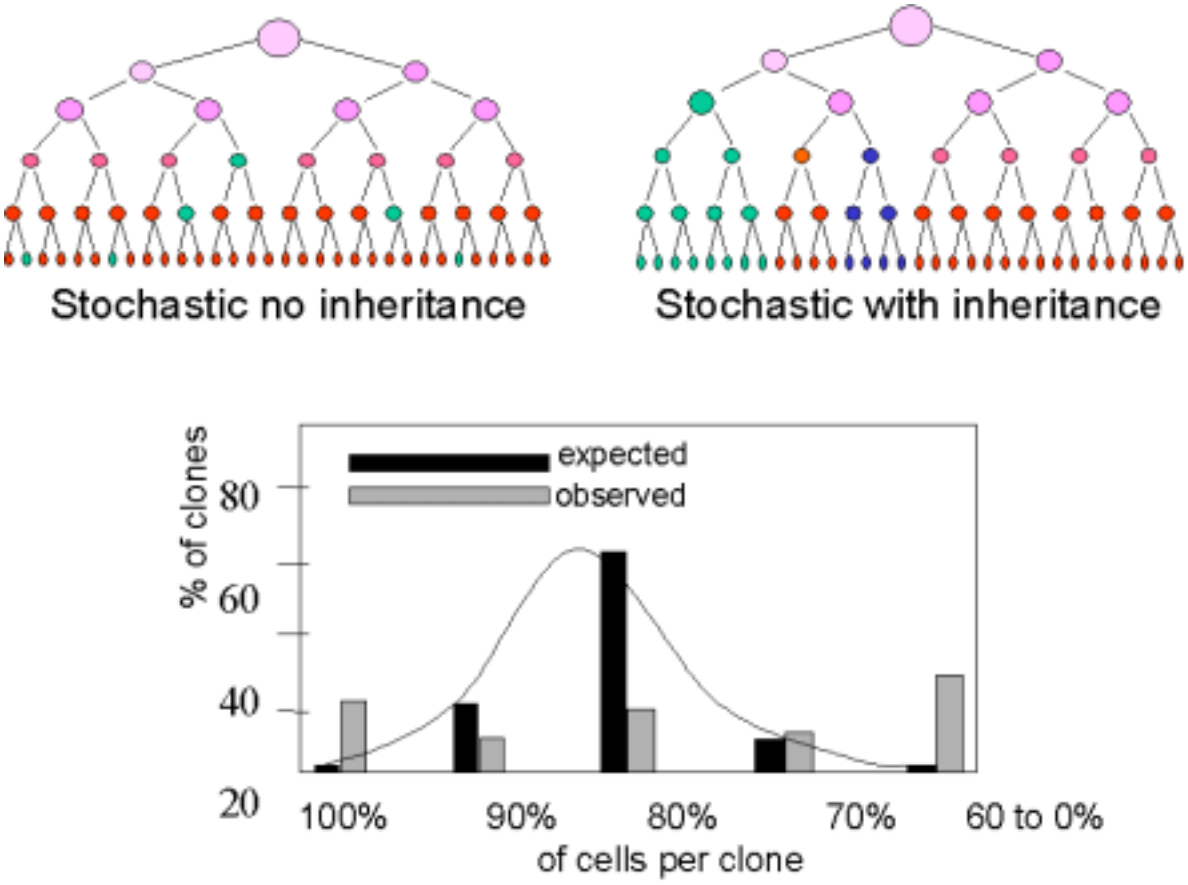



***Figure 3:*** *Clonal Analysis of Mouse α- and β-Globin Primary Expression Patterns*—(Upper panels) Examples of expected primary expression patterns if stochastic decisions are taken at the transcriptional activation step (left panel) or at a step prior to that with fixation of a choice in daughter cells (right panel). White, uncommitted cells; black, cells expressing $2\alpha2\beta$; gray, cells expressing a pattern other than $2\alpha2\beta$. (Bottom panel) All cells in each clone were scored for expression, and the percentage of clones was plotted against the expression pattern per clone. These patterns were grouped as described in Experimental Procedures. Black bars and curve represent expected distribution for transcription patterns made according to (left upper panel). Gray bars are the observed distribution, suggesting fixing (Upper right panel) around the 3rd to 4th replication of cells.

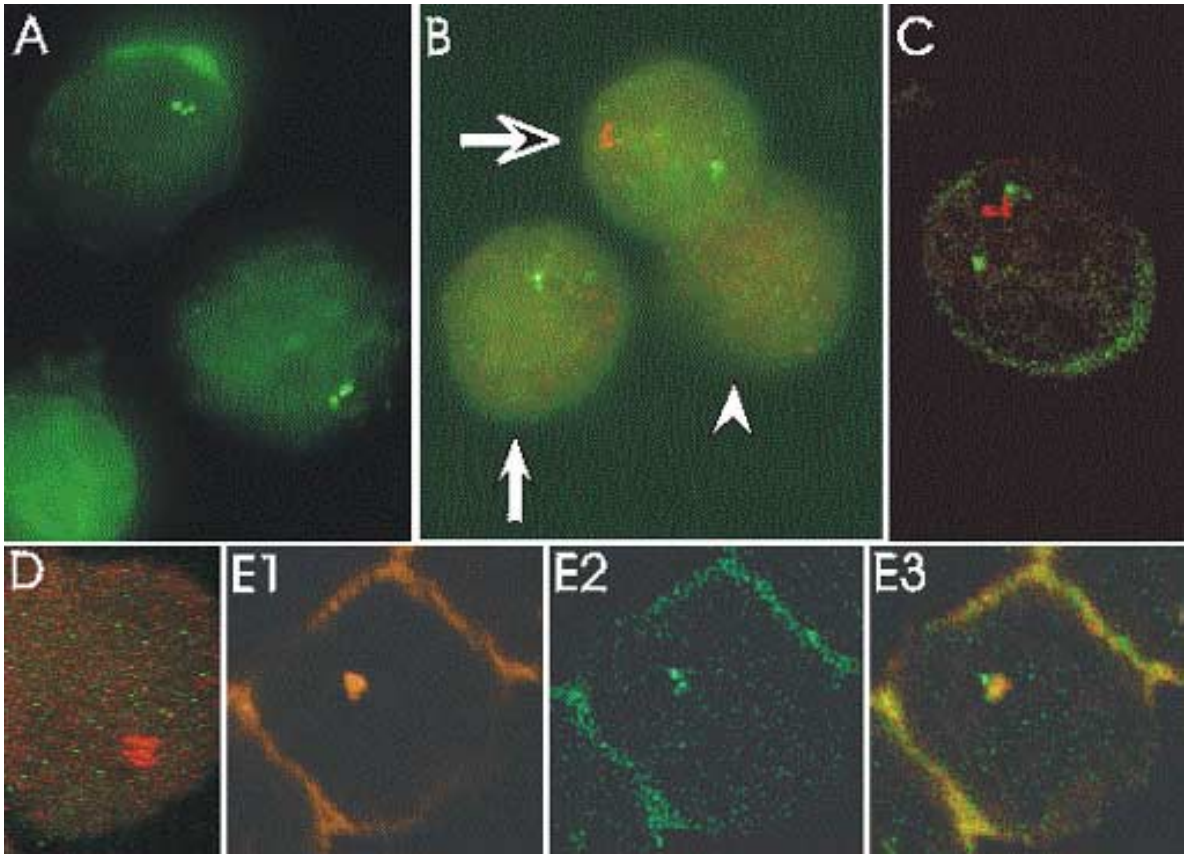



***Figure 4:*** *In Situ Hybridisation of Foetal Liver Cells from a Transgenic Mouse Heterozygous for a Three-Copy Human β-Globin Locus*—(A) Hybridization with a 12 kb LCR probe (green). (B and C) Hybridization with probes for α (green) and β (red) introns detecting nuclear primary transcripts. Photographs by CCD camera (B) or by confocal microscopy with two focal planes superimposed (C). (D) Deconvolution of confocal images as in (C) and rotation to visualize RNA tracks in one focal plane. (E1–E3) *In situ* hybridisation to detect human β-globin primary transcripts with an intron probe (green, [E2]) and the β-globin loci with a PAC probe (red, [E1]) to show colocalization (E3) of human β transcription and the human β loci by superimposing (E1) and (E2).

# *CHAPTER 3*
# *MODIFICATION OF HUMAN β-GLOBIN LOCUS PAC CLONES BY HOMOLOGOUS RECOMBINATION IN ESCHERICHIA COLI*

Ali Imam, George Patrinos, Mariken de Krom, Stefania Bottardi, Rick Janssens, Eleni Katsantoni, Albert Wai, David Sherratt and Frank Grosveld

*Nucleic Acids Research* Vol **28** e65 (2000)

## Summary

We report here modifications of human β-globin PAC clones by homologous recombination in *Escherichia coli* DH10B, utilising a plasmid temperature sensitive for replication, the *recA* gene and a wild-type copy of the *rpsL* gene which allows for an efficient selection for plasmid loss in this host. High frequencies of recombination are observed even with very small lengths of homology and the method has general utility for introducing insertions, deletions and point mutations. No rearrangements were detected with the exception of one highly repetitive genomic sequence when either the *E.coli* RecA- or the lambdoid phage encoded RecT and RecE-dependent recombination systems were used.

## Introduction

The ability to obtain high level, tissue specific and integration site-independent expression of genes requires the use of large genomic fragments. This often precludes the use of plasmids, λ phage vectors and cosmids. The use of YAC (yeast artificial chromosome) vectors has largely overcome this problem as they can propagate large DNA fragments. YACs also allow the manipulation of cloned sequences by the use of homologous recombination in yeast (Spencer *et al.*, 1993). However, the system has limitations. YACs have a high degree of chimerism and clonal instability (Green *et al.*, 1991), and the isolation and preparation of YAC clone DNAs is slow with low yields. An alternative to cloning large sequences in YACs would be the use of *E.coli* based cloning vectors. Both, BAC (bacterial artificial chromosome) and PAC (P1 derived artificial chromosome) vectors are being increasingly used as cloning vehicles for genomic libraries (Shizuya *et al.*, 1992 and Ioannou and de Jong, 1996). These vectors can be stably maintained and propagated with DNA insert sizes of up to 300 kb in length as one or two copy plasmids in bacteria. Their advantages over YAC vectors lies in the ease of library construction and the higher yield of supercoiled plasmid DNA that is more resistant to breakage. Sequences cloned in BAC and PAC vectors also show a high level of stability and low chimerism. No rearrangements have been observed in sequences cloned in these vectors following several generations of growth (Ioannou *et al.*, 1994).

Two methods for the genetic modification and manipulation of BAC and PAC clones are available. Yang *et al*. reported the modification of a BAC clone by homologous recombination in *E.coli* (Yang *et al*, 1997). Subsequently, Zhang *et al.* (Zhang *et al.*, 1998) reported a method that involves the use of RecE and RecT, encoded by a lambdoid phage, to modify cloned sequences in *E.coli* by homologous recombination.

In this paper we describe a general, highly efficient two-step method that has enabled the precise modifications of a PAC clone of the human β-globin locus by *E.coli* RecA-dependent homologous recombination and the use of a simple and powerful selection procedure in the *E.coli* DH10B host. The method employs positive and counter selection markers that independently ensure the monitoring of each step. Following integration of the recombination plasmid by a single reciprocal homologous exchange, the *E.coli* DH10B host not only acquires antibiotic resistance provided by the plasmid-derived positive selection marker; but also loses the host genome-borne resistance to another antibiotic. The stringent counter selection step ensures loss of the excised plasmid following a second homologous recombination event thereby generating the desired modification of PAC clones in a proportion of the excisants.

## Results

To perform modifications of PAC clone insert DNAs we took advantage of RecA-mediated homologous recombination in *E.coli* using the recombination vector pDF25 (Fig.1a). This plasmid has a Cm-resistance marker ($Cm^R$) for positive selection of the vector, an *rpsL*+ allele for counter selection, (gain of the plasmid-borne wild type *rpsL* + allele in a strain with the a chromosomal mutant *rpsL* allele, which specifies $str^R$ in the haploid state, converts the strain to $str^S$ from $str^R$), a polylinker

cloning site, a temperature-sensitive replication initiation protein ($RepA^{ts}$), an origin of replication and the *recA* gene of *E.coli*. Target sequences were cloned either in the *Nae*I site or one of the other unique cloning sites of the vector and used for transformation of the bacterial host carrying PAC recombinant clones. Transformants were grown overnight at 30°C on L-agar plates supplemented with Km and Cm. Thirty to 50 single colonies were picked and transferred to a fresh plate with the same antibiotics and grown overnight at the non-permissive temperature of 43°C. Under these conditions the recombination plasmid cannot replicate and resistance to Cm is conferred by the plasmid integrating either into the host genome or the region of homology contained within the PAC plasmid. Recombination can take place in the region of homology giving a type I or a type II integration upstream of the $^{G}\gamma$-globin gene (Fig.1b). The predicted sizes of bands from an *Nco*I digestion of co-integrant DNAs probed with the 2 kb *Dra*I fragment cloned in the plasmid pD3γlox is shown (Fig.1b). Blotting analysis of a targeted integration of the plasmid pD3γlox in PAC 148 gave the expected result. Both type I and type II integrations were observed (Fig. 2a, lanes 4 and 5). The frequency of integration in experiments varied from 30% to 90% (Table I). Single colonies of a mixed genotype were also observed. Fig.2a shows a clone comprising both type I and type II recombination (lane 7). The 1.3 kb band (Fig.2a, lanes 2-7) is due to cross-hybridisation with the probe and is observed depending on hybridisation conditions and exposure times (see Fig. 2b). Clones giving the correct bands were grown overnight at 43°C on L-agar with Km alone allowing excision and loss of the integrated vector. In this recombination either the original copy of the targeted sequence or the modified copy is left behind. Excised vector clones were selected overnight at 43°C on a fresh plate with Km and high Str. Only clones that have lost the excised recombination vector (ie the *rpsL*+ allele) grow in high concentration of streptomycin (Cornet *et al.*, 1996 and Feinberg and Vogelstein, 1983). Blotting of these clones showed the desired modification resulting in the generation of 1.5 kb and 1 kb *Nco*I bands due to the introduction of the *Nco*I site in the *loxP* oligo instead of the original, unmodified 2.5 kb *Nco*I band (Fig.2b, lanes 4-7 and 9). The frequency of modification was 40% following excision of the plasmid. Loss of the excised plasmid was at a frequency of 100%. Occasionally, both the modified and the unmodified copies of the PAC were seen in the same clone (Fig.2b, lane 9 and Table I). These were separated by replating. Mapping the modified PAC 148 clones (PAC 148γlox), showed that the region corresponding to the β-globin locus was unaltered and showed only the required change (Fig. 2c).

This method was used for modifying several positions in PAC 148γlox and also employed to modify a 120 kb PAC clone of the MCP (membrane cofactor protein CD46) locus (Yannoutsos *et al.*, 1996). We observed frequencies of recombination between 10-97 % (Table I). Lengths of homology at the 5' and 3' ends of the desired changes varied respectively from 157 bp and 136 bp (pDA2) to 1.6 kb and 1.3 kb (pDmcp3f). Although the shorter homology regions (pDA2), resulted in lower recombination frequencies (10 %) they were high enough to efficiently modify the PAC. Longer regions of homology gave higher frequencies, ranging from 30 % and 96 % (pD3γlox ) to 83 % (pDA5) and 97 % (pDmcp3f) although there are some unexplained variations in the observed frequencies eg. pD3γlox of 30 % (PAC 148) and 96 % (PAC 100).

We were also interested in examining the stability of highly repetitive sequences when propagated either in the presence of RecA or the recombinases RecE and RecT. Both PAC 100, which contains a large region of highly repetitive sequence 3' to the human β-globin locus, and PAC 148 which lacks this region were propagated in *E.coli* DH10B which is normally without these recombinases. This did not result in any rearrangements following several generations of growth (Fig.3 lanes 1 and 7). But on introducing either the plasmid pDF25 (Fig.3, lanes 2,5,8 and 11), which harbours the *recA* gene, or the plasmid pBAD-ETγ (Fig.3, lanes 3,6,9 and 12), which has cloned copies of the RecE and RecT recombinases, we noticed a rearrangement in PAC 100. Following digestion with *Xho*I or *Cla*I, a 13 or 10 kb band (Fig.3, lanes 1 and 4), containing highly repetitive DNA sequences (Bloomfield *et al.*, 1991), rearranges to a 7.5 or a 5.2 kb band (Fig.3, lanes 2 and 3, 5 and 6). PAC 148 did not show any changes (Fig.3, lanes 7-12).

## Discussion

The means to modify large cloned sequences has long been possible in YACs, thus enabling the introduction of point mutations, deletions, inversions etc. and a method describing the modification of cosmid clones by homologous recombination in *E.coli* has been reported (O'Conner *et al.*, 1989). This facility is now being extended to the *E.coli* BAC and PAC vectors, which have a bigger cloning capacity than cosmids and are easier and faster to prepare than YACs. We have used the method described here to introduce both a copy of the *loxP* site upstream of the $^{G}\gamma$-globin gene and to modify the human β-globin locus further at several other positions. The use of RecA-mediated homologous recombination to modify both the *E.coli* genome and episomes has been reported earlier (Yang *et al.*, 1997; Taramelli *et al.*, 1986; Poutska *et al.*, 1984 and Hamilton *et al.*, 1989). Yang *et al* (Yang *et al.*, 1997) earlier reported the modification of a BAC clone using a plasmid temperature-sensitive for replication, a cloned copy of the *recA* gene and the $Tet^{R}$ gene for both positive and counter selection. However, counter selections using the $Tet^{R}$ gene and fusaric acid have been reported to be inefficient (Crouzet *et al.*, 1997). Yang and co-workers observed an integration frequency of 10 % and a modification frequency of 4 % using homology regions of 1 kb and 1.6 kb. In our experiments using similar lengths of homology (pD3γlox, pDA5 and pDA8) we consistently observed higher frequencies of integration (83-97 %) and modification (40-80 %). Only when very short 5' and 3' lengths of homology are used do we observe a decrease in the integration frequency (10 %) and a concomitant drop in the frequency of modification (5 %). The use of the wild type *rpsL*+ allele in our recombination vector pDF25 allows for a very efficient counter selection procedure in *E.coli* DH10B; a host normally used to propagate PAC and BAC clones, as it carries a mutated genomic copy of the *rpsL* gene that confers resistance to streptomycin. The presence of two alleles of the gene conferring sensitivity to streptomycin selection in *E.coli* is well established (Lederberg, 1951) and has been used earlier to modify the *E.coli* genome (Feinberg and Vogelstein, 1983). This counter selection also obviates the need to use specialised reagents. An alternative is the use of the *sacB* gene of *B.*subtilis for counter selection (Zhang *et al.*, 1998 and Podolsky *et al.*, 1996). However, this inevitably results in deletions in the *sacB* gene following growth in sucrose resulting in a high background (Zhang *et al.*, 1998).

The presence of mixed genotypes was observed particularly when very high frequencies of integrations are obtained (Table I and Fig. 2a, lane 7). Given that there is more than one copy of the PAC molecule per cell and that multiple copies of the recombination vector pDF25 are present in the same cell, both type I and type II recombinations could proceed in that cell giving this result. We have also observed clones with either type I or type II co-integrants together with the unrecombined PAC (data not shown).

In the literature there is no reference to the fate of highly repetitive sequences propagated in YAC, BAC and PAC clones in the presence of recombinases. Mapping of the human β-globin PAC 100 clone showed a reproducible rearrangement when either RecA or the recombinases RecT and RecE were expressed in the host. Previously, a region ~50 kb downstream from the human β-globin gene, present in PAC 100, has been found to contain highly repetitive sequences (Bloomfield *et al.*, 1991). PAC 148 lacks this region and remains unaltered under the same conditions. The existence of repeat sequences *per se* , such as Alu and L1, mapped within the 70 kb region of the β-globin locus do not result in any instability. The possibility of highly repetitive sequences in cloned regions of the eukaryote genome undergoing changes, such as those seen here in PAC 100, would seem to limit the use of this method, but the use of alternative clones and careful mapping following the desired recombination event can overcome this limitation.

The use of bacterial homologous recombination procedures, such as the method described in this paper, allow for efficient introduction of insertions, point mutations and deletions in large DNA fragments cloned in the *E.coli* PAC and BAC vectors. This technology will facilitate the analysis and organisation of complex genomes and the re-introduction of defined mutations into the eukaryotic genome.

## Material and Methods

### *Bacterial host and plasmids*

DH10B was grown in L-broth supplemented with the appropriate antibiotics. This strain is streptomycin resistant as a consequence of a mutation in the *rpsL* gene. The recombination vector pDF25 was constructed by cloning a 3.3 kb *Bam*HI fragment, containing the *E.coli recA* gene, into the *Bgl*II site of the vector pLN135 (Cornet *et al.*, 1996) derived from the plasmid pSC101, and carrying a mutation in the replication protein RepA that renders the plasmid temperature-sensitive for DNA replication. The human β-globin PAC clones, PAC 100 (insert size 175 kb) and PAC 148 (insert size 185 kb) were a gift from Dr P. Ioannou, (Institute of Neurology, Nicosia, Cyprus).

### *Media*

L-broth and L-agar plates were prepared as described in Sambrook *et al* (Sambrook *et al.*, 1989). Chloramphenicol (Cm) was used at 34μg/ml, kanamycin (Km) at 25 μg/ml, ampicillin (Ap) at 20 μg/ml and streptomycin (Str) at either 20 μg/ml or at 200μg/ml.

### *Sensitivity to high streptomycin concentration and preparation of E.coli DH10B electrocompetent cells*

Since our method requires a selection step which relies on the DH10B host acquiring sensitivity to streptomycin after transformation with the recombination vector pDF25, we first tested the sensitivity of the untransformed host to low (20μg/ml) and high (200μg/ml) concentrations of streptomycin. The host was resistant to both concentrations of the antibiotic as a consequence of its mutation in the *rpsL* gene. However, on transformation by the recombination vector, which carries the *rpsL+* allele, the host became sensitive to the higher concentration of streptomycin (Cornet *et al.*, 1996 and Feinberg and Vogelstein, 1983). Electrocompetent *E.coli* DH10B bacteria were prepared as described (Cornet *et al.*, 1994) but supplementing the water with 10% glycerol throughout the procedure.

### *Construction of the plasmid pD3γlox*

A *loxP* sequence was cloned at position –1659 of a human $^{G}$γ-globin gene plasmid in the *Bgl*II site using the oligonucleotide: 5'-TCGGGGCCATGGATAACCTTCGTATAGCATACATTATACGAAG TTATGGATC-3'. This sequence includes a diagnostic *Nco*I site. A 2 kb *Dra*I fragment (-2431 to -393), which harbours the *loxP* sequence, was isolated from the plasmid and cloned into the unique *Nae*I site of the recombination vector pDF25. The resulting construct, pD3γlox was used to target *lox P* into the corresponding sequence of the $^{A}$γ-globin gene in PAC 100 and PAC 148 clones.

### *Preparation of PAC clone DNAs, plasmid DNAs and Southern blot hybridisations*

PAC DNA was prepared using the modified alkaline lysis method (Dower *et al.*, 1988). All other plasmid DNAs were prepared by the standard alkaline lysis procedure (Sambrook *et al.*, 1989). DNA fragments used for hybridisations were labelled using the DNA oligo labelling method of Feinberg and Vogelstein (Feinberg and Vogelstein, 1983). Southern blots were as described earlier (Sambrook *et al.*, 1989).

### *Induction of recombinase RecE from pBAD-ETγ and mapping gel*

The vector pBAD-ETγ (Zhang *et al.*, 1998) with cloned copies of *recT* and *recE* genes, was transformed into *E.coli* DH10B with either PAC 100 or PAC 148 and the cells grown under Ap and Km selection. A single colony was grown overnight at 37$^{o}$C in 5ml L-broth cultures supplemented with the same antibiotics. An aliquot was grown to $OD_{600}$ 0.7 and further supplemented with 0.1% L-arabinose for 1 hr before harvesting, resulting in the induction of RecE gene from the *ara*-promoter. The expression of RecE recombinase together with RecT facilitates RecA-independent homologous recombination in *E.coli* (Zhang *et al.*, 1998). Cultures were then plated on L-agar plates with Km

alone and grown at 37$^{o}$C. Absence of Ap selection leads to a loss of the plasmid pBAD-ETγ. Km resistant but Ap sensitive clones were used for PAC DNA analysis on a 0.4% agarose gel run in TBE buffer (Sambrook *et al.*, 1989).

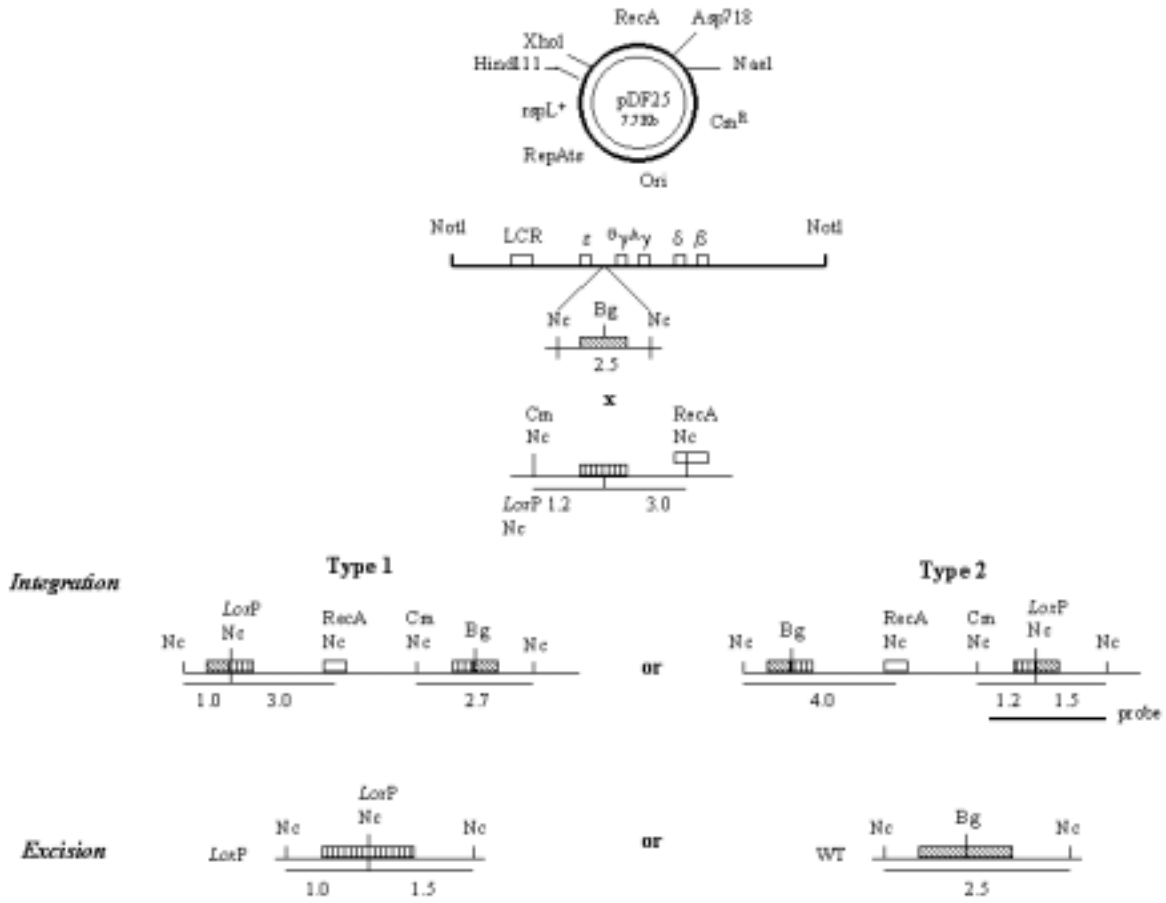



***Figure 1:*** *A map of the recombination vector pDF25. A 3.3 kb BamHI fragment of the E.coli recA gene was cloned into the polylinker BglII site of the plasmid pLN135*—The plasmid carries an origin of replication (Ori), a temperature-sensitive mutation in the replication initiation protein RepA ($RepA^{ts}$), the gene for resistance to chloramphenicol ($Cm^R$) and the *rpsL*+ gene (the wild-type allele of the *rpsL* gene). Some of the unique cloning sites are indicated. A schematic representation of the sizes of *Nco*I fragments hybridising to the 2 kb *Dra*I insert following integration of the plasmid pD3γ-lox into PAC 148 insert and its subsequent excision. The original 2.5 kb *Nco*I fragment is now interrupted by the new *Nco*I site giving fragments of 1.5 kb and 1.0 kb. The abbreviations used in the figure are: Bg (*Bgl*II), Nc (*Nco*I), RecA Nc (*Nco*I site in the *recA* gene), Cm Nc (*Nco*I site in the chloramphenicol gene), *loxP* Nc (*Nco*I site in the *loxP* oligo), *Wt* (unmodified PAC) and *LoxP* (modified PAC with a targeted insertion of the *loxP* oligo).

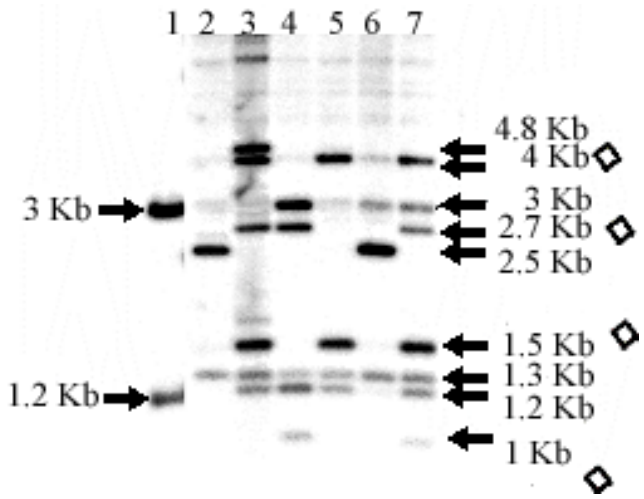



***Figure 2a:*** *Southern blot hybridisation of NcoI digested miniprep DNAs of PAC clones following integration of pD3γlox into PAC 148 insert and probing with the fragment indicated in Fig.1.*— Unintegrated pD3γlox DNA digested with *Nco*I (lane 1), PAC 148 DNA (lane 2) , DNAs of clones selected for the integration of pD3γlox into PAC 148 insert (lanes 3-7). Bands marked with a diamond are diagnostic for the integration. Both the 3 kb and the 1.2 kb bands are also seen in the unintegrated plasmid (lane 1), consequently they are not marked as diagnostic. The 1.3 kb band is due to cross-hybridisation with the probe.

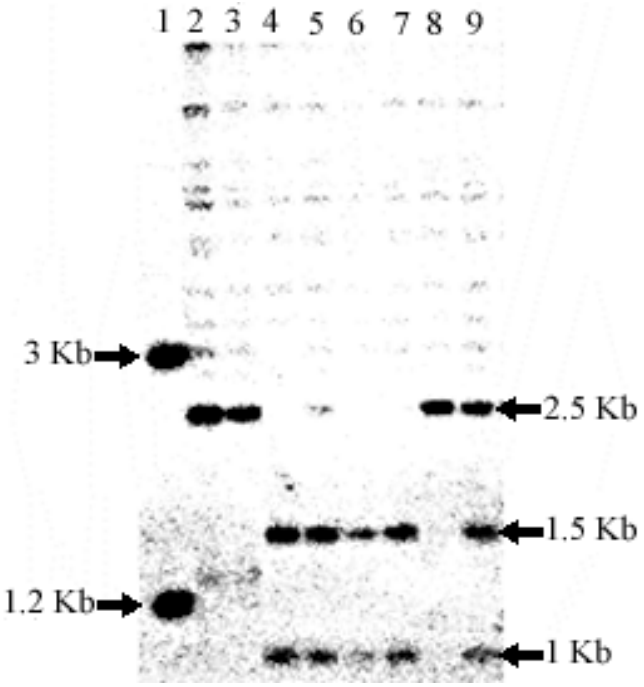



***Figure 2b:*** *Introduction of the loxP sequence and a diagnostic NcoI site into PAC 148 insert after excision of the plasmid pD3γlox*—All DNAs were digested with *Nco*I and probed as in Fig.2a. Unintegrated pD3γlox digested with *Nco*I (lane 1), PAC 148 (lane 2), DNAs from clones selected following excision of the recombination vector construct pD3γlox (lanes 3-9). The 1.5 kb and 1.0 kb bands are diagnostic for a successful homologous recombination. The 1.3 kb cross-hybridising band (Fig. 2a) is observed only on longer exposures.

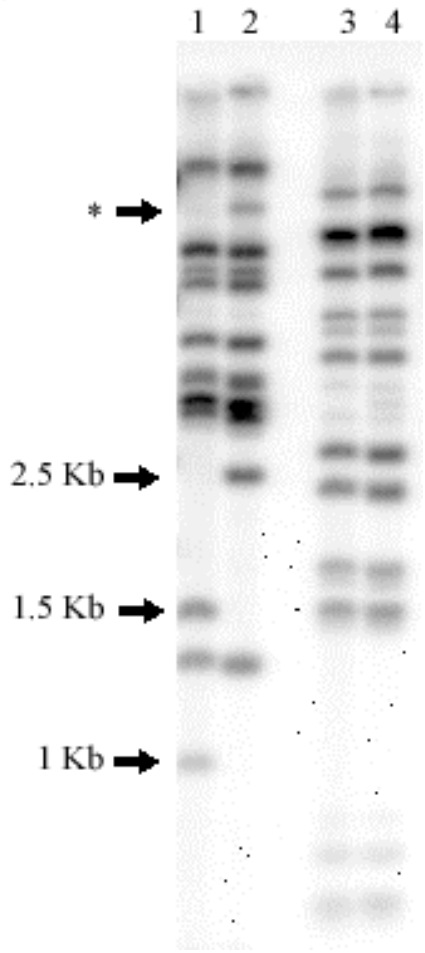



***Figure 2c:*** *Mapping of PAC 148 (lanes 2 and 4) and PAC 148γlox (lanes 1 and 3)*—DNAs were digested with either *Nco*I (lanes 1 and 2) or with *Eco*RI (lanes 3 and 4) and the blot probed with the cosmid clone Cosγγδβ (22). The band marked with an asterisk is a partial digest. The 1.5 kb and 1.0 kb bands are the result of homologous recombination.

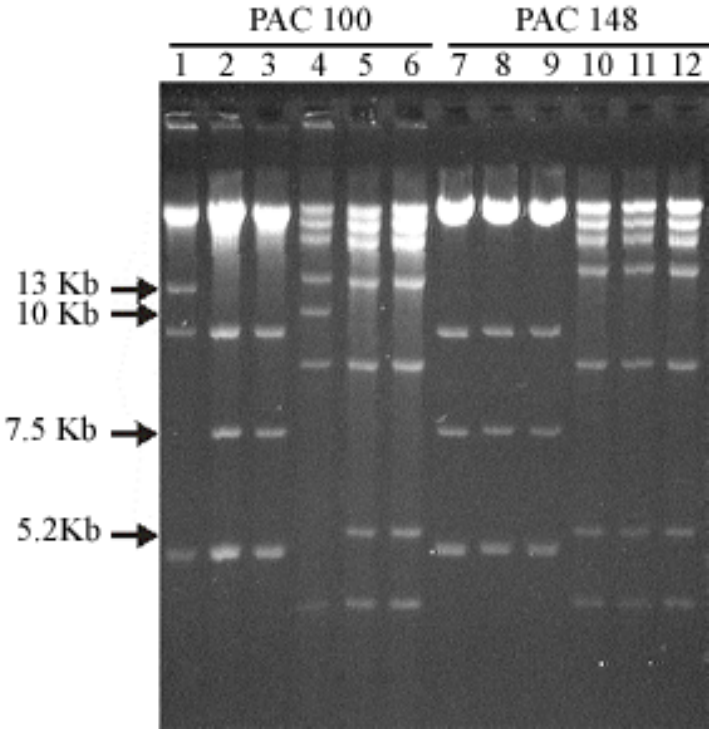



***Figure 3:*** *PAC100 (lane 1) and PAC148 (lane 7) grown in the absence of RecA and the recombinases RecE and RecT*—The effect of either RecA (lanes 2,5,8 and 11) or the recombinases RecE and RecT (lanes 3,6,9 and 12) on PAC 100 and PAC 148 inserts. DNAs were digested with either *Xho*I (lanes 1-3 and 7-9) or *Cla*I (lanes 4-6 and 10-12), run on a 0.4% agarose gel in TBE buffer and visualised by staining with 0.5ug/ml ethidium bromide.

***Table 1:*** *Plasmid pD3γlox and plasmids pDA1-7 were used for modifying different sequences along the human β-globin locus*—Plasmid pDmcp3f was used to modify the PAC clone of the MCP locus (PAC-MCP). A total of 50-100 colonies were analysed for each experiment. The percentage of co-integrants shown in the table represents the sum total of all the clones which are either of a pure or of a mixed genotype, comprising clones which have either the type I and type II co-integrants or have either of these together with the unrecombined PAC molecule. ( NA = Data not available.)

| Recombination plasmid | Length of homology | | Recombination frequency | | Targeted PAC clone |
|---|---|---|---|---|---|
| | 5' | 3' | % Co-integrants | % Mixed genotype | |
| pD3γlox | 700 bp | 1.2 kb | 96 | 47 | PAC 100 |
| | | | 30 | 10 | PAC 148 |
| pDA1 | 620 bp | 600 bp | 60 | 40 | PAC 148 |
| pDA2 | 157 bp | 136 bp | 10 | N/A | PAC 148 |
| pDA3 | 2 kb | 1.5 kb | 80 | 20 | PAC 148 |
| pDA4 | 630 bp | 950 bp | 60 | N/A | PAC 148 |
| pDA5 | 1.2 kb | 1.6 kb | 85 | 30 | PAC 148 |
| pDA6 | 1 kb | 475 bp | 20 | N/A | PAC 148 |
| pDA7 | 600 bp | 1 kb | 85 | N/A | PAC 148 |
| pDmcp3f | 1.6 kb | 1.3 kb | 97 | N/A | PAC-MCP |

# *CHAPTER 4*

# *A SILENCER LOCATED 3' TO THE Aγ-GLOBIN GENE INFLUENCES TRANSCRIPTION OF THE HUMAN β-GLOBIN LOCUS AT THE EMBRYONIC YOLK SAC STAGE IN TRANSGENIC MICE*

Eleni Katsantoni, Mariken de Krom, John Kong-a-San, Frank Grosveld, Nicholas Anagnou and John Strouboulis

This chapter has been submitted for publication

## Summary

Naturally occurring deletions associated with persistent γ-globin expression in the adult stage, have indicated the presence of important *cis*-regulatory elements in the region between the $^{A}$γ- and δ-globin genes. Our previous work had identified two elements in this region, termed Enh and F, that appear to act as silencers in transient transfection assays. In the present study we deleted the Enh and F elements in the context of a 185 kb human β-globin locus PAC and tested the effects of this deletion on the *in vivo* regulation of the locus in transgenic mice. We find that the Enh/F deletion results in an increase in ε- and γ-globin mRNA levels in the embryonic stage. However, the number of cells transcribing ε- and/or γ-globin did not increase, suggesting that removal of these elements results in an increase in the transcriptional rate of the genes. Human globin switching in transgenic mice was not affected by this deletion, thus excluding a role for Enh and F in γ gene silencing in the adult stage. Taken together, these results identify the Enh and F elements as being capable of down-regulating transcription of the human β-globin locus in an embryonic stage-specific manner.

## Introduction

The human β-globin locus spans a region of approximately 75 kb in chromosome 11 and contains five genes arranged in the order in which they are expressed during development, i.e. 5'-ε$^{G}$γ$^{A}$γδβ-3'. Activation and high level expression of the locus depends on the Locus Control Region, or LCR (Grosveld *et al*., 1987), located upstream of the genes in the locus. Two developmental switches occur in the expression of the genes in the human globin locus. Expression of the ε-and γ-globin, first activated during primitive erythropoiesis in the embryo (nucleated red cells), switches to expression of primarily γ-globin genes (and a low level of β) at the start of definitive erythropoiesis in the foetal liver (enucleated red cells). This pattern switches perinatally to expression of the adult stage δ- and β-globin genes (with δ making a minor contribution), while expression of the γ-globin genes is suppressed. It is not presently known how expression of the different genes is suppressed at the various developmental stages, but different mechanisms appear to be in operation. For example, the embryonic ε- and γ-globin genes are autonomously suppressed by (unknown) protein factors acting through gene-proximal sequences (Behringer *et al*., 1990; Dillon and Grosveld, 1991; Enver *et al*., 1990; Liebhaber *et al*., 1996; Raich *et al*., 1992 and Shih *et al*., 1990). By contrast, suppression of the β-globin gene in the early developmental stages appears to require competition by the early genes for interaction with the LCR (Behringer *et al*., 1990; Dillon *et al*., 1997; Enver *et al*., 1990 and Hanscombe *et al*., 1991) and may involve changes in chromatin structure (Gribnau *et al*., 2000).

The study of the foetal γ-globin gene regulation is of particular interest, since γ gene expression represents a distinct foetal stage found in primates. Understanding the basis for γ gene silencing in the adult stage is also of considerable clinical importance as it may provide an alternative route to the treatment of haemoglobinopathies, since even low level γ-globin expression in the adult stage can ameliorate the symptoms of thalassemias arising from a deficiency in β-globin expression. A number of naturally occurring deletions in the locus lead to persistent γ-globin gene expression in the adult stage (Wood, 1993) indicating that γ gene suppression is likely to be a complex process, but one that can be easily perturbed. One class of deletions in the locus results in substantial (15-25%) pancellular γ-globin gene expression in a condition termed hereditary persistence of foetal haemoglobin, or HPFH (Stamatoyannopoulos and Grosveld, 2001). Another set of deletions, known as $(\delta\beta)^{0}$-thalassemias, give rise to lower levels (5-15%) of heterocellular γ-globin expression in the adult. In a particular set of deletions, where the 3' end of the locus is lost, the 5' breakpoints map within a region between the $^{A}$γ- and δ-globin genes. They have therefore long been thought to harbour negative *cis*-acting elements involved in the regulation of the foetal-to-adult switch (Huisman *et al*., 1974).

