2006

Conserved locus control region sequences controlling beta-globin gene expression in mice.

Andrea A. Svendsen
University of Alabama at Birmingham

CONSERVED LOCUS CONTROL REGION SEQUENCES CONTROLLING BETA-GLOBIN GENE EXPRESSION IN MICE

by

ANDREA A. SVENDSEN

TIM M. TOWNES, COMMITTEE CHAIR
PETER DETLOFF
N. PATRICK HIGGINS
SCOTT M. WILSON
PHILIP A. WOOD

A DISSERTATION

Submitted to the graduate faculty of The University of Alabama at Birmingham
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

BIRMINGHAM, ALABAMA

2006

CONSERVED LOCUS CONTROL REGION SEQUENCES CONTROLLING BETA-GLOBIN GENE EXPRESSION IN MICE

ANDREA A. SVENDSEN

ABSTRACT

The mouse β-globin locus is composed of four globin genes and a series of DNaseI hypersensitive sites upstream of the first globin gene. The main goal of this research was to determine the function of conserved protein-binding sites within the mouse β-globin Locus Control Region (LCR). The conserved sequences were mutated in the endogenous murine LCR and the effects of these mutations were analyzed in vivo. By studying alteration of the expression of distant cis-linked globin genes, the mutated LCR sequences will lead to a better understanding of basic gene regulation. I made two different DNA targeting constructs that contain mutations of different classes of transcription factor binding sites. The first construct contained 4 mutated NF-E2-like binding sites, while the second contained 2 mutated CACBP binding sites. Each construct was used to mutate the murine genome by homologous recombination in ES cells. The ES cells were injected into blastocysts to produce two classes of cloned mice, each containing mutations of a particular transcription factor binding site. This allowed the elucidation of the role of specific transcription factor binding sites in the context of the entire LCR. The phenotype of the cloned mice was characterized for gene expression and disease pathology. I found that the NF-E2-like mutations resulted in a 25% decrease of beta-globin mRNA compared to the wildtype. This decrease resulted in slight drop in

hemoglobin protein levels in mice homozygous for the mutation which did not correspond with an overt thalassemia. The second class of mice had conserved CACBP like binding sites mutated, these mice had a 10% decrease in beta-globin mRNA levels. These results show that the conserved transcription factor binding sites play a significant but non-essential role in LCR enhancement of β-globin gene expression.

ACKNOWLEDGEMENTS

I thank Dr. Tim Townes for the opportunities he has given me by allowing me to work in his lab. I thank the members of my committee for all the time and thought they put in- thank you Dr. Detloff, Dr. Higgins, Dr. Wilson, and Dr. Wood. I thank all the members of the Townes lab for their help and support when things just wouldn't work right, and for sharing the joy when things worked! I thank my parents for all the support they gave me through the years. Love you mom and dad. Finally, but most assuredly not least, I thank my husband, Timothy A. Cronican, for his love and understanding through these years.

TABLE OF CONTENTS

ABSTRACT ii

ACKNOWLEDGMENTS iv

LIST OF TABLES vii

LIST OF FIGURES viii

INTRODUCTION 1

Hemoglobin 1
Hemoglobin Switching 2
The β-globin Genes 4
A Brief History of the Evolution of the β-globin Genes 5
Globin Gene Regulation and the Locus Control Region 7
NF-E2-like Sites 10
CACBP Sites 11
The Mouse as a Model for β-globin Gene Regulation 12
Recombineering 13
Introduction to the Chapters 14

CONSERVED NF-E2 BINDING SITES ARE NOT ESSENTIAL FOR HIGH-LEVEL β-GLOBIN LOCUS CONTROL REGION ACTIVITY 28

HIGHLY CONSERVED CACBP SITES IN HS3 AND HS2 ARE DISPENSABLE FOR HIGH-LEVEL β-GLOBIN LOCUS CONTROL REGION ACTIVITY 73

CONCLUSIONS 101

A Brief Discussion on the Methods Used 102
NF-E2 Site Mutations 105
CACBP Site Mutations 106
Additive Effects in the LCR 106
Redundancy in the LCR 108
Holocomplex Formation 108
LCR Mutations are Possible β-Thalassemia Severity Modifiers 111