We have previously identified four elements, termed Enh, F, O and P, located within the $^{A}$γ- to δ-globin intergenic region (Fig. 1A). These elements exhibited silencer activity in transient transfection assays in erythroid and non-erythroid cells and were therefore candidate elements for suppressing γ-

globin expression in the adult stage (Kosteas *et al*., 1993 and 1994, Katsantoni and Anagnou, unpublished data). Transgenic mouse studies using some of these elements have been contradictory as to their *in vivo* function. For example, constructs that included the LCR linked to a γ-globin gene fragment that contained Enh and F were autonomously silenced in the adult stage of transgenic mice (Dillon and Grosveld, 1991). At the same time, deletion in a human β-globin locus YAC of a 12.5 kb region between the $^{A}$γ- and δ-globin genes, which includes the Enh, F, O and P elements, has been reported to have no effect on human globin gene switching in transgenic mice (Zhang *et al*., 1997). In addition, deletion of Enh from a human β-globin locus YAC, showed no observable effects on the regulation of the β-globin locus in transgenic mice, thus suggesting that this element provides either no unique function(s), or the lack of observable effects represents a functional redundancy by other intact elements in the locus (Liu *et al*., 1998). On the other hand, removal of the O element as part of a 2.5 kb deletion in a human β-globin locus YAC resulted in position effects and reduced expression of the human β-globin gene in the adult stage (Calzolari *et al*., 1999). However, developmental silencing of the γ genes in the adult stage remained unaffected in these mice (Calzolari *et al*., 1999). Taken together, the findings obtained with the human β locus constructs do not directly support the presence of elements involved in the developmental regulation of globin gene switching in the $^{A}$γ- to δ-globin region, but leave open the possibility that these elements do have transcriptional regulatory activity in specific developmental stages.

In this paper we sought to clarify the *in vivo* function of the Enh and F elements and assess their putative involvement in the developmental regulation of the foetal γ-globin genes. To these ends, we have deleted both Enh and F elements in the context of a 185 kb human β-globin locus PAC by employing homologous recombination in *E. coli* and introduced the modified locus in transgenic mice. In the absence of Enh and F we observe an increase in the expression of both the human ε- and γ-globin genes in the embryonic yolk sac. No effects on human globin gene expression in the foetal liver and adult blood stages or in the silencing of the γ-globin genes were observed.

## Results

### *Deletion of the Enh and F elements from a β-globin locus PAC by recombination in E. coli.*

In order to establish whether the Enh and F play a role in the regulation of γ-globin gene expression *in vivo*, we deleted them in the context of a PAC containing the entire human β-globin genecluster as a 185 kb insert and introduced the modified PAC in transgenic mice. This approach offers a number of advantages (Imam *et al*., 2000). Human β-globin locus BACs have been previously introduced successfully in transgenic mice (Huang *et al*., 2000 and Kaufman *et al*., 1999), however, the PAC employed in our studies has a larger insert size (185 kb). In addition, we have modified the 185 kb globin locus PAC by introducing a lox P site at position –1659 of the $^{G}$γ gene, via homologous recombination in *E. coli* (Imam *et al*., 2000) to facilitate the generation of single copy globin PAC transgenic mice from multi-copy animals by cross-breeding with transgenic lines ubiquitously expressing Cre recombinase.

A 1.6 kb fragment containing Enh and F located 3' to the $^{A}$γ gene, was deleted from the human β-globin locus PAC by homologous recombination in *E.coli* (Fig. 1A). The targeting construct was generated by PCR (see Materials and Methods) and consisted of 620 bp of sequence immediately flanking the 5' of the 1.6 kb deletion and 600 bp of 3' flanking sequence. Correct excision events were analysed by Southern blot analysis of Nco I digests using as probe the 3' homology region flanking the silencer deletion (Fig. 1B). This probe detects a 3.1 kb Nco I fragment in the wild type PAC, or a 4.6 kb fragment in the modified PAC with the Enh and F deletions (Fig. 1B).

Five colonies with the correct excision pattern were selected and further mapped in greater detail in order to make sure that the modified globin PAC did not bear additional, unrelated re-arrangements following the homologous recombination event. Mapping was carried out by Southern blot analysis using a number of different restriction digests. Blots were hybridised with the LCRε and γγδβ cosmids (Fig. 2), which cover the entire 75 kb human β-globin genomic sequence (Strouboulis *et al*., 1992). As

shown by the EcoRI pattern in Figure 2, no gross re-arrangements in any other region of the 75 kb human β-globin locus were detected in the modified PAC that was used for microinjection. These results were further confirmed using additional digests of the modified PAC DNA (not shown).

### *Microinjection of the 185 kb human β-globin locus PAC insert bearing the Enh/F deletion*

The β-globin locus insert was released as a 185 kb fragment from both wild type and Enh/F-deleted (ΔEnh/F) PACs and purified for microinjection into mouse fertilised eggs, as described in Materials and Methods. Three founders transgenic for the ΔEnh/F locus and ten founders transgenic for the wild type globin locus were obtained. A single copy wild type locus mouse that transmitted the intact locus was selected as a control for the analyses presented here. The globin locus in this mouse had a peri-centromeric (i.e. close to the centromere) site of integration (data not shown). The detailed characterisation of all wild type human β-globin PAC locus mice will be presented elsewhere (Chapter 5).

The first ΔEnh/F globin locus founder mouse with a centromeric site of integration, never passed on the transgene. S1 nuclease protection analysis of RNA from the adult blood of this founder showed no γ-globin expression, while human β-globin RNA was clearly detectable (Fig. 3, last lane) although it is difficult to obtain an expression level per copy of the transgene due to the mosaicism of transgenic founders. It is clear however that γ-globin gene expression has switched off in the adult stage of this founder. The high levels of β-globin expression and the Southern blot analysis suggest that at least one copy of the human globin locus is intact making it unlikely that the lack of γ-globin expression in adult blood is due to gross rearrangements of the transgenic locus in this mouse.

The other two ΔEnh/F locus founder mice successfully transmitted the transgene to establish transgenic lines ΔEnh/F-A and ΔEnh/F-B. Further analysis by DNA-FISH showed both lines to be single copy for the transgene with pericentromeric (line A) and non-centromeric/non-telomeric (line B) sites of chromosomal integration (data not shown). The integrity of the 75 kb globin gene locus was checked by suppression hybridisation using the LCRε and γγδβ cosmids (Strouboulis *et al*., 1992), and in greater detail using additional probes, as described in Materials and Methods (data not shown). No re-arrangements were observed in the 75 kb human β gene locus in these mice.

### *Deletion of Enh and F leads to elevated ε- and γ-globin gene expression in the embryonic stage*

In order to assess the effects of the deletion of the Enh and F elements on γ gene expression, we first analysed globin gene expression by S1 nuclease protection in RNA isolated from adult blood of the two modified globin locus PAC lines and the wild type β locus line. Expression of the γ gene has normally switched off by this stage in mice (Dillon and Grosveld, 1991; Gaensler *et al*., 1993; Liu *et al*., 1998; Peterson *et al*., 1993 and Strouboulis *et al*., 1992) with β-globin being the human gene that is predominantly expressed in blood. As expected, γ-globin expression was not detected in adult blood of the wild type β locus mouse, or in the blood of the two ΔEnh/F lines (Fig. 3, adult blood lanes). These findings suggest that deleting the Enh and F from the human locus has little effect on the silencing of the γ-globin genes in the adult stage of these mice.

In order to analyse the effects of the 3' $^{A}$γ-element deletion on globin gene regulation during mouse embryonic development, animals from the established lines A and B and from the wild type β locus line were bred to non-transgenic females and erythropoietic tissues were isolated from embryos dissected at the following developmental time-points: 10.5 days (embryonic yolk sac), 12.5 and 16.5 days (foetal liver). Expression profiles for all human globin genes in all developmental time-points were analysed by S1 nuclease protection against those of the endogenous mouse β-like globin genes (Fig. 3). Quantitation of expression levels of the human globin gene expression as a percentage of mouse globin expression levels is shown in Table 1.

The expression profiles of the human globin genes in the wild type 185 kb human locus line appear identical to those previously described for a smaller 70 kb cosmid-derived human β-globin locus construct (Strouboulis *et al*., 1992). When comparing the wild type locus control line with the ΔEnh/F lines, expression patterns do not appear to be significantly affected by the deletion of Enh and F, in

that the γ genes are predominantly expressed in the embryonic yolk sac stage, with expression continuing in the early foetal liver stage and declining rapidly in later foetal liver stages, to be completely extinguished by the adult blood stage (Fig. 3).

However, quantitation of the levels of globin gene expression revealed a difference between the control β locus and the ΔEnh/F lines in human globin expression levels in the embryonic yolk sac (Table 1A, Fig. 4). In the control β locus line, γ-globin expresses at approximately 23% of the endogenous mouse embryonic globins, whereas in line ΔEnh/F-A γ-globin is expressed at around 50% (52% in embryo 1 and 46% in embryo 2, Table 1A) and in ΔEnh/F-B at around 37% (34% and 40% for embryos 1 and 2, respectively) of the mouse embryonic genes. The levels of human ε-globin gene expression also appear to increase from 6% of mouse globin in the β locus PAC line to levels approaching 10% of mouse globin in lines ΔEnh/F -A and -B (Table 1A, Fig. 4). Deletion of Enh and F therefore appears to increase expression of both ε- and γ- globin genes by at least 50%, with the effect being more pronounced in γ-globin expression (Fig. 4). This could be due to the proximity of the γ-globin genes to the Enh and F deletion and/or due to the greater stability previously suggested for the γ-gene-LCR interactions, as compared to ε-globin/LCR interactions (Wijgerde *et al*., 1995). Total human globin output remains relatively constant at all developmental stages when compared to total mouse globin output (Strouboulis *et al*., 1992 and the wild type PAC). The fact that the total human globin output (as a percentage of total mouse globin output, Table 1A and Fig. 4) is higher in the two ΔEnh/F lines compared to the control β locus line, strongly suggests that the effects of the deletion are not due to differences in developmental timing between the embryos at the time of dissection (Table 1A, last column).

We also analysed expression at 10.5 dpc embryonic yolk sac stage of another single copy wild type β-locus line to ensure that the differences we observed between the first wild type locus and the two ΔEnh/F lines were not limited to the first control β locus line we used in the analysis. Essentially no differences in expression levels were found between the two wild type locus lines (data not shown), thus suggesting that the differences observed with the ΔEnh/F lines are specific. Thus it appears that Enh and F are indeed negative regulators of human globin gene transcription at the embryonic yolk sac stage.

In contrast to expression in the embryonic yolk sac, in the foetal liver stage γ-globin expression levels in the ΔEnh/F lines are not significantly different to those of the control β locus line (Table 1B). In addition, β-globin gene expression levels are not affected by the Enh and F deletion in the foetal liver or in the adult blood stages (Table 1B). The apparently higher levels of γ-globin expression in the ΔEnh/F-B 12.5 day time-point are due to differences in developmental timing as the embryos analysed in this experiment appeared to be at a developmental stage earlier than 12.5 days (smaller embryo size and foetal livers). The fact that both γ- and β-globin levels in the B line are "corrected" to control levels by the 16.5d time-point (Table 1B), strongly suggest that developmental (mis)timing accounts for the higher γ-globin levels in the 12.5 dpc time-point of this line. We therefore conclude that deletion of Enh and F has little effect on γ- and β-globin gene expression in the foetal liver and adult blood stages.

### *Primary transcript in situ hybridisation reveals no differences in human globin expression in embryonic blood*

Primary transcript *in situ* hybridisation using intronic probes allows the detection of actively transcribing loci in single cells (van de Corput and Grosveld, 2001and Wijgerde *et al*., 1995). Primary transcript analysis in 10.5 dpc embryonic blood previously showed a heterogeneity in ε- and γ-globin transcribing loci, with a significant number of cells (~ 40%) showing overlapping ε- and γ-globin signals in the same chromosomal locus (Wijgerde *et al*., 1995). We wanted to investigate whether the increased ε- and γ-globin mRNA levels in the ΔEnh/F lines at the embryonic yolk sac stage are accompanied by an increase in the number of human globin loci actively transcribing the genes. We therefore carried out primary transcript *in situ* hybridisation of embryonic blood cells from the control β locus line and the two ΔEnh/F lines,

using intron-specific probes against ε- and γ-globin. In agreement with previous results, in the control β locus line we found approximately 42% of cells having both ε- and γ-globin signals (Fig. 5B). In lines ΔEnh/F-A and -B we found 39% and 37.8% of cells having both ε- and γ-globin signals, respectively (Fig. 5B), suggesting no significant differences between the three lines. Similarly, the numbers of γ-only and ε-only expressing cells do not differ significantly between the β locus line and the two ΔEnh/F lines (Fig. 5B) and are in agreement with those previously observed (Wijgerde *et al*., 1995). It therefore appears that the increased ε- and γ-globin mRNA levels observed in the embryonic stage are not accompanied by an increase in the number of actively transcribing loci in this stage. We further tested this by comparing the number of γ-globin primary transcripts against those for mouse α-globin in 10.5 dpc embryonic blood (Fig. 5A). We found no significant differences in the relative numbers of γ-globin and mouse α-globin transcribing cells between the control β locus line the two ΔEnh/F lines (Fig. 5B), providing further support for the notion that the Enh and F deletion does not affect primary transcript expression patterns.

*The $^{G}\gamma$: $^{A}\gamma$ RNA ratio in embryonic yolk sac is not affected by deletion of the 3'$^{A}\gamma$ element*

We next wanted to investigate how the Enh and F element deletion might affect transcription of the $^{A}\gamma$- and $^{G}\gamma$-globin genes relative to each other in the embryonic stage. Due to its proximity to $^{A}\gamma$-globin, it might be expected that the deletion may have a predominant effect on $^{A}\gamma$ gene expression in the embryonic stage. Alternatively, both $^{A}\gamma$ and $^{G}\gamma$ genes might be affected by the deletion. Analysis of γ-globin gene expression by S1 nuclease protection or by primary transcript *in situ* hybridisation in single cells, does not distinguish between $^{G}\gamma$ and $^{A}\gamma$ transcription. We therefore developed a primer extension assay that allows distinction between $^{G}\gamma$ and $^{A}\gamma$ transcripts. This assay takes advantage of a short region of non-homology near the 3' end of the $^{G}\gamma$ and $^{A}\gamma$ transcripts (Fig. 6A). We designed two primers that are complementary to sequences immediately downstream of the short region of non-homology between the $^{G}\gamma$ and $^{A}\gamma$ transcripts (Fig. 6A). These primers map to the same region of the two transcripts and differ in sequence in a single nucleotide so as to provide perfect complementarity to the $^{G}\gamma$ and $^{A}\gamma$ transcript sequences. In the primer extension assay, an equal mixture of both radiolabelled primers is annealed to RNA from erythroid tissues and extended by reverse transcriptase using a nucleotide mix that contains dideoxy-thymidine triphosphate (ddTTP). Under these conditions, when the primer extension reaction reaches an adenosine residue in the transcript sequence, the ddTTP becomes incorporated and blocks further extension. For the $^{G}\gamma$ transcript, termination will occur within the short region of non-homology to generate a 35 nucleotide extended product, whereas for the $^{A}\gamma$ transcript termination will occur just beyond the region of non-homology to generate a 40 nucleotide extended product (Fig. 6A). The two extended products can be resolved and visualised by polyacrylamide gel electrophoresis (Fig.6B). The quantitation of 10.5 dpc yolk sac RNA by primer extension showed essentially no significant differences in the $^{G}\gamma$:$^{A}\gamma$ RNA ratios between the wild type β locus line and lines ΔEnh/F-A and -B (Fig. 6B). Therefore, the increase in γ-globin expression observed in the embryonic yolk sac after deletion of Enh and F cannot be accounted for by an increase preferentially in the transcription of one of the two γ-globin genes.

## Discussion

On the basis of naturally occurring deletions associated with persistent γ-globin expression in the adult stage, it has been proposed that the region between the $^{A}\gamma$- and δ-globin genes harbours negative regulatory elements that may be involved in suppressing γ-globin gene expression in the adult stage (Huisman *et al*., 1974). We have previously described elements Enh and F located 3' to the $^{A}\gamma$ gene as silencers in transient transfection assays (Kosteas *et al*., 1993 and 1994 and Katsantoni and Anagnou, unpublished data). In order to confirm and further clarify the *in vivo* role of Enh and F in globin gene regulation, we deleted them in the context of a 185 kb human β-globin locus PAC which was

introduced in transgenic mice. The deletion from the human β-globin locus of the Enh and F elements resulted in an increase in the levels of ε- and γ-globin gene expression in the embryonic stage. At the same time, the deletion did not affect globin expression levels or the γ- to β-globin switch in the foetal liver and adult erythropoietic stages in mice. This is in agreement with previous studies in mice, in which the entire 12.5 kb region between the $^{A}$γ- and δ-globin genes, inclusive of the Enh and F elements, was deleted from a β-globin locus YAC with no observable effects on human globin switching in transgenic mice (Zhang *et al*., 1997).

These results confirm our previous observations that Enh and F act as silencers in transiently transfected K562 and HeLa cells (Kosteas *et al*., 1993 and 1994 and Katsantoni and Anagnou, unpublished data) and refute earlier reports that described Enh as an enhancer in transient reporter assays (Bodine and Ley, 1987). Our *in vivo* observations are also in contrast to the work of Liu *et al*., (1998) in which the Enh element was deleted from a human β-globin locus YAC in transgenic mice. No observable effects were found on the regulation of the globin locus upon deletion of Enh, thus leading the authors to conclude that this element provided no unique function in β-globin gene regulation (Liu *et al*., 1998). Our 1.6 kb deletion, in addition to the Enh and F elements, included the sequences in between Enh and F. It is possible therefore that the effects we see on ε- and γ-globin expression in the embryonic stage could be due to the deletion of the F element and/or the intervening sequences between Enh and F. This is reminiscent of the deletion in the Indian $(\delta\beta)^{0}$ thalassemia, the 5' breakpoint of which removes the F element and the sequence between Enh and F, while leaving Enh intact (Mishima *et al*., 1989). This deletion is associated with 5-15% foetal hemoglobin (HbF) in the adult stage and could be attributed to the deletion of F, in combination with other undefined regulatory elements located downstream of F. The silencing activity of the F element is associated with a 366 bp fragment in transient assays (Kosteas *et al*., 1994 and Katsantoni and Anagnou, unpublished data) and contains several binding sites for the YY1, GATA1 and CP1 transcription factors (Soultanas *et al*., 1996).The clustering of YY1 and GATA1 binding sites within this element, combined with previous reports documenting that both proteins exhibit a dual activity of activation and/or repression on gene expression (for example, Raich *et al*., 1995), suggest a putative role for these factors in the embryonic specific silencing activity of F.

The finding that deleting Enh and F from the β locus leads to an increase in both ε- and γ-globin expression specifically in the embryonic stage is reminiscent of two previous studies. In the first study by Liu *et al*. (1997), deletion of a small putative silencer immediately upstream of the ε-globin gene in a human β locus YAC, led to a decrease in expression of both ε- and γ-globin genes in the embryonic yolk sac stage of transgenic mice. Human globin expression levels and switching were not significantly affected in the foetal liver and adult blood stages of these mice (Liu *et al*., 1997). By contrast, in the second study by Calzolari *et al*. (1999) a 2.5 kb deletion upstream of the δ-globin gene led to a decrease in both γ- and β-globin expression in the foetal liver and adult blood stages, but no significant effects on globin expression levels in the embryonic yolk sac stage or in globin switching were observed. Taken together, these observations identify a new class of gene-proximal regulatory elements within the human β-globin locus that are involved in regulating the levels of globin gene transcription, positively (Calzolari *et al*., 1999 and Liu *et al*., 1997) and negatively (this study), in a developmental-stage specific manner.

## Materials and Methods

### *DNA constructs*

The human β-globin EPAC/148β (Narayanan *et al*., 1999) was modified by homologous recombination in *E. coli* in two steps (Imam *et al*., 2000 and see also below). Firstly, a Not I linker was cloned in the 5' end of the globin locus insert. This was done by subcloning into the targeting vector pDF25 (Imam *et al*., 2000) an Nru I/Nco I fragment from EPAC/148β, containing the very 5' end of the globin insert. The Not I linker was cloned into a unique Hind III site within this fragment and introduced back into EPAC/148β by homologous recombination in *E. coli* (see below). The Not I-fitted globin PAC was further modified by the insertion of a lox P site at position –1659 upstream of

the human $^{G}\gamma$ gene (Imam *et al*., 2000), to give what is referred to in this paper as wild type PAC β locus. The targeting construct bearing the deletion of the two Enh and F silencers was generated by PCR amplification with Deep Vent DNA polymerase (New England Biolabs) using the PAC β locus DNA as template. The primers for the 5' region of homology (HUMHBB co-ordinates 40741-41360, GenBank accession U01317) were fitted with Asp718 (forward) and Bgl II (reverse) restriction sites. The 3' region of homology (HUMHBB co-ordinates 42961-43561) was amplified using primers fitted with Bgl II (forward) and NgoM IV (reverse) sites. The 5' and 3' regions of homology were first subcloned by triple fragment ligation into the Asp718 and NgoM IV sites of pBluescript II-SK (Stratagene), thus resulting in the deletion of elements Enh (HUMHBB co-ordinates 41360-42113) and F (HUMHBB co-ordinates 42593-42961) that resided in the sequences between the amplified 5' and 3' regions of homology. The 5' and the 3' homology regions were re-cloned from pBluescript II-SK into the Asp718 and NgoM IV sites of the pDF27 recombination vector to generate the final targeting vector. pDF27 is a derivative of the pDF25 recombination vector (Imam *et al*., 2000) and was constructed by the introduction of a BamHI-Not I-Sal I linker in the unique HindIII site of pDF25 (R. Janssens, personal communication). The fidelity of all cloning steps involving PCR amplification was verified by DNA sequencing.

### *Homologous recombination in E. coli*

Homologous recombination in the β-globin locus PAC was performed according to Imam *et al*. (2000). Briefly, the pDF27 targeting vector was electroporated into *E. coli* DH10B cells carrying the β-globin locus PAC. Selection for integration of the targeting vector into the PAC was carried out by plating on media containing chloramphenicol (Cm 25μg/ml) and kanamycin (Kan 25μg/ml) followed by incubation at $30^0$C. Integration was further selected by re-plating 80-100 positive colonies on Cm/Kan plates and incubating at $43^0$C. At this non-permissive temperature the pDF27 vector cannot replicate and resistance to Cm can only be acquired if the targeting vector integrates into the PAC insert or the *E.coli* genome. Integration of the pDF27 vector into the PAC globin sequences via homologous recombination will lead to duplication of the homologous sequence to yield Type I and Type II integrants depending on the precise breakpoint of the integration event (Imam *et al*., 2000). Colonies were picked and grown in liquid cultures for plasmid minipreparations according to standard methods and analysed for the correct integration event in the PAC as described in the Results section. Clones with the correct integration patterns were re-plated on Kan plates (25μg/ml) and incubated at $43^0$C for the excision step. This results from the RecA-mediated recombination between the duplicated copies of the target sequence that flanks the integrated vector, leaving behind either the original copy in the PAC insert or the modified copy. Incubation at the non-permissive temperature of $43^0$C results in loss of the excised vector. To ensure that this is the case, colonies were plated, in arrays, on Kan (25μg/ml)/High Streptomycin (Str 200μ/ml) media at $43^0$C. Under these conditions, colonies with two copies of the Str resistance conferring rpsL+ allele (one in the targeting vector and one in the DH10B genome) are sensitive to high Str concentrations, whereas colonies that have lost one copy of the rpsL+ gene by excision and loss of the integrated vector will be resistant. Correct excision and PAC modification were checked by Southern blot analysis, as described in the Results section. In addition, the region with the silencer deletion in the modified PAC was sequenced in order to check the exact breakpoints of the deletion that were generated by the homologous recombination. The resulting PAC bearing the silencer element deletion was named PAC β locus ΔEnh/F.

### *Transgenic Mice*

The 185 kb PAC insert was isolated by Not I digestion and purified from vector sequences by salt gradient centrifugation, essentially as described by Dillon and Grosveld (1993). Briefly, the digested PAC was layered on top of a 5-25% NaCl gradient and centrifuged at 40,000rpm, room temperature for 50 minutes in a SW41 swing-out rotor. 0.5ml fractions were collected and analysed by agarose gel electrophoresis. Fractions containing only the PAC insert were pooled and dialysed against a large volume of TE (10mM Tris-HCl pH 8.0, 1mM EDTA)/0.1M NaCl for 5 h at $4^0$C in UH 100-75 dialysis tubing (Schleicher & Schuell). Dialysis was continued overnight at $4^0$C after replacing the buffer. The PAC insert was concentrated by vacuum dialysis and subsequently dialysed against a large volume of

microinjection buffer (10mM Tris-HCl pH 7.4, 0.1mM EDTA) containing 0.1M NaCl in order to protect the high molecular weight PAC insert DNA from shearing during microinjection. The purified PAC fragment was checked for DNA integrity and concentration by pulsed field gel electrophoresis in a 1% agarose gel in 0.25XTAE buffer using a Biometra RotaphorType V apparatus, under the following conditions: 8-2sec pulse interval logarithmic ramp, 120-110° rotor angle linear ramp, 200-180volt logarithmic ramp, rotor speed 6 at 13°C for 21hours. The purified insert was injected at approximately 0.5ng/µl into the pronucleus of fertilised eggs of FVB/N mice. The injected eggs were transferred into the oviducts of pseudo-pregnant BCBA foster females as previously described (Kollias *et al*., 1986). Transgenic founder animals were identified by Southern blot analysis using the LCR's HS5 3.3 kb EcoRI fragment and the 2.3 kb EcoRI fragment 3' to the $^{A}\gamma$ gene as probes. Transgenic lines were established by breeding transgenic founders to non-transgenic FVB/N mice. F1 or F2 males were used for mating with non-transgenic females for the collection of embryos at various developmental stages for expression analysis.

The integrity of the 75 kb human β-globin locus in the PAC transgenes was checked by cosmid suppression hybridisation using the LCRε and γγδβ cosmids as probes, according to Strouboulis *et al*. (1992). Additional transgene mapping was done using the probes HS5 (3.3 kb EcoRI fragment), HS2 (4.2 kb Hind III fragment), 5' $^{A}\gamma$ (1.7 kb EcoRI-BamHI fragment), β intron II (0.9 kb BamHI-EcoRI fragment), 3' $^{A}\gamma$ (2.3 kb EcoRI fragment), 5'δ (0.98 kb Xba I-Bgl II fragment). Transgene copy numbers were determined using as probes the 5' $^{A}\gamma$-globin probe, the β-globin intron II probe and a 0.9 kb Pvu I fragment from the endogenous mouse carbonic anhydrase II (CA-II) gene. The ratios of intensities of the 5'γ /CA-II bands obtained for the PAC transgenic mice were compared to those obtained for the single copy human β-globin locus transgenic lines 2 and 72 (Strouboulis *et al*., 1992). Analysis was performed by PhosphorImager using ImageQuant software (Molecular Dynamics).

### *DNA FISH analysis*

Peripheral blood cells were cultured for 72 hours in RPMI 1640 medium. Chromosome preparations were made according to standard procedures. FISH was carried out as described by Mulder *et al*. (1995) (Mulder *et al*., 1995). The specific probe used was the biotin-labelled human β-globin LCR to detect the transgene, which was immunochemically detected with fluorescein. Chromosomal DNA was counter-stained with DAPI, which stains centromeric domains more intensely.

### *S1 nuclease protection assays*

S1 nuclease protection analysis was carried out with total RNA from 10.5 dpc embryonic yolk sac, 12.5 and 16.5 dpc foetal livers and blood from adult animals. RNA was isolated using the Trizol reagent according to the manufacturer's instructions (Life Technologies). Conditions for S1 nuclease protection assays and polyacrylamide gel electrophoresis were essentially as previously described (Fraser *et al*., 1990 and Kollias *et al*., 1986). The probes used were as follows:
Human probes: ε-globin 5' probe, 340 bp BamHI/BspMI fragment, protected fragment size 135 bp; γ-globin 5' probe, 320 bp Ava II fragment, protected fragment size 165 bp; β-globin 5' probe, 525 bp Acc I fragment, protected fragment size 155 bp.
Mouse probes: βH1-globin 5' probe, 255 bp Hinf I fragment, protected fragment size 180 bp; εy-globin 3' probe 369 bp EcoRI fragment, protected fragment size 195 bp; $\beta_{maj}$-globin 5' probe, 700 bp Hind III/Nco I fragment, protected fragment size100 bp.
Specific activities of probes were determined as previously described (Lindenbaum and Grosveld, 1990) and are indicated in the Figure legends. Quantitation was done on a PhosphorImager using the ImageQuant software (Molecular Dynamics).

### *Primary transcript in situ hybridisation*

Embryonic blood cells from 10.5 dpc embryos and 12.5 and 16.5 dpc foetal liver cells were disrupted by pipetting in PBS, spotted onto poly-lysine-coated slides (Sigma) and fixed according to van de Corput and Grosveld (2001). For detection of each globin gene transcript, a mixture of three or four different 50-mer oligonucleotide probes was used. Each oligo probe was labelled in the middle and at

the ends with biotin or digoxigenin (DIG) haptens. Probes for primary transcript *in situ* hybridisations were designed from the intron sequences of the different globin genes, such that they are at least 25 nucleotides apart. For 10.5 dpc yolk sac, biotinylated γ-globin intron probe was used in combination with DIG-labelled mouse εy-or mouse α-globin probes. The sequences of probes used have been previously described (Gribnau *et al*., 1998 and Trimborn *et al*., 1999). Fluorescence was detected by epifluorescence microscopy and photographs recorded with a CCD camera. At least 100 cells per slide were scored for primary transcripts in each experiment.

*Primer extensions*

$^{G}\gamma$ and $^{A}\gamma$ transcripts were distinguished by the use of two $^{G}\gamma$- and $^{A}\gamma$-specific primers. Primers were end-labelled with T4 polynucleotide kinase (New England Biolabs) at $37^0$C for 1 hour and unincorporated radionucleotides were removed by Sephadex G-25 chromatography. Specific activities were determined for both primers, as previously described (Lindenbaum and Grosveld, 1990) and in all experiments were found to be of very similar levels. Each labelled primer was phenol extracted, ethanol precipitated and re-dissolved in DEPC-treated TE (10mM Tris-HCl, 1mM EDTA pH 8.0) to a final concentration of 2ng/μl. Approximately 10μg of 10.5 dpc yolk sac or 12.5 dpc foetal liver RNA was annealed to 2ng of both $^{G}\gamma$ and $^{A}\gamma$ primers in a 33μl final reaction volume also containing 4μl of 5x Superscript buffer (Life Technologies) and 1μl of 0.1M DTT. Annealing of the primers to the RNA template was carried out by incubation at $65^0$C for 10 minutes followed by incubation at $37^0$C for 1 hour. The extension step was carried out by adding to each reaction 1μl of nucleotide mix containing 10mM each of dGTP, dATP, dCTP and 1.25mM of ddTTP, 0.5μl RNase inhibitor (RNAguard, Amersham-Pharmacia) and 0.5μl of Superscript reverse transcriptase (Life Technologies), followed by incubation at $42^0$C for 1 hour. Samples were ethanol precipitated, resuspended in 5μl of 97.5% formamide loading dye, heat-denatured and resolved by electrophoresis on a 12% denaturing polyacrylamide sequencing gel.

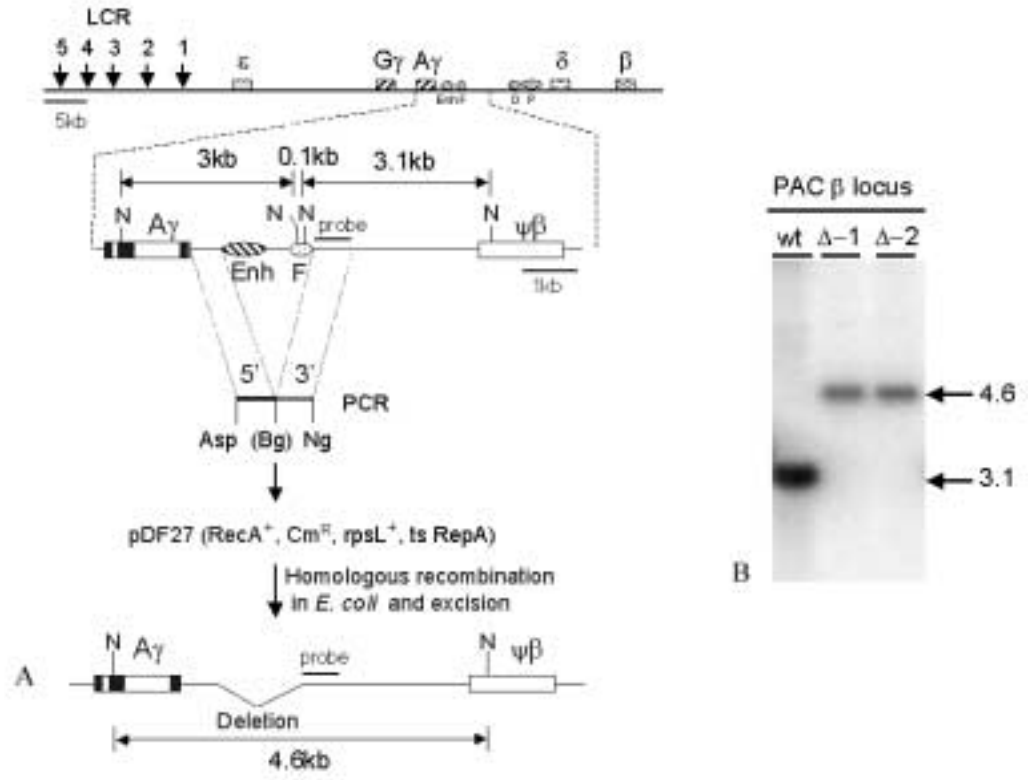



***Figure 1****: Deletion of Enh and F from the human β-globin locus PAC by homologous recombination in E. coli—* **(**A**)** The 75 kb human β-globin locus is shown at the top with the region 5' to the δ-globin gene magnified above it to show the O and P elements (ovals). The region 3' of the $^{A}\gamma$ gene has also been magnified below the locus to show the Enh and F elements (ovals) (ψβ: pseudo-β-globin gene). The 5' and 3' regions of homology amplified for use in the targeting construct are indicated (dotted lines). These were cloned into the Asp 718/NgoM IV sites of the targeting vector pDF27 which carries a chloramphenicol resistance gene ($Cm^R$), a temperature sensitive replication initiation protein gene (ts RepA), the recA gene and a wild type copy of the rpsL gene required for *E. coli* DH10B sensitivity to high doses of streptomycin (Imam *et al*., 2000). Homologous recombination followed by excision of the targeting construct results in the deletion of the Enh and F elements 3' of the $^{A}\gamma$ gene in a subset of the modified PAC clones (Imam *et al*., 2000). As a result, two Nco I sites are deleted giving rise to a new 4.6 kb fragment from the original 3.0 kb and 3.1 kb bands. **(**B**)** Southern blot hybridisation of Nco I-digested PAC DNA obtained from the original unmodified human β locus PAC (wt lane) and two modified PAC clones bearing the 3' $^{A}\gamma$ deletion (lanes Δ-1 and Δ-2). The probe (shown in panel A) detects a 3.1 kb fragment in the unmodified PAC and a 4.6 kb fragment in the modified PAC resulting from the deletion of two Nco I sites (panel A). Abbreviations: N-Nco I, Asp-Asp 718, Bg-Bgl II, Ng-NgoM IV.

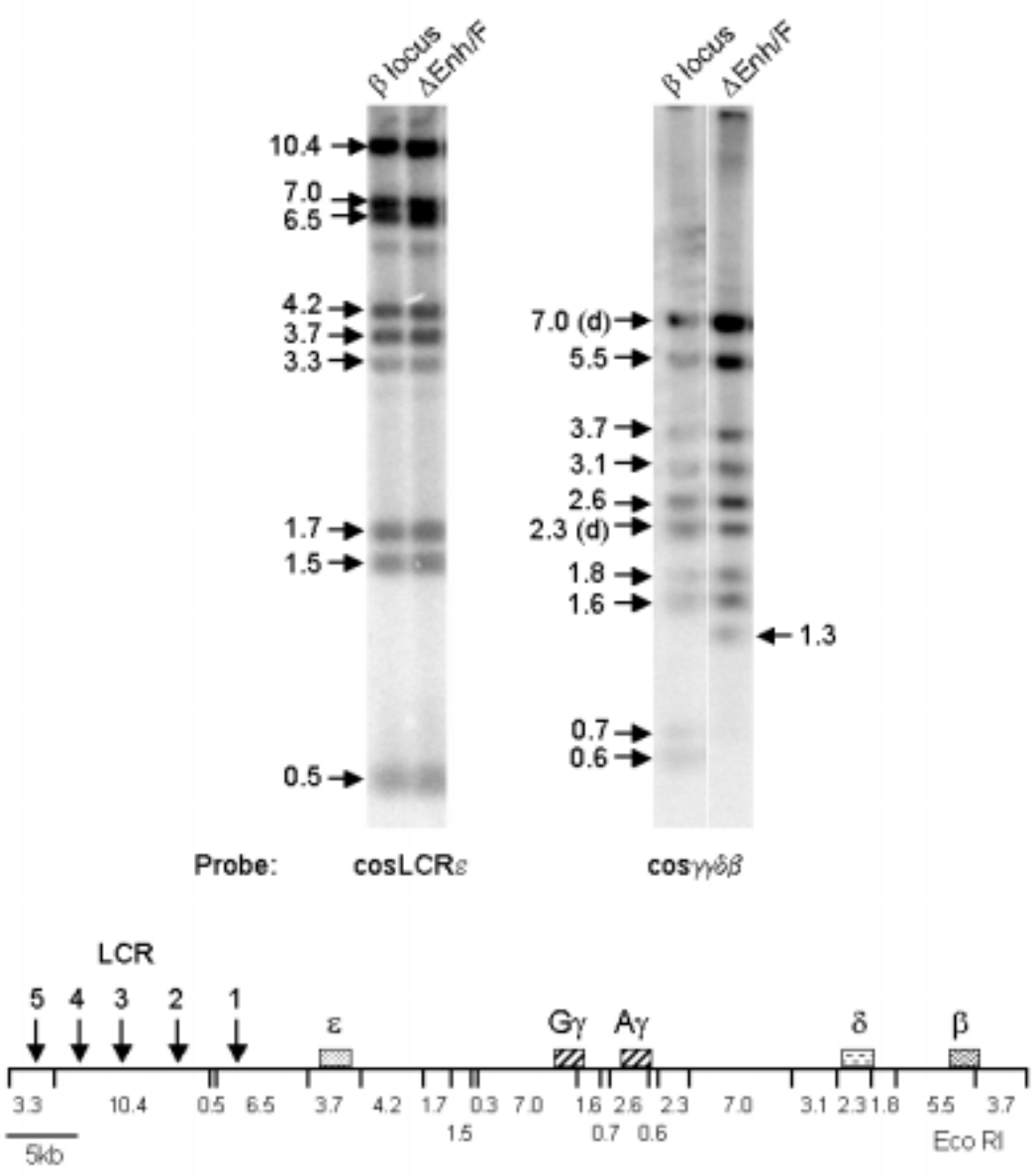



***Figure 2****: Mapping the integrity of the 75 kb human β-globin locus in the unmodified (wt lane) and Enh/F-deleted (ΔEnh/F) globin locus PACs*—Eco RI digested PAC DNA was probed with cosmids LCRε (left panel) and γγδβ (right panel). The modified PAC clone DNA that was used for transgenesis is shown. The Eco RI map of the locus is shown below the blots. Deletion of the Enh and F elements results in a 2.3 kb Eco RI fragment that is part of a doublet (d) in the unmodified PAC, becoming a new 1.3 kb band.