GENERAL LIST OF REFRENCES 118

APPENDIX: IACUC APPROVAL LETTER.....132

LIST OF TABLES

Table *Page*

CONSERVED NF-E2 BINDING SITES ARE NOT ESSENTIAL FOR HIGH-LEVEL β-GLOBIN LOCUS CONTROL REGION ACTIVITY

1 Hematological values of WT and LCR NF-E2 -/- mice.....50

HIGHLY CONSERVED CACBP SITES IN HS3 AND HS2 ARE DISPENSABLE FOR HIGH-LEVEL β-GLOBIN LOCUS CONTROL REGION ACTIVITY

1 Hematological values of WT and LCR CACBP -/- mice.....88

LIST OF FIGURES

Figure *Page*

INTRODUCTION

1 Structure of Hemoglobin 16

2 Mouse Globin Chain Synthesis during development 18

3 The β-Globin Locus of Mouse Chromosome 7 20

4. A Schematic Overview of Hemoglobin Gene Evolution 22

5 A Diagram of LCR Transgenic Studies 24

6 Conserved Transcription Factor Binding Sites in the HS of the β-Globin LCR 26

CONSERVED NF-E2 BINDING SITES ARE NOT ESSENTIAL FOR HIGH-LEVEL β-GLOBIN LOCUS CONTROL REGION ACTIVITY

1 Mouse β-Globin LCR and Mutagenic Oligos 51

2 Schematic of Knockin Strategy 53

3 PCR Genotyping of ES Cells 55

4 PCR Genotyping of Mice 57

5 Sequence of HS4, HS3 and HS2 in Homozygous NF-E2 Site Mutant Mice 59

6 Allele Specific RT-PCR of Adult Mouse Blood 61

7 Allele Specific RT-PCR of 10.5 dpc Mouse Yolk Sacs 63

8 Blood Smears of LCR NF-E2 +/+, +/- and -/- Mice 65

Supplemental Figure 1 BAC Recombineering and Capture of the Mutant LCR Fragment into the Targeting Vector 67

Supplemental Figure 2 Allele Specific RT-PCR of 10.5 dpc Mouse Yolk Sacs.....69

Supplemental Figure 3 Generation of pCV6.....71

HIGHLY CONSERVED CACBP SITES IN HS3 AND HS2 ARE DISPENSABLE FOR HIGH-LEVEL β-GLOBIN LOCUS CONTROL REGION ACTIVITY

1 Mouse β-Globin LCR and Mutagenic Oligos 89

2 Schematic of Knockin Strategy.....91

3 PCR Genotyping of ES Cells 93

4 PCR Genotyping of Mice.....95

5 Sequence of HS3 and HS2 in Homozygous CACBP Site Mutant Mice 97

6 Allele Specific RT-PCT of Adult Mouse Blood 99

CONCLUSIONS

1 A Proposed Series of Mutations of Conserved Sites in the HSs of the β-Globin LCR.....113

2 Models of LCR holocomplex formation 115

INTRODUCTION

Hemoglobin may very well be the most studied protein in history. Since the discovery of hemoglobin it has been on the forefront of scientific discovery. Research on hemoglobin has driven forward the fields of protein biochemistry, molecular biology, and gene therapy. Hemoglobin was the first protein crystallized, to have its amino acid sequence revealed, and to have its structure elucidated. In addition, the β-globin gene was the first to be identified, cloned and sequenced. The initial studies of mRNA splicing utilized the beta-globin gene. From the identification of its structure to the discovery of the globin genes, the knowledge gleaned from hemoglobin has aided scientists studying other systems. The study of β-globin gene regulation has developed into the model system for developmental switching and for long range enhancer activity. Research on globin gene regulation is still leading to the uncovering of new mechanisms of tissue specific and temporal gene expression.

Hemoglobin

In the late 1700's Antoine Lavoisier determined that oxygen was critical for the support of life (1). This discovery drove scientists to find the protein that transported oxygen through our blood. Over the next two centuries, many studies were performed, and it was determined that hemoglobin was responsible for oxygen delivery in the blood. These studies culminated with the elucidation of hemoglobin's structure by x-ray

crystallography (2). The structure of Perutz et. al. conclusively demonstrated that hemoglobin is a tetrameric protein composed of two beta-globin polypeptides, two alpha-globin polypeptides and four heme groups (fig 1). Hemoglobin is the one of the most abundant proteins in the blood.

Its role as an oxygen carrier makes it essential for life, and many hemoglobinopathes in humans result in much suffering and an early death. In humans there are at least 400 variants of hemoglobin. The first hemoglobin variant causing hemoglobinopathy to be identified was the M-Saskatoon variant (3). Soon after, sickle cell anemia was found to be due to a single amino acid substitution in the β chain (4). There are four major classes of clinically important hemoglobin variants: sickle variants, unstable variants, abnormal oxygen affinity variants, and structural variants that result in a thalassemic phenotype (5). The majority of these variants were identified through studies of hemoglobin properties and not through the sequencing of the globin genes.

Hemoglobin Switching

Hemoglobin is not just a simple oxygen tank; instead, hemoglobin is a sophisticated delivery system to provide oxygen to all parts of the body. Four oxygen molecules can bind to hemoglobin in a reversible and cooperative manner. The first O_2 molecule binds to deoxyhemoglobin with low affinity, initiating a chain reaction of conformational changes between the polypeptides. These changes result in a higher affinity of O_2 binding for each subsequent molecule of oxygen bound to hemoglobin. Hemoglobin binds its fourth O_2 with a 100-fold higher affinity than its first. Hemoglobin's multiple subunits and O_2 binding sites permit a very sensitive response to

small changes in oxygen concentration. This sensitivity leads to the binding of oxygen in the lungs and the release of oxygen in the tissues that allows for life.

Throughout development, different O_2 binding affinities are necessary for hemoglobin to effectively deliver oxygen, due to the different oxygen concentrations that occur during development and adult life. As an embryo's oxygen first comes from the mother's blood it is available at a lower concentration than would be found in the lungs of adults. To overcome this difficulty, the polypeptides that compose hemoglobin undergo a developmental switch (figure 2), allowing hemoglobin to act as a sensitive delivery system in embryonic, fetal, and adult life.

The mouse has been extensively utilized as a model for the study of hematopoiesis and erythropoiesis. Hematopoiesis is the development of the entire blood system cells, and erythropoiesis is the development of red blood cells (RBC), or erythrocytes. Mice are advantageous to use as a model system as they are easy to maintain and that blood samples for analysis are relatively easy to obtain. In addition to there is an extensive body of literature on mouse hematopoiesis and erythropoiesis (6-9). However, the switch between hemoglobin forms in mice is not as clear cut as it is in humans.

In mice, hematopoiesis first occurs in the yolk sac at 7-8 days post coitus (dpc), and then shifts to the fetal liver between 12 and 14 dpc. A final shift in hematopoiesis to the bone marrow occurs shortly before birth and bone marrow continues to be the main site of normal hematopoiesis for the rest of adult mouse life. At day 9 of mouse development, embryonic βh1 and fetal εy2 beta-like globins genes are expressed and an embryonic ζ alpha-like globin chain is produced (10,11). The levels of βh1 expression

slowly decrease from this point, while εy2 expression levels rise and then fall. The ζ-globin chains slowly decrease in abundance until 17 dpc when they are completely absent (11). Before the site of hematopoiesis moves to the fetal liver, the levels of εy2 expression increase reaching a peak around 13 dpc, and the adult forms of α-globin increase in abundance (11). By 12 dpc α is as common as ζ, and by 17 dpc it is nearly 100% of the alpha-like polypeptide found in hemoglobin (11). Coinciding with the switch in the site of hematopoiesis, the adult β-globin forms increase in abundance, and the fetal hepatic erythropoietic cells predominately produce adult β-globin chains (12,13). From this point forward adult β-globin polypeptides become the predominant peptides, and fetal globins are completely absent by 17 dpc (11). The end result is by birth mouse hemoglobin is composed of 2 polyeptides of β-globin and 2 polypeptides of α-globin.