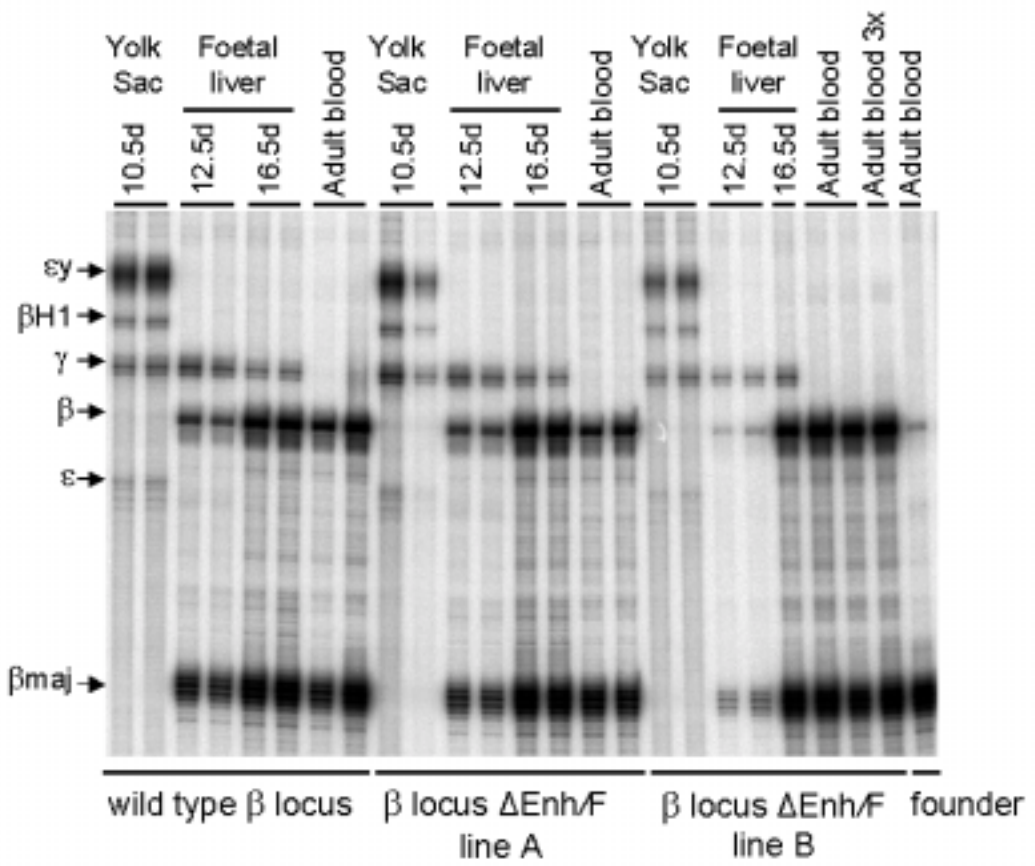



***Figure 3****: S1 nuclease protection analysis of expression of human versus mouse β-like globin genes in embryonic yolk sac, foetal liver and adult blood*—Duplicate RNA samples obtained from embryos (littermates) and adult blood from two mouse lines transgenic for a single copy of the Enh and F-deleted β locus PAC (ΔEnh/F line A and B) were analysed (only one embryo was analysed for the 16.5d time point in line ΔEnh/F-B). As control, one line transgenic for a single copy of the unmodified (wild type) human β globin locus PAC was analysed. Adult blood RNA from a transgenic founder that did not transmit the deleted PAC transgene, was also included in the analysis (last lane). The protected fragments are indicated to the left of the panel. The relative specific activities of probes used in this experiment are given in the legend for Table 1.

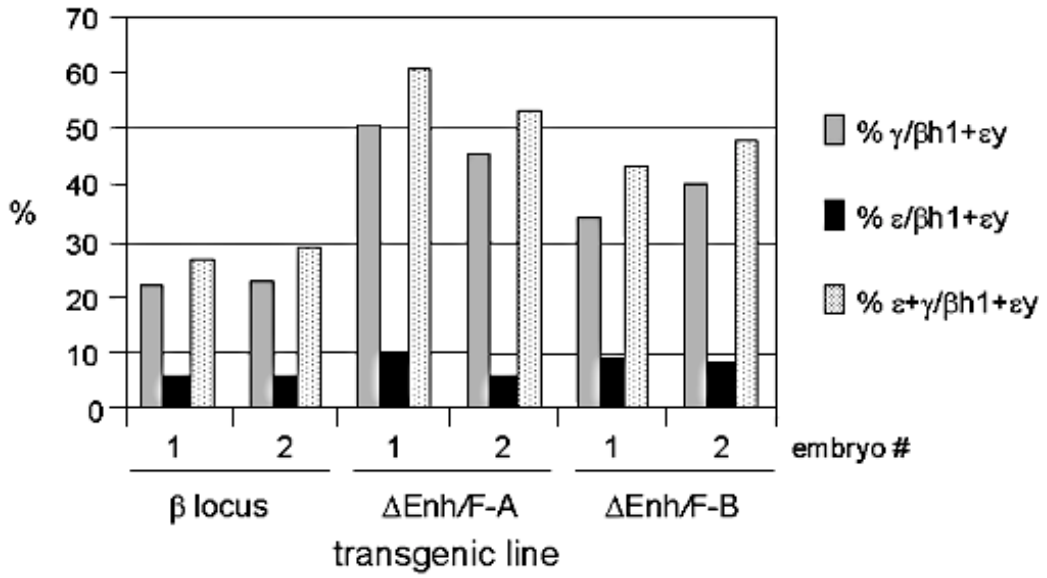
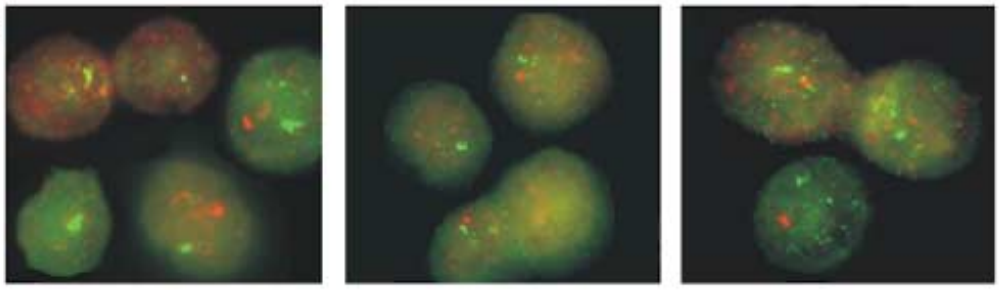



***Figure 4****: Bar chart representation of human ε- and γ-globin expression levels at 10.5 day yolk sac, as a percentage of mouse εy-and βh1-expression*—The values are those presented in Table 1 and were calculated from the S1 nuclease protection assay in Figure 3.

A



β locus  Δ Enh/F line A  Δ Enh/F line B

B

| Transgenic line | % γ+ε | % γ | % ε | % α+γ | % α | % γ |
|---|---|---|---|---|---|---|
| β locus | 42 | 56 | 2 | 61 | 35 | 4 |
| Δ Enh/F line A | 39 | 57 | 4 | 60 | 30.5 | 4.5 |
| Δ Enh/F line B | 37.8* | 59.5* | 2.7* | 56 | 39 | 6 |

***Figure 5****: Primary transcript in situ hybridisation in 10.5d mouse embryonic blood*—**(**A**)** Representative fields of cells hybridised with mouse α-globin (green signal) and human γ-globin (red signal) intron-specific probes. **(**B**)** Quantitation of ε- and γ-globin primary transcripts (columns 2-4) and γ-globin and mouse α-globin primary transcripts (columns 5-7) by *in situ* hybridisation of 10.5d embryonic blood cells. At least 100 cells were counted in each case, with the exception of (*) where 36 cells were counted.

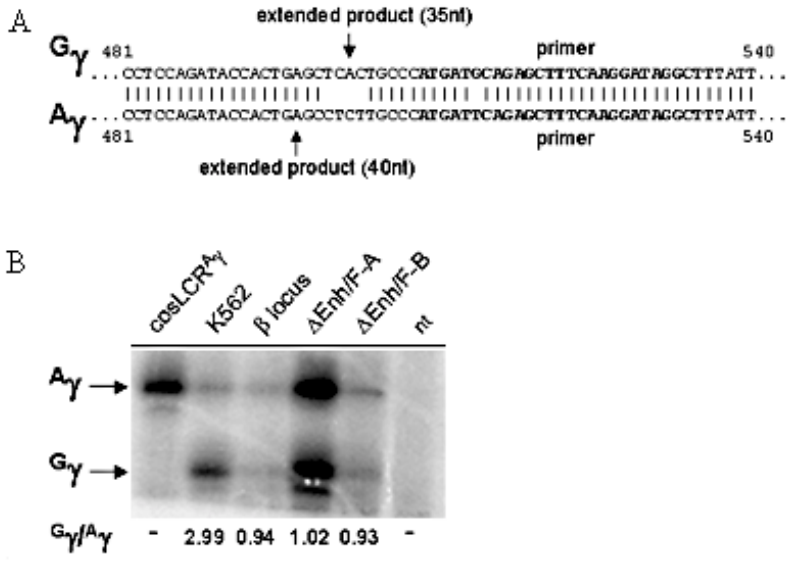



***Figure 6**: Analysis of $^{G}\gamma$-versus $^{A}\gamma$-globin expression by primer extension—* **(**A**)** alignment of sequences from the 3' of the $^{G}\gamma$-and $^{A}\gamma$-cDNAs (nucleotides 481-540) which shows the 5nt region of non-homology between the two sequences. The sequences of the $^{G}\gamma$- and $^{A}\gamma$-specific primers used in the assay are shown in bold. The use of dideoxy-TTP in the extension mix results in 35nt and 40nt extended products from the $^{G}\gamma$ and $^{A}\gamma$ mRNAs, respectively (arrows). **(**B**)** resolution of $^{G}\gamma$ and $^{A}\gamma$ primer extension products on a 10% polyacrylamide gel. 10.5d yolk sac RNA was analysed from the control single copy unmodified β locus PAC line (β locus lane) and the two lines carrying the Enh/F-deleted locus PAC (ΔEnh/F-A and -B lanes). Controls include a 12.5 day foetal liver RNA sample from an embryo transgenic for a single copy of an LCR $^{A}\gamma$ construct (cosLCR $^{A}\gamma$ lane), RNA from K562 cells that express γ-globin and a 10.5 day yolk sac RNA from a non- transgenic embryo (nt lane).

***Table 1****:S1 nuclease protection assay quantitations*—(A**)** Quantitation of human ε- and γ-globin expression levels at 10.5day embryonic yolk sac as a percentage of mouse embryonic εy and βH1 globin expression. Values were calculated from the S1 nuclease protection shown in Figure 3 using ImageQuant and were corrected for probe specific activities according to the following ratios: 1 : 1.56 : 1.7 : 2.4 for ε : γ : βH1 : εy probes. (B) Quantitation of human γ- and β-globin expression in 12.5d and 16.5d foetal liver as well as adult blood as a percentage of mouse $\beta_{maj}$ globin expression. Two embryos per line were analysed for each time-point, with the exception of the 16.5d time-point for line ΔEnhF-B where only one embryo was available for analysis. In addition, the embryos analysed for the 12.5d time-point for line ΔEnhF-B were a little earlier in development than the rest of the embryos analysed at this time-point (see results). Values were calculated from the S1 nuclease protection shown in Figure 3 using ImageQuant and were corrected for probe specific activities according to the following ratios: 1 :1.2 : 1.3 for γ : $\beta_{maj}$ : β probes.

A.

| | | 10.5d | | | | | |
|---|---|---|---|---|---|---|---|
| Transgenic line | Embryo Number | γ/βH1 | ε/εy | γ/βH1+εy | ε/βH1+εy | γ+ε/βH1+εy | εy/βH1 |
| wt β locus | 1 | 2.2 | 0.06 | 0.22 | 0.06 | 0.27 | 9.1 |
| | 2 | 2.3 | 0.07 | 0.23 | 0.06 | 0.29 | 8.8 |
| ΔEnh/F line A | 1 | 3.9 | 011 | 0.51 | 0.10 | 0.61 | 6.5 |
| | 2 | 4.0 | 0.07 | 0.46 | 0.06 | 0.53 | 7.7 |
| ΔEnh/F line B | 1 | 2.3 | 0.10 | 0.34 | 0.09 | 0.43 | 5.7 |
| | 2 | 2.6 | 0.09 | 0.40 | 0.08 | 0.48 | 5.4 |

B.

| | | 12.5d | | | 16.5d | | | Adult blood |
|---|---|---|---|---|---|---|---|---|
| Transgenic line | Embryo Number | $\gamma/\beta_{maj}$ | $\beta/\beta_{maj}$ | $\gamma+\beta/\beta_{maj}$ | $\gamma/\beta_{maj}$ | $\beta/\beta_{maj}$ | $\gamma+\beta/\beta_{maj}$ | $\beta/\beta_{maj}$ |
| wt β locus | 1 | 0.22 | 0.27 | 0.49 | 0.03 | 0.48 | 0.51 | 0.67 |
| | 2 | 0.21 | 0.27 | 0.49 | 0.03 | 0.61 | 0.65 | 0.76 |
| ΔEnh/F line A | 1 | 0.25 | 0.36 | 0.61 | 0.03 | 0.70 | 0.72 | 0.28 |
| | 2 | 0.18 | 0.37 | 0.55 | 0.02 | 0.54 | 0.56 | 0.62 |
| ΔEnh/F line B | 1 | 0.44 | 0.13 | 0.57 | 0.03 | 0.49 | 0.51 | 0.76 |
| | 2 | 0.28 | 0.15 | 0.44 | - | - | - | 0.81 |

# *CHAPTER 5*
# *THE EFFECTS OF DELETING HS2 OR HS3 FROM THE LCR IN THE CONTEXT OF A 185KB HUMAN β-GLOBIN LOCUS IN SINGLE COPY TRANSGENIC MICE: A REQUIREMENT FOR HS2 IN ε-GLOBIN EXPRESSION AND A GLOBAL ROLE FOR HS3 IN LOCUS ACTIVATION*

Mariken de Krom, John Kong-a-San, Frank Grosveld and John Strouboulis

This chapter will be submitted for publication

## Summary

The human β-globin LCR is required for the high level transcriptional activation of the globin genes in the locus. The LCR consists of five DNase I hypersensitive sites which appear to interact together in forming a single functional unit, termed the holocomplex. We have previously deleted individual HS sites from the LCR and analysed the effects of these deletions on the regulation of the complete 70 kb human β-globin locus in transgenic mice. We found that HS1, HS2 and HS3 deletions rendered the locus susceptible to classical position effect variegation and a novel cell timing position effect. We have extended these studies by repeating the deletions of HS2 and HS3 in the context of a 185 kb human β-globin locus PAC. We analysed single copy transgenic mice with defined chromosomal sites of transgene integration. With the exception of one line that exhibits severe position effects particularly in the foetal liver and adult stages, we found that deletion of HS2 results in a more pronounced reduction of ε-globin expression in the embryonic yolk sac, whereas the effects on γ-globin expression in the yolk sac and γ- and β-globin expression in the foetal liver and adult blood stages were milder. These results confirm that HS2 contributes to LCR function and also reveal a potential requirement of HS2 for ε-globin expression. The deletion of HS3 gives rise to severe position effect variegation in the two transgenic lines tested. These results further confirm the key role that HS3 plays in LCR function. Taken together, these results provide additional support for the holocomplex model of LCR function and additional suggest discrete functions of individual HS sites within the holocomplex.

## Introduction

The human β-globin locus contains five developmentally regulated genes organised in the order in which they are activated during development, i.e. 5'-$\varepsilon^{G}\gamma^{A}\gamma\delta\beta$-3'. The locus is regulated by the Locus Control Region (LCR) located upstream of the gene domain. Naturally occurring deletions in the locus that give rise to thalassemia, as well as transgenic mouse studies, have suggested that the LCR is a critical element for the long-range chromatin organisation and transcriptional activation of the β-globin locus (Grosveld, 1999). The LCR is comprised of five DNase I hypersensitive sites (denoted HS1-5) spanning a region of approximately 21kb immediately upstream of ε-globin, the 5'-most gene in the locus (Tuan *et al.*, 1985 and Dhar *et al.*, 1990). Highly localised DNase I hypersensitivity in this region underlies a high concentration of binding sites for transcription factors, both haematopoietic-restricted and ubiquitous (Talbot *et al.*, 1990; Philipsen *et al.*, 1990; Pruzina *et al.*, 1991and Zafarana *et al.*, 1995). Several studies have been previously carried out in attempting to delineate the functional properties of individual HS sites and the contribution of each HS site to full LCR function. For example, HS2 was shown to act as a classical enhancer in transient transfection assays as well as in stable transfection assays (Caterina, 1994 and Talbot *et al.*, 1990). By contrast, HS3 and HS4 can activate transcription only in a chromatin context (Philipsen *et al.*, 1990; Pruzina *et al.*, 1991 and Tuan *et al.*, 1992 ). In transgenic mice, HS2, HS3 and HS4 were each shown to be independently capable of driving high level transcription of linked human β-globin transgenes, though at lower levels than those obtained when using the full LCR (Fraser *et al.*, 1990; Talbot and Grosveld, 1991; Pruzina *et al.*, 1991 and Philipsen *et al.*, 1993). In addition, when individually linked to a cosmid construct bearing the human γ- and β-globin genes, HS2, 3 and 4 were able to activate high levels of transcription with different developmental-stage specificities (Fraser *et al.*, 1993). Finally, when analysed in single-copy transgenic mice, only HS3 was found to be capable of consistently driving high levels of expression that were related to transgene copy-number (Ellis *et al.*, 1996). These observations led to the suggestion that the individual HS sites in the LCR interact with each other to form a holocomplex which acts as a single functional entity in the long-range chromatin organisation and transcriptional activation of the human β-globin locus by the LCR (Fraser *et al.*, 1993). This model gained further support by the demonstration that the LCR interacts with only one gene at a time, as though it presents a single target for interaction, but can activate more than one gene by "flip-flopping" between genes in the locus (Wijgerde *et al.*, 1995).

More recent approaches aimed at gaining insight into the role of each HS site in LCR function have involved the deletion of individual HS sites from the LCR and their analysis in the context of the complete human β-globin locus regulation in transgenic mice (Bungert *et al.*, 1995; Milot *et al.*, 1996; Peterson *et al.*, 1996; Navas *et al.*, 1998; Bungert *et al.*, 1999 and Navas *et al.*, 2001). Our own previous work, involved the deletion of each of HS1, HS2, HS3 and HS4 in the context of two ligated cosmids that together spanned ~70 kb of the human β-globin locus (Milot *et al.*, 1996). With the exception of HS4 which did not affect regulation of the human locus, these deletions rendered expression of the human globin genes sensitive to chromosomal positions effects, where levels of expression for all genes were low and not always directly related to transgene copy number (Milot *et al.*, 1996). These effects were most profound in transgenic mice with pericentromeric integrations of the HS1-, 2- or 3-deleted constructs and were evident in all developmental stages. A closer examination of expression in single erythroid cells, revealed two types of position effects in these mice. Firstly, classical position effect variegation (PEV) was observed in which only a subset of cells (variable between transgenic lines) transcribed the transgene. A second type of position effect was also observed, in which all cells transcribed the transgene but at a lower level per cell. This observation suggested that transcription of the human β-globin genes was taking place in all cells but for a shorter period of time during the cell cycle. This type of position effect had not been previously described and was termed cell timing position effect (CTPE; Milot *et al.*, 1996).

Our results, however, were in contrast with a number of other studies. For instance, deletion of HS2 in the context of a 246 kb human β-globin locus YAC by Peterson *et al.* (1996), showed only a small decrease in the expression of all globin genes at all developmental stages. Bungert *et al*. (1999) reported a severe decrease in expression of all genes when HS2 was deleted in the context of a 155 kb YAC. In addition, deletion of HS3 in the study of Peterson *et al.* (1996) showed a milder effect on the expression of human globin genes, especially in the foetal liver and adult stages, whereas deletion of a smaller HS3 fragment by the same group (Navas *et al.*, 1998) showed a more severe effect on human globin gene expression, particularly in the foetal liver and adult stages. By contrast, Bungert *et al.* (1995) found expression of all human genes to be severely impaired in all developmental stages following deletion of HS3, which is in agreement with our observations. However, in contrast to our observations, Bungert *et al.* (1995) found that the deletion of HS4 also had a severe effect on expression of the human genes. These observations on the effects of the HS4 deletion are also at odds with the findings of Navas *et al.* (2001) who found that a similar deletion of HS4 had a more pronounced effect in the foetal liver and adult stages with a much milder effect in the embryonic stage. The differences between the studies could be due to the use of different constructs, different deletions, different methods of analysis of transgene integrity and globin expression levels, differences in copy-numbers and in integration sites of the transgenes. It should be mentioned that nobody (except our group), did check the integration site of the transgene, which does make a difference in outcome of a study.

In this paper, we repeated the deletions of HS2 and HS3 from the human β-globin LCR. These deletions are identical to the ones carried by Milot *et al.* (1996) but were carried out in the context of a 185 kb human β-globin locus PAC. The main impetus behind this work has been to obtain more transgenic lines with these HS deletions since the observations of Milot *et al.* (1996) raised the intriguing possibility that different types of position effects are associated with the deletion of specific HS sites (e.g. CTPE was observed with the HS3- and HS1- but not with the HS2-deleted mice, raising the potential that it was a HS1- or HS3-specific effect). In addition, in repeating these deletions, we incorporated a number of important differences with the aim of facilitating the interpretation of results. Firstly, the deletions were carried out in the context of a 185 kb human β-globin locus PAC containing much more flanking sequence compared to the 70 kb locus obtained by ligating two cosmids together (Milot *et al.*, 1996), the extra flanking sequence may buffer the transgene from severe chromosomal position effects. These constructs are also structurally closer to the human β-globin locus YACs that have been used in the HS deletion studies by Bungert *et al.* (1995 and 1999), Peterson *et al.* (1996) and Navas *et al.* (1998 and 2001). Secondly, the β-globin locus PAC was fitted with a loxP site upstream for the $^{G}\gamma$ gene so that the effects of the HS deletions could be studied in single copy transgenic animals, generated by crossing multi copy mice with transgenic mice expressing Cre recombinase in the germline. The analysis of HS deletions in single copy mice should remove the

added complication arising from interactions between transgenes in multi copy mice, as has been previously observed (Ellis *et al.*, 1996). Finally, in contrast to the previous studies using globin locus YACs, the chromosomal sites of transgene integration were identified in all transgenic lines generated and analyses of expression were carried out at the single cell level.

## Results

### *Generation of LCR deletion transgenic lines*

HS2 and HS3 were deleted from the LCR in the context of a 185 kb PAC containing the entire human β-globin locus with extensive flanking sequence (approximately 50 kb on either side of the locus; A. Imam, personal communication). The deletions were performed by homologous recombination in *E. coli* as described by Imam *et al.* (2000). The human β-globin locus PAC has been previously fitted with a loxP site upstream of the $^{G}\gamma$ gene (Imam *et al.*, 2000). The fragments used to delete the core region of the hypersensitive sites were the same as those previously described by Milot *et al.*, 1996 (Fig. 1A). The targeting fragments, a 2.8 kb HindIII fragment containing the 700 bp core deletion for HS2 and a 4.2 kb HindIII fragment containing the 1.4 kb core deletion for HS3, were cloned into the HindIII site of the pDF25 recombination vector. After homologous recombination, the correct excision events were checked by Southern blot analysis of HindIII digests for both HS2 and HS3 deletions.

In order to ensure that no re-arrangements had occurred in the human β-globin gene locus other than the targeted loxP integration and HS deletions, the wild type and HS-deleted PACs were checked by Southern blotting following homologous recombination and excision, using the LCRε and γγδβ cosmids as probes which, together, cover the ~75 kb human β-globin gene locus and LCR (Fig. 1B). Unrearranged PAC clones for each construct were selected and digested with Not I to release the globin locus insert. The correct size of the released fragment was checked by PFGE and subsequently isolated by salt gradient and prepared for microinjection into the pronuclei of fertilised mouse eggs.

From these microinjections we obtained 12 founders for the wild type β-locus, 8 founders for the HS2 deletion (ΔHS2) and 7 founders for the HS3 deletion (ΔHS3). Genomic DNA from tail biopsies from all of the founder mice was checked for the presence of the PAC transgene, using a probe detecting intron 2 of the β-globin gene at the 3' end of the locus and a probe which detects HS5 at the 5' end of the locus. The integrity of the 70 kb human β-locus and the presence of the HS2 and 3 deletions in the transgenes were checked in two ways after the transmission of the transgene to the F1 generation. Firstly, EcoRI-digested DNA was probed for integrity of the ~75 kb β-globin locus by Southern blotting and probed with the LCRε and γγδβ cosmids, as previously described (Strouboulis *et al.*, 1992). This assay reveals deletions and rearrangements within the 75 kb locus (data not shown). In addition, genomic DNA was digested with a number of additional restriction enzymes that together generate overlapping fragment spanning a region of approximately 100 kb across the β-globin locus and hybridised with different probes. In this way, linkage of the various fragments detected across the human β-globin locus can be demonstrated (data not shown). These assays do not test for integrity of the ~185 kb PAC-derived transgene in regions outside the ~100 kb region covered by the analysis. From these analyses, 9 wild type, 8 ΔHS2 and 6 ΔHS3 founders showed the human β-locus to be intact and were further bred in order to establish transgenic lines. Of those, 8 wild type, 4 ΔHS2 and 4 ΔHS3 founders transmitted the transgene to the F1 generation (Table 1).

For all transgenic lines we determined the chromosomal integration site of the transgene DNA FISH of metaphase spreads. We made a distinction between euchromatic (i.e. any integration other than pericentromeric or telomeric), (peri)-centromeric and (peri)- telomeric integration types. All three types were observed in both wild type and deletion lines (Fig. 2 and Table 1).

Copy numbers for each line were determined using the human β-globin intron 2 probe and a probe against the endogenous mouse carbonic anhydrase II (CA-II) gene as control (see materials and methods). The multiple copy transgenic lines were bred to Zp3-cre transgenic mice to obtain single copy transgenic mice. These mice express the Cre recombinase during oocyte maturation, thus resulting in the recombination to single copy of a loxP-fitted transgene in the eggs of these mice

(Lewandoski *et al.*, 1997). Except for two of the ΔHS3 transgenic lines, all other multi-copy lines could be bred to single copy, as shown by copy number determination. The recombination of these lines to yield single copy transgenes was also checked by comparing transgene expression levels between the multi copy and single copy animals (see below). All single copy lines were again checked for integrity of the human β-globin locus and were found to be intact (data not shown). The fact that we couldn't breed two ΔHS3 lines to single copy could be due to the integrated copies not being in a head-to-tail arrangement which is required for the correct alignment of the LoxP sites for Cre-mediated recombination. Our analyses for integrity of the human β-locus as described above, do not address the orientation of integrated transgene copies relative to each other.

We chose three wild type transgenic single copy lines and all single copy ΔHS2 and ΔHS3 transgenic lines for further analysis.

### *Expression analysis of the transgenic lines*

RNA was isolated from the adult blood of the single copy transgenic lines and their multi copy parent lines and globin gene expression was analysed by S1 nuclease protection. All transgenic lines showed an equivalent fold lower expression of the human β-globin gene in the single copy lines compared to the multiple copy lines (data not shown). This confirms the copy number-dependent expression of the human β-globin gene in our transgenic lines, as has been described previously (Grosveld *et al.*, 1987 and Blom van Assendelft *et al.*, 1989).

To analyse the effects of the HS site deletions on expression of the human globin genes during mouse embryonic development, we isolated yolk sac from 10.5dpc embryos, foetal livers from 14.5dpc embryos and blood from adult mice. Total RNA was isolated from these tissues and human globin gene expression was analysed against that of the endogenous mouse β-like globin genes by S1 nuclease protection assays (Strouboulis *et al.*, 1992).

The three wild type transgenic lines showed similar expression patterns at all time points (Fig. 3 and Fig. 4). The levels and developmental expression patterns obtained with the three wild type lines are comparable to those previously reported for transgenic mice containing cosmid, BAC and YAC based human β-globin locus transgenes (Strouboulis *et al.*, 1992; Gaensler *et al.*, 1993; Peterson *et al.*, 1993; Bungert *et al.*, 1995; Porcu *et al.*, 1997; Kaufman *et al.*, 1999 and Alami *et al.*, 2000 ). Human globin gene switching also took place in these mice as previously described, although the γ to β switch occurred a bit later than observed with the cosmid based β locus transgenic mice (Strouboulis *et al.*, 1992), but in agreement with the timing reported for YAC transgenic mice (Kaufman *et al.*, 1999; Peterson *et al.*, 1993 and Bungert *et al.*, 1995).

Expression analysis of the ΔHS2 transgenic lines showed different effects for different human globin genes at different developmental stages (Fig. 3, Fig. 4). The most severe effect resulting from the deletion of HS2 was in the expression of human ε-globin in the embryonic yolk sac. Lines 4 and 5 (centromeric and euchromatic integrants, respectively) show an almost complete lack of ε-globin expression with lines 6 and 7 (both telomeric integrants) also showing markedly reduced ε expression (Fig. 3, Fig. 4). Expression of γ-globin in the same stage is also markedly reduced in all ΔHS2 lines, but the effect is milder compared to ε-globin (Fig. 3, Fig. 4). The overall transcriptional output of the human globin genes (ε + γ) compared to total mouse globin output (εy + βH1) is reduced, primarily due to the sharp reduction in human ε expression.

Expression analysis in the foetal liver and adult blood stages, shows γ-globin expression to be only mildly affected by the deletion of HS2 compared to the wild type locus control mice (Fig. 3, Fig. 4). Expression of human β-globin appears to be affected to a greater extent compared to γ expression, but the effects are again more moderate compared to ε-globin expression in the yolk sac (Fig. 3, Fig. 4). Overall transcriptional output of the human globin genes (γ + β) compared to $\beta_{maj}$ is reduced in the absence of HS2 in the foetal liver and adult blood stages. An exception is line 5 which exhibits severely reduced levels of human γ- and β-globin expression in the foetal liver and adult blood stages (Fig. 3, Fig. 4). Despite its apparent euchromatic integration site, the human globin locus in this line appears to suffer from a severe chromosomal position effect which becomes very evident in the foetal liver and adult stages.

Analysis of the ΔHS3 transgenic lines showed for both lines (lines 8 and 9, euchromatic and pericentromeric integrants, respectively) a severe reduction in expression of all human globin genes at all developmental time points (Fig 3, Fig 4). It is striking that in the adult blood stage there is hardly any human β-globin expression detectable in both lines. The deletion of HS3 therefore leads to severe position effects in these lines, as has been previously observed (Bungert *et al.*, 1995 and Milot *et al.*, 1996 (only observed in some lines)).

### *Primary transcript in situ hybridisation*

The basis of chromosomal position effects becomes evident by assaying for transgene expression at the single cell level (Milot *et al.*, 1996 and McMorrow *et al.*, 2000). To test whether the decrease in human globin gene expression in the HS-deleted mice is reflected in the number of transcribing loci in the nucleus, we carried out primary transcript *in situ* hybridisation on 10.5 day embryonic blood and 14.5 day foetal liver cells from the same samples that were analysed by S1 nuclease protection.

The transcription patterns in 10.5dpc embryonic blood were analysed using gene-specific intron probes for mouse α-, human ε- and human γ-globin genes (Fig. 5A-B, Table 2). In agreement with the S1 nuclease protection data, we find a decrease in the number of human globin gene loci actively transcribing ε-globin (Fig 5A and Table 2). At the same time we observe an almost imperceptible reduction in the number of human globin loci actively transcribing the γ genes, compared to the wild type locus control mice (Fig 5B and Table 2). This is in contrast to the S1 nuclease protection results which indicated a marked reduction in γ-globin expression levels. Taken together, these observations suggest that deletion of HS2 does not appear to affect the number of γ-globin transcribing loci, however, transcriptional output from the γ genes (which cannot be measured in the *in situ* hybridisation assays) appears to be compromised.

For the ΔHS3 lines, we observe an almost complete lack of ε-globin transcription in embryonic blood and a severe reduction in the number of γ-globin transcribing loci (Fig 5B and Table 2) in agreement with the S1 assays.

Primary transcript patterns in 14.5 day foetal livers were analysed using intron probes specific for the mouse α-and human γ-and β-genes. The transcription patterns observed for γ-globin in the ΔHS2 mice are in agreement with the results obtained from the S1 analysis in that the drop in steady-state mRNA levels is accompanied by a decrease in the number of human γ transcribing loci (Table 3). For ΔHS2 line 5 and both ΔHS3 transgenic lines, the severe reduction in human globin gene expression observed in the S1 protection analysis is also evident in the primary transcript *in situ* hybridisation.

### *In situ mRNA analysis*

In order to address the question whether the low expression levels of the human genes in the different deletion lines (line5, 8 and 9) was caused by classical position effect variegation (PEV) (Milot et al., 1996), a reduction of the number of cells expressing the transgene, we performed mRNA *in situ* hybridisation on 14.5dpc foetal liver cells of the same embryos used for S1-nuclease protection assays and primary transcript *in situ* hybridisations. Expressing erythroid cells were visualised with an exon specific oligo for mouse-$\beta_{maj}$ and cells expressing the transgene were identified with either an exon specific oligo for human-γ or for human-β.

The mRNA *in situ* hybridisation showed for all three deletion lines a clear reduction of number of cells expressing the transgene in comparison to the wild type transgenic lines. We conclude from this that these three deletion lines (one ΔHS2, line 5, and both ΔHS3, line8 and 9) show position effect variegation.

## Discussion

In this study we have used a 185 kb PAC containing the human β-globin locus to investigate the influence of the deletion of either HS2 or HS3 of the LCR on the regulation of human β-globin gene locus in transgenic mice. We generated a number of transgenic lines bearing the wild type and the

HS2- and HS3-deleted human globin loci. Expression of human globin genes was analysed in single copy mice with chromosomal integration sites identified. We found that all wild type PAC transgenic lines analysed were regulated similarly to human β-globin locus transgenic mouse models previously generated by a number of groups using ligated cosmids, YACs and BACs (Strouboulis *et al.*, 1992; Peterson *et al.*, 1993; Gaensler *et al.*, 1993; Bungert *et al.*, 1995; Porcu *et al.*, 1997; Kaufman *et al.*, 1999 and Alami *et al.*, 2000).

### *The effects of deleting HS2 and HS3 on human β-globin locus regulation*

We analysed four single copy transgenic mouse lines bearing the HS2 deletion. We found that the gene most severely affected by the HS2 deletion in all four lines was that of human ε-globin in the embryonic yolk sac. The mRNA expression analysis and the primary transcript *in situ* hybridisation data suggested that the effects on ε gene expression were due to a drop in the number of human globin gene loci actively transcribing ε-globin. The decrease of ε expression in all ΔHS2 lines indicates that interaction of HS2 with the ε-promoter is of importance for normal expression of the gene. These observations may be related to recent work which has shown that the presence of HS2 is required for the specific histone H3 hyperacetylation of a TATA box-proximal nucleosome in the ε-globin promoter in episomally maintained minichromosomes (Gui and Dean, 2001). On the basis of these observations it was suggested that HS2 is responsible for the recruitment of a specific acetylase activity responsible for the high-level modification of histone H3 at the proximal promoter in ε-globin activation (Gui and Dean, 2001).

By contrast, γ-globin gene expression in the embryonic yolk sac is only moderately affected by the HS2 deletion. Interestingly, the reduction in γ-globin mRNA levels does not appear to be reflected to the same extent in the number of γ-globin transcribing gene loci. This suggests that while the frequency and/or stability of interaction of the LCR with the γ-globin genes may not be significantly affected by the HS2 deletion, the transcriptional output of the genes is reduced. The same does not appear to be true for γ-globin expression in the foetal liver stage. In this case the (moderate) reduction in γ gene mRNA levels is accompanied by a reduction in the number of human globin loci actively transcribing γ-globin. Interestingly, β-globin gene expression in the foetal liver and adult blood appears to behave in a manner similar to that of γ-globin in the embryonic yolk sac, i.e. the reduction in β-globin mRNA levels is not accompanied by a similar reduction in the number of actively transcribing β gene loci. Take together, these data suggest that HS2 may have more specialised function(s) in the embryonic yolk sac stage, in addition to its more general functions in the transcriptional activation of all genes in the locus. The latter function may be related to the transcriptional enhancer activity previously reported for HS2 in transient transfection assays (e.g. Caterina *et al.*, 1994).

Our observations described here on the HS2 and HS3 deletions, present differences as well as similarities to previous reports on the effects of HS site deletions on the regulation of the human β-globin locus in transgenic mice (Bungert *et al.*, 1995; Milot *et al.*, 1996; Peterson *et al.*, 1996; Navas *et al.*, 1998; Bungert *et al.*, 1999 and Navas *et al.*, 2001). It is difficult to directly compare and reconcile these differences since there are significant variations in the constructs employed and in the ways the deletions were made, the transgenes were analysed and expression patterns determined. What is perhaps more pertinent is the comparison of the HS2 and HS3 deletions presented here to those carried out by Milot *et al.* (1996) in the context of the ~70 kb locus obtained through the ligation of two cosmids. In both studies the deletions for HS2 and HS3 were identical. However, in the present study the β-locus was part of a 185 kb transgene, thus carrying extensive additional flanking sequence (~50 kb on either side of the β-globin locus) compared to the 70 kb locus transgene used by Milot *et al.* In addition, our present analysis on the HS2 and HS3 deletions was restricted to single copy mice. The study by Milot *et al.* (1996) included two single copy lines for the HS2 deletion and only multi copies for the HS3 deletion.

The most significant difference observed between our study and that of Milot *et al.* was related to the effects of the HS2 deletion. Milot *et al.* found that deleting HS2 had a severe effect on the expression of all genes, with the locus being subject to chromosomal position effects. By contrast, we find that the deletion of HS2 has a more severe effect on expression of ε-globin, with γ- and β-globin

gene expression being only mildly affected, with the exception of ΔHS2 line 5 which shows a severe position effect in the foetal liver and adult stages.

By contrast, our study agrees with that of Milot *et al.* on the severe effects that deleting HS3 exerts on the regulation of the locus. Similar effects were also observed in the context of a YAC by Bungert *et al.* (1995). These observations suggest that the presence of extra flanking sequences cannot provide additional function(s) that can protect the locus from chromosomal position effects or that can partly substitute for HS3 activity. We did not, however, observe any CTPE phenomena in our new ΔHS3 lines. This is probably due to the fact that only two lines were analysed, but could also be because the additional flanking sequence may protect against this type of position effect.

The analysis of the effects of deleting HS2 and HS3 on the regulation of the human β-globin in transgenic mice provides further support for the holocomplex model, with different contributions of each HS site towards full LCR function. Our analysis presented here suggests a functional specialization of HS2 in ε-globin regulation, as well as a more general transcriptional activation function. The effects of deleting of HS3 are in agreement with its suggested role as a key element in the long-range chromatin organization in the globin locus.

## Materials and Methods

### *Human β-globin locus PAC constructs and transgenic mice*

Hypersensitive site two and three of the locus control region of the human β-globin locus were deleted in the PAC 148loxp containing the human β-globin locus (70 kb) plus additional 3' and 5' sequences (115 kb), using the homologous recombination protocol as described by Imam *et al.*, 2000. The homologous recombination for the introduction of the loxP site in the locus has been described in Chapter 4. The deletions span a region of 742 bp for HS2 and 1384 bp for HS3 and were previously described in Milot *et al.* (1996). A 2.8 kb HindIII fragment and a 4.2 kb HindIII fragment containing the HS2 and HS3 core deletions respectively, where cloned in the HindIII site of the pDF25 recombination vector. After recombination, the ΔHS2 and ΔHS3 PACs were checked for integrity by Southern blot using two cosmid probes, one containing the LCRε region and a second containing the γγδβ region (Strouboulis *et al.*, 1992). The 185 kb PAC insert was isolated by Not I digestion and purified from vector sequences by salt gradient centrifugation, essentially as described by Dillon and Grosveld (1993). Briefly, the digested PAC was layered on top of a 5-25% NaCl gradient and centrifuged at 40,000rpm, room temperature for 50 minutes in a SW41 swing-out rotor. 0.5ml fractions were collected and analysed by agarose gel electrophoresis. Fractions containing only the PAC insert were pooled and dialysed against a large volume of TE (10mM Tris-HCl pH 8.0, 1mM EDTA)/0.1M NaCl for 5 h at $4^0$C in UH 100-75 dialysis tubing (Schleicher & Schuell). Dialysis was continued overnight at $4^0$C after replacing the buffer. The PAC insert was concentrated by vacuum dialysis and subsequently dialysed against a large volume of microinjection buffer (10mM Tris-HCl pH 7.4, 0.1mM EDTA) containing 0.1M NaCl in order to protect the high molecular weight PAC insert DNA from shearing during microinjection. The purified PAC fragment was checked for DNA integrity and concentration by pulsed field gel electrophoresis in a 1% agarose gel in 0.25XTAE buffer using a Biometra RotaphorType V apparatus, under the following conditions: 8-2sec pulse interval logarithmic ramp, 120-110° rotor angle linear ramp, 200-180volt logarithmic ramp, rotor speed 6 at 13°C for 21hours. The purified insert was injected at approximately 0.5ng/μl into the pronucleus of fertilised eggs of FVB/N mice. The injected eggs were transferred into the oviducts of pseudo-pregnant BCBA foster females as previously described (Kollias *et al.*, 1986).

Transgenic founders were identified via Southern blot analysis using as probes the 970 bp BamHI-EcoRI βIVS2 fragment and a 3.3 kb EcoRI fragment containing HS5. After transmission of the transgene to the F1, the β-globin locus was checked for integrity using the LCRε and γγδβ cosmid probes and specific probes within the LCR and the different genes. The genomic DNA was digested with: EcoRI, BamHI, BglII, HindII, HindIII, EcoRV, SacI, PstI, XhoI-KpnI, XhoI-ApaLI and SacII-KpnI. Southern blots were hybridised with LCRε and γγδβ cosmid probes, HS5 (3.3 kb EcoRI fragment), ε (340 bp BamHi-BspMI fragment), 5' $^{A}\gamma$ (1.7 kb EcoRI-BamHI fragment), β intron II (0.9

kb BamHI-EcoRI fragment) and 5'δ (0.98 kb Xba I-Bgl II fragment). Transgene copy numbers were determined using as probes the βIVS2 fragment and a 0.9 kb Pvu I fragment from the endogenous mouse carbonic anhydrase II (CA-II) gene. The ratios of intensities of the βIVS2 /CA-II bands obtained for the PAC transgenics were compared to those obtained for the single copy human β-globin locus transgenic lines 2 and 72 (Strouboulis *et al.*, 1992). Analysis was performed by PhosphorImager using ImageQuant software (Molecular Dynamics). Multi copy lines were bred to single copy by crossing them with Zp3-Cre transgenic mice expressing the Cre recombinase during oocyte maturation (Lewandoski *et al.*, 1997). Mice obtained from these crosses were checked for copy numbers as above. In addition, comparison of human β-globin expression levels between the parental multi copy lines and the Cre-recombined mice was also carried out by S1 protection analysis in testing for single copy mice. The single copy lines thus obtained, were also checked for integrity of the human β-globin locus as above.