The β-globin Genes

The β-globin gene locus of mice is composed of four functional genes and three pseudogenes. The genes are located on chromosome 7 in the following order 5'-εy2–βh0–βh1–βh2–βh3–β1–β2-3' (figure 3). The first genes in this set to be fully characterized in the mouse are responsible for the adult β-globin polypeptides. Konkel *et al.* sequenced and compared the two genes (β1 and β2) responsible for the adult β-globin polypeptide in mice (14). This group demonstrated that these two genes are closely related and seemed to have arisen from a duplication of a single beta-globin gene during the course of evolution. In 1980, Jahn *et al.* identified 5 beta-like coding regions that were linked to β1 and β2 (15). Using partial sequences, they identified a potential gene εy2 responsible for the embryonic/fetal globin polypeptide in mouse development and a

set of potential genes named βh0, βh1, βh2, βh3, which are homologous to the β-globin genes β1 and β2. The βh3 gene sequence revealed aberrations in the 5' sequence resulting in no translation to produce a normal β-globin, and as such was defined as a pseudogenes. Subsequently, complete sequencing and characterization of βh2 revealed that is also was a pseudogenes (16). Finally, complete characterization of βh1(17) and εy2 (18) revealed that these genes are the functional genes responsible for the embryonic/fetal beta-like globin polypeptides.

To complicate the picture, inbred strains of mice have two different haplotypes or allelic alternatives of the β-globin locus. Strains of the single type (HbbS), such as C57BL, synthesize only one type of adult β-globin chain from both β1 and β2. Sequencing of HbbS revealed that the β1 and β2 genes are identical and were named β^s and β^t, respectively (19). Other strains such as 129 have the diffuse haplotype (HbbD). These strains produce two β-globin chains β^{maj} and β^{min} in unequal amounts from the β1 (β^{maj}) and β2 (β^{min}) genes. Comparison of the sequences of the β1 and β2 genes of the HbbD and HbbS haplotypes revealed distinct differences in the genes from each haplotype and within the HbbD haplotype (19).

A Brief History of the Evolution of the β-globin Genes

Though hemoglobin was first identified from the blood of mammals due to the incredibly high levels found there, hemoglobins are in fact found in all groups of organisms from prokaryote to fungi, plants and finally in the familiar animals (20). Hemoglobin has more functions than just being a O_2 delivery system in these other groups. In bacteria, plants, and fungi hemoglobins may act as a transporter of oxygen to

electron transport chain, to sequester oxygen, to scavenge oxygen, as well as having some catalytic activity (20). The various hemoglobins appear to have all descended from an ancient common ancestral gene (figure 4). The ancestral gene precedes the divergence of prokaryotes and eukaryotes approximately 1800 million years ago (mya) (20). Next an ancestral gene to hemoglobin was in the last ancestor of plants and animals around 1500 mya (20). It is interesting that the common plant hemoglobins differ by 80% of the amino acids and yet folds into the same three dimensional structures as animal globins. The primary structure of vertebrate hemoglobins is highly related to inverterbrate hemoglobins suggesting that these genes diverged around 670 mya (20). As a side note it appears the myoglobin and hemoglobin diverged around 700 mya (21). Approximately 450 mya a duplication of the globin genes in vertebrates resulted in the precursors to α-like and β-like globin genes. This is supported by the fact that regardless of verterbrate species the α-like and β-like globin genes are 50% identical at the amino acid sequence level (20-22). The duplication of the α-like globin gene was fixed prior to the divergence of birds and mammals (22). This means that the α-like genes in each are orthologous. On the other hand the duplications resulting in the current β-globin locus in mammals happened after the divergence of the avian and mammalian lines (22). The formation of the current β-globin locus with fetal and adult globins started 180-200 mya when a duplication event resulted in two β-globin genes (21). The fetal genes then duplicated again around 110-130 mya resulting in ε-like and γ-like globin genes (21). A final duplication of the adult β-like globin gene resulted in δ-like and β-like globin genes around 85-100 mya (21). Over the course of the final 100 million years to the present

multiple gene conversions and additional gene duplications resulted in the current β-globin locus in mice.

Globin Gene Regulation and the Locus Control Region

Upstream of the εy2 gene there are a series of DNAseI hypersensitive sites (HS). These sites were first characterized in humans (23,24) and subsequently their presence in the mouse beta globin locus was demonstrated (25,26). Deletions of this region severely down regulate β-globin expression in humans (27-29). Multiple studies have also demonstrated the necessity of these HS sites for high levels of β-globin gene expression (Figure 5), and so have they have collectively been named a locus control region (LCR).

The β-globin gene and the tyrosinase gene locus (30,31), the Th2 cytokine locus (32), and the T-cell receptor α/δ locus (33) are just a few of the over 38 known loci that contain LCRs. LCRs can function both as enhancers and insulators of the loci they control. When a LCR sequence is linked to a transgenic gene it confers tissue-specific and physiological levels of expression of the linked genes in a position-independent and copy number dependent manner (34-37).

Initial studies of human beta globin transgenic mice revealed that when the human gene was introduced into mouse embryos, it was expressed at the correct stage of development but its expression was very low and in many cases no expression could be detected (38-40). When the LCR was linked to the human globin gene, all mouse lines which contained an intact mini-locus expressed the globin gene at levels equivalent to the endogenous mouse beta-globin genes (34,40). This data suggested that the LCR acts as

an enhancer of the beta-globin genes. Further understanding of the function of individual HS sites was obtained through multiple deletion and transgenic studies.

HS1 is located 6kb upstream of the εy gene. When the human HS1 is linked to a β-globin gene in transgenic mice, it confers position independent expression of the linked gene but does not enhance the expression of the gene, suggesting a minor role for HS1 in the activity of the LCR(41,42). On the other hand, when the endogenous murine HS1 is deleted, the linked genes have a decrease in expression of approximately 24% (43). The deletion of the endogenous murine HS2 results in an approximate 30% decrease of the linked gene (44-46).When the human HS2 is linked to a beta globin gene in transgenic mice, high levels of globin expression are observed in a position independent manner (40,41). HS3 has been shown to have very similar effects to HS2 on beta globin expression in both transgenic studies(41,47) and in endogenous knockouts.(45,48) The deletion of HS4 decreases beta globin expression by approximately 30% (49), while the linkage of human HS4 to a globin gene in transgenic mice results in a 30% level of expression compared to the same gene linked to the entire LCR (42). Collectively, HS1-4 contains the enhancer activity of the LCR. HS5 has been implicated as an insulator element(50,51), and its deletion has minimal effects on beta globin expression (52). Finally, Bender et al. showed that the HS sites act additively rather than synergistically in enhancing the expression of genes of the β-globin locus (43).

The mechanism by which the LCR activates transcription of β-globin genes is not fully elucidated. Looping (53-55) and linking (56,57) models have been proposed to explain the necessity of LCR sequences for high level globin gene expression. In the

looping model, individual HSs form a larger LCR holocomplex through the cooperative binding of many proteins to each HS. This LCR holocomplex interacts with proteins bound to downstream globin gene promoters to form an active transcription complex (58-60) that directs globin gene expression. DNA loops are created as a consequence of these long-distance interactions. In the linking model, the LCR holocomplex interacts with "transcription factories" (61-65) that are outside the β-globin locus chromosomal domain. DNA is then spooled into the transcription factory until a promoter complex is encountered, which activates gene expression. Both models are supported by recent studies demonstrating that LCR sequences interact directly with downstream β-globin genes in erythroid cells (66,67).

The entire beta-globin locus is highly conserved throughout the mammalian class. This conservation includes not only the genes but also the LCR. Hardison et al. aligned the HS sites of human, galago, rabbit, goat, and mouse LCRs to identify regions of conservation (68). Within each HS, blocks with very high conservation were observed. These highly conserved regions corresponded with known putative transcription factor binding sites, such as CACCC, GATA1, and NF-E2-like binding sites (figure 6). HS2 was the most highly conserved region in the LCRs of the five species(68). HS3 is not as highly conserved as HS2 but still contains blocks of conservation between the five species(68). HS4 has blocks of conservation between human, galago, rabbit, and mouse (68). HS1 also displays regions of conservation, the highest of which corresponds with two GATA1 sites. The high conservation of these transcription factor sites suggests that they are important in the activation of β-globin gene expression by the LCR. The bulk of

the work of this dissertation will cover an investigation of the role of the conserved NF-E2 and CACBP sites in LCR activation of endogenous globin genes.