### *DNA FISH analysis*

Peripheral blood cells were cultured for 72 hours in RPMI 1640 medium. Chromosome preparations were made according to standard procedures. FISH was carried out as described by Mulder *et al.* (1995). The specific probe used was the biotin-labelled human β-globin LCR to detect the transgene and was immunochemically detected with fluorescein. Chromosomal DNA was counter-stained with DAPI, which stains centromeric domains more intensely.

### *RNA fluorescent in situ hybridisations*

Mouse 14.5dpc foetal liver and 10.5dpc embryonic blood cells were isolated, fixed and hybridised essentially as described previously (Wijgerde *et al.*, 1995 and van de Corput and Grosveld, 2001). For detection of each globin gene transcript, a mixture of two, three or four different 50-mer oligodeoxynucleotide probes was used. Each oligo probe was labelled with dinitrophenol (DNP), digoxigenin (dig) or biotin (bio) in the middle and at the 3' and 5' ends. Probes for primary transcript and mRNA *in situ* hybridisations were designed from the intron and exon sequences, respectively, of the different globin genes, such that they are at least 25 nucleotides apart. In mRNA *in situ* hybridisations on 14.5 dpc foetal liver cells we used a combination of mouse $\beta_{maj}$-DNP and human β-bio and a combination of $\beta_{maj}$-DNP and human γ-bio. In the primary transcript *in situ* hybridisations on 14.5 dpc foetal liver cells we used a combination of mouse α-dig and human β-bio. In the primary transcript *in situ* hybridisations on 10.5 dpc embryonic blood we used a combination of mouse α-dig and human γ-bio, a combination of mouse α-dig and human ε-bio. Overnight incubation at 37°C of fixed cells with the oligodeoxynucleotide probe mixtures was followed by detection of the DNP, dig and bio labels with specific antibodies (van de Corput and Grosveld, 2001). Antibody detection was done using a five-layer avidin texas red tree (avidin texas red, goat-α-avidin, avidin texas red, goat-α-avidin and avidin texas red) for bio, a three-layer FITC (green) tree (sheep-α-dig, rabbit-α-sheep fitc, goat-α-rabbit fitc) for dig and a three-layer FITC (green) tree (rat-α-DNP, rabbit-α-rat fitc, goat-α-rabbit fitc) for DNP. Fluorescence was detected by epifluorescence/CCD.

### *S1 nuclease protection assays*

S1 nuclease protection analysis was carried out with total RNA from 10.5dpc embryonic yolk sac, 14.5dpc foetal livers and blood from adult animals. RNA was isolated using the Trizol reagent according to the manufacturer's instructions (Life Technologies). Conditions for S1 nuclease protection assays and polyacrylamide gel electrophoresis were essentially as previously described (Fraser *et al.*, 1990; Kollias *et al.*, 1986; Strouboulis *et al.*, 1992; Milot *et al.,* 1996 and Chapter 4 of this thesis). Specific activities of probes were determined as previously described (Lindenbaum and Grosveld, 1990) and are indicated in the Figure legends. Quantitation was done on a PhosphorImager using the ImageQuant software (Molecular Dynamics).

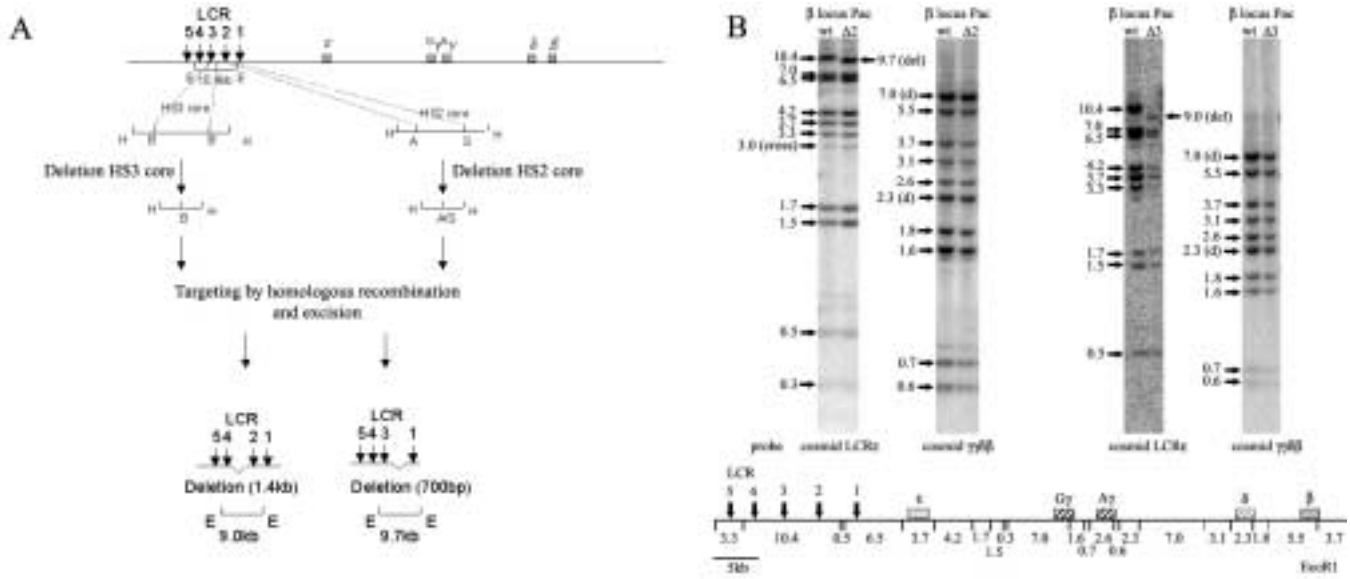



***Figure 1****: Construction of the pac148loxp hypersensitive site deletion constructs*—(A): On top a representation of the PAC-insert containing the human β-globin locus. The hypersensitive sites of the LCR are indicated with five black arrows and small boxes represent the globin genes. Homologous recombination followed by temperature sensitive excision (Imam *et al.*, 2000) resulted in deletion of the HS3 and HS2 core regions. (EcoRI fragments as indicated below the LCR). E: EcoRI; B: BamHI; H: HindIII; A: ApaLI and S: SnaBI (B): Mapping of integrity of the human β-locus in the PAC constructs following recombination in E. coli. Southern blots of wt and HS2/3 deleted PAC DNA digested with EcoRI. Bands specific for the human β-locus are detected with LCRε and γγδβ cosmid probes. Arrows indicate the bands present in the EcoRI map of the human β-locus, as shown at the bottom. The diagnostic 9.7 kb fragment arising from the HS2 deletion and the 9.0 kb fragment of the HS3 deletion are indicated to the right of the LCRε panels. Cross: mouse cross-hybridising fragment often observed with the LCRε probe; d: fragments appearing as doublets; del: fragments arising from the HS2 or HS3 deletions.

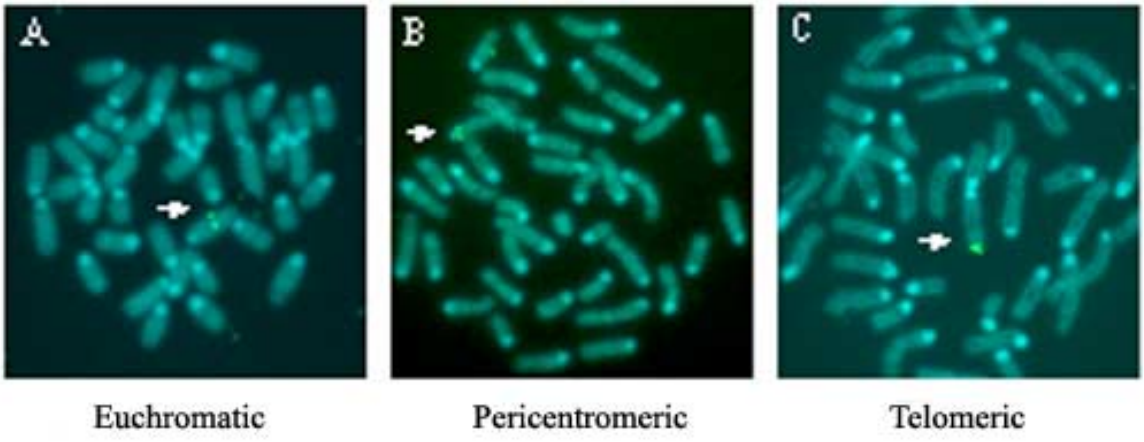



***Figure 2****: FISH analysis of metaphase spreads from circulating lymphocytes in the adult blood of transgenic lines*—The human β-globin locus is detected with a dig labelled plasmid containing the LCR (green spot). The chromosomes are visualised with DAPI staining. Arrows indicate site of integration on the chromosome. Three examples are shown. A: euchromatic integration, B: (peri) centromeric integration, C: telomeric integration. All three types were observed in both wild type and deletion lines.

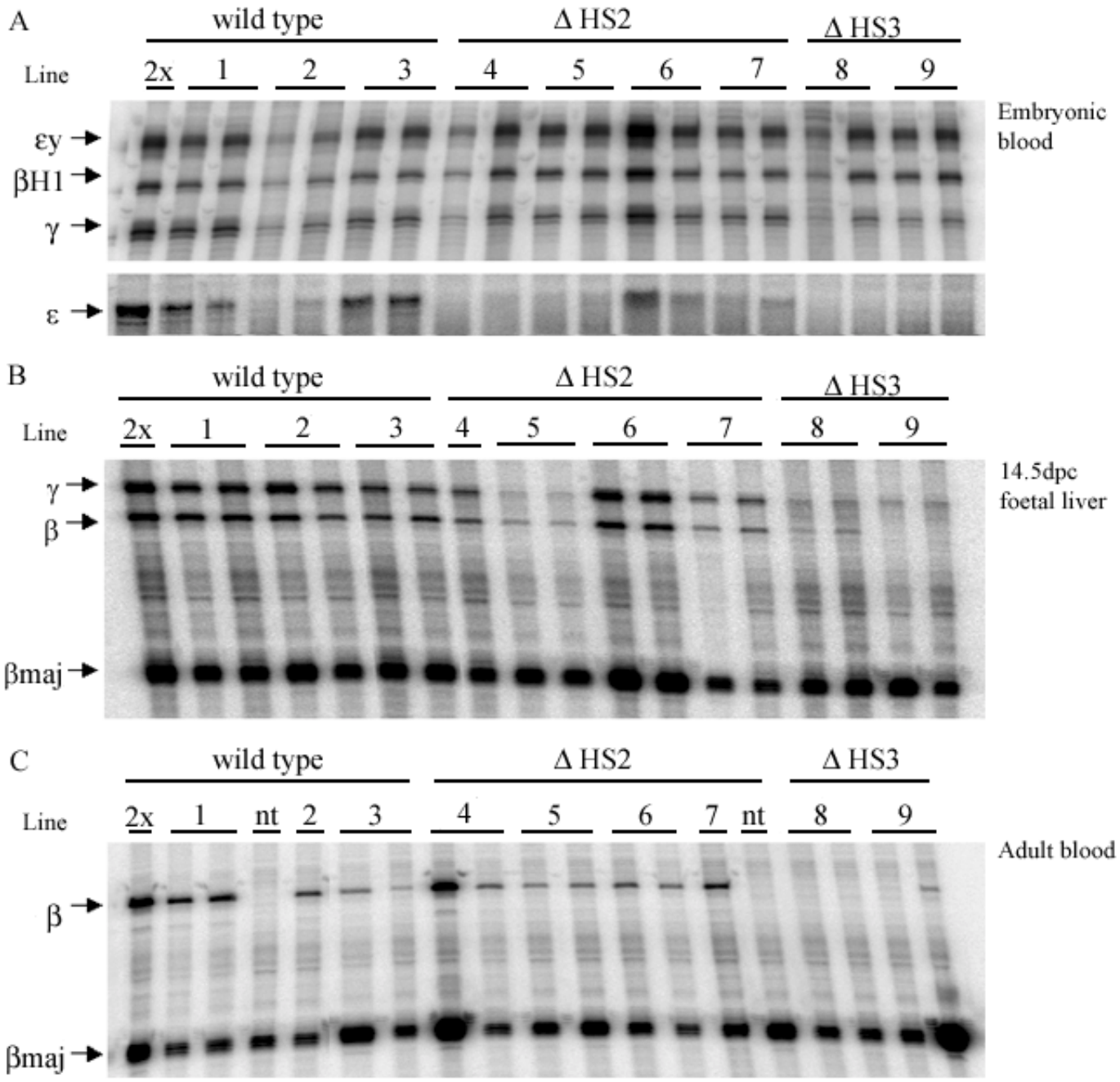



***Figure 3****: S1 nuclease protection assays*—Panel A: S1 nuclease protection assay on RNA isolated from 10.5dpc yolk sacs of all single copy transgenic lines. Two littermates for each line are shown as well as a 2xRNA control from wild type line 1 for probe excess. The samples were assayed with radioactively labeled probes, for mouse εy and βH1 and for human ε and γ. The corresponding protected fragments are indicated with arrows. The lower panel shows a longer exposure of the human ε signal. The relative ratios of specific activities of εy:βh1:ε:γ probes were 5.2:1.9:1:2.7. Panels B and C: S1 nuclease protection assay on RNA isolated from 14.5 dpc foetal liver (B) and adult blood (C) of all single copy transgenic lines. Two littermates of each line were also tested, except for adult blood of lines 2 and 7. A 2x RNA control for probe excess is also shown. All samples were assayed with the same probe mix of radioactively labeled probes for mouse $\beta_{major}$ and for human γ and β, with specific activitiy ratios of 3:5:1, respectively. Corresponding protected fragments are indicated with arrows. nt: non-transgenic controls.

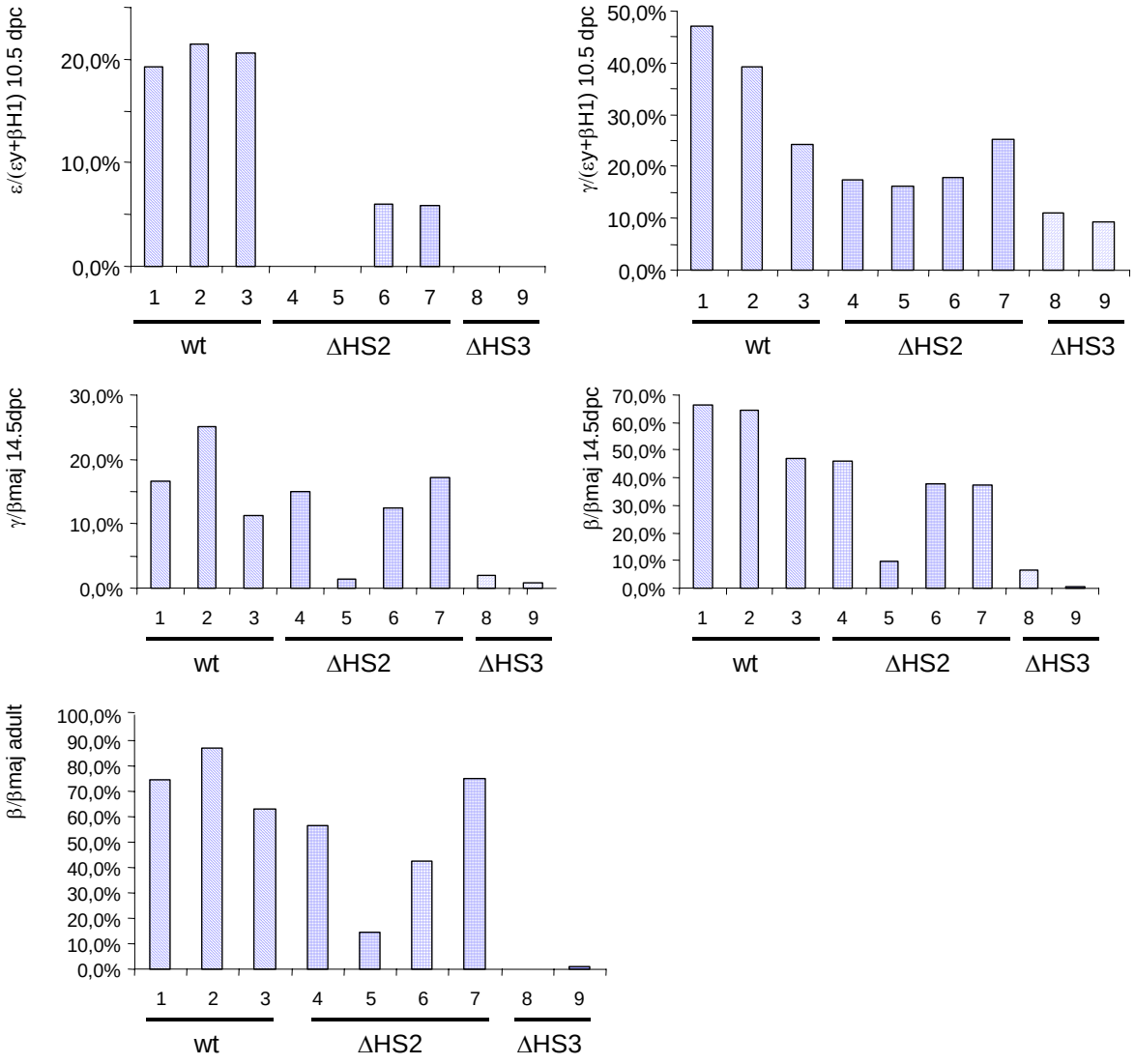



***Figure 4****: Bar chart representation of the human globin expression levels at all developmental time points assayed as a percentage of the total mouse globin expression per transgenic line*—Line 1, 2 and 3 are wild type control lines, line 4, 5, 6 and 7 the HS2 deletion lines and line 8 and 9 the HS3 deletion lines.

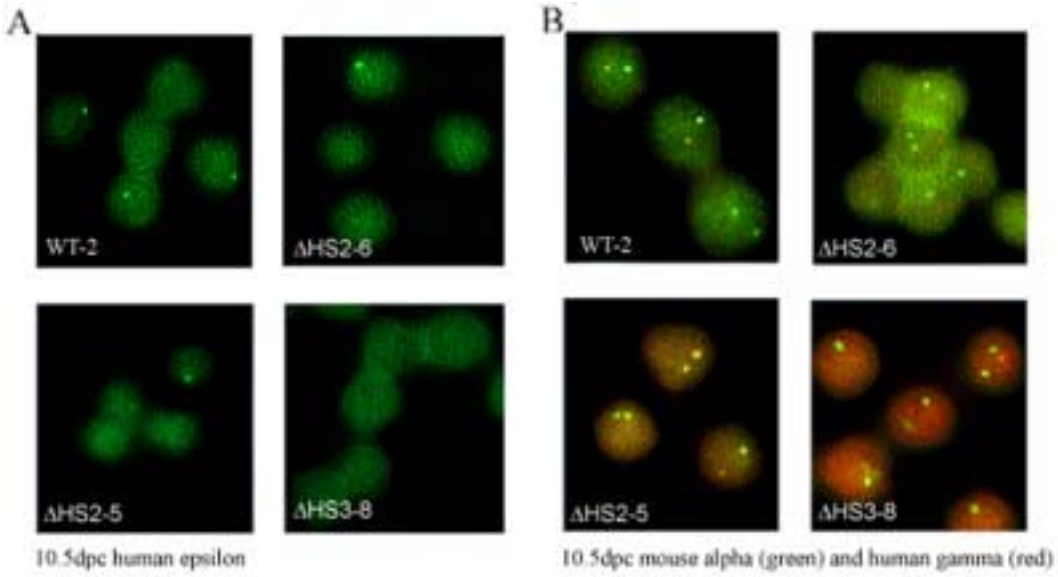



***Figure 5****: In situ hybridisations on 10.5dpc embryonic blood*—A primary transcript *in situ* hybridisation using a human ε-globin intron specific probe (green signal) on wild type; ΔHS2 and ΔHS3. The deletion lines show less cells with a human ε signal. B primary transcript *in situ* hybridisation using a mouse α-globin intron specific probe (green signal) and a human γ-globin intron specific probe (red signal) on wild type; ΔHS2 (line 6); ΔHS2 (line 5) and ΔHS3. The wild type and ΔHS2-6 show comparable amounts of cells with human γ signal, ΔHS2-5 and ΔHS3-8 show clearly less cells positive for human γ.

***Table 1****: Overview of transgenic mice*—In the table are listed; wild type line or deletion line, the initial copy number of the line, whether the line could be bred to single copy, the site of integration of the transgene and which line has been used in the analysis.

| Trangenic line | copy nr | bred to single copy | integration site | used for analysis |
|---|---|---|---|---|
| wt β locus line 1 | 2 | no | euchromatic | no |
| wt β locus line 2 | 3 | no | telomeric | no |
| wt β locus line 3 | 3 | no | euchromatic | no |
| wt β locus line 4 | 2 | yes | centromeric | yes, line 1 |
| wt β locus line 5 | 2 | yes | subcentromeric | no |
| wt β locus line 6 | 1 | _ | subcentromeric | yes, line 2 |
| wt β locus line 7 | 2 | yes | telomeric | no |
| wt β locus line 8 | 2 | yes | euchromatic | yes, line 3 |
| ΔHS2 line 1 | 1 | _ | centromeric | yes, line 4 |
| ΔHS2 line 2 | 4 | yes | euchromatic | yes, line 5 |
| ΔHS2 line 3 | 2 | yes | telomeric | yes, line 6 |
| ΔHS2 line 4 | 1 | _ | telomeric | yes, line 7 |
| ΔHS3 line 1 | 3 | no | subtelomeric | no |
| ΔHS3 line 2 | 2 | no | centromeric | no |
| ΔHS3 line 3 | 2 | yes | euchromatic | yes, line 8 |
| ΔHS3 line 4 | 1 | _ | subcentromeric | yes, line 9 |

***Table 2**: primary transcript data 10.5dpc transgenic embryos*—The percentages of cells transcribing human ε-globin (n=200) and of cells transcribing both mouse α-and human γ-globin (n=200) are listed per transgenic line.

| | **primary transcript 10.5dpc** | |
|---|---|---|
| transgenic line | cells expressing ε | cells expressing α and γ |
| wt β locus line1 | 17.5% | 70% |
| wt β locus line2 | 13.5% | 65.5% |
| wt β locus line3 | 16.6% | 58.5% |
| ΔHS2 line 4 | 4.9% | 66% |
| ΔHS2 line 5 | 6.1% | 49.5% |
| ΔHS2 line 6 | 9.8% | 62% |
| ΔHS2 line 7 | 9.7% | 58% |
| ΔHS3 line 8 | 0% | 36% |
| ΔHS3 line 9 | 0.8% | 21% |

***Table 3**: primary transcript data 14.5dpc transgenic embryos*—The percentages of cells transcribing both mouse $\beta_{maj}$-and human γ-globin (n=300) and of cells transcribing both mouse $\beta_{maj}$-and human β-globin (n=300) are listed per transgenic line.

| | **pimary transcript 14.5 dpc** | |
|---|---|---|
| transgenic line | cells expressing γ and $\beta_{maj}$ | cells expressing β and $\beta_{maj}$ |
| wt β locus line1 | 16.6% | 67.3% |
| wt β locus line2 | 25.1% | 52% |
| wt β locus line3 | 11.3% | 54% |
| ΔHS2 line 4 | 15% | 49.5% |
| ΔHS2 line 5 | 1.4% | 17.5% |
| ΔHS2 line 6 | 12.5% | 49.3% |
| ΔHS2 line 7 | 17.2% | 54.5% |
| ΔHS3 line 8 | 1.9% | 3.5% |
| ΔHS3 line 9 | 0.9% | 5% |

# *CHAPTER 6*
# *GENERAL DISCUSSION*

## Discussion

Globin genes are arranged in the order in which they are expressed during development i.e. 5'-ε-$^{G}\gamma^{A}\gamma$-δ-β-3'. Switching in expression between the different genes occurs during development, i.e. ε→γ in the embryo and γ→β at the time of birth. The regulation of the human β-globin locus has been the subject of intense investigations in the last decades. These have led to the identification of important regulatory elements, protein factors and mechanisms and given insight into the regulation, activation and expression of the human β-globin genes. Regulatory elements upstream of the gene locus, such as the LCR, as well as elements proximal to the genes, such as promoter, silencer and enhancer sequences, are required for the correct developmental expression of the globin genes (Stamatoyannopoulos and Grosveld, 2000).

In the 1980s, a number of studies using transgenic mice with human β-globin genes, showed very low, though developmentally regulated, levels of expression (Townes *et al.*, 1985 and Kollias *et al.*, 1986). Studies on deletions giving rise to Dutch and Hispanic thalassemias (van der Ploeg *et al.*, 1980 and Driscoll *et al.*, 1989), resulted in the identification of the Locus Control Region (LCR). The LCR consists of five DNase I hypersensitive sites that have since been shown to be important for copy-number dependent, integration-site independent and tissue-specific expression of globin genes in transgenic mouse assays (Grosveld *et al.*, 1987).

The LCR is located upstream of the globin genes, at a distance varying between 5 kb, for the ε-globin gene, and more than 50 kb for the β-globin gene. The LCR therefore exhibits long-range activation properties for the expression of the globin genes. Three models have been proposed as the basis for this long-range activation: an accessibility model, a scanning/tracking model and a looping model (Martin *et al.*, 1996; Tuan *et al.*, 1992; Kong *et al.*, 1997; Stamatoyannopoulos *et al.*, 1991; Epner *et al.*, 1992; Dillon *et al.*, 1993; Grosveld *et al.*, 1993 and Orkin *et al.*, 1990). Which model(s) is actually the correct one remains to be directly demonstrated, though recent data indicate that the looping model appears to be operating *in vivo* in mediating long-range LCR-gene interactions (Tolhuis *et al.* submitted).

A great deal of research has also been done in order to gain a better understanding of the molecular basis of LCR function, the role of each DNase I hypersensitive site, the identification and the role of the proximal regulatory elements of the genes and the interplay of the LCR and these regulatory elements (see introduction for further details and references). Chapters 4 and 5 of this thesis have presented two such studies and the implication of these studies on globin gene regulation will be discussed in this chapter.

To study the elements that play a role in the regulation, activation and expression of the globin genes, it is important to do so in the context of the intact human β-globin locus and to be able to manipulate the locus in its complete context. The use of artificial vectors, like YAC, PAC and BAC, which can carry large inserts, and the use of homologous recombination methods in manipulating these inserts in host yeast and bacterial cells, are very useful tools for these purposes.

### *The method of homologous recombination for manipulating large DNA constructs*

Several methods for homologous recombination have been developed and used in the generation of constructs to study large genes and multiple gene loci (Yang *et al.*, 1997; Zang *et al.*, 1998; Pavan *et al.*, 1990 and Duff and Huxley, 1996). For studies on the human β-globin locus methods for homologous recombination to modify both YAC and PAC/BAC inserts have been described (Narayanan *et al.*, 1999; Peterson *et al.*, 1997 (review) and chapter 3) Here, I will focus on the methods used to modify PAC/BAC since we used a construct based on a PAC vector to manipulate the human β-globin locus (see Chapter 3).

PAC and BAC vectors are artificial chromosomes based on the P1 phage and the *Escherichia* coli F-factor, which is a low copy plasmid. Both types of vectors can be used to clone and maintain large DNA fragments (100-300 kb) in *E. coli* cells. In this way, large gene constructs and gene loci can be isolated and studied in their complete context.

All homologous recombination methods for manipulating BAC/PAC vectors are based on RecA or RecE/RecT mediated recombination. Yang *et al.* 1997 were the first to publish a general, simple method for the manipulation of BAC vectors, which could also be applied to PAC vectors. The method consists of two sequential steps: an integration step of a targeting vector bearing the desired modification and an excision step that leaves behind the desired modification in the insert. A recombination plasmid containing a RecA gene, a $Tet^R$ marker and the DNA sequence homologous to the BAC containing the desired mutation, is introduced into a $RecA^-$ *E. coli* host strain containing the BAC which is to be modified. Upon integration of the recombination plasmid by a single homologous exchange, selection using $Tet^R$ takes place. The clones with an insertion of the recombination plasmid are then subjected to a counter-selection step, also based on $Tet^R$ . This counter-selection results in the loss of the recombination plasmid giving rise to the desired modification staying behind in the BAC/PAC insert in a fraction of the cell clones.

A second method for homologous recombination in BAC/PAC uses the RecE/RecT mediated recombination pathway in *E. coli* host strains (Zhang *et al.*, 1998). This method differs, besides the used recombination pathway, from the other two methods in the usage of a linear DNA fragment, instead of a recombination plasmid. A PCR fragment containing homologous sequences to the wild type BAC or PAC and the desired mutation and a selection marker are transformed into the *E. coli* host strain containing RecE/RecT with the wild type BAC/PAC. After introduction of the PCR fragment, homologous recombination takes place and the BAC/PAC is modified in a single step. However, the homologous recombination also results in the integration of the resistance marker. In most cases this is not desired and the resistance marker has to be removed. This is most often accomplished by the use of FRT sites and Flp-mediated recombination. Therefore, this technique is in essence a two-step recombination method and has been called ET-recombination.

The same group that first described ET-recombination later published a revision of the method (Muyrers *et al.*, 1999 and 2000), in which the addition of a second selection marker, SacB, circumvented the FRT/FLP recombinase step. During the first step of the new protocol, a PCR fragment containing both resistance markers and the homologous sequences to the BAC/PAC insert flanking both sides of the resistance markers, is introduced into the *E. coli* host strain with the wild type BAC/PAC. After selection for the homologous recombination event, a second PCR fragment is introduced, this fragment containing the same homology sequences but also the desired mutation of the BAC/PAC, so that the final modification can take place. The presence of the same homologous sequences in both PCR fragments results in the second recombination event and in the excision of the resistance markers and the insertion of the desired mutation.

The methods described above had been published prior to the completion of the method described in chapter 3. We, however, proceeded with the development of our method for homologous recombination because of a number of disadvantages presented by the other methods. The method of Yang *et al.* is almost identical to ours, however, this method shows very low efficiency of integration and correct recombination events. This is because excision is mediated through the use of $Tet^R$ as counter-selection, which has been reported to work inefficiently (Podolsky *et al.*, 1996). We used the rpsL gene for counter-selection, which mediates either streptomycin resistance or sensitivity depending on the correct homologous recombination event and the number of copies of rpsL genes thus introduced into the host strain. The use of this counter-selection resulted in a much higher percentage of correct recombination events.

The ET-recombination has some advantages but also disadvantages. It has the advantage of using linear PCR fragments together with the standard ET-plasmid and thus there is no need for constructing a new recombination plasmid for each recombination event. Furthermore, when the presence of the resistance marker is not interfering, the method requires just one recombination step and is very fast. However, when loss of the resistance marker is required, the method takes two recombination steps with each step requiring a new (specific) PCR fragment to be transformed into *E. coli* cells containing the BAC/PAC that has to be modified, meaning that electrocompetent cells have to be produced twice in the same experiment. This makes the ET-recombination method more laborious and time-consuming. A second disadvantage is the use of the SacB gene as selection marker. This gene has been shown to be easily mutated, resulting in inefficient counter-selection and, as a consequence, high backgrounds. In our method, which consist of two steps, the recombination vector, including the resistance marker, is lost upon counter selection and no second step of transformation is

necessary, thus making our method less laborious and faster. Furthermore our counter-selection gives rise to lower background.

The choice of a PAC vector, instead of YAC, for the manipulation of the human β-globin locus and the use of homologous recombination method in our studies, is based on the advantages that the PAC system offers compared to YACs. YACs were the first to be used for studying large genes and multiple gene loci in their full context. They have the advantage over conventional plasmids and cosmids of being able to carry very large inserts (up to 1-2 Mb) which can be easily modified because of the high efficiency of endogenous homologous recombination in yeast. However, DNA fragments cloned into YAC vectors show genetic instability and it is difficult to obtain high yields of purified YAC DNA. Finally in transgenesis there is low frequency of YAC integration into the mouse genome, presumably due to shearing during microinjection of the (larger) YAC DNA inserts (Green *et al.*, 1991 and Schedl *et al.*, 1992). PAC and BAC vectors also have a large capacity of insert DNA (up to 200-250 kb) and homologous recombination in *E. coli* can also be used to manipulate the insert DNA. Furthermore PACs/BACs circumvent the disadvantages of the YAC as they can be treated as conventional plasmids in DNA preparations, resulting in high yields of intact purified DNA and exhibit high insert stability.

YAC, BAC, PAC and cosmid-based constructs have been used to study the regulation, activation and expression of the human β-globin gene locus in its complete context in transgenic mice (Peterson *et al.*, 1993; Bungert *et al.*, 1995; Gaensler *et al.*, 1993; Porcu *et al.*, 1997; Narayanan *et al.*, 1999; Kaufman *et al.*, 1999; Strouboulis *et al.*, 1992 and Chapters 4 and 5).

### *Studies on elements required for globin gene expression*

The homologous recombination method in *E. coli* has been used to delete hypersensitive sites 2 and 3 of the LCR and the Enh and F elements located downstream of the $^{A}\gamma$-globin gene, in the context of the complete human β-globin locus. Transgenic mice were generated with these constructs and analysed (see Chapters 4 and 5). The analysis of these mice shows that these deletions have an effect on the regulation and expression of the human β-globin genes.

Deletion of HS2 showed a significant decrease in ε expression in all lines analysed and the occurrence of PEV in the foetal liver and adult blood stages in one line. Expression of γ- and β-globin genes was also affected but to a lesser degree than ε expression. The transgene in the PEV line had integrated in the middle of a chromosome arm and is thus in a domain that is not obviously heterochromatic. Deletion of HS3 resulted in the decrease of expression of all the globin genes, with the decrease becoming worse as development proceeds. All lines analysed for the HS3 deletion showed PEV, with one line having a peri-centromeric integration site and the second line an integration similar to that of the HS2 PEV-line.

The deletion of the Enh and F elements, previously postulated to be involved in the silencing of γ gene expression in the adult stage, did not have an effect on globin gene switching and expression patterns in the foetal and adult stages. However, at the embryonic stage an elevation of both ε and γ expression levels was observed in the absence of these elements.

These results again demonstrate the interplay of multiple regulatory elements in the correct developmental regulation of expression of globin genes. The results from the hypersensitive site deletion studies show that the LCR is not acting just as a holocomplex (Ellis *et al.*, 1993; Fraser *et al.*, 1993 and Dillon *et al.*, 1993) and that its function seems to be more complex affecting differently individual globin genes. The deletions of the Enh and F elements have indicated another example of a new class of developmental stage-specific regulatory elements affecting expression of all globin genes in the locus at that particular developmental stage (see also Calzolari *et al.*, 1999 and Liu *et al.*, 1998).

Although the LCR holocomplex is necessary for the position independent, high-level expression of all the human β-globin genes (Fraser *et al.*, 1993 and Milot *et al.*, 1996), the interaction of specific hypersensitive sites with the individual globin gene promoters seems to be important for the correct gene expression during development (Navas *et al.*, 1998; Navas *et al.*, 2002; Bungert *et al.*, 1998; Peterson *et al.*, 1996 and Gui and Dean, 2001). The studies described in Chapter 5 indicate that the deletion of a hypersensitive site can have both a general effect, as in the PEV observed in the expression of all globin genes in the HS3 deletion, and/or a gene-specific effect on globin gene

transcription, as is observed in the decrease or even complete absence of ε-globin expression in all the HS2 deletion lines.

A number of other studies have also reported gene-specific effects caused by the deletion of either HS2 or HS3 (Navas *et al.*, 1998; Navas *et al.*, 2002; Bungert *et al.*, 1998 and Peterson *et al.*, 1996). HS3 has been shown in two different studies to be important for γ expression in the foetal stage (Navas *et al.*, 1998 and Bungert *et al.*, 1999). The deletion of HS3 in the context of a 248 kb human β-globin YAC construct in transgenic mice showed a decrease in γ expression levels in the foetal, but not in the embryonic stage (Navas *et al.*, 1998). The same group recently reported that the expression of the γ gene in adult transgenic mice carrying the –117 HPFH mutation (Berry *et al.*, 1992) was abolished when they deleted HS3 in the context of the 248 kb YAC construct (Navas *et al.*, 2002). Finally, a comparison between a transgenic mouse line with a deletion of HS2 and a transgenic mouse line in which HS2 was replaced by HS3 (Bungert *et al.*, 1999) showed that while in the HS2 deletion line expression of all the globin genes was affected, in the HS3-substitution line expression of all the genes, except γ, remained affected. Expression of the γ genes was partially rescued by the replacement. The HS3 for HS2 replacement therefore partially rescued γ-globin gene expression suggesting an interaction between HS3 and γ-globin potentially effected by the extra copy of HS3 in the LCR in this construct.

Questions arise related to the differences on globin gene expression observed upon deletion of HS2 between the Bungert *et al*., 1995 study and our study. These probably can be explained by differences in the sizes of the HS2 core deletions used to make the deletion constructs (for further discussion on the differences see discussion in Chapter 5).

The studies presented give some indication that a specific interaction of HS3 is important for human γ-globin gene expression at the foetal stage. The reason why our deletion of HS3 did not show such a specific effect is probably because of the position effect variegation and the subsequent decrease in expression of all the globin genes. The PEV probably masks the effect of HS3 deletion on foetal γ gene expression and the specific effect of HS3 on the foetal γ expression is therefore not observed in our studies. This indicates that an effect on the general function of the LCR as a holocomplex overrules the effects of the individual hypersensitive sites. This however, does not appear to be the case with the HS2 deletion since, in the apparent absence of PEV in the embryonic yolk sac, we are able to see a specific effect on ε-globin gene expression, as well as a more generalised reduction in expression of all globin genes in all developmental stages. HS3 has also been suggested to be the hypersensitive site necessary for ε-globin gene expression, since deletion of HS3 in the context of a complete β-globin locus resulted in a specific decrease of ε expression in three transgenic studies (Milot *et al*, 1996; Navas *et al.*, 1998 and Peterson *et al.*, 1996). This reduction in ε expression, however, could also be due to the overall reduction of the globin gene expression observed in HS3 deletion lines. Since ε-globin is already expressed at low levels in the wild type transgenic lines compared to γ-globin, a general reduction of globin expression can already make ε expression difficult to detect. Indeed, recent evidence indicates that it is HS2 that is the prime hypersensitive site necessary for ε-globin expression (Gui and Dean, 2001). Our studies clearly show a decrease and even an absence of ε expression in all the HS2 deletion lines, whereas the other globin genes are hardly affected, except for the one line showing PEV in the foetal liver and adult stages. The study in which HS2 is replaced by HS3 in transgenic mice, shows an absence of ε expression, thus suggesting that HS3 cannot functionally replace HS2 for ε-globin expression. At the same time the hypersensitive site replacement could partly overcome the decrease of the γ expression, indicating that the HS3 used to replace HS2 is functional (Bungert *et al.*, 1999). Furthermore, a specific function in histone acetylation has been recently described for the core of HS2 (Gui and Dean, 2001). This function has been proposed to be necessary for normal levels of ε expression (Gui and Dean, 2001).

Some of the hypersensitive sites within the LCR holocomplex thus seem to have different specialised affinities for different globin gene promoters. This would suggest that the LCR has to change its conformation during development, such that the right hypersensitive site can interact with the globin gene promoter that it has a higher affinity for, at the right time during development. Such a conformational change hypothesis has already been put forward by Navas *et al.*, 1998 to explain why a

deletion of HS3 had an effect on γ expression during the foetal stage, but not during the embryonic stage.

The suggestion that hypersensitive sites have different affinities for the individual β-globin gene promoters could be addressed by comparing DNase I sensitivity and histone acetylation status in the promoter regions of individual genes, for example in comparing ε-and γ-globin promoters in the HS2-versus the HS3 deleted mice. Using ChIP assays, it has already been shown that peaks of histone hyperacetylation can be observed in activated globin promoters and the LCR (review Bulger *et al.*, 2002). Furthermore, it has been shown that the deletion of the LCR resulted in the loss of peak histone hyperacetylation on the β-globin promoter (Reik *et al.*, 1998). This assay can thus be used to correlate hypersensitive site activity and gene activity or the absence of gene activity with the deletion of specific hypersensitive sites.

The deletion studies on the Enh-F elements show that, besides the interaction of the LCR and the hypersensitive sites with the individual globin gene promoters, the presence of specific gene-proximal regulatory elements is also necessary for normal globin gene expression levels. Enh and F elements are two of four elements (Enh, F, O and P) identified in the $^{A}\gamma$-δ intergenic region. This region has been implicated as being important for the γ→β switch, since naturally-occurring deletions mapping within this region in HPFH and δβ-thalassemias, give rise to elevated γ-globin expression levels in the adult (Wood, 1993).

As described in Chapter 4, deletion of the Enh and F elements resulted in an increase in ε and γ gene expression in the embryonic stage, indicating that in the wild type situation these elements have a regulatory function in repressing the embryonic genes during development. However, no effect was observed on the γ→β switch, and no expression of the γ genes was detected in the adult. This argues that the elements do not have the previously suggested silencing effect on γ gene expression in the foetal to adult switch. Other studies, in which the role of the $^{A}\gamma$-δ region in the γ→β switch has been investigated (Liu *et al.*, 1998; Zhang *et al.*, 1997 and Calzolari *et al.*, 1999) also came to the same conclusion, in that this region is not involved in the γ→β switch and that no elevation of γ gene expression in the adult was seen when this region was deleted. For example, two studies that deleted either the Enh element alone, or a 12.5 kb fragment from the $^{A}\gamma$-δ region that included the Enh and F elements as well as other putative regulatory elements reported to map within this region, did not show any observable effect on globin gene expression (Liu *et al.*, 1998 and Zhang *et al.*, 1997). By contrast, the work described in Chapter 4 reports an effect on levels of embryonic ε and γ gene expression upon Enh and F deletion. Another study in which the O and P elements, located further downstream from the Enh and F elements in the $^{A}\gamma$-δ intergenic region, were deleted showed a decrease in β expression as well as PEV (Calzolari *et al.*, 1999).