NF-E2-like Sites

Previous studies have provided some insight into the role of the NF-E2-like sites in the individual HS sites. Within HS2, 2 NF-E2-like sites are highly conserved with a single base pair difference within the alignments for each (68). These highly conserved sites in HS2 are in a tandem array, and when they are deleted in transgenic mice with HS2 linked to the β-globin gene, the transgene is expressed at very low levels(69-71). The drastic drop in expression linked to the absence of the NF-E2-like sites in HS2 suggests that they play a critical role in HS2's enhancer activity. One major block of conservation in HS3 is a single NF-E2-like site that is perfectly conserved between the five species (68). When a small region of DNA containing the conserved NF-E2-like binding site was added to a smaller HS3 fragment in transgenic mice, the expression of the reporter gene increased (72). Within the blocks of conservation in HS4, a single NF-E2-like site is conserved in all the species except mice which has a block of four base pairs with no homology to the other species in that region(68). Mutations of the region of conservation containing the NF-E2-like binding site in human HS4 results a loss in DNaseI sensitivity (73).

The above studies suggested that the proteins that bind to the NF-E2 sites in the LCR have an important role in mediating globin gene activation. In order to identify the proteins involved in this process, many groups began looking for factors that could bind to these sites. Initially, activating protein 1 (AP-1) was a prime candidate. AP-1 is a

heterodimer of the jun and fos families of proteins. To determine if AP-1 was the factor that bound to the NF-E2 site, several experiments where performed. The most important showed that these sites are bound by an erythroid specific factor and not by the ubiquitously expressed AP-1 (74). A second candidate NF-E2 was proposed to bind to the conserved sites in the LCR. Studies to determine if NF-E2 could bind to these sites generated conflicting results. Ney *et al.* (74)showed that NF-E2 appears to bind to these sites in K562 cells, where as Moi *et al.* (70) suggested that an NF-E2 related protein in facts what binds to these sites. Another factor that was identified using the NF-E2 sequences, LCR-F1 (75) was subsequently shown to have no globin phenotype when deleted in mice(76).

CACBP Sites

Within HS2 and HS3 there are a set of highly conserved CACBP sites. A single CACBP site was highly conserved in HS2 (68). When this site is deleted in an HS2/ β-globin gene construct and tested in transgenic mice, a modest reduction in transgene expression was observed compared to wild-type HS2/ β-globin transgene expression (69,71). HS3 also contains blocks of conservation between the five species, including a single CACBP site(68). In addition to these highly conserved sites, four other CACBP sites are found in HS1-4 of the mouse LCR. Whereas the position and sequence of the highly conserved sites in HS3 and HS2 are nearly the same in all five species, the position and sequence of the four additional sites are not as well conserved.

Currently the most likely candidate for binding to the CACBP sites in the LCR during adult erythropoiesis is erythroid krupple like factor (EKLF). In 1993, EKLF was

identified as an erythroid specific factor in mice with high homology to the Drosphila body determining gap gene *Krupple* (77). EKLF's consensus binding sequence was predicted to be CCN CNC CCN (77) and further experiments confirmed this binding sequence and refined it further to CCA CAC CCT and also showed that single base mutations of this sequence severely impairs EKLF's transcriptional activation of reporter genes (78). The above data strongly suggested that EKLF is important for expression of adult β-globin genes as they contain CACBP sites in the promoters that match to EKLF's binding sequence. Interestingly, the single base mutations in the CACBP site tested in the above experiment corresponded with mutations found in the promoter of β-globin of some non-deletion β-thalassemic human patients (79,80). EKLF expression was demonstrated to be regulated in a developmentally specific manner in all types of erythroblasts in the fetal liver and adult bone marrow (81). Knock-out of EKLF in mice resulted in death before birth, due to severe defects in adult β-globin gene expression (81,82). Surprisingly, EKLF knock-outs express the fetal/embryonic globin genes (βh1 and εy2) in a normal manner (81,82).

The Mouse as a Model for β-globin Gene Regulation

The common mouse is an attractive model system for the study of long range gene enhancement. It has been extensively studied and so information on its husbandry, physiology and genetics are easily obtained. In addition the many parallels between human and mouse genetics are highly studied. The Townes laboratory has produce mice in which human globin genes replace the mouse globin genes and express only human hemoglobin within normal levels (TMT and TR unpublished). The mouse is easy to feed

and house and it is also relatively affordable to do so. Common diseases of mice are known and can be protected against eliminating disease as a variable in any experiment. Females reach reproductive age at 3 to 6 weeks of age and remain fertile for up to 18 months of their 2- year life span. Females enter estrus every 3 to 5 days and pregnancy last for 18-21 days. This rapid turnover allows for the production of many generations very quickly. In addition litters are of a larger size, around 6-12 pups per litter, allowing rapid expansion of a mouse line for experimentation. Finally, the availability of many inbred strains, whose DNA sequences have been fairly well characterized, allows for allele specific assays such as those used to determine the effects of HS2 in the LCR (44).