These two studies indicate, that the Enh-F and O-P as pairs of elements have regulatory functions in the expression of the β-globin genes. The data from these studies are in contrast with the observation that the deletion of all the elements from the locus does not have an effect on globin gene expression (Zang *et al.*, 1997). The latter study, however, has only been published as a meeting abstract with no real data published. Comparison of the studies on the deletion of just the Enh element and our studies on the deletion of Enh and F suggests that the intervening sequences between Enh and F and/or the F element itself are the sequences which affect the ε and γ expression in the embryo.

Although the $^{A}\gamma$-δ region does not seem to be important for the γ→β switch, the studies by Calzolari *et al*. and the work presented in Chapter 4 suggest that elements that may be important for the transcriptional regulation of globin genes at specific stages of development map within this region.

Other developmental-stage specific elements globally affecting globin gene expression, include the 5' ε-globin silencer. This region 5' of the ε gene may be quite complex in its function, since its deletion resulted in a decrease of both ε and γ expression in the embryo (Liu *et al.*, 1997), whereas mutation studies on transcription factor binding sites within the silencer affected ε silencing, however, the effect these mutations had on γ expression was not determined (Raich *et al.*, 1995).

It has become clear from these and other studies, that for the regulated developmental expression of the globin locus, the coordinate action of many elements is required, including the LCR holocomplex, as well as the action of individual hypersensitive sites and gene-proximal regulatory elements. β-

globin gene regulation is complex and still many studies need to be done to be able to understand the regulation of the locus completely.

### *Stochastic gene activation; general or not?*

The work presented in Chapter 2 on the basis of transcriptional activation of the murine globin gene loci, showed that a large percentage of cells (23%) had an imbalance in the α versus β nuclear transcription patterns, which was maintained in the cytoplasmic mRNA levels. These data indicate a stochastic basis of transcription for the murine globin loci. Furthermore, we observed that the α-locus is activated prior to the β-locus during erythroid cell differentiation and that the transcription patterns of the globin loci, once activated, become fixed after two or three cell cycles.

One question that arises from these studies is at what stage is the stochastic decision taken for a locus to transcribe. A simple scenario would be that the decision is taken at the actual transcriptional activation step, or at a step before that e.g. in the opening up of the chromatin domain. For a triple copy human β-globin PAC transgene, as described in chapter 2 of this thesis, we showed that the three copies of the human β-globin are expressed in an all-or-none fashion, suggesting that the decision is taken at a step prior to transcriptional activation itself.

Examples of decisions being taken either before or at the stages of transcriptional activation have come from previous studies on mono-allelically expressed genes, like the Il-4 and the Ly49 gene loci (Agarwal and Rao, 1998 and Tanamachi *et al.*, 2001).

In the case of a decision being taken after the chromatin opening level, the chromatin structure in which a gene resides becomes accessible prior to the start of transcription. In this case, the number of accessible alleles that will be expressed in the end will also depend on the availability of transcription factors, which may be limiting. For example, for the Il-4 gene it has been shown that only cells competent to express Il-4 show an open chromatin conformation at the Il-4 locus. Although all these cells are all able to express Il-4, as has been shown in re-stimulation assays, only a small fraction of accessible Il-4 alleles indeed express Il-4, probably because of transcription factor limitations, resulting in mono-allelic expression (Agarwal and Rao, 1998 and Hu-Li *et al.*, 2001).

If activation of a gene is determined at the chromatin opening level, this may also be reflected upon linked genes present in the same genecluster. Those will be co-expressed and will all show the same stochastic activation pattern. This has also been observed for the Il-4 gene which is linked to the Il-13 and Il-5 genes. Il-13 and Il-5 were indeed found to be co-expressed from the same allele as Il-4 (Agarwal and Rao, 1998 and Hu-Li *et al.*, 2001).

Stochastic decisions at the level of actual transcriptional activation would predict that linked genes in the same locus will be expressed independently to each other, resulting in all kinds of allelic expression patterns. Such a stochastic activation pattern has been observed for the Ly49 locus, in which the expression of linked genes was not coordinated and many combinations of allelic expression patterns were observed (Tanamachi *et al.*, 2001).

At which level stochastic decisions for the activation (or not) of the murine α- and the β-loci are made, has not yet been addressed, but can be speculated upon. It is clear that chromatin plays a role in the stochastic activation pattern of the globin gene loci. This is reflected in the calculated probabilities for expressing either an α- or a β-allele (93% and 88%, respectively) and in the activation of α- globin prior to β-globin in the total population of erythroid cells. The globin loci reside in different chromatin environments. The α-locus is found in a constitutively open chromatin domain rich in housekeeping genes, whereas the β-locus is in a closed chromatin domain in non-erythroid cells. Because of the open chromatin configuration the α-locus is probably more accessible to transcription factors, resulting in a lower threshold for transcriptional activation and thus a higher activation probability.

A question arises therefore, as to what is the level at which the decision for transcription (or not) is taken for the α-globin locus, since it does not appear to be at the chromatin opening step. It could be investigated whether the decision is taken at the level of transcription by checking, for example, whether the linked housekeeping genes are also co-expressed in a locus expressing α-globin, or whether they remain silent in a locus not expressing α-globin.

The stochastic decision for β-globin activation is most likely to be at the chromatin level. This assumption is based on the experiments with the triple copy human β-globin PAC transgene. Since the

mouse β-globin locus resembles the human β-globin locus closely it is likely that the determinants of the stochastic decision will be the same.

A recalculation of the activation probabilities of α- and β-globin, leaving out the early expressing α-cells, shows that once the β-globin locus is accessible for transcriptional activation both loci have a more or less equal activation probability (93% and 94%, respectively). This shows that the difference in chromatin status of the both loci has a clear effect on the activation probability for globin transcription in the total cell population.

Chromatin also seems to play a role in the fixation of stochastic choices. As described in Chapter 2, the decision to express a triple copy human β-globin PAC transgene is taken prior to the actual transcriptional activation step. The all-or-none decision in transcriptional activation observed with the triple copy human β-globin locus transgene is also observed in the expression of genes influenced by PEV, which also has a stochastic basis (Lock *et al.*, 1988).

A stochastic basis for gene activation has been reported before, but only for genes involved in very specific processes like lineage commitment, differentiation decisions and for genes involved in responses to specific stimuli. The studies on the mouse globin loci, however, indicate that the mechanism of stochastic gene activation could be a more general phenomenon for gene activation than previously assumed, since the globin genes are expressed in a cell type that is fully committed and the genes are not involved in specific cellular processes. The stochastic activation of the globin alleles is even a disadvantage for the erythroid cells, since cells with an imbalance in globin chain synthesis have a shorter life time (Weatherhall, 2001). It is therefore suggested from our data that stochastic gene activation is a general phenomenon and is not restricted to the activation of genes involved in specific cellular decisions.

After publication of our work on stochastic globin gene activation, two more studies (Ozbudak *et al.*, 2002 and Elowitz *et al.*, 2002) were published addressing the question whether stochastic gene activation has a general role in gene expression. Both groups address the question whether observed variations of gene expression levels between cells in a clonal cell population were caused by stochastic gene expression. This was determined by measuring "noise" in gene expression in bacterial cells using GFP-reporter genes under the control of promoters regulated by the Lac repressor. Elowitz *et al.*, made use of a bacterial system in which they expressed two identical copies of the same gene in the same cell, thus in the same intracellular environment. This way two types of "noise" could be determined, intrinsic and extrinsic noise. Intrinsic noise was defined as the difference in gene expression between the two identical genes under the exact same conditions, and extrinsic noise as the cell-to-cell variation in expression levels of each of the reporter genes. Obzudak *et al.*, used only one gene under control of a GFP reporter and thus could only measure total phenotypic noise. Both groups indeed observe noise in gene expression and changes in noise under different cellular conditions. They both conclude that stochasticity plays a role in giving rise to overall variations in gene expression between genotypically identical cells.

If gene activation has indeed a general stochastic nature, this will have wide consequences. For example the observations that genotypically identical bacterial cells have phenotypic differences (Elowitz *et al.*, 2002 and Obzudak *et al.*, 2002), because of stochastic differences in gene activation, could also hold true for eukaryotic cells. This could lead to diversity allowing evolution and selection. Furthermore, stochastic gene activation may also allow the fine-tuning of expression levels of gene cascades in response to changes in conditions. However, the stochastic basis of gene expression as a general phenomenon also implies that, sometimes, important genes will not be expressed, or genes that should not be activated will be expressed. To circumvent these problems nature must have evolved back-up systems. For instance, if a cell fails to express an essential gene, this cell will be arrested during the cell cycle at a cell cycle check-point and undergo apoptosis. The opposite, when expression is not desired, can be solved by the evolution of gene activation as a multi-step and multi-level process. Thus the existence of stochastic gene activation as a general principle, in combination with the required back-up systems, ensures that there is order in gene expression, but also that the necessary diversity within a population of cells or even within an organism can be generated and maintained.

# *SUMMARIES*

## Summary

DNA contains all genetic information which makes people unique individuals. Each cell in the human body contains 23 chromosomes which together contain about 40,000 genes. All these genes encode for different proteins which are important for the biological processes in the body. Many genes are expressed in all the cells of the human body, whereas a smaller number of genes is expressed only in specific cell types. One such cell type is the red blood cell or erythrocyte. The erythrocyte originates from a so called pluripotent cell, the HSC. This cell undergoes several steps of differentiation in a process termed erythropoiesis, with the resulting erythrocyte being committed to the production and accumulation of haemoglobin, the oxygen carrier in the blood. Haemoglobin contains two α-globin chains and two β-globin chains, which are encoded for independently by the α-globin genecluster and the β-globin genecluster.

The human β-globin locus is composed of five developmentally regulated genes, which are arranged in the order of developmental expression, and a LCR, containing five DNase I hypersensitive sites, upstream of the genes, which plays an important role in de regulation of the expression of the globin genes and is required for high levels of gene expression.

The work presented in this thesis has been primarily concerned with the regulatory elements present in the human β-globin locus, but also an important part of the work has addressed the basis of the transcriptional activation of the murine globin gene loci. We have shown that the activation of the murine α- and β-globin genes takes place in a stochastic fashion and we have provided evidence that the decision to activate transcription is made at a step prior to the actual transcriptional activation, probably at the level of chromatin organisation. Furthermore, these studies have shown that the activation of the α-locus takes place prior to that of the β-locus during erythroid differentiation.

The studies on the regulatory elements of the locus have concentrated on the hypersensitive sites of the LCR and the Enh and F putative silencers downstream of the $^{A}\gamma$-globin gene. We have deleted both HS2 and HS3 and the Enh and F elements via homologous recombination in a PAC vector, which contains the entire human β-globin locus as a 180 kb insert. Our studies show that each element has an individual role, but also complex interactions between the different elements are required for globin gene activation and developmental regulation.

HS2 and HS3 were deleted to investigate the role of LCR function in chromatin organisation and gene activation. We show that the deletion of HS2 has little effect on the LCR chromatin opening properties, since in only one out of four transgenic lines with this site deleted PEV is observed. We do, however, observe a severe effect on ε-globin expression, indicating that HS2 may play a role in the ε-globin gene activation in the embryo. The HS3 deletion shows a severe impairment of gene expression, which increases with development. This appeared to be caused by PEV expression of the transgenic locus, indicating that in contrast to HS2, HS3 does seem to play a role in the chromatin opening activity of the LCR. A direct effect on a specific globin gene was not observed with this deletion.

The Enh and F elements were deleted in order to assess their role in the developmental regulation of γ-globin gene expression. These elements reside in the $^{A}\gamma$-δ intergenic region, which has been suggested to play a role in the γ→β switch and the regulation of γ gene silencing in the adult, since deletions within this region are associated with the hereditary persistence of foetal haemoglobin (HPFH). The deletion of the Enh and F elements, however, did not affect the silencing of the γ genes in the adult, but instead resulted in an increase in the levels of ε- and γ-globin expression in the embryonic stage.

In conclusion, the studies presented here show:

1. that there is evidence for a stochastic basis for activation of globin gene expression;
2. that the LCR does function as a holocomplex, but also that the individual hypersensitive sites act differently on each globin gene;
3. that the regulation of the individual globin genes is very complex and probably involves the interaction of various combinations of elements during different stages of development.

## Samenvatting

DNA bevat de genetische informatie die mensen tot unieke individuen maakt. Elke cel in een menselijk lichaam telt 23 chromosomen, die bij elkaar zo'n 40.000 genen bezitten. Elk van deze genen codeert een ander eiwit, dat van belang is voor de biologische processen in het lichaam. Het merendeel van de genen komt tot expressie in alle cellen van het menselijk lichaam, een minderheid echter uitsluitend in een specifiek celtype. Een voorbeeld van een dergelijk celtype is de rodebloed cel of erytrocyt. De erytrocyt stamt af van een zogenoemde pluripotente cel, de HSC. Zo'n cel maakt een ontwikkeling in stappen door, die erytropoëse wordt genoemd. De uiteindelijke erytrocyt is verantwoordelijk voor de productie en opslag van het hemoglobine-eiwit, dat zorgt voor het zuurstoftransport in het bloed. Het hemoglobine-eitwit bestaat uit twee α-globineketens en twee β-globineketens, die onafhankelijk van elkaar gecodeerd worden door de α- en de β- globinegenclusters.

Het humane β-globinecluster bestaat uit vijf genen, gerangschikt naar de volgorde van expressie gedurende de ontwikkeling, en een *locus control region* (LCR), gelegen vóór de genen, die vijf gebieden bevat, die overgevoelig zijn voor Dnase I. De LCR speelt een belangrijke rol in de regulatie van de expressie van de globinegenen en is noodzakelijk voor hoge genexpressieniveaus.

Het werk dat in dit proefschrift wordt gepresenteerd, heeft zich voornamelijk gericht op de regulatieve elementen in het humane β-globinecluster. Daarnaast is veel aandacht besteed aan de manier waarop muizenglobineclusters worden geactiveerd. We hebben aangetoond, dat de α- en β-globinegenen van de muis op een stochastische wijze worden geactiveerd en dat de beslissing om een gen te activeren wordt genomen vóór het begin van de daadwerkelijke transcriptie. We hebben aanwijzingen, dat deze beslissing wordt genomen op het niveau van de chromatine-organisatie. Daarnaast heeft het onderzoek uitgewezen, dat het α- gencluster wordt geactiveerd vóór het β-gencluster.

Met betrekking tot de regulatieve elementen van het humane β-globinecluster heeft het onderzoek zich geconcentreerd op de overgevoelige gebieden van de LCR en de vermoedelijke onderdrukkers Enh en F, gelegen achter het $^{A}\gamma$ gen. We hebben zowel HS2, HS3 als de Enh- en F-elementen door middel van homologe recombinatie in een PAC-vector verwijderd. Deze PAC-vector bevat het hele humane β-globinecluster als een 180 kb-insertie. Ons onderzoek toont aan, dat elk bestudeerd element een eigen rol heeft, maar dat de interactie tussen de diverse elementen ook noodzakelijk is voor de activering van de globinegenen en de regulering van de genen gedurende de ontwikkeling.

Het verwijderen van HS2 en HS3 had tot doel de rol van de LCR functie in de chromatine-organisatie en in de activering van de globinegenen te bestuderen. We laten zien, dat het verwijderen van HS2 vrijwel geen effect heeft op het openen van het chromatine door de LCR, aangezien in slechts een van de vier transgene lijnen met een HS2 deletie een PEV werd waargenomen. We hebben echter wel een sterk effect op de ε-globine-expressie waargenomen, dat duidt op een mogelijke rol voor HS2 in de activering van het ε-globinegen in het embryo. Het verwijderen van HS3 had een sterke afname in globine-expressie tot gevolg, die zich nadrukkelijker manifesteerde naarmate het embryo zich ontwikkelde. De afname in expressie bleek te worden veroorzaakt door een PEV-expressie van het transgen, wat meteen aangeeft dat, anders dan HS2, HS3 wél van belang is voor het openen van het chromatine door de LCR. Het verwijderen van HS3 had geen specifiek effect op de expressie van een van de individuele globinegenen.

De Enh- en F-elementen zijn verwijderd om hun rol in de regulatie van de γ-globine-genexpressie gedurende de ontwikkeling te bestuderen. De elementen bevinden zich in het $^{A}\gamma$-δ intergenengebied, waarvan is gesuggereerd, dat het een rol speelt in de γ→β-wisseling en in de regulering van de γ-expressie-onderdrukking bij volwassenen. Deze suggesties zijn gedaan naar aanleiding van de connectie tussen de verwijdering van dit gebied en erfelijke permanente aanwezigheid van foetale hemoglobine (HPFH). De verwijdering van de Enh- en F-elementen in ons onderzoek had echter geen effect op de onderdrukking van de γ-genexpressie in de volwassene, maar resulteerde daarentegen in een verhoging van de ε- en γ-genexpressie in de embryonale fase.

Concluderend: de hier gepresenteerde studies laten zien:

1. dat de globinegenen op een stochastische wijze worden geactiveerd;

2. dat de LCR weliswaar functioneert als een holocomplex, maar dat daarnaast ook de individuele overgevoelige gebieden elk een eigen interactie hebben met de individuele globinegenen;
3. dat de regulering van de individuele globinegenen een complex geheel is, waarin waarschijnlijk de interactie van uiteenlopende combinaties van elementen gedurende de verschillende stadia van de ontwikkeling een rol speelt.

# *REFERENCES*

**Aalfs, J.D.; Kingston, R.E.** What does 'chromatin remodeling' mean? *Trends Biochem Sci* ***25 (11)*** 2000. pp. 548-555.
**Agarwal, S.; Rao, A.** Modulation of chromatin structure regulates cytokine gene expression during T cell differentiation. *Immunity* ***9 (6)*** 1998. pp. 765-775.
**Aladjem, M.I.; Rodewald, L.W.; Kolman, J.L.; Wahl, G.M.** Genetic dissection of a mammalian replicator in the human beta-globin locus. *Science* ***281 (5379)*** 1998. pp. 1005-1009.
**Alami, R.; Greally, J.M.; Tanimoto, K.; Hwang, S.; Feng, Y.; Engel, J.D.; Fiering, S.; Bouhassira, E.E.** Beta-globin YAC transgenes exhibit uniform expression levels but position effect variegation in mice. *Hum Mol Gen* ***9 (4)*** 2000. pp. 631-636.
**Alberts, B.; Bray, D.; Lewis, J.; Raff, M.; Roberts, K.; Watson, J.D.** *Molecular Biology of the Cell, third edition* 1994. pp. 368-369.
**Albitar, M.; Katsumata, M.; Liebhaber, S.A.** Human alpha-globin genes demonstrate autonomous developmental regulation in transgenic mice. *Mol Cell Biol* ***11 (7)*** 1991. pp. 3786-3794.
**Allison, A.C.** Parasitological reviews: malaria in carriers of the sickle-cell trait and in newborn children. *Exp Parasitol* ***6*** 1957. pp. 418-447.
**Anderson, J.D.; Widom, J.** Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. *J Mol Biol* ***296 (4)*** 2000. pp. 979-987.
**Andrews, N.C.; Erdjument-Bromage, H.; Davidson, M.B.; Tempst, P.; Orkin, S.H.** Erythroid transcription factor NF-E2 is a haematopoietic-specific basic-leucine zipper protein. *Nature* ***362 (6422)*** 1993. pp. 722-728.
**Andrews, N.C.; Kotkow, K.J.; Ney, P.A.; Erdjument-Bromage, H.; Tempst, P.; Orkin, S.H.** 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 (24)*** 1993. pp. 11488-11492.
**Andrulis, E.D.; Neiman, A.M.; Zappulla, D.C.; Sternglanz, R.** Perinuclear localization of chromatin facilitates transcriptional silencing. *Nature* ***395 (6693)*** 1998. pp. 525-525.
**Antequera, F.; Bird, A.** *DNA methylation: Molecular Biology and Biological significance* 1993. pp. 169-185.
**Antequera, F.B.A.; Bird, A.P.** Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci U S A* ***90 (24)*** 1993. pp. 11995-11999.
**Antoniou, M.; Geraghtry, F.; Hurst, J.; Grosveld, F.** Efficient 3'-end formation of human beta-globin mRNA in vivo requires sequences within the last intron but occurs independently of the splicing reaction. *Nucleic Acids Res* ***26 (3)*** 1998. pp. 721-729.
**Antoniou, M.; Grosveld, F.** beta-globin dominant control region interacts differently with distal and proximal promoter elements. *Genes Dev* ***4 (6)*** 1990. pp. 1007-1013.
**Antoniou, M.; de Boer, E.; Habets, G.; Grosveld, F.** The human beta-globin gene contains multiple regulatory regions: identification of one promoter and two downstream enhancers. *EMBO J* ***7 (2)*** 1988. pp. 377-384.
**Arcasoy, M.O.; Romana, M.; Fabry, M.E.; Skarpidi, E.; Nagel, R.L.; Forget, B.G.** High levels of human gamma-globin gene expression in adult mice carrying a transgene of deletion-type hereditary persistence of fetal hemoglobin. *Mol Cell Biol* ***17 (4)*** 1997. pp. 2076-2089.
**Arents, G; Burlingame, R.W.; Wang, B.C.; Moudrianakis, E.N.** The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. *Proc Natl Acad Sci U S A* ***88 (22)*** 1991. pp. 10148-10152.
**Armstrong, J.A.; Bieker, J.J.; Emerson, B.M.** A SWISNF-related chromatin remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLF in vitro *Cell* ***95 (1)*** 1998. pp. 93-104.
**Asano, H.; Li, X.S.; Stamatoyannopoulos, G.** FKLF, a novel Kruppel-like factor that activates human embryonic and fetal beta-like globin genes. *Mol Cell Biol* ***19 (5)*** 1999. pp. 3571-3579.
**Asano, H.; Li, X.S.; Stamatoyannopoulos, G.** FKLF-2: a novel Kruppel-like transcriptional factor that activates globin and other erythroid lineage genes. *Blood* ***95 (11)*** 2000. pp. 3578-3584.
**Banerji, J.; Rusconi, S.; Schaffner, W.** Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* ***27*** 1981. pp. 299-308.
**Bannister, A.J.; Zegerman, P.; Patridge, J.F.; Miska, E.A.; Thomas, J.O.; Allshire, R.C.; Kouzarides, T.** Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* ***410 (6824)*** 2001. pp. 120-124.
**Bannister, A.J.; Miska, E.; Gorlich, D.; Kouzarides, T.** Acetylation of importin-alpha nuclear import factors by CBPp300. *Curr Biol* ***10 (8)*** 2000. pp. 467-470.
**Bateson, G** *Mind and Nature* 1997. pp. ?-?.
**Bauer, U-T.; Daujat, S.; Nielsen, S.J.; Nightingale, J.; Kouzarides, T.** Methylation at arginine 17 of histone H3 is linked to gene activation. *EMBO rep* ***1 (3)*** 2002. pp. 39-44.
**Bean, T.L.; Ney, P.A.** Multiple regions of p45 NF-E2 are required for beta-globin gene expression in erythroid cells. *Nucleic Acids Res* ***25 (12)*** 1997. pp. 2509-2515.
**Behringer, R.R.; Hammer, R.E.; Brinster, R.L.; Palmiter, R.D.; Townes, T.M.** Two 3' sequences direct adult erythroid-specific expression of human beta-globin genes in transgenic mice. *Proc Natl Acad Sci U S A* ***84 (20)*** 1987. pp. 7056-7060.
**Behringer, R.R.; Ryan, T.M.; Palmiter, R.D.; Brinster, R.L.; Townes, T.M.** Human gamma- to beta-globin gene switching in transgenic mice. *Genes Dev* ***4 (3)*** 1990. pp. 380-389.
**Belikov, S.; Gelius, B.; Wrange, O.** Hormone-induced nucleosome positioning in the MMTV promoter is reversible. *EMBO J* ***11 (20)*** 2001. pp. 2802-2811.
**Bell, A.C.; West, A.G.; Felsenfeld, G.** Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science* ***291 (5503)*** 2001. pp. 447-450.
**Bell, A.C.; West, A.G.; Felsenfeld, G.** The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* ***98 (3)*** 1999. pp. 387-396.
**Bender, M.A.; Bulger, M.; Halow, J.; Telling, A.; Close, J.; Groudine, M.** 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 (2)*** 2000. pp. 387-393.
**Bender, M.A.; Reik, A.; Close, J.; Telling, A.; Epner, E.; Fiering, S.; Hardison, R.C.; Groudine, M.** Description and targeted deletion of 5' hypersensitive site 5 and 6 of the mouse beta-globin locus control region. *Blood* ***92 (11)*** 1998. pp. 4394-4403.
**Berk, A.J.** Regulation of eukaryotic transcription factors by post-translational modification. *Biochim Biophys Acta* ***1009 (2)*** 1989. pp. 103-109.
**Bernet, A.; Sabatier, S.; Picketts, D.J.; Ouazana, R.; Morle, F.; Higgs, D.R.; Godet, J.** Targeted inactivation of the major positive regulatory element (HS-40) of the human alpha-globin gene locus. *Blood* ***86 (3)*** 1995. pp. 1202-1211.
**Berry, M.; Grosveld, F.; Dillon, N.A.** A single point mutation is the cause of the Greek form of hereditary persistence of fetal haemoglobin. *Nature* ***358 (6386)*** 1992. pp. 499-502.
**Bhaumik, S.R.; Green, M.R.** SAGA is an essential in vivo target of the yeast acidic activator Gal4p. *Genes Dev* ***15 (15)*** 2001. pp. 1935-1945.
**Bhoite, L.T.; Yu, Y.; Stillman, D.J.** The Swi5 activator recruits the Mediator complex to the HO promoter without RNA polymerase II. *Genes Dev* ***15 (18)*** 2001. pp. 2457-2469.
**Bienz, M.; Muller, J.** Transcriptional silencing of homeotic genes in Drosophila. *Bioessays* ***17 (9)*** 1995. pp. 775-784.
**Bird, A.P.; Wolffe, A.P.** Methylation-induced repression--belts, braces, and chromatin. *Cell* ***99 (5)*** 1999. pp. 451-454.

**Bird, A.P.; Taggert, M.; Nicholls, R.; Higgs, D.** Non-methylated CpG-rich islands at the human alpha-globin locus: implications for evolution of the alpha-globin pseudogene. *EMBO J* ***6 (4)*** 1987. pp. 999-1004.
**Blobel, G.A.; Nakajima, T.; Eckner, R.; Montminy, M.; Orkin, S.H.** CREB-binding protein cooperates with transcription factor GATA-1 and is required for erythroid differentiation. *Proc Natl Acad Sci U S A* ***95 (5)*** 1998. pp. 2061-2066.
**Blom van Assendelft, G.; Hanscombe, O.; Grosveld, F.; Greaves, D.R.** The beta-globin dominant control region activates homologous and heterologous promoters in a tissue-specific manner. *Cell* ***56 (6)*** 1989. pp. 969-977.
**Blomquist, P.; Belikov, S.; Wrange, O.** Increased nuclear factor 1 binding to its nucleosomal site mediated by sequence-dependent DNA structure. *Nucleic Acids Res* ***27 (2)*** 1999. pp. 517-525.
**Bloom, W.; Bartelmez, G.W.** - *Am J Anat* ***67*** 1940. pp. 21-?.
**Bloomfield, I.C.; Vaughn, V.; Rest, R.S.; Eisenstein, B.I.** Allelic exchange in Escherichia coli using the Bacillus subtilis sacB gene and a temperature-sensitive pSC101 replicon. *Mol Microbiol* ***5 (6)*** 1991. pp. 1447-1457.
**Bodine, D.M.; Ley, T.J.** An enhancer element lies 3' to the human A gamma globin gene. *EMBO J* ***6 (10)*** 1987. pp. 2997-3004.
**Bohm, L.; Crane-Robinson, C.** Proteases as structural probes for chromatin: the domain structure of histones. *Biosci Rep* ***4 (5)*** 1984. pp. 365-86.
**Boyes, J.; Byfield, P.; Nakatani, Y.; Ogryzko, V.** Regulation of activity of the transcription factor GATA-1 by acetylation. *Nature* ***396 (6711)*** 1998. pp. 594-598.
**Boyes, J.; Bird, A.** DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell* ***64 (6)*** 1991. pp. 1123-1134.
**Boyes, J.; Bird, A.** Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein *EMBO J* ***11 (1)*** 1992. pp. 327-333.
**Brand, A.H.; Breeden, L.; Abraham, J.; Sternglanz, R.; Nasmyth, K.** Characterization of a silencer in yeast: a DNA sequence with properties opposite to those of a transcriptional enhancer. *Cell* ***41 (1)*** 1985. pp. 41-48.
**Brehm, A.; Kouzarides, T.** Retinoblastoma protein meets chromatin. *Trends Genet* ***24 (4)*** 1999. pp. 142-145.
**Breiling, A.; Bonte, E.; Ferrari, S.; Becker, P.; Paro, R.** The Drosophila polycomb protein interacts with nucleosomal core particles In vitro via its repression domain. *Mol Cell Biol* ***19 (12)*** 1999. pp. 8451-8460.
**Brock, H.W.; van Lohuizen, M.** The Polycomb group--no longer an exclusive club? *Curr Opin Genetics Dev* ***11 (2)*** 2001. pp. 175-181.
**Brown, K.E.; Guest, S.S.; Smale, S.T.; Hahm, K.; Merkenschlager, M.; Fisher, A.G.** Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell* ***91 (6)*** 1997. pp. 845-854.
**Brown, K.E.; Baxter, J.; Graf, D.; Merkenschlager, M.; Fisher, A.G.** Dynamic repositioning of genes in the nucleus of lymphocytes preparing for cell division. *Mol Cell* ***3 (2)*** 1999. pp. 207-217.
**Brown, K.E.; Amoilis, S.; Horn, J.M.; Buckle, V.J.; Higgs, D.R.; Merkenschlager, M.; Fisher, A.G.** Expression of alpha- and beta-globin genes occurs within different nuclear domains in haemopoietic cells. *Nature Cell Biol* ***3 (6)*** 2001. pp. 602-606.
**Brownell, J.E.; Zhou, J.; Ranalli, T.; Kobayashi, R.; Edmonson, D.G.; Roth, S.Y.; Allis, C.D.** Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* ***84 (6)*** 1996. pp. 843-851.
**Bulger, M.; Sawado, T.; Schubeler, D.; Groudine, M.** ChIPs of the beta-globin locus: unraveling gene regulation within an active domain. *Curr Opin Genetics Dev* ***12 (2)*** 2002. pp. 170-177.
**Bungert, J.; Dave, U.; Lim, K.C.; Lieuw, K.H.; Shavit, J.A.; Liu, Q.; Engel, J.D.** Synergistic regulation of human beta-globin gene switching by locus control region elements HS3 and HS4. *Genes Dev* ***9 (24)*** 1995. pp. 3083-3096.
**Bungert, J.; Tanimoto, K.; Patel, S.; Liu, Q.; Fear, M.; Engel, J.D.** Hypersensitive site 2 specifies a unique function within the human beta-globin locus control region to stimulate globin gene transcription. *Mol Cell Biol* ***19 (4)*** 1999. pp. 3062-3072.
**Bunn, H.F.; Forget, B.G.** *Hemoglobin: Molecular, genetic and clinical aspects.* 1986. pp. ?-?.
**Buratowski, S.; Hahn, S.; Guarente, L.; Sharp, P.A.** Five intermediate complexes in transcription initiation by RNA polymerase II *Cell* ***56 (4)*** 1989. pp. 549-561.
**Burns, L.G.; Peterson, C.L.; Tamkun, J.W.; Mendel, D.B.; Crabtree, G.R.** The yeast SWI-SNF complex facilitates binding of a transcriptional activator to nucleosomal sites in vivo. *Mol Cell Biol* ***17 (8)*** 1997. pp. 4811-4819.
**Cai, H.N.; Shen, P.** Effects of cis arrangement of chromatin insulators on enhancer-blocking activity. *Science* ***291 (5503)*** 2001. pp. 493-495.
**Cairns, B.R.** Chromatin remodeling machines: similar motors, ulterior motives. *Trends Biochem Sci* ***23 (1)*** 1998. pp. 20-25.
**Calzolari, R.; McMorrow, T.; Yannoutsos, N.; Langeveld, A.; Grosveld, F.** Deletion of a region that is a candidate for the difference between the deletion forms of hereditary persistence of fetal hemoglobin and deltabeta-thalassemia affects beta- but not gamma-globin gene expression. *EMBO J* ***18 (4)*** 1999. pp. 949-958.
**Cao, S.X.; Gutman, P.D.; Davie, H.P.; Schechter, A.N.** Identification of a transcriptional silencer in the 5'-flanking region of the human epsilon-globin gene. *Proc Natl Acad Sci U S A* ***86 (14)*** 1989. pp. 5306-5309.
**Carruthers, L.M; Hansen, J.C.** The core histone N termini function independently of linker histones during chromatin condensation. *J Biol Chem* ***275 (47)*** 2000. pp. 37285-37290.
**Caterina, J.J.; Ryan, T.M.; Pawlik, K.M.; Palmiter, R.D.; Brinster, R.L.; Behringer, R.R.; Townes, T.M.** Human beta-globin locus control region: analysis of the 5' DNase I hypersensitive site HS 2 in transgenic mice. *Proc Natl Acad Sci U S A* ***88 (5)*** 1991. pp. 1326-1330.
**Caterina, J.J.; Donze, D.; Sun, C.W.; Ciavatta, D.J.; Townes, T.M.** Cloning and functional characterization of LCR-F1: a bZIP transcription factor that activates erythroid-specific, human globin gene expression. *Nucleic Acids Res* ***22 (12)*** 1994. pp. 2383-2391.
**Caterina, J.J.; Ciavatta, D.J.; Donze, D.; Behringer, R.R.; Townes, T.M.** 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 (6)*** 1994. pp. 106-111.
**Cavalli, G.; Paro, R.** Epigenetic inheritance of active chromatin after removal of the main transactivator. *Science* ***286 (5441)*** 1999. pp. 955-958.
**Chada, K.; Magram, J.; Costantini, F.** An embryonic pattern of expression of a human fetal globin gene in transgenic mice. *Nature* ***319 (6055)*** 1986. pp. 685-689.
**Chan, F; Robinson, J.; Browlie, A.** Characterization of adult alpha- and beta-globin genes in the zebrafish. *Blood* ***89 (2)*** 1997. pp. 688-700.
**Chang, Y.L.; Peng, Y.H.; Pan, I.C.; Sun, D.S.; King, B.; Huang, D.H.** Essential role of Drosophila Hdac1 in homeotic gene silencing. *Proc Natl Acad Sci U S A* ***98 (17)*** 2001. pp. 9730-9735.
**Charnay, P.; Treisman, R.; Mellon, P.; Chao, M.; Axel, R.; Maniatis, T.** Differences in human alpha- and beta-globin gene expression in mouse erythroleukemia cells: the role of intragenic sequences. *Cell* ***38 (1)*** 1984. pp. 251-263.
**Chen, C.; Yang, T.P.** Nucleosomes are translationally positioned on the active allele and rotationally positioned on the inactive allele of the HPRT promoter. *Mol Cell Biol* ***21 (22)*** 2001. pp. 7682-7695.
**Chen, D.; Ma, H.; Honh, H.; Koh, S.S.; Huang, S.M.; Schurter, B.T.; Aswad, D.W.; Stallcup, M.R.** Regulation of transcription by a protein methyltransferase. *Science* ***284 (5423)*** 1999. pp. 2174-2177.
**Chen, H.; Lin, J.R.; Xie, W.; Wilpitz, D.; Evans, M.R.** Regulation of hormone-induced histone hyperacetylation and gene activation via acetylation of an acetylase. *Cell* ***98 (5)*** 1999. pp. 675-686.
**Chen, S.; Corces, V.G.** The gypsy insulator of Drosophila affects chromatin structure in a directional manner *Genetics* ***159 (4)*** 2001. pp. 1649-1658.

**Chen, Z.J.; Pikaard, C.S.** Epigenetic silencing of RNA polymerase I transcription: a role for DNA methylation and histone modification in nucleolar dominance. *Genes Dev* ***11 (16)*** 1997. pp. 2124-2136.
**Chess, A.** Expansion of the allelic exclusion principle? *Science* ***279 (5359)*** 1998. pp. 2067-2068.
**Cheung, P.; Allis, C.D.; Sassone-Corsi, P.** Signaling to chromatin through histone modifications. *Cell* ***103 (2)*** 2000. pp. 263-271.
**Cheung, P.; Tanner, K.G.; Cheung, W.L.; Sassone-Corsi, P.; Denu, J.M.; Allis, C.D.** Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. *Mol Cell* ***5 (6)*** 2000. pp. 905-915.
**Chi, T.; Carey, M.F.; Kakidani, H.; Roeder, R.G.** Assembly of the isomerized TFIIA--TFIID--TATA ternary complex is necessary and sufficient for gene activation. *Genes Dev* ***10 (20)*** 1996. pp. 2540-2550.
**Cho, H.G.; Orphanides, G.; Sun, X.; Yang, X.J.; Ogryzko, V.; Lees, E.; Nakatani, Y.; Reinberg, D.** A human RNA polymerase II complex containing factors that modify chromatin structure. *Mol Cell Biol* ***18 (9)*** 1998. pp. 5355-5363.
**Chodosh, L.A.; Fire, A.; Samuels, M.; Sharp, P.A.** 5,6-Dichloro-1-beta-D-ribofuranosylbenzimidazole inhibits transcription elongation by RNA polymerase II in vitro. *J Biol Chem* ***264 (4)*** 1989. pp. 2250-2257.
**Choi, O.R.; Engel, J.D.** Developmental regulation of beta-globin gene switching. *Cell* ***55 (1)*** 1988. pp. 17-26.
**Chung, J.H.; Whiteley, M.; Felsenfeld, G.** 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 (3)*** 1993. pp. 505-514.
**Cirillo, L.A.; Zaret, K.S.** An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol Cell* ***4 (6)*** 1999. pp. 961-969.
**Cirillo, L.A.; Lin, F.R.; Cuesta, I.; Friedman, D.; Jarnik, M.; Zaret, K.S.** Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell* ***9 (2)*** 2002. pp. 279-289.
**Clapier, C.R.; Langst, G.; Corona, D.F.; Becker, P.B.** Critical role for the histone H4 N terminus in nucleosome remodeling by ISWI. *Mol Cell Biol* ***21 (3)*** 2001. pp. 875-883.
**Clark, A.R.; Docherty, K.** Negative regulation of transcription in eukaryotes. *Biochem J* ***296*** 1993. pp. 521-541.
**Clark, D.J.; Hill, C.S.; Martin, S.R.; Thomas, J.O.** Alpha-helix in the carboxy-terminal domains of histones H1 and H5. *EMBO J* ***7 (1)*** 1988. pp. 69-75.
**Clark, D.J.; Kimura, T** Electrostatic mechanism of chromatin folding. *J Mol Biol* ***211 (4)*** 1990. pp. 883-896.
**Clark, D.J.; O'Neill, L.P.; Turner, B.M.** Selective use of H4 acetylation sites in the yeast Saccharomyces cerevisiae. *Biochem J* ***294*** 1993. pp. 557-561.
**Cockerill, P.N.; Garrard, W.T.** Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. *Cell* ***44 (2)*** 1986. pp. 273-282.
**Collins, F.S.; Weissman, S.M.** The molecular genetics of human hemoglobin. *Prog Nucleic Acid Res Mol Biol* ***31*** 1984. pp. 315-462.
**Collis, P.; Antoniou, M.; Grosveld, F.** Definition of the minimal requirements within the human beta-globin gene and the dominant control region for high level expression *EMBO J* ***9 (1)*** 1990. pp. 233-240.
**Conaway, R.C.; Conaway, J.W.** General initiation factors for RNA polymerase II. *Annu Rev Biochem* ***62*** 1993. pp. 161-190.
**Corces, V.G.; Geyer, P.K.** Interactions of retrotransposons with the host genome: the case of the gypsy element of Drosophila. *Trends Genet* ***7 (3)*** 1991. pp. 86-90.
**Cornet, F.; Louarn, J.; Patte, J.; Louarn, J-M.** Restriction of the activity of the recombination site dif to a small zone of the Escherichia coli chromosome. *Genes Dev* ***10 (9)*** 1996. pp. 1152-1161.
**Cornet, F.; Mortier, I.; Patte, J.; Louarn, J-M.** Plasmid pSC101 harbors a recombination site, psi, which is able to resolve plasmid multimers and to substitute for the analogous chromosomal Escherichia coli site dif. *J Bacteriol* ***176 (11)*** 1994. pp. 3188-3195.
**Corona, D.F.; Langst, G.; Clapier, C.R.; Bonte, E.J.; Ferrari, S.; Tamkun, J.W.; Becker, P.B.** ISWI is an ATP-dependent nucleosome remodeling factor. *Mol Cell* ***3 (2)*** 1999. pp. 239-245.
**Corral, J.; Lavenir, I.; Impey, H.; Warren, A.J.; Forster, A.; Larson, T.A.; Bell, S.; McKenzie, A.N.J.; King, G.; Rabbitts, T.H.** An Mll-AF9 fusion gene made by homologous recombination causes acute leukemia in chimeric mice: a method to create fusion oncogenes. *Cell* ***85 (6)*** 1996. pp. 853-861.
**Cosma, M.P.; Tanaka, T.; Nasmyth, K.** Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter *Cell* ***97 (3)*** 1999. pp. 299-311.
**Cote, J.; Qiunn, J.; Workman, J.L.; Peterson, C.L.** Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWISNF complex. *Science* ***265 (5168)*** 1994. pp. 53-60.
**Craddock, C.F.; Vyas, P.; Sharpe, J.A.; Ayyub, H.; Wood, W.G.; Higgs, D.R.** Contrasting effects of alpha and beta globin regulatory elements on chromatin structure may be related to their different chromosomal environments. *EMBO J* ***14 (8)*** 1995. pp. 1718-1726.
**Crossley, M.; Whitelaw, E.; Perkins, A.; Williams, G.; Fujiwara, Y.; Orkin, S.H.** Isolation and characterization of the cDNA encoding BKLFTEF-2, a major CACCC-box-binding protein in erythroid cells and selected other cells. *Mol Cell Biol* ***16 (4)*** 1996. pp. 1695-1705.
**Crossley, M.; Tsang, A.P.; Bieker, J.J.; Orkin, S.H.** Regulation of the erythroid Kruppel-like factor (EKLF) gene promoter by the erythroid transcription factor GATA-1. *J Biol Chem* ***269 (22)*** 1994. pp. 15440-15444.
**Crotta, S.; Nicolis, S.; Ronchi, A.; Ottolenghi, S.; Ruzzi, L.; Shimada, Y.; Migliaccio, A.R.; Migliaccio, G.** Progressive inactivation of the expression of an erythroid transcriptional factor in GM- and G-CSF-dependent myeloid cell lines. *Nucleic Acids Res* ***18 (23)*** 1990. pp. 6863-6869.
**Crouzet, J.; Naudin, L.; Orsini, C.; Vigne, E.; Ferrero, L.; Le Roux, A.; Benoit, P.; Latta, M.; Torrent, C.; Barnellec, D.; Denefle, P.; Mayaux, J-F; Perricaudet, M.; Yeh, P.** Recombinational construction in Escherichia coli of infectious adenoviral genomes. *Proc Natl Acad Sci U S A* ***94 (4)*** 1997. pp. 1414-1419.
**Cryderman, D.E.; Cuaycong, M.H.; Elgin, S.C.R.; Wallrath, L.L.** Characterization of sequences associated with position-effect variegation at pericentric sites in Drosophila heterochromatin. *Chromosoma* ***107 (5)*** 1998. pp. 277-285.
**Csordas, A.** On the biological role of histone acetylation. *Biochem J* ***265 (1)*** 1990. pp. 23-38.
**Czermin, B.; Schotta, G.; Hulsman, B.B.; Brehm, A.; Becker, P.B.; Reuter, G.; Imhof, A.** Physical and functional association of SU(VAR)3-9 and HDAC1 in Drosophila. *EMBO rep* ***2 (10)*** 2001. pp. 915-919.
**Dasso, M; Dimitrov, S.; Wolffe, A.P.** Nuclear assembly is independent of linker histones. *Proc Natl Acad Sci U S A* ***91 (26)*** 1994. pp. 12477-481.
**Davey, C.A.; Sargent, D.F.; Luger, K.; Maeder, A.W.; Richmond, T.J.** Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *J Mol Biol* ***319 (5)*** 2002. pp. 1097-1113.
**de Boer, E.; Antoniou, M.; Mignotte, V.; Wall, L.; Grosveld, F.** The human beta-globin promoter, nuclear protein factors and erythroid specific induction of transcription. *EMBO J* ***7 (13)*** 1988. pp. 4203-4212.
**De Lange, R.J.; Farnbrough, D.M.; Smith, E.L.; Bonner, J.** Identification of episilon-N-methyllysine in Spirillum serpens flagella and of episilon-N-dimethyllysine in Salmonella typhimurium flagella. *J Biol Chem* ***244 (1)*** 1969. pp. 319-334.
**de Villiers, J.; Olson, L.; Banerji, J.; Schaffner, W.** *Cold Spring Harb. Symp. Quant. Biol* 1983. pp. 911-919.
**Deisseroth, A.; Velez, R.; Nienhuis, A.W.** Hemoglobin synthesis in somatic cell hybrids: independent segregation of the human alpha- and beta-globin genes. *Science* ***191 (4233)*** 1976. pp. 1262-1263.

**Deisseroth, A.; Nienhuis, A.W.; Turner, P.; Velez, R.; Anderson, W.I.; Ruddle, F.; Lawrence, J.; Creagan, R.; Kuchelapati, R.** Localization of the human alpha-globin structural gene to chromosome 16 in somatic cell hybrids by molecular hybridization assay. *Cell* ***12 (1)*** 1977. pp. 205-218.
**Deisseroth, A.; Nienhuis, A.W.; Lawrence, J.; Giles, R.; Turner, P.; Ruddle, F.H.** Chromosomal localization of human beta globin gene on human chromosome 11 in somatic cell hybrids. *Proc Natl Acad Sci U S A* ***75 (3)*** 1978. pp. 1456-1460.
**Dhalluin, C.; Carlson, J.E.; Zeng, L.; He, C.; Aggerwal, A.; Zhou, M.** Structure and ligand of a histone acetyltransferase bromodomain. *Nature* ***399 (6735)*** 1999. pp. 491-496.
**Dhar, V.; Nandi, A.; Schildkraut, C.L.; Skoultchi, A.** Erythroid-specific nuclease-hypersensitive sites flanking the human beta-globin domain. *Mol Cell Biol* ***10 (8)*** 1990. pp. 4324-4333.
**Dickerson, R.E.; Geis, I.** *Hemoglobin: Structure, function, evolution and pathology* 1983. pp. ?-?.
**Dillon, N.; Grosveld, F.** Human gamma-globin genes silenced independently of other genes in the beta-globin locus. *Nature* ***350 (6315)*** 1991. pp. 252-254.
**Dillon, N.; Trimborn, T.; Strouboulis, J.; Fraser, P.; Grosveld, F.** The effect of distance on long-range chromatin interactions. *Mol Cell* ***1 (1)*** 1997. pp. 131-139.
**Dillon, N.; Grosveld, F.** Transcriptional regulation of multigene loci: multilevel control. *Trends Genet* ***9 (4)*** 1993. pp. 134-137.
**Dong, J.M.; Lim, L.** The human neuronal alpha 1-chimaerin gene contains a position-dependent negative regulatory element in the first exon *Neurochem Res* ***21 (9)*** 1996. pp. 1023-1030.
**Donze, D.; Jeancake, P.H.; Townes, T.M.** Activation of delta-globin gene expression by erythroid Krupple-like factor: a potential approach for gene therapy of sickle cell disease. *Blood* ***88 (10)*** 1996. pp. 4051-4057.
**Donze, D.; Townes, T.M.; Bieker, J.J.** Role of erythroid Kruppel-like factor in human gamma- to beta-globin gene switching. *J Biol Chem* ***270 (4)*** 1995. pp. 1955-1959.
**Dorsett, D.** Distance-independent inactivation of an enhancer by the suppressor of Hairy-wing DNA-binding protein of Drosophila. *Genetics* ***134 (4)*** 1993. pp. 1135-1144.
**Dower, W.J.; Miller, J.F.; Ragsdale, C.W.** High efficiency transformation of E. coli by high voltage electroporation. *Nucleic Acids Res* ***16 (13)*** 1988. pp. 6127-6145.
**Driscoll, M.C.; Dobkin, C.S.; Alter, B.P.** 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 (19)*** 1989. pp. 7470-7474.
**Duan, Z.J.; Stamatoyannopoulos, G.; Li, Q.** Role of NF-Y in in vivo regulation of the gamma-globin gene. *Mol Cell Biol* ***21 (9)*** 2001. pp. 3083-3095.
**Duff, K.; Huxley, C.** Targeting mutations to YACs by homologous recombination. *Methods Mol Biol* ***54*** 1996. pp. 187-198.
**Dynan, W.** Modularity in promoters and enhancers. *Cell* ***58 (1)*** 1989. pp. 1-4.
**Dynlacht, B.D.; Hoey, T.; Tjian, R.** Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. *Cell* ***66 (3)*** 1991. pp. 563-576.
**Eberharter, A.; Ferrari, S.; Langst, G.; Straub, T.; Imhof, A.; Varga-Weisz, P.D.; Wilm, M.; Becker, P.B.** Acf1, the largest subunit of CHRAC, regulates ISWI-induced nucleosome remodelling. *EMBO J* ***20 (14)*** 2001. pp. 3781-3788.
**Eden, S.; Hashishony, T.; Keshet, I.; Cedar, H.; Thorne, A.W.** DNA methylation models histone acetylation. *Nature* ***394 (6696)*** 1998. pp. 842-842.
**Efstratiadis, A.; Posakony, J.W.; Maniatis, T.; Lawn, R.M.; O'Connell, C.; Spritz, R.A.; de Riel, J.K.; Forget, B.G.; Weissman, S.M.; Slightom, J.L.; Blechl, E.A.; Smithies, O.; Baralle, F.E.; Shoulders, C.C.; Proudfoot, N.J.** The structure and evolution of the human beta-globin gene family. *Cell* ***21 (3)*** 1980. pp. 653-668.
**Eikbush, T.H; Moudrianakis, E.N.** The histone core complex: an octamer assembled by two sets of protein-protein interactions. *Biochemistry* ***17 (23)*** 1978. pp. 4955-4964.
**Eisenfeld, K.; Candau, R.; Truss, M.; Breato, M.** Binding of NF1 to the MMTV promoter in nucleosomes: influence of rotational phasing, translational positioning and histone H1. *Nucleic Acids Res* ***25 (18)*** 1997. pp. 3733-3742.
**Eissenberg, J.C; Elgin, S.C.R.** Boundary functions in the control of gene expression. *Trends Genet* ***7 (10)*** 1991. pp. 335-340.
**Eissenberg, J.C; James, T.C.; Foster-Hartnett, D.M.; Hartnett, T.; Ngan, V.; Elgin, S.C.** Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect ariegation in Drosophila melanogaster. *Proc Natl Acad Sci U S A* ***87 (24)*** 1990. pp. 9923-9927.
**Eissenberg, J.C; Morris, G.D.; Reuter, G.; Hartnett, T.** The heterochromatin-associated protein HP-1 is an essential protein in Drosophila with dosage-dependent effects on position-effect variegation. *Genetics* ***87 (131)*** 1992. pp. 9923-9927.
**Ekwall, K.; Olsson, T.; Turner, B.M.; Cranston, G.; Allshire, R.C.** Transient inhibition of histone deacetylation alters the structural and functional imprint at fission yeast centromeres. *Cell* ***91 (7)*** 1997. pp. 1021-1032.
**Ellis, J.; Talbot, D.; Dillon, N.; Grosveld, F.** Synthetic human beta-globin 5'HS2 constructs function as locus control regions only in multicopy transgene concatamers. *EMBO J* ***12 (1)*** 1993. pp. 127-134.
**Ellis, J.; Tan-Un, K.C.; Harper, A.; Michalovich, D.; Yannoutsos, N.; Philipsen, S.; Grosveld, F.** A dominant chromatin-opening activity in 5' hypersensitive site 3 of the human beta-globin locus control region. *EMBO J* ***15 (3)*** 1996. pp. 562-568.
**Elnitski, L.; Li, J.; Noguchi, C.T.; Miller, W.; Hardison, R.** A negative cis-element regulates the level of enhancement by hypersensitive site 2 of the beta-globin locus control region. *J Biol Chem* ***276 (9)*** 2001. pp. 6289-6298.
**Elowitz, M.B.; Levine, A.J.; Siggia, E.D.; Swain, P.S.** Stochastic gene expression in a single cell. *Science* ***297 (5584)*** 2002. pp. 1183-1186.
**Emerson, B.M.; Lewis, C.D.; Felsenfeld, G.** Interaction of specific nuclear factors with the nuclease-hypersensitive region of the chicken adult beta-globin gene: nature of the binding domain. *Cell* ***41 (1)*** 1985. pp. 21-30.
**Enver, T.; Ebens, A.J.; Forrester, W.C.; Stamatoyannopoulos, G.** The human beta-globin locus activation region alters the developmental fate of a human fetal globin gene in transgenic mice. *Proc Natl Acad Sci U S A* ***86 (18)*** 1989. pp. 7033-7037.
**Enver, T.; Raich, N.; Ebens, A.J.; Papayannopoulou, T.; Costantini, F.; Stamatoyannopoulos, G.** Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. *Nature* ***344 (6264)*** 1990. pp. 309-313.
**Epner, E.; Rifkin, R.A.; Marks, P.A.** Replication of alpha and beta globin DNA sequences occurs during early S phase in murine erythroleukemia cells. *Proc Natl Acad Sci U S A* ***78 (5)*** 1981. pp. 3058-3062.
**Epner, E.; Reik, A.; Cimbora, D.; Telling, A.; Bender, M.A.; Fiering, A.; Enver, T.; Martin, D.I.; Kennedy, M.; Keller, G.; Groudine, M.** 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 (4)*** 1998. pp. 447-455.
**Farace, M.G.; Brown, B.A.; Raschella, G.; Alexander, J.; Cambari, R.; Fantoni, A.; Hardies, S.C.; Hutchison III, C.A.; Edgell, M.H.** The mouse beta h1 gene codes for the z chain of embryonic hemoglobin. *J Biol Chem* ***259 (11)*** 1984. pp. 7123-7128.
**Farrell, C.M.; Grinberg, A.; Huang, S.P.** 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 (26)*** 2000. pp. 14554-14559.
**Feinberg, A.W.; Vogelstein, B.** A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* ***132 (1)*** 1983. pp. 6-13.

**Feingold, E.A.; Forget, B.G.** The breakpoint of a large deletion causing hereditary persistence of fetal hemoglobin occurs within an erythroid DNA domain remote from the beta-globin gene cluster. *Blood* ***74 (6)*** 1989. pp. 2178-2186.
**Felsenfeld, G.** Chromatin unfolds. *Cell* ***86 (1)*** 1996. pp. 13-19.
**Feng, W.C.; Southwood, C.M.; Bieker, J.J.** Analyses of beta-thalassemia mutant DNA interactions with erythroid Kruppel-like factor (EKLF), an erythroid cell-specific transcription factor. *J Biol Chem* ***269 (2)*** 1994. pp. 1493-1500.
**Fessas, P.; Stamatoyannopoulos, G.** Hereditary Persistence of foetal hemoglobin in Greece : a study and a comparison. *Blood* ***24*** 1964. pp. 223-223.
**Festenstein, R.; Tolaini, M.; Corbella, P.; Mamalaki, C.; Parrington, J.; Fox, M.; Miliou, A.; Jones, M.; Kioussis, D.** Locus control region function and heterochromatin-induced position effect variegation. *Science* ***271 (5252)*** 1996. pp. 123-125.
**Festenstein, R.; Kioussis, D.** Locus control regions and epigenetic chromatin modifiers. *Curr Opin Genetics Dev* ***10 (2)*** 2000. pp. 199-203.
**Fiering, S.; Epner, E.; Robinson, K.; Zhuang, Y.; Telling, A.; Hu, M.; Martin, D.I.K.; Enver, T.; Ley, T.J.; Groudine, M.** 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 (18)*** 1995. pp. 2203-2213.
**Fiering, S.; Northrop, J.P.; Nolan, G.P.; Mattila, P.S.; Crabtree, G.R.; Herzenberg, L.A.** Single cell assay of a transcription factor reveals a treshold in transcription activated by signals emanating from the T-cell antigen receptor *Genes Dev* ***4*** 1990. pp. 1823-1834.
**Fiering, S.; Whitelaw, E.; Martin, D.I.K.** To be or not to be active: the stochastic nature of enhancer function *Bioessays* ***22*** 2000. pp. 381-387.
**Filipe, A.; Li, Q.; Deveaux, S.; Godin, I.; Romeo, P.H.; Stamatoyannopoulos, G.; Mignotte, V.** Regulation of embryonicfetal globin genes by nuclear hormone receptors: a novel perspective on hemoglobin switching. *EMBO J* ***18 (3)*** 1999. pp. 687-687.
**Fischel-Ghodsian, N.; Nicholls, R.D.; Higgs, D.R.** Long range genome structure around the human alpha-globin complex analysed by PFGE. *Nucleic Acids Res* ***15 (15)*** 1987. pp. 9215-9225.
**Fischle, W.; Emiliani, S.; Hendzel, M.J.; Nagase, T.; Nomura, N.; Voelter, W.; Verdin, E.** A new family of human histone deacetylases related to Saccharomyces cerevisiae HDA1p. *J Biol Chem* ***274 (17)*** 1999. pp. 11713-11720.
**Fletcher, T.M; Hansen, J.C.** The nucleosomal array: structurefunction relationships. *Crit Rev Eukaryoti Gene Expr* ***6*** 1996. pp. 88-91.
**Flint, J.; Thomas, K.; Micklem, G.; Raynham, H.; Clark, K.; Doggett, N.A.; Higgs, D.R.** The relationship between chromosome structure and function at a human telomeric region. *Nature Genet* ***15 (3)*** 1997. pp. 252-257.
**Forrester, W.C.; Takegawa, S.; Papayannopoulou, T.; Stamatoyannopoulos, G.; Groudine, M.** Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids. *Nucleic Acids Res* ***15 (24)*** 1987. pp. 10159-10177.
**Forrester, W.C.; Thompson, C.; Elder, J.T.; Groudine, M.** A developmentally stable chromatin structure in the human beta-globin gene cluster. *Proc Natl Acad Sci U S A* ***83 (5)*** 1986. pp. 1359-1363.
**Forrester, W.C.; Epner, E.; Driscoll, M.C.; Enver, T.; Brice, M.; Papayannopoulou, T.; Groudine, M.** 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 (10)*** 1990. pp. 1637-1649.
**Forsberg, E.C.; Bresnick, E.H.** Histone acetylation beyond promoters: long-range acetylation patterns in the chromatin world. *Bioessays* ***23 (9)*** 2001. pp. 820-830.
**Francastel, C.; Walters, M.C.; Groudine, M.; Martin, D.I.** A functional enhancer suppresses silencing of a transgene and prevents its localization close to centrometric heterochromatin. *Cell* ***99 (3)*** 1999. pp. 259-269.
**Francis, N.J.; Saurin, A.J.; Shao, Z.; Kingston, R.E.** Reconstitution of a functional core polycomb repressive complex *Mol Cell* ***8 (3)*** 2001. pp. 545-556.
**Fraser, P.; Grosveld, F.** Locus control regions, chromatin activation and transcription. *Curr Opin Cell Biol* ***10 (3)*** 1998. pp. 361-365.
**Fraser, P.; Hurst, J.; Collis, P.; Grosveld, F.** DNaseI hypersensitive sites 1, 2 and 3 of the human beta-globin dominant control region direct position-independent expression. *Nucleic Acids Res* ***18 (12)*** 1990. pp. 3503-3508.
**Fraser, P.; Pruzina, S.; Antoniou, M.; Grosveld, F.** 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 (1)*** 1993. pp. 106-113.
**Fujiwara, Y.; Browne, C.P.; Cunniff, K.; Goff, S.C.; Orkin, S.H.** Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc Natl Acad Sci U S A* ***93 (22)*** 1996. pp. 12355-12358.
**Furst, A.; Brown, E.H.; Braunstein, J.D.; Schildkraut, C.L.** alpha-Globulin sequences are located in a region of early-replicating DNA in murine erythroleukemia cells. *Proc Natl Acad Sci U S A* ***78 (2)*** 1981. pp. 1023-1027.
**Furukawa, T.; Navas, P.A.; Josephson, B.M.; Peterson, K.R.; Papayannopoulou, T.; Stamatoyannopoulos, G.** Coexpression of epsilon, G gamma and A gamma globin mRNA in embryonic red blood cells from a single copy beta-YAC transgenic mouse *Cells Blood Cells Mol Dis* ***21 (2)*** 1995. pp. 168-178.
**Gaensler, K.M.; Kitamura, M.; Kan, Y.W.** Germ-line transmission and developmental regulation of a 150-kb yeast artificial chromosome containing the human beta-globin locus in transgenic mice. *Proc Natl Acad Sci U S A* ***90 (23)*** 1993. pp. 11381-11385.
**Galy, V.; Olivo-Marin, J.C.; Scherthan, H.; Doye, V.; Rascalou, N.; Nehrbass, U.** Nuclear pore complexes in the organization of silent telomeric chromatin. *Nature* ***403 (6765)*** 2000. pp. 108-112.
**Gasser, S.M.** Positions of potential: nuclear organization and gene expression. *Cell* ***104 (5)*** 2001. pp. 639-642.
**Georgakopoulos, T.; Thireos, G.** Two distinct yeast transcriptional activators require the function of the GCN5 protein to promote normal levels of transcription. *EMBO J* ***11 (11)*** 1992. pp. 4145-4152.
**Gerasimova, T.I.; Corces, V.G.** Polycomb and trithorax group proteins mediate the function of a chromatin insulator. *Cell* ***92 (4)*** 1998. pp. 511-521.
**Geyer, P.K.** The role of insulator elements in defining domains of gene expression. *Curr Opin Genetics Dev* ***7 (2)*** 1997. pp. 242-248.
**Giglioni, B.; Casisni, C.; Mantovani, R.; Merli, S.; Comi, P.; Ottolenghi, S.; Saglio, G.; Camaschella, C.; Mazza, U.** A molecular study of a family with Greek hereditary persistence of fetal hemoglobin and beta-thalassemia. *EMBO J* ***3 (11)*** 1984. pp. 2641-2645.
**Gillemans, N.; Tewari, R.; Lindeboom, F.; Rottier, R.; de Wit, T.; Wijgerde, M.; Grosveld, F.; Philipsen, S.** 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 (18)*** 1998. pp. 2863-2873.
**Gindhart, J.G.; Kaufman, T.C.** Identification of Polycomb and trithorax group responsive elements in the regulatory region of the Drosophila homeotic gene Sex combs reduced. *Genetics* ***139 (2)*** 1995. pp. 797-814.
**Gong, Q.H.; Dean, A.** Enhancer-dependent transcription of the epsilon-globin promoter requires promoter-bound GATA-1 and enhancer-bound AP-1NF-E2. *Annu Rev Cell Biol* ***13 (2)*** 1993. pp. 911-917.
**Gong, Q.H.; Stern, J.; Dean, A.** Transcriptional role of a conserved GATA-1 site in the human epsilon-globin gene promoter. *Mol Cell Biol* ***11 (5)*** 1991. pp. 2558-2566.
**Gong, Q.H.; McDowell, J.C.; Dean, A.** Essential role of NF-E2 in remodeling of chromatin structure and transcriptional activation of the epsilon-globin gene in vivo by 5' hypersensitive site 2 of the beta-globin locus control region. *Mol Cell Biol* ***16 (11)*** 1996. pp. 6055-6064.
**Goodrich, J.A.; Tjian, R.** Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. *Cell* ***77 (1)*** 1994. pp. 145-156.

**Goto, T.; Monk, M.** Regulation of X-chromosome inactivation in development in mice and humans. *Microbiol Mol Biol Rev* ***62 (2)*** 1998. pp. 362-378.
**Gotta, M.; Laroche, T.; Formenton, A.; Maillet, L.; Scherthan, H.; Gasser, S.M.** The clustering of telomeres and colocalization with Rap1, Sir3, and Sir4 proteins in wild-type Saccharomyces cerevisiae. *J Cell Biol* ***134 (6)*** 1996. pp. 1349-1363.
**Grant, P.A.; Berger, S.L.** Histone acetyltransferase complexes. *Sem Cell Dev Biol* ***10 (2)*** 1999. pp. 169-177.
**Greaves, D.R.; Wilson, F.D.; Lang, G.; Kioussis, D.** Human CD2 3'-flanking sequences confer high-level, T cell-specific, position-independent gene expression in transgenic mice. *Cell* ***56 (6)*** 1989. pp. 979-986.
**Green, E.D.; Reithman, H.C.; Dutchik, J.E.; Olson, M.V.** Detection and characterization of chimeric yeast artificial-chromosome clones. *Genomics* ***11 (3)*** 1991. pp. 658-669.
**Gregory, P.D.; Schmid, A.; Zavari, M.; Munsterkotter, M.; Horz, W.** Chromatin remodelling at the PHO8 promoter requires SWI-SNF and SAGA at a step subsequent to activator binding. *EMBO J* ***18 (22)*** 1999. pp. 6407-6414.
**Gribnau, J.; de Boer, E.; Trimborn, T.; Wijgerde, M.; Milot, E.; Grosveld, F.; Fraser, P.** Chromatin interaction mechanism of transcriptional control in vivo. *EMBO J* ***17 (20)*** 1998. pp. 6020-6027.
**Gribnau, J.; Diderich, K.; Pruzina, S.; Calzolari, R.; Fraser, F.** Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus. *Mol Cell* ***5 (2)*** 2000. pp. 377-386.
**Gross, D.S.; Garrard, W.T.** Nuclease hypersensitive sites in chromatin. *Annu Rev Biochem* ***57*** 1988. pp. 159-197.
**Grosveld, F.; van Assendelft, G.B.; Greaves, D.R.; Kollias, G.** Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* ***51 (6)*** 1987. pp. 975-985.
**Grosveld, F.** Activation by locus control regions? *Curr Opin Genetics Dev* ***9 (2)*** 1999. pp. 152-157.
**Groudine, M.; Kohwi-Shigematsu, T.; Gelinas, R.** Human fetal to adult hemoglobin switching: changes in chromatin structure of the beta-globin gene locus. *Proc Natl Acad Sci U S A* ***80 (24)*** 1983. pp. 7551-7555.
**Grozinger, C.M.; Hassig, C.A.; Schreiber, S.L.** Three proteins define a class of human histone deacetylases related to yeast Hda1p. *Proc Natl Acad Sci U S A* ***96 (9)*** 1999. pp. 4868-4873.
**Grunstein, M.; Durrin, L.K.; Man, R.K.; Fisher-Adams, G.; Johnson, L.M.** *Transcriptional regulation* 1992. pp. 1295-1315.
**Grunstein, M.** Yeast heterochromatin: regulation of its assembly and inheritance by histones. *Cell* ***93 (3)*** 1998. pp. 325-328.
**Gu, W.; Roeder, R.G.** Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell* ***90 (4)*** 1997. pp. 595-606.
**Gui, C.Y.; Dean, A.** Acetylation of a specific promoter nucleosome accompanies activation of the epsilon-globin gene by beta-globin locus control region HS2. *Mol Cell Biol* ***21 (4)*** 2001. pp. 1155-1163.
**Gumucio, D.L.; Shelton, D.A.; Bailey, W.J.; Slightom, J.L.; Goodman, M.** Phylogenetic footprinting reveals unexpected complexity in trans factor binding upstream from the epsilon-globin gene. *Proc Natl Acad Sci U S A* ***90 (13)*** 1993. pp. 6018-6022.
**Gumucio, D.L.; Rood, K.L.; Gray, T.A.; Riordan, M.F.; Sartor, C.I.; Collins, F.S.** Nuclear proteins that bind the human gamma-globin gene promoter: alterations in binding produced by point mutations associated with hereditary persistence of fetal hemoglobin. *Mol Cell Biol* ***8 (12)*** 1988. pp. 5310-5322.
**Hagstrom, K.; Muller, M.; Schedl, P.** A Polycomb and GAGA dependent silencer adjoins the Fab-7 boundary in the Drosophila bithorax complex. *Genetics* ***146 (4)*** 1997. pp. 1365-1380.
**Hamiche, A.; Sandaltzopoulos, R.; Gdula, D.A.; Wu, C.** ATP-dependent histone octamer sliding mediated by the chromatin remodeling complex NURF. *Cell* ***97 (7)*** 1999. pp. 833-842.
**Hamilton, C.M.; Aldea, M.; Washburn, B.K.; Babitzke, P.; Kushner, S.R.** New method for generating deletions and gene replacements in Escherichia coli. *J Bacteriol* ***171 (9)*** 1989. pp. 4617-4622.
**Hanscombe, O.; Vidal, M.; Kaeda, J.; Luzzatto, L.; Greaves, D.R.; Grosveld, F.** High-level, erythroid-specific expression of the human alpha-globin gene in transgenic mice and the production of human hemoglobin in murine erythrocytes. *Genes Dev* ***3 (10)*** 1989. pp. 1572-1581.
**Hanscombe, O.; Whyatt, D.; Fraser, P.; Yannoutsos, N.; Greaves, D.; Dillon, N.; Grosveld, F.** Importance of globin gene order for correct developmental expression. *Genes Dev* ***5 (8)*** 1991. pp. 1387-1394.
**Hardison, R.C.** A brief history of hemoglobins: plant, animal, protist, and bacteria. *Proc Natl Acad Sci U S A* ***93 (12)*** 1996. pp. 5675-5679.
**Hartzog, G.A.; Myers, R.M.** Discrimination among potential activators of the beta-globin CACCC element by correlation of binding and transcriptional properties. *Mol Cell Biol* ***13 (1)*** 1993. pp. 44-56.
**Haswell, E.S.; O Shea, E.K.** An in vitro system recapitulates chromatin remodeling at the PHO5 promoter. *Mol Cell Biol* ***19 (4)*** 1999. pp. 2817-2827.
**Haupt, Y.; Alexander, W.S.; Barri, G.; Klinken, S.P.; Adams, J.M.** Novel zinc finger gene implicated as myc collaborator by retrovirally accelerated lymphomagenesis in E mu-myc transgenic mice. *Cell* ***65 (5)*** 1991. pp. 753-763.
**Hayes, J.J.; Tullius, T.D.; Wolffe, A.P.** The structure of DNA in a nucleosome. *Proc Natl Acad Sci U S A* ***87 (19)*** 1990. pp. 7405-7409.
**Hayes, J.J.; Clark, D.J.; Wolffe, A.P.** Histone contributions to the structure of DNA in the nucleosome. *Proc Natl Acad Sci U S A* ***88 (15)*** 1991. pp. 6829-6833.
**Hebbes, T.R.; Clayton, A.L.; Thorne, A.W.; Crane-Robinson, C.** Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. *EMBO J* ***13 (8)*** 1994. pp. 1823-1830.
**Heintz, E** - *Botanik* ***69*** 1928. pp. 762-818.
**Held, W.; Kunz, B.; Ioannidis, V.; Lowin-Kropf, B.** Mono-allelic Ly49 NK cell receptor expression. *Semin Immunol* ***11 (5)*** 1999. pp. 349-355.
**Henikoff, S.** Nuclear organization and gene expression: homologous pairing and long-range interactions. *Curr Opin Cell Biol* ***9 (3)*** 1997. pp. 388-395.
**Herendeen, D.R.; Kassavetis, G.A.; Geiduschek, E.P.** A transcriptional enhancer whose function imposes a requirement that proteins track along DNA. *Science* ***256 (5061)*** 1992. pp. 1298-1303.
**Herr, W.** *Transcriptional regulation* 1992. pp. ?-?.
**Higgs, D.; Wood, W.G.; Jarman, A.; Sharpe, J.; Lida, J.; Pretorius, J.M.; Ayyub, H.** A major positive regulatory region located far upstream of the human alpha-globin gene locus. *Genes Dev* ***4 (9)*** 1990. pp. 1588-1601.
**Higgs, D.** Do LCRs open chromatin domains? *Cell* ***95 (3)*** 1998. pp. 299-302.
**Higgs, D.R.; Sharpe, J.A.; Wood, W.G.** Understanding alpha globin gene expression: a step towards effective gene therapy. *Semin Hematol* ***35 (2)*** 1998. pp. 93-104.
**Hollander, G.A.; Zuklys, S.; Morel, C.; Mizoguchi, E.; Mobisson, K.; Simpson, S.; Terhorst, C.; Wishart, W.; Golan, D.E.; Bhan, A.K.; Burakoff, S.J.** Monoallelic expression of the interleukin-2 locus. *Science* ***279 (5359)*** 1998. pp. 2118-2121.
**Holstege, F.C.; Jennings, E.G.; Wyrick, J.J.; Lee, T.I.; Hengarter, C.J.; Green, M.R.; Golub, T.R.; Lander, E.S.; Young, R.A.** Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* ***95 (5)*** 1998. pp. 717-728.
**Horikoshi, M.; Carey, M.F.; Kakidani, H.; Roeder, R.G.** Mechanism of action of a yeast activator: direct effect of GAL4 derivatives on mammalian TFIID-promoter interactions. *Cell* ***54 (5)*** 1988. pp. 665-669.

**Hosbach, H.A.; Wyler, T.; Weber, R.** The Xenopus laevis globin gene family: chromosomal arrangement and gene structure. *Cell* ***32 (1)*** 1983. pp. 45-53.
**Hu-Li, J.; Pannetier, C.; Guo, L.; Lohning, M.; Gu, H.; Watson, C.; Assenmacher, M.; Radbruch, A.; Paul, W.E.** Regulation of expression of IL-4 alleles: analysis using a chimeric GFPIL-4 gene. *Immunity* ***14 (1)*** 2001. pp. 1-11.
**Huang, H.C.; Schwyter, D.H.; Shirokawa, J.M.; Courey, A.J.** The interplay between multiple enhancer and silencer elements defines the pattern of decapentaplegic expression. *Genes Dev* ***7 (4)*** 1993. pp. 694-704.
**Huang, Y.; Liu, D.P.; Wu, L.; Li, T.C.; Feng, D.X.; Liang, C.C.** Proper developmental control of human globin genes reproduced by transgenic mice containing a 160-kb BAC carrying the human beta-globin locus. *Blood Cells Mol Dis* ***26 (6)*** 2000. pp. 598-610.
**Huehns, F.; Dance, N.; Beaven, G.H.; Keil, J.V.; Hecht, F.; Motulsky, A.G.** - *Nature* ***201*** 1964. pp. 1095-?.
**Hug, B.; Moon, A.; Ley, T.** Structure and function of the murine beta-globin locus control region 5' HS-3. *Nucleic Acids Res* ***20 (21)*** 1992. pp. 5771-5778.
**Hug, B.A.; Wesseschmidt, R.L.; Fiering, S.; Bender, M.A.; Epner, E.; Groudine, M.; Ley, T.J.** Analysis of mice containing a targeted deletion of beta-globin locus control region 5' hypersensitive site 3. *Mol Cell Biol* ***16 (6)*** 1996. pp. 2906-2912.
**Hughes, S.; Stubblefield, E.; Payvar, F.; Engel, J.D.; Dodgson, J.B.; Cordell, B.; Schmike, R.T.; Varmus, H.E.** Gene localization by chromosome fractionation: globin genes are on at least two chromosomes and three estrogen-inducible genes are on three chromosomes. *Proc Natl Acad Sci U S A* ***76 (3)*** 1979. pp. 1348-1352.
**Huisman, T.H.; Carver, M.F.** The beta- and delta-thalassemia repository (Ninth Edition: Part I). *Hemoglobin* ***22 (2)*** 1998. pp. 169-195.
**Huisman, T.H.; Schroeder, W.A.; Efremov, G.D.; Duma, H.; Mladenovski, B.; Hyman, C.B.; Rachemilewitz, E.A.; Bouver, N.; Miller, A.; Brodie, A.; Shelton, J.R.; Apell, G.** The present status of the heterogeneity of fetal hemoglobin in beta-thalassemia: an attempt to unify some observations in thalassemia and related conditions. *Ann NY Acad Sci* ***232 (0)*** 1974. pp. 107-124.
**Hume, D.A.** Probability in transcriptional regulation and its implication for leukocyte differentiation and inducible gene expression *Blood* ***96*** 2000. pp. 2323-2328.
**Humphries, R.K.; Ley, T.; Turner, P.; Moulton, A.D.; Nienhuis, A.W.** Differences in human alpha-, beta- and delta-globin gene expression in monkey kidney cells. *Cell* ***30 (1)*** 1982. pp. 173-183.
**Hunter, T.; Karin, M.** The regulation of transcription by phosphorylation. *Cell* ***70 (3)*** 1992. pp. 375-387.
**Iguchi-Ariga, S.M.; Schaffner, W.** CpG methylation of the cAMP-responsive enhancerpromoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes Dev* ***3 (5)*** 1989. pp. 612-619.
**Ikuta, T.; Papayannopoulou, T.; Stamatoyannopoulos, G.; Kan, Y.W.** Globin gene switching. In vivo protein-DNA interactions of the human beta-globin locus in erythroid cells expressing the fetal or the adult globin gene program. *J Biol Chem* ***271 (24)*** 1996. pp. 14082-14091.
**Ikuta, T.; Kan, Y.W.** In vivo protein-DNA interactions at the beta-globin gene locus. *Proc Natl Acad Sci U S A* ***88 (22)*** 1991. pp. 10188-10192.
**Imai, S.; Armstrong, C.M.; Kaeberlein, M.; Guarente, L.** Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature* ***403 (6771)*** 2000. pp. 795-800.
**Imam, A.; Patrinos, G.; de Krom, M.; Bottardi, S.; Janssens, R.; Katsantoni, E.; Wai, A.; Sheratt, D.; Grosveld, F.** Modification of human beta-globin locus PAC clones by homologous recombination in Escherichia coli. *Nucleic Acids Res* ***28 (12)*** 2000. pp. E65-E65.
**Ioannou, P.; De Jong, P.** *Current Protocols in Human Genetics. John* 1996. pp. ?-?.
**Ioannou, P.A.; Amemiya, C.T.; Garnes, K. J.; Kroisel, P.M.; Shizuya, H.; Chen, C.; Batzer, B.A.; de Jong, P.J.** A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genet* ***6 (1)*** 1994. pp. 84-89.
**Ito, E.; Toki, T.; Ishihara, H.; Ohtani, H.; Gu, L.; Yoloyama, M.; Engel, J.D.; Yamamoto, M.** Erythroid transcription factor GATA-1 is abundantly transcribed in mouse testis. *Nature* ***362 (6419)*** 1993. pp. 466-468.
**Ito, T.; Bulger, M.; Pazin, M.J.; Kobayashi, R.; Kadonaga, J.T.** ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell* ***90 (1)*** 1997. pp. 145-155.
**Ito, T.; Levenstein, M.E.; Fyodorov, D.V.; Kutach, A.K.; Kobayashi, R.; Kadonaga, J.T.** ACF consists of two subunits, Acf1 and ISWI, that function cooperatively in the ATP-dependent catalysis of chromatin assembly. *Genes Dev* ***13 (12)*** 1999. pp. 1529-1539.
**Jacobs, J.J.; van Lohuizen, M.** Polycomb repression: from cellular memory to cellular proliferation and cancer. *Biochim Biophys Acta* ***1602 (2)*** 2002. pp. 151-161.
**Jacobs, J.J.; Scheijen, B.; Voncken, J.W.; Kieboom, K.; Berns, A.; van Lohuizen, M.** Bmi-1 collaborates with c-Myc in tumorigenesis by inhibiting c-Myc-induced apoptosis via INK4aARF. *Genes Dev* ***13 (20)*** 1999. pp. 2678-2690.
**Jahn, C.L.; Hutchison III, C.A.; Philips, S.J.; Weaver, S.; Haigwood, N.L.; Voliva, C.F.; Edgell, M.H.** DNA sequence organization of the beta-globin complex in the BALBc mouse. *Cell* ***21 (1)*** 1980. pp. 159-168.
**Jane, S.M.; Gumucio, D.L.; Ney, P.A.; Cunningham, J.M.; Nienhuis, A.W.** Methylation-enhanced binding of Sp1 to the stage selector element of the human gamma-globin gene promoter may regulate development specificity of expression. *Mol Cell Biol* ***13 (6)*** 1993. pp. 3272-3281.
**Jane, S.M.; Ney, P.A.; Vanin, E.F.; Gumucio, D.L.; Nienhuis, A.W.** Identification of a stage selector element in the human gamma-globin gene promoter that fosters preferential interaction with the 5' HS2 enhancer when in competition with the beta-promoter. *EMBO J* ***11 (8)*** 1992. pp. 2961-2969.
**Jarman, A.P.; Higgs, D.R.** Nuclear scaffold attachment sites in the human globin gene complexes. *EMBO J* ***7 (11)*** 1988. pp. 3337-3344.
**Jarman, A.P.; Wood, W.G.; Sharpe, J.A.; Gourdon, G.; Ayyub, H.; Higgs, D.R.** Characterization of the major regulatory element upstream of the human alpha-globin gene cluster. *Mol Cell Biol* ***11 (9)*** 1991. pp. 4679-4689.
**Jeppesen, P.; Turner, B.M.** The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. *Cell* ***74 (2)*** 1993. pp. 281-289.
**Johnson, P.F.; McKnight, S.L.** Eukaryotic transcriptional regulatory proteins. *Annu Rev Biochem* ***58*** 1989. pp. 799-839.
**Jones, P.L.; Veenstra, G.J.; Wade, P.A.; Vermaak, D.; Kass, S.U.; Landsberger, N.; Strouboulis, J.; Wolffe, A.P.** Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genet* ***19 (2)*** 1998. pp. 187-191.
**Junewein, T.; Allis, C.D.** Translating the histone code. *Science* ***293 (5532)*** 2001. pp. 1074-1080.
**Kadam, S.; McAlpine, G.S.; Phelan, M.L.; Kingston, R.E.; Jones, K.A.; Emerson, B.M.** Functional selectivity of recombinant mammalian SWISNF subunits *Genes Dev* ***14 (19)*** 2000. pp. 2441-2451.
**Kal, A.J; Mahmoudi, T.; Zak, N.B.; Verrijzer, P.** The Drosophila brahma complex is an essential coactivator for the trithorax group protein zeste. *Genes Dev* ***14 (9)*** 2000. pp. 1058-1071.
**Kass, S.U.; Landsberger, N.; Wolffe, A.P.** DNA methylation directs a time-dependent repression of transcription initiation. *Curr Biol* ***7 (3)*** 1997. pp. 157-165.
**Kaufman, R.M.; Pham, C.T.N.; Ley, T.J.** Transgenic analysis of a 100-kb human beta-globin cluster-containing DNA fragment propagated as a bacterial artificial chromosome. *Blood* ***94 (9)*** 1999. pp. 3178-3184.
**Kazminova, E.; Selker, E.U.** dim-2 encodes a DNA methyltransferase responsible for all known cytosine methylation in Neurospora. *EMBO J* ***20 (15)*** 2001. pp. 4309-4320.