The mouse is an attractive model for more then just the above reasons. The ability to create mice with a discrete change is a major reason for selecting mice as a model system for the study of β-globin gene regulation. Embryonic stem (ES) cells from mice are relatively easy to maintain in culture. Using the ES cell's natural homologous recombination ability it is possible to make distinct changes that are defined by the experimenter to the mouse genome. A wide variety of selection strategies allows for the rapid identification of ES cell with the desired change. The ES cells can then be injected into blastocysts and incorporated into the developing embryo. These blastocysts can then be implanted into females and chimeric pups can be generated. Male chimeric mice can be breed to produce mice which are heterozygous for the change desired.

Recombineering

Previous studies have only examined the role of each NF-E2-like site or CACBP site in the context of one HS in transgenic mice. Though transgenic mice are a very

powerful tool for understanding gene expression, only a study in the context of the endogenous LCR can give a true representation of the role of these conserved sites. With the advent of efficient systems to generate mutations within bacterial artificial chromosomes (BACs), one can efficiently generate discrete mutations within larger regions of interest, and a study of these sites in the endogenous locus becomes possible.

Recombineering was developed in 2000 by Yu *et al.* (83) as an efficient system for chromosome engineering in *E. coli*. This system was rapidly adapted for recombinogenic targeting of bacterial artificial chromosomes (BAC) (84,85). In this system, a BAC is maintained in a *recA*$^-$ strain of *E. coli* to prevent genomic rearrangement. Within this strain, a defective lambda prophage was developed which provides the recombination genes *exo*, *bet*, and *gam* under the control of a temperature-sensitive λ cI-repressor (83). Upon induction of these recombination genes a single stranded oligonucleotide with a total of 70-140 bp of homology and a desired mutation is efficiently targeted and recombines with the BAC (84,85). Mutations such as single-base changes, deletions and insertions can be easily introduced into a BAC with this system, whereas with traditional subcloning methodologies this can be a very difficult feat and at times impossible.

Introduction to the Chapters

The next two chapters of this dissertation are papers I have coauthored during my graduate studies. The first paper "Conserved NF-E2 binding sites are not essential for high-level β-globin locus control region activity" demonstrated the role of the conserved NF-E2 sites in LCR activity. Mutation of these sites resulted in a relatively mild decrease in β-globin gene expression from the mutant allele compared to the wild type allele at all

developmental stages. The second paper "Highly conserved CACBP sites in HS3 and HS2 are dispensable for high-level β-globin locus control region activity" showed the mutation of the highly conserved CACBP sites. The mutation of these sites resulted in a 10% decrease in expression of the linked β-globin genes. Finally the conclusion proposes further experiments for understanding the role of conserved sites in the LCR.

Figure 1. Structure of Hemoglobin. The α-globin chains are indicated by white, and the β-globin chains are indicated by black. The heme groups are indicated by grey disks. (Adapted from Perutz Nobel Lecture 1962 (86))

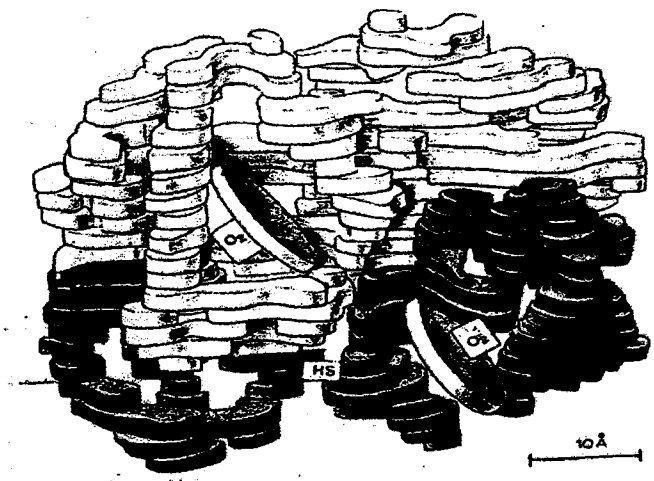


Figure 2. Mouse globin chain synthesis during development. The individual α- and β-globin protein chains, expressed as a percentage of the total globin chains, are plotted against time. The α-like globin chains are drawn with a thick line, while the β-like globin chains are drawn with a thin line. (Adapted from Ryan 1990 (87))