**Kellum, R.; Schedl, P.** A position-effect assay for boundaries of higher order chromosomal domains. *Cell* ***64 (5)*** 1991. pp. 941-950.
**Kellum, R.; Schedl, P.** A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Mol Cell Biol* ***12 (5)*** 1992. pp. 2424-2431.
**Kennison, J.A.** The Polycomb and trithorax group proteins of Drosophila: trans-regulators of homeotic gene function. *Annu Rev Genet* ***29*** 1995. pp. 289-303.
**Kennison, J.A.; Tamkun, J.W.** Dosage-dependent modifiers of polycomb and antennapedia mutations in Drosophila. *Proc Natl Acad Sci U S A* ***85 (21)*** 1988. pp. 8136-8140.
**Keshet, I.; Lieman-Hurwitz, J.; Cedar, H.** DNA methylation affects the formation of active chromatin. *Cell* ***44 (4)*** 1986. pp. 535-543.
**Khavari, P.A.; Peterson, C.L.; Tamkun, J.W.; Mendel, D.B.; Crabtree, G.R.** BRG1 contains a conserved domain of the SWI2SNF2 family necessary for normal mitotic growth and transcription. *Nature* ***366 (6451)*** 1993. pp. 170-174.
**Kieffer, L.J.; Greally, J.M.; Landres, I.; Nag, S.; Nakajima, Y.; Kohwi-Shigematsu, T.; Kavathas, P.B.** Identification of a candidate regulatory region in the human CD8 gene complex by colocalization of DNase I hypersensitive sites and matrix attachment regions which bind SATB1 and GATA-3. *J Immunol* ***168 (8)*** 2002. pp. 3915-3922.
**Kim, C.G.; Swendeman, S.L.; Barnhart, K.M.; Sheffery, M.** Promoter elements and erythroid cell nuclear factors that regulate alpha-globin gene transcription in vitro. *Mol Cell Biol* ***10 (11)*** 1990. pp. 5958-5966.
**Kim, M.K.; Lesoon-wood, L.A.; Weintraub, B.D.; Chung, J.H.** A soluble transcription factor, Oct-1, is also found in the insoluble nuclear matrix and possesses silencing activity in its alanine-rich domain. *Mol Cell Biol* ***16 (8)*** 1996. pp. 4366-4377.
**Kioussis, D.; Vanin, E.; de Lange, T.; Flavell, R.A.; Grosveld, F.G.** Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. *Nature* ***306 (5944)*** 1983. pp. 662-666.
**Kioussis, D.; Festenstein, R.** Locus control regions: overcoming heterochromatin-induced gene inactivation in mammals. *Curr Opin Genetics Dev* ***7 (5)*** 1997. pp. 614-619.
**Kitsberg, D.; Selig, S.; Keshet, I.; Cedar, H.** Replication structure of the human beta-globin gene domain. *Nature* ***366 (6455)*** 1993. pp. 588-590.
**Knoll, W.; Pingel, E.** - *Acta Haematol* ***2*** 1949. pp. 369-369.
**Ko, M.S.H.** A stochastic model for gene induction. *J Theor Biol* ***153 (2)*** 1990. pp. 181-194.
**Kollias, G.; Wrighton, N.; Hurst, J.; Grosveld, F.** Regulated expression of human A gamma-, beta-, and hybrid gamma beta-globin genes in transgenic mice: manipulation of the developmental expression patterns. *Cell* ***46 (1)*** 1986. pp. 89-94.
**Kong, S.; Bohl, D.; Li, C.; Tuan, D.** 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 (7)*** 1997. pp. 3955-3965.
**Kornberg, R.** Chromatin structure: a repeating unit of histones and DNA. *Science* ***184 (139)*** 1974. pp. 868-871.
**Kosteas, G.; Manifava, M.; Moschonas, N.; Anagnou, N.P.** Functional analysis of the Agamma to pseudo-beta globin gene region of the human beta-locus: evidence for negative regulatory elements. *Clin Res* ***41*** 1993. pp. 308-308.
**Kosteas, G.; Manifava, M.; Moschonas, N.; Anagnou, N.P.** Functional analysis of the Agamma to delta globin gene region of the beta-cluster: evidence for multiple negative regulatory elements *Blood* ***84*** 1994. pp. 506-506.
**Kouzarides, T.** Histone methylation in transcriptional control. *Curr Opin Genetics Dev* ***12 (2)*** 2002. pp. 198-209.
**Kulessa, H.; Frampton, J.; Graf, T.** GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblasts, and erythroblasts. *Genes Dev* ***9 (10)*** 1995. pp. 1250-1262.
**Kulozik, A.E.; Bellan-Koch, A.; Bail, S.; Kohne, E.; Kleihauer, E.** Thalassemia intermedia: moderate reduction of beta globin gene transcriptional activity by a novel mutation of the proximal CACCC promoter element. *Blood* ***77 (9)*** 1991. pp. 2054-2058.
**Kulozik, A.E.; Bail, S.; Bellan-Koch, A.; Bartram, C.R.; Kohne, E.; Kleihauer, E.** The proximal element of the beta globin locus control region is not functionally required in vivo. *J Clin Invest* ***87 (6)*** 1991. pp. 2142-2146.
**Kuo, M.H.; Brownell, J.E.; Sobel, R.E.; Ranalli, T.A.; Cook, R.G.; Edmonson, D.G.; Roth, S.Y.; Allis, C.D.** Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines. *Nature* ***383 (6597)*** 1996. pp. 269-272.
**Kuo, M.H.; Zhou, J.; Jambeck, P.; Churchill, M.E.; Allis, C.D.** Histone acetyltransferase activity of yeast Gcn5p is required for the activation of target genes in vivo. *Genes Dev* ***12 (5)*** 1998. pp. 627-639.
**Kwon, H.; Imbalzano, A.N.; Khavari, P.A.; Kingston, R.E.; Green, M.R.** Nucleosome disruption and enhancement of activator binding by a human SW1SNF complex. *Nature* ***370 (6489)*** 1994. pp. 477-481.
**L'Hernault, S.W.; Rosenbaum, J.L.** Chlamydomonas alpha-tubulin is posttranslationally modified by acetylation on the epsilon-amino group of a lysine. *Biochemistry* ***24 (2)*** 1985. pp. 473-478.
**Lachner, M.; O Carroll, D.; Rea, S.; Mechtler, K.; Junewein, T.** Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* ***410 (6824)*** 2001. pp. 116-120.
**Laemmli, U.K.; Kas, E.; Poljak, L.; Adachi, Y.** Scaffold-associated regions: cis-acting determinants of chromatin structural loops and functional domains. *Curr Opin Genetics Dev* ***2 (2)*** 1992. pp. 275-285.
**Langst, G.; Bonte, E.J.; Corona, D.F.; Becker, P.B.** Nucleosome movement by CHRAC and ISWI without disruption or trans-displacement of the histone octamer. *Cell* ***97 (7)*** 1999. pp. 843-852.
**Langst, G.; Becker, P.B.; Wu, C.** Nucleosome mobilization and positioning by ISWI-containing chromatin-remodeling factors. *J Cell Sci* ***114*** 2001. pp. 2561-2568.
**Larschan, E.; Winston, F.** The S. cerevisiae SAGA complex functions in vivo as a coactivator for transcriptional activation by Gal4 *Genes Dev* ***15 (15)*** 2001. pp. 1946-1956.
**Leder, A.; Weir, L.; Leder, P.** Characterization, expression, and evolution of the mouse embryonic zeta-globin gene. *Mol Cell Biol* ***5 (5)*** 1985. pp. 1025-1033.
**Lederberg, J.** Streptomycin resistance: a genetically recessive mutation *J Bacteriol* ***61*** 1951. pp. 549-550.
**Lee, C-H.; Murphy, M.R.; Lee, J-S.; Chung, J.H.** Targeting a SWISNF-related chromatin remodeling complex to the beta-globin promoter in erythroid cells. *Proc Natl Acad Sci U S A* ***96 (22)*** 1999. pp. 12311-12315.
**Lee, D.Y.; Hayes, J.J.; Pruss, D.; Wolffe, A.P.** A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell* ***72 (1)*** 1993. pp. 73-84.
**Lee, J.S.; Ngo, H.; Kin, D.; Chung, H.** Erythroid Kruppel-like factor is recruited to the CACCC box in the beta-globin promoter but not to the CACCC box in the gamma-globin promoter: the role of the neighboring promoter elements. *Proc Natl Acad Sci U S A* ***97 (6)*** 2000. pp. 2468-2473.
**Lei, H.; Oh, S.P.; Okano, M.; Jutterman, R.; Goss, K.A.; Jeanisch, R.; Li, E.** De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* ***122 (10)*** 1996. pp. 3195-3205.
**Leipe, D.D.; Landsman, D.** Histone deacetylases, acetoin utilization proteins and acetylpolyamine amidohydrolases are members of an ancient protein superfamily. *Nucleic Acids Res* ***25 (18)*** 1997. pp. 3693-3697.
**Lewis, J.D.; Meehan, R.R.; Henzel, W.J.; Maurer-Vogy, I.; Jeppesen, P.; Klein, G.; Bird, A.P.** Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* ***96 (6)*** 1992. pp. 905-914.
**Li, Q.; Stamatoyannopoulos, G.** - *Mol Cell Biol* ***14*** 1994. pp. 6087-6087.

**Li, Q.; Blau, C.A.; Clegg, C.H.; Rohde, A.; Stamatoyannopoulos, G.** Multiple epsilon-promoter elements participate in the developmental control of epsilon-globin genes in transgenic mice. *J Biol Chem* ***273 (28)*** 1998. pp. 17361-17367.
**Li, Q.; Zhang, M.; Stamatoyannopoulos, G.** Structural analysis and mapping of DNase I hypersensitivity of HS5 of the beta-globin locus control region. *Genomics* ***61 (2)*** 1999. pp. 183-193.
**Li, Q.; Stamatoyannopoulos, G.** Hypersensitive site 5 of the human beta locus control region functions as a chromatin insulator. *Blood* ***84 (5)*** 1994. pp. 1399-1401.
**Liebhaber, S.A.; Wang, Z.; Cash, F.E.; Monks, B.; Russell, J.E.** Developmental silencing of the embryonic zeta-globin gene: concerted action of the promoter and the 3'-flanking region combined with stage-specific silencing by the transcribed segment. *Mol Cell Biol* ***16 (6)*** 1996. pp. 2637-2646.
**Lim, L.C.; Swendeman, S.L.; Sheffery, M.** Molecular cloning of the alpha-globin transcription factor CP2. *Mol Cell Biol* ***12 (2)*** 1992. pp. 828-835.
**Lim, L.C.; Fang, L.; Swendeman, S.L.; Sheffery, M.** Characterization of the molecularly cloned murine alpha-globin transcription factor CP2. *J Biol Chem* ***286 (24)*** 1993. pp. 18008-18017.
**Lin, Y.S.; Green, M.R.** Mechanism of action of an acidic transcriptional activator in vitro. *Cell* ***64 (5)*** 1991. pp. 971-981.
**Lindenbaum, M.H.; Grosveld, F.** An in vitro globin gene switching model based on differentiated embryonic stem cells. *Genes Dev* ***4 (12)*** 1990. pp. 2075-2085.
**Litt, M.D.; Simpson, M.; Gaszner, M.; Allis, C.D.; Felsenfeld, G.** Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus. *Science* ***293 (5539)*** 2001. pp. 2453-2455.
**Liu, B.; Maul, R.S.; Kaetzel, D.M.** Repression of platelet-derived growth factor A-chain gene transcription by an upstream silencer element. Participation by sequence-specific single-stranded DNA-binding proteins. *J Biol Chem* ***271 (42)*** 1996. pp. 26281-26290.
**Liu, Q.; Bungert, J.; Engel, J.D.** 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 (1)*** 1997. pp. 169-174.
**Liu, Q.; Chang, J.C.; Moi, P.; Liu, W.; Kan, Y.W.; Curtin, P.T.** Dissection of the enhancer activity of beta-globin 5' DNase I-hypersensitive site 2 in transgenic mice. *Proc Natl Acad Sci U S A* ***89 (9)*** 1992. pp. 3899-3899.
**Liu, Q.; Tanimoto, K.; Bungert, J.; Engel, J.D.** The A gamma-globin 3' element provides no unique function(s) for human beta-globin locus gene regulation. *Proc Natl Acad Sci U S A* ***95 (17)*** 1998. pp. 9944-9949.
**Lo, W.S.; Trievel, R.C.; Rojas, J.R.; Duggan, L.; Hsu, J.Y.; Allis, C.D.; Marmorstein, R.; Berger, S.L.** Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to Gcn5-mediated acetylation at lysine 14. *Mol Cell* ***5 (6)*** 2000. pp. 917-926.
**Locke, J.; Kotarski, M.A.; Tartof, K.D.** Dosage-dependent modifiers of position effect variegation in Drosophila and a mass action model that explains their effect. *Genetics* ***120 (1)*** 1988. pp. 181-198.
**Lohuizen, M.V.** The trithorax-group and polycomb-group chromatin modifiers: implications for disease. *Curr Opin Genetics Dev* ***9 (3)*** 1999. pp. 355-361.
**Lowrey, C.H.; Bodine, D.M.; Nienhuis, A.W.** Mechanism of DNase I hypersensitive site formation within the human globin locus control region. *Proc Natl Acad Sci U S A* ***89 (3)*** 1992. pp. 1143-1147.
**Lu, S.J.; Rowan, S.; Bani, M.R.; Ben-David, Y.** 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 (18)*** 1994. pp. 8398-8402.
**Luger, K.; Mader, A.W.; Richmond, R.K.; Sargent, D.F.; Richmond, T.J.** Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* ***389 (6648)*** 1997. pp. 251-260.
**Lund, A.H.; van Lohuizen, M.** RUNX: a trilogy of cancer genes. *Cancer Cell* ***1 (3)*** 2002. pp. 213-215.
**Lundgren, M.; Chow, C.M.; Sabbattini, P.; Georgiou, A.; Minaee, S.; Dillon, N.A.** Transcription factor dosage affects changes in higher order chromatin structure associated with activation of a heterochromatic gene. *Cell* ***103 (5)*** 2000. pp. 733-743.
**Lyko, F.; Ramsahoye, B.H.; Kashevsky, H.; Tudor, M.; Mastrangelo, M.A.; Orr-Weaver, T.L.; Jaenisch, R.** Mammalian (cytosine-5) methyltransferases cause genomic DNA methylation and lethality in Drosophila. *Nature Genet* ***23 (3)*** 1999. pp. 363-366.
**Ma, H.; Bauman, C.T.; Li, I.H.; Strahl, B.D.; Rice, R.; Jelnek, M.A.; Aswad, D.W.; Allis, C.D.; Hager, G.L.; Stallcup, M.R.** Hormone-dependent, CARM1-directed, arginine-specific methylation of histone H3 on a steroid-regulated promoter. *Curr Biol* ***11 (24)*** 2001. pp. 1981-1985.
**Mantovani, R.; Malgaretti, N.; Nicolis, S.; Ronchi, A.; Giglioni, B.; Ottolenghi, S.** The effects of HPFH mutations in the human gamma-globin promoter on binding of ubiquitous and erythroid specific nuclear factors. *Nucleic Acids Res* ***16 (16)*** 1988. pp. 7783-7797.
**Marin, M.; Karis, A.; Visser, P.; Grosveld, F.; Philipsen, S.** Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. *Cell* ***89 (4)*** 1997. pp. 619-628.
**Marizio, G.; Wagener, C.; Gutierrez, M.I.; Cartwright, P.; Helin, K.; Giacca, M.** E2F family members are differentially regulated by reversible acetylation. *J Biol Chem* ***275 (15)*** 2000. pp. 10887-10892.
**Martens, J.H.; Verlaan, M.; Kalkhoven, E.; Dorsman, J.C.; Zantema, A.** Scaffoldmatrix attachment region elements interact with a p300-scaffold attachment factor A complex and are bound by acetylated nucleosomes. *Mol Cell Biol* ***22 (8)*** 2002. pp. 2598-5606.
**Martin, D.I.K.; Tsai, S.F.; Orkin, S.H.** Increased gamma-globin expression in a nondeletion HPFH mediated by an erythroid-specific DNA-binding factor. *Nature* ***338 (6214)*** 1989. pp. 435-438.
**Martin, D.I.K.; Zon, L.I.; Mutter, G.; Orkin, S.H.** Expression of an erythroid transcription factor in megakaryocytic and mast cell lineages. *Nature* ***344 (6265)*** 1990. pp. 444-447.
**Martin, D.I.K.; Fiering, S.; Groudine, M.** Regulation of beta-globin gene expression: straightening out the locus. *Curr Opin Genetics Dev* ***6 (4)*** 1996. pp. 488-495.
**Martin, S.L.; Vincent, K.A.; Wilson, A.C.** Rise and fall of the delta globin gene. *J Mol Biol* ***164 (4)*** 1983. pp. 513-528.
**Martinez-Balbas, M.; Bauer, U.M.; Nielsen, S.J.; Brehm, A.; Kouzarides, T.** Regulation of E2F1 activity by acetylation. *EMBO J* ***19 (4)*** 2000. pp. 662-671.
**Matsuo, K.; Caly, O.; Takahashi, T.; Silke, J.; Schaffner, W.** Evidence for erosion of mouse CpG islands during mammalian evolution. *Somat Cell Mol Genet* ***19 (6)*** 1993. pp. 543-555.
**Maxon, M.; Goodrich, J.A.; Tjian, R.** Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. *Genes Dev* ***8 (5)*** 1994. pp. 515-524.
**McMorrow, T.; van den Wijngaard, A.; Wollenschlaeger, A.; van de Corput, M.; Monkhorst, K.; Trimborn, T.; Fraser, P.; van Lohuizen, M.; Junewein, T.; Djabali, M.; Philipsen, S.; Grosveld, F.; Milot, E.** Activation of the beta globin locus by transcription factors and chromatin modifiers. *EMBO J* ***19 (18)*** 2000. pp. 4986-4996.
**McPherson, C.E.; Shim, E.Y.; Friedman, D.S.; Zaret, K.S.** An active tissue-specific enhancer and bound transcription factors existing in a precisely positioned nucleosomal array. *Cell* ***75 (2)*** 1993. pp. 387-398.
**Medvinsky, A.; Dzierzak, E.** Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell* ***86 (6)*** 1996. pp. 897-906.
**Medvinsky, A.; Samoylina, N.L.; Muller, A.M.; Dzierzak, E.A.** An early pre-liver intraembryonic source of CFU-S in the developing mouse. *Nature* ***364 (6432)*** 1993. pp. 64-67.

**Merika, M.; Orkin, S.H.** Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Kruppel family proteins Sp1 and EKLF. *Mol Cell Biol* ***15 (5)*** 1995. pp. 2437-2447.
**Metcalf, D.** Lineage commitment and maturation in hematopoietic cells: the case for extrinsic regulation. *Blood* ***92 (2)*** 1998. pp. 345-347.
**Michaelson, J.** Cellular selection in the genesis of multicellular organization. *Lab Invest* ***69 (2)*** 1993. pp. 136-151.
**Mignotte, V.; Wall, L.; de Boer, E.; Grosveld, F.; Romeo, P.H.** Two tissue-specific factors bind the erythroid promoter of the human porphobilinogen deaminase gene. *Nucleic Acids Res* ***17 (1)*** 1989. pp. 37-54.
**Miller, I.J.; Bieker, J.J.** 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 (5)*** 1993. pp. 2776-2786.
**Milot, E.; Strouboulis, J.; Trimborn, T.; Wijgerde, M.; de Boer, E.; Langeveld, A.; Tan-Un, K.; Vergeer, W.; Yannoutsos, N.; Grosveld, F.; Fraser, P.** Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell* ***87 (1)*** 1996. pp. 105-114.
**Mirkovitch, J.; Mirault, M.E.; Laemmli, U.K.** Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. *Cell* ***39 (1)*** 1984. pp. 223-232.
**Mishima, N.; Landman, H.; Huisman, T.H.; Gilman, J.G.** The DNA deletion in an Indian delta beta-thalassaemia begins one kilobase from the A gamma globin gene and ends in an L1 repetitive sequence. *Br J Haematol* ***73 (3)*** 1989. pp. 375-379.
**Misteli, T.** Protein dynamics: implications for nuclear architecture and gene expression. *Science* ***291 (5505)*** 2001. pp. 843-847.
**Mizuguchi, G.; Tsukiyama, T.; Wisniewski, J.; Wu, C.** Role of nucleosome remodeling factor NURF in transcriptional activation of chromatin. *Mol Cell* ***1 (1)*** 1997. pp. 141-150.
**Mizuguchi, G.; Vassilev, A.; Tsukiyama, T.; Nakatani, Y.; Wu, C.** ATP-dependent nucleosome remodeling and histone hyperacetylation synergistically facilitate transcription of chromatin. *J Biol Chem* ***276 (18)*** 2001. pp. 14773-14783.
**Moi, P.; Kan, Y.W.** Synergistic enhancement of globin gene expression by activator protein-1-like proteins. *Proc Natl Acad Sci U S A* ***87 (22)*** 1990. pp. 9000-9004.
**Molete, J.M.; Petrykowska, H.; Bouhassira, E.E.; Feng, Y.; Miller, W.; Hardison, R.C.** Sequences flanking hypersensitive sites of the beta-globin locus control region are required for synergistic enhancement. *Mol Cell Biol* ***21 (9)*** 2001. pp. 2969-2980.
**Moon, A.M.; Ley, T.J.** 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 (19)*** 1990. pp. 7693-7697.
**Moore, M.A.; Metcalf, D.** Ontogeny of the haemopoietic system: yolk sac origin of in vivo and in vitro colony forming cells in the developing mouse embryo. *Br J Haematol* ***18 (3)*** 1970. pp. 279-296.
**Moreau, P.; Hen, R.; Wasylyk, B.; Everett, R.; Gaub, M.P.; Chambon, P.** The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res* ***9 (22)*** 1981. pp. 6047-6068.
**Morley, B.J.; Abbot, C.A.; Sharpe, J.; Lida, P.S.; Chan-Thomas, X.; Wood, W.G.** A single beta-globin locus control region element (5' hypersensitive site 2) is sufficient for developmental regulation of human globin genes in transgenic mice. *Mol Cell Biol* ***12 (5)*** 1992. pp. 2057-2066.
**Muchardt, C.; Yaniv, M.** A human homologue of Saccharomyces cerevisiae SNF2SWI2 and Drosophila brm genes potentiates transcriptional activation by the glucocorticoid receptor. *EMBO J* ***12 (11)*** 1993. pp. 4279-4290.
**Mueller-Storm, H.P.; Sogo, J.M.; Schaffner, W.** An enhancer stimulates transcription in trans when attached to the promoter via a protein bridge. *Cell* ***58 (4)*** 1989. pp. 767-777.
**Mukouyama, Y.S.; Hara, T.; Xu, M.J.; Tamura, K.; Donovan, P.J.; Nakahata, T.; Miyajima, A.** In vitro expansion of murine multipotential hematopoietic progenitors from the embryonic aorta-gonad-mesonephros region. *Immunity* ***8 (1)*** 1998. pp. 105-114.
**Mulder, M.P.; Wilke, M.; Langeveld, A.; Whilming, L.G.; Hagemeijer, A.; van Drunen, E.; Zwarthoff, E.C.; Riegman, P.H.; Deelen, W.H.; van den Ouweland, A.M.** Positional mapping of loci in the DiGeorge critical region at chromosome 22q11 using a new marker (D22S183). *Hum Genet* ***96 (2)*** 1995. pp. 133-141.
**Muller, H.** Types of visible variations induced by X-rays in Drosophila *Drosophila J Genet* ***22*** 1930. pp. 299-334.
**Muravyova, E.; Golovnin, A.; Gracheva, E.; Parshikov, A.; Belenkaya, T.; Pirrotta, V.; Georgiev, P.** Loss of insulator activity by paired Su(Hw) chromatin insulators. *Science* ***291 (5503)*** 2001. pp. 495-498.
**Muyrers, J.P.P.; Zhang, Y.; Testa, G.; Stewart, A.F.** Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res* ***27 (6)*** 1999. pp. 1555-1557.
**Muyrers, J.P.P.; Zhang, Y.; Benes, V.; Testa, G.; Ansorge, W.; Stewart, A.F.** Point mutation of bacterial artificial chromosomes by ET recombination. *EMBO rep* ***1 (3)*** 2000. pp. 239-243.
**Myers, R.M.; Tilly, K.; Maniatis, T.** Fine structure genetic analysis of a beta-globin promoter. *Science* ***232 (4750)*** 1986. pp. 613-618.
**Nakao, M.** Epigenetics: interaction of DNA methylation and chromatin. *Gene* ***278 (1)*** 2001. pp. 25-31.
**Nan, X.; Meehan, R.R.; Bird, A.P.** Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res* ***21 (21)*** 1993. pp. 4886-4892.
**Nan, X.; Tate, P.; Bird, A.P.** DNA methylation specifies chromosomal localization of MeCP2. *Mol Cell Biol* ***16 (1)*** 1996. pp. 414-421.
**Nan, X.; Ng, H.H.; Johnson, C.A.; Laherty, C.D.; Turner, B.M.; Eisenman, R.N.; Bird, A.P.** Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* ***393 (6683)*** 1998. pp. 386-389.
**Narayanan, K.; Wiliamson, R.; Zhang, Y.; Stewart, A.F.; Ioannou, P.A.** Efficient and precise engineering of a 200 kb beta-globin humanbacterial artificial chromosome in E. coli DH10B using an inducible homologous recombination system. *Gene Therapy* ***6 (3)*** 1999. pp. 442-447.
**Narliker, G.J.; Fan, H.Y.; Kingston, R.E.** Cooperation between complexes that regulate chromatin structure and transcription. *Cell* ***108 (4)*** 2002. pp. 475-487.
**Natarajan, K.; Jackson, B.M.; Zhou, H.; Winston, F.; Hinnebush, A.G.** Transcriptional activation by Gcn4p involves independent interactions with the SWISNF complex and the SRBmediator. *Mol Cell* ***4 (4)*** 1999. pp. 657-664.
**Navas, P.A.; Peterson, K.R.; Li, Q.; Skarpidi, E.; Rohde, A.; Shaw, S.E.; Clegg, C.H.; Asano, H.; Stamatoyannopoulos, G.** 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 (7)*** 1998. pp. 4188-4196.
**Navas, P.A.; Li, Q.; Peterson, K.R.; Swank, R.A.; Rohde, A.; Roy, J.; Stamatoyannopoulos, G.** Activation of the beta-like globin genes in transgenic mice is dependent on the presence of the beta-locus control region. *Hum Mol Gen* ***11 (8)*** 2002. pp. 893-903.
**Neely, K.E.; Hassan, A.H.; Wallberg, A.E.; Steger, D.J.; Cairns, B.R.; Wright, A.P.; Workman, J.L.** Activation domain-mediated targeting of the SWISNF complex to promoters stimulates transcription from nucleosome arrays. *Mol Cell* ***4 (4)*** 1999. pp. 649-655.
**Nemazee, D.** Receptor selection in B and T lymphocytes. *Annu Rev Immunol* ***18*** 2000. pp. 19-51.
**Ney, P.A.; Sorrentino, B.P.; McDonagh, K.T.; Nienhuis, A.W.** Tandem AP-1-binding sites within the human beta-globin dominant control region function as an inducible enhancer in erythroid cells. *Genes Dev* ***4 (6)*** 1990. pp. 993-1006.
**Ng, K.W.; Ridgway, P.; Cohen, D.R.; Tremethick, D.J.** The binding of a FosJun heterodimer can completely disrupt the structure of a nucleosome. *EMBO J* ***16 (8)*** 1997. pp. 2072-2085.

**Nielsen, S.J.; Schneider, R.; Bauer, U-M.; Bannister, A.J.; Morrison, A.; O Carroll, D.; Firestein, R.; Cleary, M.; Jenuwein, T.; Herrera, R.E.; Kouzarides, T.** Rb targets histone H3 methylation and HP1 to promoters. *Nature* ***412 (6846)*** 2001. pp. 561-565.
**Nishioka, Y.; Leder, P.** The complete sequence of a chromosomal mouse alpha--globin gene reveals elements conserved throughout vertebrate evolution. *Cell* ***18 (3)*** 1979. pp. 875-882.
**Norma, K-I.; Allis, C.D.; Grewall, S.I.S.** Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science* ***293 (5532)*** 2001. pp. 1150-1155.
**Nuez, B.; Michalovich, D.; Bygrave, A.; Ploemacher, R.; Grosveld, F.** Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature* ***375 (6529)*** 1995. pp. 316-318.
**O Carroll, D.; Scherthan, H.; Peters, A.H.; Opravil, S.; Haynes, A.R.; Laible, G.; Rea, S.; Schmid, M.; Lebersorger, A.; Jeratsch, M.** Isolation and characterization of Suv39h2, a second histone H3 methyltransferase gene that displays testis-specific expression. *Mol Cell Biol* ***20 (24)*** 2000. pp. 9423-9433.
**O'Connor, M.; Peifer, M.; Bender, W.** Construction of large DNA segments in Escherichia coli. *Science* ***244 (4910)*** 1989. pp. 1307-1312.
**O'Neill, L.P.; Turner, B.M.** Histone H4 acetylation distinguishes coding regions of the human genome from heterochromatin in a differentiation-dependent but transcription-independent manner. *EMBO J* ***14 (16)*** 1995. pp. 3946-3957.
**Ogawa, M.** Differentiation and proliferation of hematopoietic stem cells. *Blood* ***81 (11)*** 1993. pp. 2844-2853.
**Ogbourne, S.; Antalis, T.M.** Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem J* ***331*** 1998. pp. 1-14.
**Ohki, I.; Shimotake, N.; Fujita, N.; Jee, J.; Ikegami, T.; Nakao, M.; Shirakawa, M.** Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell* ***105 (4)*** 2001. pp. 487-497.
**Okano, M.; Xie, S.; Li, E.** Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature Genet* ***19 (3)*** 1998. pp. 219-220.
**Okano, M.; Bell, D.W.; Haber, D.A.; Li, E.** DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* ***99 (3)*** 1999. pp. 247-257.
**Olins, A.L.; Olins, D.E.** Spheroid chromatin units (v bodies). *Science* ***183 (122)*** 1974. pp. 330-332.
**Oliva, R.; Bazett-Jones, D.P.; Locklear, L.; Dixon, G.H.** Histone hyperacetylation can induce unfolding of the nucleosome core particle. *Nucleic Acids Res* ***18 (9)*** 1990. pp. 2739-2747.
**Omichinski, J.G.; Clore, G.M.; Schaad, O.; Felsenfeld, G.; Trainor, C.; Appella, E.; Stahl, S.J.; Gronenborn, A.M.** NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science* ***261 (5120)*** 1993. pp. 438-446.
**Orkin, S.H.** *Disorders of Hemoglobin Synthesis: The Thalassemias* 1986. pp. ?-?.
**Orkin, S.H.** GATA-binding transcription factors in hematopoietic cells. *Blood* ***80 (3)*** 1992. pp. 575-581.
**Orkin, S.H.** *Hemoglobin switching: the molecular basis of blood diseases* 2000. pp. 80-102.
**Orlando, V.** Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends Biochem Sci* ***25 (3)*** 2000. pp. 99-104.
**Ozbudak, E.M.; Thattai, M.; Kurtser, I.; Grossman, A.D.; van Oudenaarden, A.** Regulation of noise in the expression of a single gene. *Nature Genet* ***31 (1)*** 2002. pp. 69-73.
**Papoulas, O.; Beek, S.J.; Moseley, S.L.; McCallum, C.M>; Sarte, M.; Shearn, A.; Tamkun, J.W.** The Drosophila trithorax group proteins BRM, ASH1 and ASH2 are subunits of distinct protein complexes. *Development* ***125 (20)*** 1998. pp. 3955-3966.
**Paro, R.** Propagating memory of transcriptional states. *Trends Genet* ***11 (8)*** 1995. pp. 295-297.
**Pavan, W.J.; Hieter, P.; Reeves, R.H.** Generation of deletion derivatives by targeted transformation of human-derived yeast artificial chromosomes. *Proc Natl Acad Sci U S A* ***87 (4)*** 1990. pp. 1300-1304.
**Perkins, A.C.; Sharpe, A.H.; Orkin, S.H.** Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature* ***375 (6529)*** 1995. pp. 318-322.
**Perkins, A.C.; Gaensler, K.M.L.; Orkin, S.H.** 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 (22)*** 1996. pp. 12267-12271.
**Peters, B.; Merezhinskaya, N.; Diffley, J.F.; Noguchi, C.T.** Protein-DNA interactions in the epsilon-globin gene silencer. *J Biol Chem* ***286 (5)*** 1993. pp. 3430-3437.
**Peterson, K.R.; Clegg, C.H.; Navas, P.A.; Norton, E.J.; Kimborough, T.G.; Stamatoyannopoulos, G.** Effect of deletion of 5'HS3 or 5'HS2 of the human beta-globin locus control region on the developmental regulation of globin gene expression in beta-globin locus yeast artificial chromosome transgenic mice. *Proc Natl Acad Sci U S A* ***93 (13)*** 1996. pp. 6605-6609.
**Peterson, K.R.; Clegg, C.H.; Huxley, C.; Josephson, B.M.; Haugen, H.S.; Furukawa, T.; Stamatoyannopoulos, G.** Transgenic mice containing a 248-kb yeast artificial chromosome carrying the human beta-globin locus display proper developmental control of human globin genes *Proc Natl Acad Sci U S A* ***90 (16)*** 1993. pp. 7593-7597.
**Peterson, K.R.; Navas, P.; Stamatoyannopoulos, G.** beta-YAC transgenic mice for studying LCR function. *Ann NY Acad Sci* ***850*** 1998. pp. 28-37.
**Petruk, S.; Sedkov, Y.; Smith, S.; Tillib, S.; Kraevski, V.; Nakamura, T.; Canaani, E.; Croce, C.M.; Mazo, A.** Trithorax and dCBP acting in a complex to maintain expression of a homeotic gene. *Science* ***294 (5545)*** 2001. pp. 1331-1334.
. pp. -.
**Pettersson, S; Cook, G.P.; Bruggeman, M.; Williams, G.; Neuberger, M.S.** A second B-cell specific enhancer 3' of the immunoglobin heavy-chain locus *Nature* ***344 (6262)*** 1990. pp. 165-168.
**Pevny, L.; Simon, M.C.; Robertson, E.; Klein, W.H.; Tsai, S.F.; D'Agati, V.; Orkin, S.H.; Costantini, F.** Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* ***349 (6306)*** 1991. pp. 257-260.
**Pham, A.D.; Sauer, F.** Ubiquitin-activatingconjugating activity of TAFII250, a mediator of activation of gene expression in Drosophila. *Science* ***289 (5488)*** 2000. pp. 2357-2360.
**Philipsen, S.; Talbot, D.; Fraser, P.; Grosveld, F.** The beta-globin dominant control region: hypersensitive site 2. *EMBO J* ***9 (7)*** 1990. pp. 2159-2167.
**Philipsen, S.; Pruzina, S.; Grosveld, F.** The minimal requirements for activity in transgenic mice of hypersensitive site 3 of the beta globin locus control region. *EMBO J* ***12 (3)*** 1993. pp. 1077-1085.
**Pikaart, M.J.; Recillas-Targa, F.; Felsenfeld, G.** Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. *Genes Dev* ***12 (18)*** 1998. pp. 2852-2862.
**Pirrotta, V.** Polycombing the genome: PcG, trxG, and chromatin silencing. *Cell* ***93 (3)*** 1998. pp. 333-336.
**Platt, T.** Transcription termination and the regulation of gene expression. *Annu Rev Biochem* ***55*** 1986. pp. 339-372.
**Podolsky, T.; Fong, S.T.; Lee, B.T.O.** Direct selection of tetracycline-sensitive Escherichia coli cells using nickel salts. *Plasmid* ***36 (2)*** 1996. pp. 112-115.
**Pondel, M.; Murphy, S.; Pearson, L.; Craddock, C.; Proudfoot, N.J.** Sp1 functions in a chromatin-dependent manner to augment human alpha-globin promoter activity. *Proc Natl Acad Sci U S A* ***92 (16)*** 1995. pp. 7237-7241.

**Pondel, M.D.; Proudfoot, N.J.; Whitelaw, C.; Whitelaw, E.** The developmental regulation of the human zeta-globin gene in transgenic mice employing beta-galactosidase as a reporter gene. *Nucleic Acids Res **20 (21)*** 1992. pp. 5655-5660.
**Porcu, S.; Kitamura, M.; Witkowska, E.; Zhang, Z.; Mutero, A.; Lin, C.; Chang, J.; Gaensler, K.M.L.** The human beta globin locus introduced by YAC transfer exhibits a specific and reproducible pattern of developmental regulation in transgenic mice. *Blood **90 (11)*** 1997. pp. 4602-4609.
**Poustka, A.; Rackwitz, H.R.; Frischauf, A.M.; Hohn, B.; Lehrac, H** Selective isolation of cosmid clones by homologous recombination in Escherichia coli. *Proc Natl Acad Sci U S A **81 (13)*** 1984. pp. 4129-4133.
**Proudfoot, N.** Poly(A) signals. *Cell **64 (4)*** 1991. pp. 671-674.
**Pruzina, S.; Hanscombe, O.; Whyatt, D.; Grosveld, F.; Philipsen, S.** Hypersensitive site 4 of the human beta globin locus control region. *Nucleic Acids Res **19 (7)*** 1991. pp. 1413-1419.
**Ptashne, M.** How eukaryotic transcriptional activators work. *Nature **335 (6192)*** 1988. pp. 638-639.
**Ptashne, M.; Gann, A.** Transcriptional activation by recruitment. *Nature **386 (6625)*** 1997. pp. 569-577.
**Purucker, M.; Bodine, D.; Lin, H.; McDonagh, K.; Nienhuis, A.W.** Structure and function of the enhancer 3' to the human A gamma globin gene. *Nucleic Acids Res **18 (24)*** 1990. pp. 7407-7415.
**Raich, N.; Clegg, C.H.; Grofti, J.; Romeo, P.H.; Stamatoyannopoulos, G.** GATA1 and YY1 are developmental repressors of the human epsilon-globin gene. *EMBO J **14 (4)*** 1995. pp. 801-809.
**Raich, N.; Enver, T.; Nakamoto, B.; Josephson, B.; Papayannopoulou, T.; Stamatoyannopoulos, G.** Autonomous developmental control of human embryonic globin gene switching in transgenic mice. *Science **250 (4984)*** 1990. pp. 1147-1149.
**Raich, N.; Papayannopoulou, T.; Stamatoyannopoulos, G.; Enver, T.** Demonstration of a human epsilon-globin gene silencer with studies in transgenic mice. *Blood **79 (4)*** 1992. pp. 861-864.
**Rea, S.; Eisenhaber, F.; O Carroll, D.; Strahl, B.D.; Sun, Z.W.; Schmid, M.; Opravil, S.; Mechtler, K.; Ponting, C.P.; Allis, C.D.; Junewein, T.** Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature **403 (6796)*** 2000. pp. 593-599.
**Reik, A.; Telling, A.; Zitnik, G.; Cimbora, D.; Epner, E.; Groudine, M.** 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 (10)*** 1998. pp. 5992-6000.
**Rich, I.N.** Primordial germ cells are capable of producing cells of the hematopoietic system in vitro. *Blood **86 (2)*** 1995. pp. 463-472.
**Riviere, I.; Sunshine, M.J.; Littman, D.R.** Regulation of IL-4 expression by activation of individual alleles. *Immunity **9 (2)*** 1998. pp. 217-228.
**Robzyk, K.; Recht, J.; Osley, M.A.** Rad6-dependent ubiquitination of histone H2B in yeast. *Science **287 (5452)*** 2000. pp. 501-504.
**Rohrbaugh, M.L.; Hardison, R.C.** Analysis of rabbit beta-like globin gene transcripts during development. *J Mol Biol **164 (3)*** 1983. pp. 395-417.
**Rojas, J.R.; Trievel, R.C.; Zhou, J.; Mo, Y.; LI, X.; Berger, S.L.; Allis, C.D.; Marmorstein, R.** Structure of Tetrahymena GCN5 bound to coenzyme A and a histone H3 peptide. *Nature **401 (6748)*** 1999. pp. 93-97.
**Ronchi, A.; Berry, M.; Raguez, S.; Imam, A.; Yannoutsos, N.; Ottolenghi, S.; Grosveld, F.; Dillon, N.A.** Role of the duplicated CCAAT box region in gamma-globin gene regulation and hereditary persistence of fetal haemoglobin. *EMBO J **15 (1)*** 1996. pp. 143-149.
**Ross, I.L.; Brown, C.M.; Hume, D.A.** Transcription of individual genes in eukaryotic cells occurs randomly and infrequently *Immunol Cell Biol **72*** 1994. pp. 177-185.
**Rountree, M.R.; Bachman, K.E.; Baylin, S.B.** DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nature Genet **25 (3)*** 2000. pp. 269-277.
**Rubnitz, J.E.; Behm, F.G.; Downing, J.R.** 11q23 rearrangements in acute leukemia. *Leukemia **10 (1)*** 1996. pp. 74-82.
**Rundlett, S.E.; Carman, A.A.; Kobayashi, R.; Bavykin, S.; Turner, B.M.; Grunstein, M.** HDA1 and RPD3 are members of distinct yeast histone deacetylase complexes that regulate silencing and transcription. *Proc Natl Acad Sci U S A **93 (25)*** 1996. pp. 14503-14508.
**Ryan, T.M.; Behringer, R.R.; Townes, T.M.; Palmiter, R.D.; Brinster, R.L.** High-level erythroid expression of human alpha-globin genes in transgenic mice. *Proc Natl Acad Sci U S A **86 (1)*** 1989. pp. 37-41.
**Sabath, D.E.; Koehler, K.M.; Yang, W.Q.; Patton, K.; Stamatoyannopoulos, G.** Identification of a major positive regulatory element located 5' to the human zeta-globin gene. *Blood **85 (9)*** 1995. pp. 2587-2597.
**Sabath, D.E.; Spangler, E.A.; Rubin, E.M.; Stamatoyannopoulos, G.** Analysis of the human zeta-globin gene promoter in transgenic mice. *Blood **82 (9)*** 1993. pp. 2899-2905.
**Saitoh, S.; Wada, T.** Parent-of-origin specific histone acetylation and reactivation of a key imprinted gene locus in Prader-Willi syndrome. *Am J Hum Genet **6 (66)*** 2000. pp. 1958-1962.
**Sambrook, J.; Fritsch, E.F.; Maniatis, T.** *Molecular Cloning: A Laboratory Manual 2nd Edition.* 1989. pp. ?-?.
**Sanchez, M.J.; Holmes, A.; Miles, C.; Dzierzak, E.** Characterization of the first definitive hematopoietic stem cells in the AGM and liver of the mouse embryo. *Immunity **5 (6)*** 1996. pp. 513-525.
**Satijn, D.P.E.; Otte, A.P.** Polycomb group protein complexes: do different complexes regulate distinct target genes? *Biochim Biophys Acta **1447 (1)*** 1999. pp. 1-16.
**Saurin, A.J.; Shao, Z.; Erdjument-Bromage, H.; Tempst, P.; Kingston, R.E.** A Drosophila Polycomb group complex includes Zeste and dTAFII proteins. *Nature **412 (6847)*** 2001. pp. 655-660.
**Sawado, T.; Igarashi, K.; Groudine, M.** 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 (18)*** 2001. pp. 10226-10231.
**Schedl, A.; Beermann, F.; Thies, E.; Montoliu, L.; Kelsey, G.; Schutz, G.** Transgenic mice generated by pronuclear injection of a yeast artificial chromosome. *Nucleic Acids Res **20 (12)*** 1992. pp. 3073-3077.
**Scheidereit, C.; Heguy, A.; Roeder, R.G.** Identification and purification of a human lymphoid-specific octamer-binding protein (OTF-2) that activates transcription of an immunoglobulin promoter in vitro. *Cell **51 (5)*** 1987. pp. 783-793.
**Schoenherr, C.J.; Tilghman, S.M.** *Chromatin and gene structure* 2000. pp. 252-277.
**Schubeler, D.; Francastel, C.; Cimbora, D.; Reik, A.; Martin, D.I.; Groudine, M.** Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human beta-globin locus. *Genes Dev **14 (8)*** 2000. pp. 940-950.
**Scnitzler, G.; Sif, S.; Kingston, R.E.** Human SWISNF interconverts a nucleosome between its base state and a stable remodeled state. *Cell **94 (1)*** 1998. pp. 17-27.
**Sewack, G.F.; Ellis, T.W.; Hansen, U.** Binding of TATA binding protein to a naturally positioned nucleosome is facilitated by histone acetylation. *Mol Cell Biol **21 (4)*** 2001. pp. 1404-1415.
**Shao, Z.; Raible, F.; Mollaaghababa, R.; Guyon, J.R.; Wu, C.T.; Bender, W.; Kingston, R.E.** Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell **98 (1)*** 1999. pp. 37-46.
**Shen, X.; Mizuguchi, G.; Hamiche, A.; Wu, C.** A chromatin remodelling complex involved in transcription and DNA processing. *Nature **406 (6795)*** 2000. pp. 541-544.
**Shewchuk, B.M.; Hardison, R.C.** CpG islands from the alpha-globin gene cluster increase gene expression in an integration-dependent manner. *Mol Cell Biol **17 (10)*** 1997. pp. 5856-5866.

**Shih, D.M.; Wall, R.J.; Shapiro, S.G.** Developmentally regulated and erythroid-specific expression of the human embryonic beta-globin gene in transgenic mice. *Nucleic Acids Res* ***18 (18)*** 1990. pp. 5465-5472.
**Shilatifard, A.; Conaway, J.W.; Conaway, R.C.** Mechanism and regulation of transcriptional elongation and termination by RNA polymerase II. *Curr Opin Genetics Dev* ***7 (2)*** 1997. pp. 199-204.
**Shivdasani, R.; Orkin, S.H.** Erythropoiesis and globin gene expression in mice lacking the transcription factor NF-E2. *Proc Natl Acad Sci U S A* ***92 (19)*** 1995. pp. 8690-8694.
**Shivdasani, R.; Rosenblatt, M.F.; Zucker-Franklin, D.; Jackson, C.W.; Hunt, P.; Saris, C.J.M.; Orkin, S.H.** Transcription factor NF-E2 is required for platelet formation independent of the actions of thrombopoietinMGDF in megakaryocyte development. *Cell* ***81 (5)*** 1995. pp. 695-704.
**Shizuya, H.; Birren, B.; Kim, U-J; Mancino, V.; Slepak, T.** Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proc Natl Acad Sci U S A* ***89 (18)*** 1992. pp. 8794-8797.
**Shulman, M.J.; Wu, G.E.** Hypothesis: genes which function in a stochastic lineage commitment process are subject to monoallelic expression. *Semin Immunol* ***11 (5)*** 1999. pp. 369-371.
**Simon, J.A.; Tamkun, J.W.** Programming off and on states in chromatin: mechanisms of Polycomb and trithorax group complexes. *Curr Opin Genetics Dev* ***12 (2)*** 2002. pp. 210-218.
**Simon, J.A.** Locking in stable states of gene expression: transcriptional control during Drosophila development *Curr Opin Cell Biol* ***7 (3)*** 1995. pp. 376-385.
**Skalnik, D.G.; Strauss, E.C.; Orkin, S.H.** CCAAT displacement protein as a repressor of the myelomonocytic-specific gp91-phox gene promoter. *J Biol Chem* ***266 (25)*** 1991. pp. 16736-16744.
**Smith, R.D.; Yu, J.; Seale, R.L.** Chromatin structure of the beta-globin gene family in murine erythroleukemia cells. *Biochemistry* ***23 (4)*** 1984. pp. 785-790.
**Sorrentino, B.; Ney, P.; Bodine, D.; Nienhuis, A.W.** Inducibility of the HS II enhancer depends on binding of an erythroid specific nuclear protein. *Nucleic Acids Res* ***18 (20)*** 1990. pp. 2721-2731.
**Soultanas, P.; Kosteas, T.; Anagnou, N.P.** YY1 binds to multiple sites of a silencer element located 1.6 kb 3' to the fetal Agamma-globin gene. *Blood* ***88*** 1996. pp. 147a-147a.
**Southwood, C.M.; Downs, K.M.; Bieker, J.J.** Erythroid Kruppel-like factor exhibits an early and sequentially localized pattern of expression during mammalian erythroid ontogeny. *Dev Dyn* ***206 (3)*** 1996. pp. 248-259.
**Spencer, F.; Ketner, G.; Connelly, C.; Heiter, P.** Targeted recombination-based cloning and manipulation of large DNA segments in yeast. *Methods Enzymol* ***5*** 1993. pp. 161-175.
**Spritz, R.; de Riel, J.; Forget, B.; Weissman, S.** Complete nucleotide sequence of the human delta-globin gene. *Cell* ***21 (3)*** 1980. pp. 639-646.
**Stalder, J.; Larsen, A.; Engel, J.D.; Dolan, M.; Groudine, M.; Weintraub, H.** Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I. *Cell* ***20 (2)*** 1980. pp. 451-460.
**Stamatoyannopoulos, G.; Grosveld, F.** *The molecular Basis of Blood Diseases* 2001. pp. 135-182.
**Stamatoyannopoulos, G.; Nienhuis, A.W.** *Hemoglobin switching: the molecular basis of blood diseases* 1996. pp. ?-?.
**Stamatoyannopoulos, G.; Wood, W.G.; Papayannopoulou, T.; Nute, P.E.** A new form of hereditary persistence of fetal hemoglobin in blacks and its association with sickle cell trait. *Blood* ***46 (5)*** 1975. pp. 683-692.
**Stamatoyannopoulos, G.; Josephson, B.; Zhang, J.W.; Li, Q.** Developmental regulation of human gamma-globin genes in transgenic mice. *Mol Cell Biol* ***13 (12)*** 1993. pp. 7636-7644.
**Stamatoyannopoulos, G.; Clegg, C.H.; Li, Q.** Sheltering of gamma-globin expression from position effects requires both an upstream locus control region and a regulatory element 3' to the A gamma-globin gene. *Mol Cell Biol* ***17 (1)*** 1997. pp. 240-240.
**Stamatoyannopoulos, G.; Goodwin, A.J.; Joyce, T.M.; Lowrey, C.H.** *Molecular Biology of Hemoglobin switching* 1995. pp. 71-86.
**Sterner, D.E.; Berger, S.L.** Acetylation of histones and transcription-related factors. *Microbiol Mol Biol Rev* ***64 (2)*** 2000. pp. 435-459.
**Sterner, R.; Vidali, G.; Allfrey, V.G.** Studies of acetylation and deacetylation in high mobility group proteins. Identification of the sites of acetylation in HMG-1. *J Biol Chem* ***254 (22)*** 1979. pp. 11577-11583.
**Stoming, T.A.; Stoming, G.S.; Lanclos, K.D.; Fei, Y.I.; Altay, C.; Kultar, F.; Huisman, T.H.J.** An A gamma type of nondeletional hereditary persistence of fetal hemoglobin with a T----C mutation at position -175 to the cap site of the A gamma globin gene. *Blood* ***73 (1)*** 1989. pp. 329-333.
**Strahl, B; Allis, C** The language of covalent histone modifications *Nature* ***403 (6765)*** 2000. pp. 41-45.
**Strauss, E.C.; Orkin, S.H.** In vivo protein-DNA interactions at hypersensitive site 3 of the human beta-globin locus control region. *Proc Natl Acad Sci U S A* ***89 (13)*** 1992. pp. 5809-5813.
**Strauss, E.C.; Andrews, N.C.; Higgs, D.R.; Orkin, S.H.** In vivo footprinting of the human alpha-globin locus upstream regulatory element by guanine and adenine ligation-mediated polymerase chain reaction. *Mol Cell Biol* ***12 (5)*** 1992. pp. 2135-2143.
**Strouboulis, J.; Dillon, N.; Grosveld, F.** Developmental regulation of a complete 70-kb human beta-globin locus in transgenic mice. *Genes Dev* ***6 (10)*** 1992. pp. 1857-1864.
**Studitsky, V.M.; Clark, D.J.; Felsenfeld, G.** Overcoming a nucleosomal barrier to transcription. *Cell* ***83 (1)*** 1995. pp. 19-27.
**Sturm, R.A.; Das, G.; Herr, W.** The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain. *Genes Dev* ***2 (12)*** 1988. pp. 1582-1599.
**Sudarnasam, P.; Iyer, V.R.; Brown, P.O.; Winston, F.** Whole-genome expression analysis of snfswi mutants of Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* ***97 (7)*** 2000. pp. 3364-3369.
**Sun, F.L.; Cuaycong, M.H.; Elgin, S.C.R.** Long-range nucleosome ordering is associated with gene silencing in Drosophila melanogaster pericentric heterochromatin. *Mol Cell Biol* ***21 (8)*** 2001. pp. 2867-2879.
**Surrey, S.; Delgrosso, K.; Malladi, P.; Schwartz, E.** A single-base change at position -175 in the 5'-flanking region of the G gamma-globin gene from a black with G gamma-beta+ HPFH. *Blood* ***71 (3)*** 1988. pp. 807-810.
**Svaren, J.; Horz, W.** Regulation of gene expression by nucleosomes. *Genes Dev* ***6 (2)*** 1996. pp. 164-170.
**Svaren, J.; Horz, W.** Transcription factors vs nucleosomes: regulation of the PHO5 promoter in yeast. *Trends Biochem Sci* ***22 (3)*** 1997. pp. 93-97.
**Swendeman, S.L.; Spielholz, C.; Jenkins, N.A.; Gilbert, D.J.; Copeland, N.G.; Sheffery, M.** Characterization of the genomic structure, chromosomal location, promoter, and development expression of the alpha-globin transcription factor CP2. *J Biol Chem* ***269 (15)*** 1994. pp. 11663-11671.
**Tachibana, M.; Sugimoto, K.; Fukushima, T.; Shinkai, Y.** Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *J Biol Chem* ***276 (27)*** 2001. pp. 25309-25317.
**Taddei, A.; Maison, C.; Roche, D.; Almouzni G.,** Reversible disruption of pericentric heterochromatin and centromere function by inhibiting deacetylases. *Nature Cell Biol* ***3 (2)*** 2001. pp. 114-120.

**Takemura, R.; Okabe, S.; Umeyama, T.; Kanai, Y.; Cowan, N.J.; Hirokawa, N.** Increased microtubule stability and alpha tubulin acetylation in cells transfected with microtubule-associated proteins MAP1B, MAP2 or tau. *J Cell Sci* ***103*** 1992. pp. 953-964.
**Talbot, D.; Collis, P.; Antoniou, M.; Vidal, M.; Grosveld, F.; Greaves, D.R.** A dominant control region from the human beta-globin locus conferring integration site-independent gene expression. *Nature* ***338 (6213)*** 1989. pp. 352-355.
**Talbot, D.; Philipsen, S.; Fraser, P.; Grosveld, F.** Detailed analysis of the site 3 region of the human beta-globin dominant control region. *EMBO J* ***9 (7)*** 1990. pp. 2169-2177.
**Talbot, D.; Grosveld, F.** 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 (6)*** 1991. pp. 1391-1398.
**Tanabe, O.; Katsuoka, F.; Campbell, A.D.; Song, W.; Yamamoto, M.; Tanimoto, K.; Engel, J.D.** An embryonicfetal beta-type globin gene repressor contains a nuclear receptor TR2TR4 heterodimer. *EMBO J* ***21 (13)*** 2002. pp. 3434-3442.
**Tanamachi, D.M.; Hanke, T.; Takizawa, H.; Jamieson, A.M.; Raulet, D.H.** Expression of natural killer receptor alleles at different Ly49 loci occurs independently and is regulated by major histocompatibility complex class I molecules. *J Exp Med* ***193 (3)*** 2001. pp. 307-315.
**Tanese, N.; Pugh, B.F.; Tjian, R.** Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. *Genes Dev* ***5 (12)*** 1991. pp. 2212--224.
**Tang, D.C.; Ebb, D.; Hardison, R.C.; Rodgers, G.P.** Restoration of the CCAAT box or insertion of the CACCC motif activates [corrected] delta-globin gene expression. *Blood* ***90 (1)*** 1997. pp. 421-427.
**Tanimoto, K.; Liu, Q.; Bungert, J.; Engel, J.D.** Effects of altered gene order or orientation of the locus control region on human beta-globin gene expression in mice. *Nature* ***398 (6725)*** 1999. pp. 344-348.
**Tanimoto, K.; Liu, Q.; Grosveld, F.; Bungert, F.; Engel, J.D.** Context-dependent EKLF responsiveness defines the developmental specificity of the human epsilon-globin gene in erythroid cells of YAC transgenic mice. *Genes Dev* ***14 (21)*** 2000. pp. 2778-2794.
**Taramelli, R.; Kioussis, D.; Vanin, E.; Bartram, K.; Groffen, J.; Hurst, J.; Grosveld, F.** Gamma delta beta-thalassaemias 1 and 2 are the result of a 100 kbp deletion in the human beta-globin cluster. *Nucleic Acids Res* ***14 (17)*** 1986. pp. 7017-7029.
**Taunton, J.; Hassig, C.A.; Schreiber, S.L.** A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* ***272 (5260)*** 1996. pp. 408-411.
**Terrell, A.R.; Wongwisansri, S.; Pilon, J.L.; Laybourn, P.J.** Reconstitution of nucleosome positioning, remodeling, histone acetylation, and transcriptional activation on the PHO5 promoter. *J Biol Chem* ***10 (34)*** 2002. pp. 31038-31047.
**Tewari, R.; Gillemans, N.; Wijgerde, M.; Nuez, B.; van Lindren, M.; Grosveld, F.; Philipsen, S.** 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 (8)*** 1998. pp. 2334-2341.
**Thein, S.L.** *Baillieres Clinical Haematology* 1993. pp. 151-175.
**Tie, F.; Furuyama, T.; Prasad-Sinha, J.; Jane, E.; Harte, P.J.** The Drosophila Polycomb Group proteins ESC and E(Z) are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3. *Development* ***128 (2)*** 2001. pp. 275-286.
**Tillib, S.; Petruk, S.; Sedkov, Y.; Kuzin, A.; Fujioka, M.; Goto, T.; Mazo, A.** Trithorax- and Polycomb-group response elements within an Ultrabithorax transcription maintenance unit consist of closely situated but separable sequences. *Mol Cell Biol* ***19 (7)*** 1999. pp. 5189-5202.
**Tong, J.K.; Hassig, C.A.; Schnitzler, G.R.; Kingston, R.E.; Schreiber, S.L.** Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature* ***395 (6705)*** 1998. pp. 917-921.
**Townes, T.M.; Lingrel, J.B.; Chen, H.Y.; Brinster, R.L.; Palmiter, R.D.** Erythroid-specific expression of human beta-globin genes in transgenic mice. *EMBO J* ***4 (7)*** 1985. pp. 1715-1723.
**Trainor, C.D.; Evans, T.; Felsenfeld, G.; Boguski, M.S.** Structure and evolution of a human erythroid transcription factor. *Nature* ***343 (6253)*** 1990. pp. 92-96.
**Tramelli, R.; Kioussis, D.; Vanin, E.; Bartram, K.; Groffen, J.; Hurst, J.; Grosveld, F.G.** Gamma delta beta-thalassaemias 1 and 2 are the result of a 100 kbp deletion in the human beta-globin cluster. *Nucleic Acids Res* ***14 (17)*** 1986. pp. 7017-7029.
**Trepicchio, W.L.; Dyer, M.A.; Baron, M.H.** Developmental regulation of the human embryonic beta-like globin gene is mediated by synergistic interactions among multiple tissue- and stage-specific elements. *Mol Cell Biol* ***13 (12)*** 1993. pp. 7457-7468.
**Trimborn, T.; Gribnau, J.; Grosveld, F.; Fraser, P.** Mechanisms of developmental control of transcription in the murine alpha- and beta-globin loci. *Genes Dev* ***13 (1)*** 1999. pp. 112-124.
**Trudel, M.; Costantini, F.** A 3' enhancer contributes to the stage-specific expression of the human beta-globin gene. *Genes Dev* ***1 (9)*** 1987. pp. 954-961.
**Trudel, M.; Magram, J.; Bruckner, L.; Costantini, F.** Upstream G gamma-globin and downstream beta-globin sequences required for stage-specific expression in transgenic mice. *Mol Cell Biol* ***7 (11)*** 1987. pp. 4024-4029.
**Tsai, S.; Martin, D.; Zon, L.; D'Andrea, A.; Wong, G.; Orkin, S.** Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. *Nature* ***339 (6224)*** 1989. pp. 446-451.
**Tsang, A.P.; Visvader, J.E.; Turner, C.A.; Fujiwara, Y.; Yu, C.; Weiss, M.J.; Crossley, M.; Orkin, S.H.** FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell* ***90 (1)*** 1997. pp. 109-119.
**Tsang, A.P.; Fujiwara, Y.; Hom, D.B.; Orkin, S.H.** Failure of megakaryopoiesis and arrested erythropoiesis in mice lacking the GATA-1 transcriptional cofactor FOG. *Genes Dev* ***12 (8)*** 1998. pp. 1176-1188.
**Tsukiyama, T.; Wu, C.** Chromatin remodeling and transcription. *Curr Opin Genetics Dev* ***7 (2)*** 1997. pp. 182-191.
**Tsukiyama, T.; Wu, C.** Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell* ***83 (6)*** 1995. pp. 1011-1020.
**Tsukiyama, T.; Becker, P.B.; Wu, C.** ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature* ***367 (6463)*** 1994. pp. 525-532.
**Tsukiyama, T.; Daniel, C.; Tamkun, J.; Wu, C.** ISWI, a member of the SWI2SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. *Cell* ***83 (6)*** 1995. pp. 1021-1026.
**Tuan, D.; Solomon, W.; Li, Q.; London, I.M.** The beta-like-globin gene domain in human erythroid cells. *Proc Natl Acad Sci U S A* ***82 (19)*** 1985. pp. 6384-6388.
**Tuan, D.; Solomon, W.B.; London, I.M.; Lee, D.P.** An erythroid-specific, developmental-stage-independent enhancer far upstream of the human beta-like globin genes. *Proc Natl Acad Sci U S A* ***86 (8)*** 1989. pp. 2554-2558.
**Tuan, D.; Kong, S.; Hu, K.** Transcription of the hypersensitive site HS2 enhancer in erythroid cells. *Proc Natl Acad Sci U S A* ***89 (23)*** 1992. pp. 11219-11223.
**Turner, B.M.; Birley, A.J.; Lavender, J.** Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in Drosophila polytene nuclei. *Cell* ***69 (2)*** 1992. pp. 375-384.
**Turner, B.M.** Histone acetylation and an epigenetic code. *Bioessays* ***22 (9)*** 2000. pp. 836-845.
**Ugrinova, I.; Pasheva, E.A.; Armengaud, J.; Pashev, I.G.** In vivo acetylation of HMG1 protein enhances its binding affinity to distorted DNA structures. *Biochemistry* ***48 (40)*** 2001. pp. 14655-14660.
**Utley, R.T.; Ikeda, K.; Grant, P.A.; Cote, J.; Steger, D.J.; Eberharter, A.; John, S.; Workman, J.L.** Transcriptional activators direct histone acetyltransferase complexes to nucleosomes. *Nature* ***394 (6692)*** 1998. pp. 498-502.

**van de Corput, M.P.C.; Grosveld, F.** Fluorescence in situ hybridization analysis of transcript dynamics in cells. *Methods* ***25 (1)*** 2001. pp. 111-118.
**van der Ploeg, L.H.; Konings, A.; Oort, M.; Roos, D.; Bernini, L.; Flavell, R.A.** gamma-beta-Thalassaemia studies showing that deletion of the gamma- and delta-genes influences beta-globin gene expression in man. *Nature* ***283 (5748)*** 1980. pp. 637-642.
**van der Ploeg, L.H.T.; Flavell, R.A.** DNA methylation in the human gamma delta beta-globin locus in erythroid and nonerythroid tissues. *Cell* ***19 (4)*** 1980. pp. 947-958.
**van der Vlag, J.; Otte, A.P.** Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. *Nature Genet* ***23 (4)*** 1999. pp. 474-478.
**van Holde, K.E.** *chromatin* 1989. pp. ?-?.
**van Lohuizen, M.; Verbeek, S.; Scheijen, B.; Wientjens, E.; van der Gulden, H.; Berns, A.** Identification of cooperating oncogenes in E mu-myc transgenic mice by provirus tagging. *Cell* ***65 (5)*** 1991. pp. 737-752.
**van Raamsdonk, C.D.; Tilghman, S.M.** Optimizing the detection of nascent transcripts by RNA fluorescence in situ hybridization. *Nucleic Acids Res* ***29 (8)*** 2001. pp. E42-E42.
**Varga-Weisz, P.D.; Wilm, M.; Bonte, E.J.; Dumas, K.; Mann, M.; Becker, P.B.** Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. *Nature* ***388 (6642)*** 1997. pp. 598-602.
**Varshavsky, A.J.; Bakayev, V.V.; Georgiev, G.P.** Minichromosome of simian virus 40: presence of histone HI. *Nucleic Acids Res* ***3 (8)*** 1976. pp. 477-492.
**Verdel, A.; Khochbin, S.** Identification of a new family of higher eukaryotic histone deacetylases. Coordinate expression of differentiation-dependent chromatin modifiers. *J Biol Chem* ***274 (4)*** 1999. pp. 2440-2445.
**Vettesse-Dadey, M.; Grant, P.A.; Hebbes, T.R.; Crane-Robinson, C.; Allis, C.D.; Workman, J.L.** Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. *EMBO J* ***15 (10)*** 1996. pp. 2508-2518.
**Vignali, M.; Hassan, A.H.; Neely, K.E.; Workman, J.L.** ATP-dependent chromatin-remodeling complexes. *Mol Cell Biol* ***20 (6)*** 2000. pp. 1899-1910.
**Visvader, J.E.; Elefanty, A.G.; Strasser, A.; Adams, J.M.** GATA-1 but not SCL induces megakaryocytic differentiation in an early myeloid line. *EMBO J* ***11 (12)*** 1992. pp. 4557-4562.
**Vonlanthen, S.; Heighway, J.; Altermat, H.J.; Gugger, M.; Kappeler, A.; Borner, M.M.; van Lohuizen, M.; Betticher, D.C.** The bmi-1 oncoprotein is differentially expressed in non-small cell lung cancer and correlates with INK4A-ARF locus expression. *Br J Cancer* ***84 (10)*** 2001. pp. 1372-1376.
**Wakimoto, B.T.** Beyond the nucleosome: epigenetic aspects of position-effect variegation in Drosophila. *Cell* ***93 (3)*** 1998. pp. 321-324.
**Wall, L.; Destroismaisons, N.; Delvoye, N.; Guy, L.G.** CAATenhancer-binding proteins are involved in beta-globin gene expression and are differentially expressed in murine erythroleukemia and K562 cells. *J Biol Chem* ***271 (28)*** 1996. pp. 16477-16488.
**Wallberg, A.E.; Neely, K.E.; Hassan, A.H.; Gustafsson, J.A.; Workman, J.L.; Wright, A.P.** Recruitment of the SWI-SNF chromatin remodeling complex as a mechanism of gene activation by the glucocorticoid receptor tau1 activation domain. *Mol Cell Biol* ***20 (6)*** 2000. pp. 2004-2013.
**Wallrath, L.L.; Elgin, S.C.R.** Position effect variegation in Drosophila is associated with an altered chromatin structure. *Genes Dev* ***9 (10)*** 1995. pp. 1263-1267.
**Wallrath, L.L.** Unfolding the mysteries of heterochromatin. *Curr Opin Genetics Dev* ***8 (2)*** 1998. pp. 147-153.
**Walters, M.; Martin, D.I.K.** Functional erythroid promoters created by interaction of the transcription factor GATA-1 with CACCC and AP-1NFE-2 elements. *Proc Natl Acad Sci U S A* ***89 (21)*** 1992. pp. 104444-104448.
**Walters, M.; Kim, C.; Gelinas, R.** Characterization of a DNA binding activity in DNAse I hypersensitive site 4 of the human globin locus control region. *Nucleic Acids Res* ***19 (19)*** 1991. pp. 5385-5393.
**Walters, M.; Fiering, S.; Eidemiller, J.; Magis, W.; Groudine, M.; Martin, D.I.K.** Enhancers increase the probability but not the level of gene expression. *Proc Natl Acad Sci U S A* ***92 (15)*** 1995. pp. 7125-7129.
**Walters, M.; Magis, W.; Fiering, S.; Eidemiller, J.; Scalzo, D.; Groudine, M.** Transcriptional enhancers act in cis to suppress position-effect variegation. *Genes Dev* ***10 (2)*** 1996. pp. 185-195.
**Wang, H.; Huang, Z.Q.; Xia, L.; Feng, Q.; Erdjument-Bromage, H.; Strahl, B.D.; Briggs, S.D.; Allis, C.D.; Wong, J.; Tempst, P.; Zhang, Y.** Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor. *Science* ***296 (5531)*** 2001. pp. 853-857.
**Wasylyk, B.; Wasylyk, C.; Augereau, P.; Chambon, P.** The SV40 72 bp repeat preferentially potentiates transcription starting from proximal natural or substitute promoter elements *Cell* ***32 (2)*** 1983. pp. 503-514.
**Watt, P.; Lamb, P.; Squire, L.; Proudfoot, N.** A factor binding GATAAG confers tissue specificity on the promoter of the human zeta-globin gene. *Nucleic Acids Res* ***18 (6)*** 1990. pp. 1339-1350.
**Weatherall, D.J.** *The Molecular Basis of Blood Diseases* 2001. pp. 183-226.
**Weintraub, H.; Groudine, M.** Chromosomal subunits in active genes have an altered conformation. *Science* ***193 (4256)*** 1976. pp. 848-856.
**Weintraub, H.** Formation of stable transcription complexes as assayed by analysis of individual templates. *Proc Natl Acad Sci U S A* ***85 (16)*** 1988. pp. 5819-5823.
**Weiss, M.J.; Keller, G.; Orkin, S.H.** Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. *Genes Dev* ***8 (10)*** 1994. pp. 1184-1197.
**Weiss, M.J.; Yu, C.; Orkin, S.H.** Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. *Mol Cell Biol* ***17 (3)*** 1997. pp. 1642-1651.
**Whitehouse, I.; Flaus, A.; Cairns, B.R.; White, M.F.; Workman, J.L.; Owen-Hughes, T.** Nucleosome mobilization catalysed by the yeast SWISNF complex. *Nature* ***400 (6746)*** 1999. pp. 784-787.
**Whyatt, D.J.; Karis, A.; Harkes, I.C.; Verkerk, A.; Gillemans, N.; Elefanty, A.G.; Vairo, G.; Poemacher, R.; Grosveld, F.; Philipsen, S.** The level of the tissue-specific factor GATA-1 affects the cell-cycle machinery. *Genes Funct* ***1 (1)*** 1997. pp. 11-24.
**Whyatt, D.J.; Lindeboom, F.; Karis, A.; Ferreira, R.; Milot, E.; Hendriks, R.; de Bruijn, M.; Langeveld, A.; Gribnau, J.; Grosveld, F.; Philipsen, S.** An intrinsic but cell-nonautonomous defect in GATA-1-overexpressing mouse erythroid cells. *Nature* ***406 (6795)*** 2000. pp. 519-524.
**Wijgerde, M.; Gribnau, J.; Trimborn, T.; Nuez, B.; Philipsen, S.; Grosveld, F.; Fraser, P.** The role of EKLF in human beta-globin gene competition. *Genes Dev* ***10 (22)*** 1996. pp. 2894-2902.
**Wijgerde, M.; Grosveld, F.; Fraser, P.** Transcription complex stability and chromatin dynamics in vivo. *Nature* ***377 (6546)*** 1995. pp. 209-213.
**Wilson, C.J.; Chao, D.M.; Imbalzano, A.N.; Schnitzler, G.R.; Kingston, R.E.; Young, R.A.** RNA polymerase II holoenzyme contains SWISNF regulators involved in chromatin remodeling. *Cell* ***84 (2)*** 1996. pp. 235-244.
**Wilson, C.J.; Bellen, H.J.; Gehring, W.J.** Position effects on eukaryotic gene expression. *Annu Rev Cell Biol* ***6*** 1990. pp. 679-714.
**Winston, F.; Carlson, M.** Yeast SNFSWI transcriptional activators and the SPTSIN chromatin connection. *Trends Genet* ***8 (11)*** 1992. pp. 387-391.

**Wintrobe, M.M.; Schumacker, H.B.** - *J Clin Invest* ***14*** 1935. pp. 837-837.
**Wolffe, A.P.** *San Diego: Academic Press* 1998. pp. ?-?.
**Wood, W.G.** *Baillieres Clinical Haematology* 1993. pp. 177-213.
**Wood, W.G.** *Baillieres Clinical Haematology* 1993. pp. 177-213.
**Wood, W.I.; Felsenfeld, G.** Chromatin structure of the chicken beta-globin gene region. Sensitivity to DNase I, micrococcal nuclease, and Dnase II. *J Biol Chem* ***257 (13)*** 1982. pp. 7730-7736.
**Wu, C.; Binham, P.M.; Livak, K.J.; Holmgren, R.; Elgin, S.C.R.** The chromatin structure of specific genes: I. Evidence for higher order domains of defined DNA sequence. *Cell* ***16 (4)*** 1979. pp. 797-806.
**Xiao, H.; Sandaltzopoulos, R.; Wang, H.M.; Hamiche, A.; Ranallo, R.; Lee, K.M.; Fu, D.; Wu, C.** Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. *Mol Cell* ***8 (3)*** 2001. pp. 1-20.
**Yang, L.; Xia, L.; Wu, D.Y.; Wang, H.; Chansky, H.A.; Schubach, W.H.; Hickstein, D.D.; Zhang, Y.** Molecular cloning of ESET, a novel histone H3-specific methyltransferase that interacts with ERG transcription factor. *J Biol Chem* ***21 (1)*** 2002. pp. 148-152.
**Yang, X.W.; Model, P.; Heintz, N.** Homologous recombination based modification in Escherichia coli and germline transmission in transgenic mice of a bacterial artificial chromosome. *Nature Biotechnol* ***15 (9)*** 1997. pp. 859-865.
**Yannoutsos, N.; IJzermans, J.N.M.; Harkes, C.; Bonthuis, F.; Zhou, C-Y.; White, D.; Marquet, R.L.M.; Grosveld, F.** A membrane cofactor protein transgenic mouse model for the study of discordant xenograft rejection. *Genes Cells* ***1 (4)*** 1996. pp. 409-419.
**Yokata, H.; van den Engh, G.; Hearst, J.E.; Sachs, R.K.; Trask, B.J.** Evidence for the organization of chromatin in megabase pair-sized loops arranged along a random walk path in the human G0G1 interphase nucleus. *J Cell Biol* ***130 (6)*** 1995. pp. 1239-1249.
**Yu, C.Y.; Motamed, K.; Chen, J.; Bailey, A.; Shen, C.K.J.** The CACC box upstream of human embryonic epsilon globin gene binds Sp1 and is a functional promoter element in vitro and in vivo. *J Biol Chem* ***266 (14)*** 1991. pp. 8907-8915.
**Yu, C.Y.; Chen, J.; Lin, L.I.; Tam, M.; Shen, C.K.** Cell type-specific protein-DNA interactions in the human zeta-globin upstream promoter region: displacement of Sp1 by the erythroid cell-specific factor NF-E1. *Mol Cell Biol* ***10 (1)*** 1990. pp. 282-294.
**Zafarana, G.; Raguez, S.; Pruzina, S.; Grosveld, F.; Meijer, D.** *Developmental Control of Globin Genes* 1995. pp. 39-39.
**Zafarana, G.; Rottier, R.; Grosveld, F.; Philipsen, S.** Erythroid expression of CEBPg in transgenic mice affects g-globin expression and fetal liver erythropoiesis *EMBO J* ***19 (21)*** 2000. pp. 5856-5863.
**Zawel, L.; Reinberg, D.** Initiation of transcription by RNA polymerase II: a multi-step process. *Prog Nucleic Acid Res Mol Biol* ***44*** 1993. pp. 67-108.
**Zenke, M.; Grundstrom, T.; Matthes, H.; Wintzerith, M.; Schatz, C.; Wildeman, A.; Chambon, P.** Multiple sequence motifs are involved in SV40 enhancer function. *EMBO J* ***5 (2)*** 1986. pp. 387-397.
**Zhan, H.C.; Liu, D.P.; Liang, C.C.** Insulator: from chromatin domain boundary to gene regulation. *Hum Genet* ***109 (5)*** 2001. pp. 471-478.
**Zhang, Q.; Reddy, P.M.; Yu, C.Y.; Basiani, C.; Higgs, D.; Stamatoyannopoulos, G.; Papayannopoulou, T.; Shen, C.K.** Transcriptional activation of human zeta 2 globin promoter by the alpha globin regulatory element (HS-40): functional role of specific nuclear factor-DNA complexes. *Mol Cell Biol* ***13 (4)*** 1993. pp. 2298-2308.
**Zhang, W.; Bieker, J.J.** Acetylation and modulation of erythroid Kruppel-like factor (EKLF) activity by interaction with histone acetyltransferases. *Proc Natl Acad Sci U S A* ***95 (17)*** 1998. pp. 9855-9860.
**Zhang, Y.; LeRoy, G.; Seelig, H.P.; Lane, W.S.; Reinberg, D.** The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell* ***95 (2)*** 1998. pp. 279-289.
**Zhang, Y.; Buchholz, F.; Muyrers, J.P.P.; Stewart, A.F.** A new logic for DNA engineering using recombination in Escherichia coli. *Nature Genet* ***20 (2)*** 1998. pp. 123-128.
**Zhang, Z.; Ling, C.; Wang, S.; Gaensler, K.M.L.** Globin gene switching in beta-globin YAC transgenic with a 12.5kb deletion of the region between the Agamma and delta genes. *Blood* ***90*** 1997. pp. 129-129.
**Zhang, Y.; Buchholz, F.; Muyrers, J.P.; Stewart, A.F.** A new logic for DNA engineering using recombination in Escherichia coli. *Nature Genet* ***20 (2)*** 1998. pp. 123-128.
**Zhou, Q.; Lieberman, P.M.; Boyer, T.G.; Berk, A.J.** Holo-TFIID supports transcriptional stimulation by diverse activators and from a TATA-less promoter. *Genes Dev* ***6 (10)*** 1992. pp. 1964-1974.
**Zink, D.; Paro, R.** Drosophila Polycomb-group regulated chromatin inhibits the accessibility of a trans-activator to its target DNA. *EMBO J* ***14 (22)*** 1995. pp. 5660-5671.

## List of abbreviations

| | |
|---|---|
| AGM | aortic/gonad/mesonephros |
| αMRE | hypersensitive site 40 |
| BAC | bacterial artificial chromosome |
| BFU-e | burst forming units erythroid |
| bp | base pare |
| CFU-e | colony forming units erythroid |
| CTD | C terminal domain |
| CTPE | cell timing position effect |
| DNA | deoxy nucleic acid |
| E(var) | enhancer of variegation |
| EKLF | erythroid kruppel like factor |
| GR | glucocorticoid receptor |
| HAT | histone acetyltransferases |
| Hb | haemoglobin |
| HCS | haematopoietic stem cell |
| HDAC | histone deacetylases |
| HP1 | heterochromatin protein 1 |
| HPFH | Hereditary Persistence of Foetal Haemoglobin |
| HS | hypersensitive site |
| kb | kilo base |
| kDa | kilo dalton |
| LCR | locus control region |
| m | meter |
| MAR | matrix attachment region |
| MDa | mega dalton |
| ME | maintenance element |
| MEL cells | murine erythoid leukemia cells |
| μLCR | micro locus control region |
| μm | micro meter |
| MMTV | mammary tumour virus |
| mRNA | messenger ribo nucleic acid |
| nm | nano meter |
| NRE | negative regulatory element |
| PAC | P1 artificial chromosome |
| PcG | Polycomb group |
| PEV | position effect variegation |
| PRE | Polycomb response elements |
| RB | retinoblastoma co-repressor protein |
| RNA | ribo nucleic acid |
| SAR | scaffold attachment region |
| Su(var) | suppressor of variegation |
| TBP | TATA box binding protein |
| TF | transcription factor |
| TRE | Trithorax response elements |
| TrxG | Trithorax group |
| YAC | yeast artificial chromosome |

## Stellingen

1. Stochastische expressie speelt een belangrijke rol in de expressie van de globinegenen
   (dit proefschrift)

2. Alle hypersensitive sites hebben onafhankelijke functies, maar alleen die functies tezamen kunnen resulteren in volledig normale globine-expressiepatronen.
   (dit proefschrift)

3. Een goede wetenschapper belicht zowel de kwalitatieve als de kwantitatieve kant van zijn onderzoek.

4. Als antwoord op "Is uncertainty the only thing that is certain?". Niets is geheel zeker. En zelfs dat niet!
   (Fedorofff and Fontana, 2002, *Science* **Vol 97**, 1129-1131 en vrij naar Multatuli)

5. Niet HS3 maar HS2 is van belang voor ε-globine-expressie.
   (dit proefschrift)

6. Als je conclusies trekt ten aanzien van het effect van deleties van regulatoire elementen, moet je altijd de integratie van het transgen in de endogene omgeving in de gaten houden.
   (dit proefschrift)

7. Het valideren van gegevens door middel van statistische toetsen is van groot belang in de wetenschap.

8. Normen en waarden behoren een belangrijke rol te spelen in wetenschappelijk onderzoek.

9. De ChIP-assay is een belangrijke additionele moleculair-biologische techniek voor de *in vivo* bestudering van de activerings status van een gen.
   (Bulger *et al.*, 2002, *Curr. Opin. Genet. Dev.***Vol 12**, 170-177 en Banerjee and Zhang, 2002, *Curr Opin Microbiol* **Vol 5**, 313-317)

10. De α- en β-globine-genclusters stammen af van een gemeenschappelijke voorouder (Hardison, 1996, *Proc. Natl. Acad. Sci. U.S.A* **Vol 93**, 5675-5679), maar worden op een verschillende manier gereguleerd. (dit proefschrift)

11. Het schrijven van een proefschrift is als het beklimmen van een berg, stapje voor stapje kom je er wel.

## Curriculum Vitae

| | |
|---|---|
| Naam: | Mariken de Krom |
| Geboren: | 19 september 1976 te Schiedam |
| sept 1988 – jun 1994: | VWO te SG Spieringshoek, Schiedam |
| sept 1994 – aug 1998: | Medische Biologie aan de faculteit Biologie van de Universiteit van Amsterdam |
| sept 1998 – nov 2002: | OIO aan de afdeling Celbiologie en Genetica van de Erasmus Universiteit Rotterdam. |