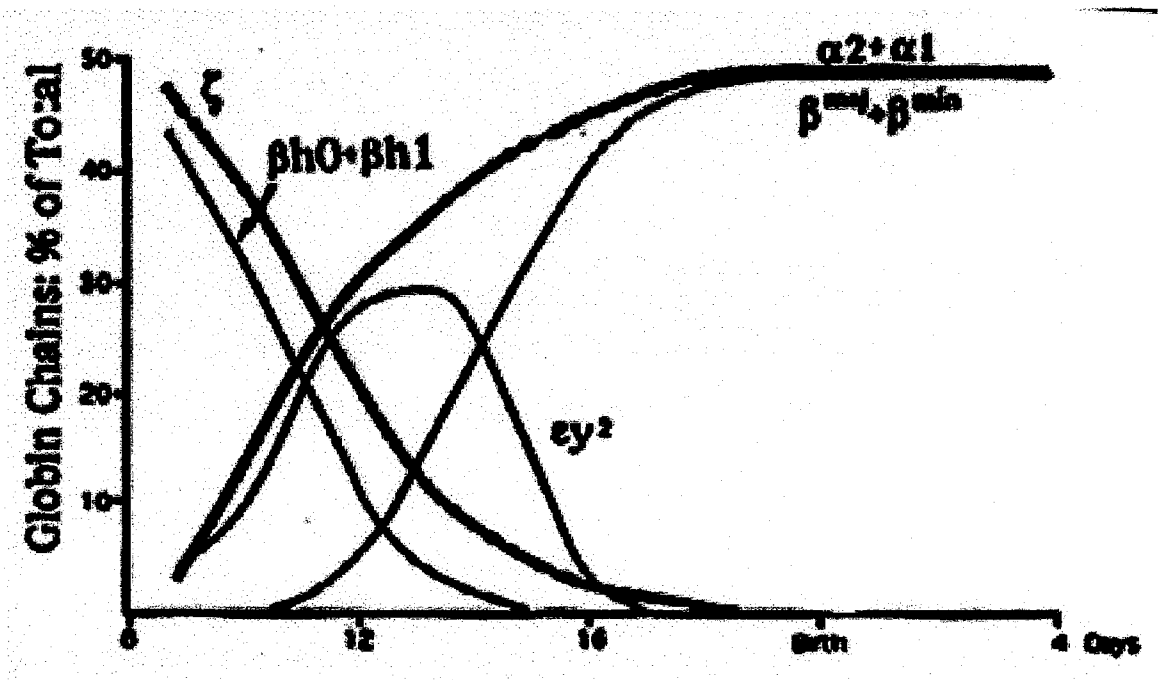


Figure 3. The β-globin locus of mice on chromosome 7. The red boxes represent the functional globin genes. The black boxes represent the pseudogenes present in the locus. The arrows represent the series of DNaseI hypersensitive sites upstream of the beta globin genes.

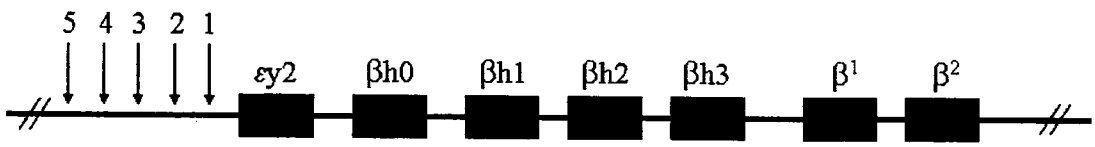


Figure 4. A schematic overview of hemoglobin gene evolution. Speciation events are depicted as black circles and gene duplication events are depicted as black boxes. Mya, millions of years ago

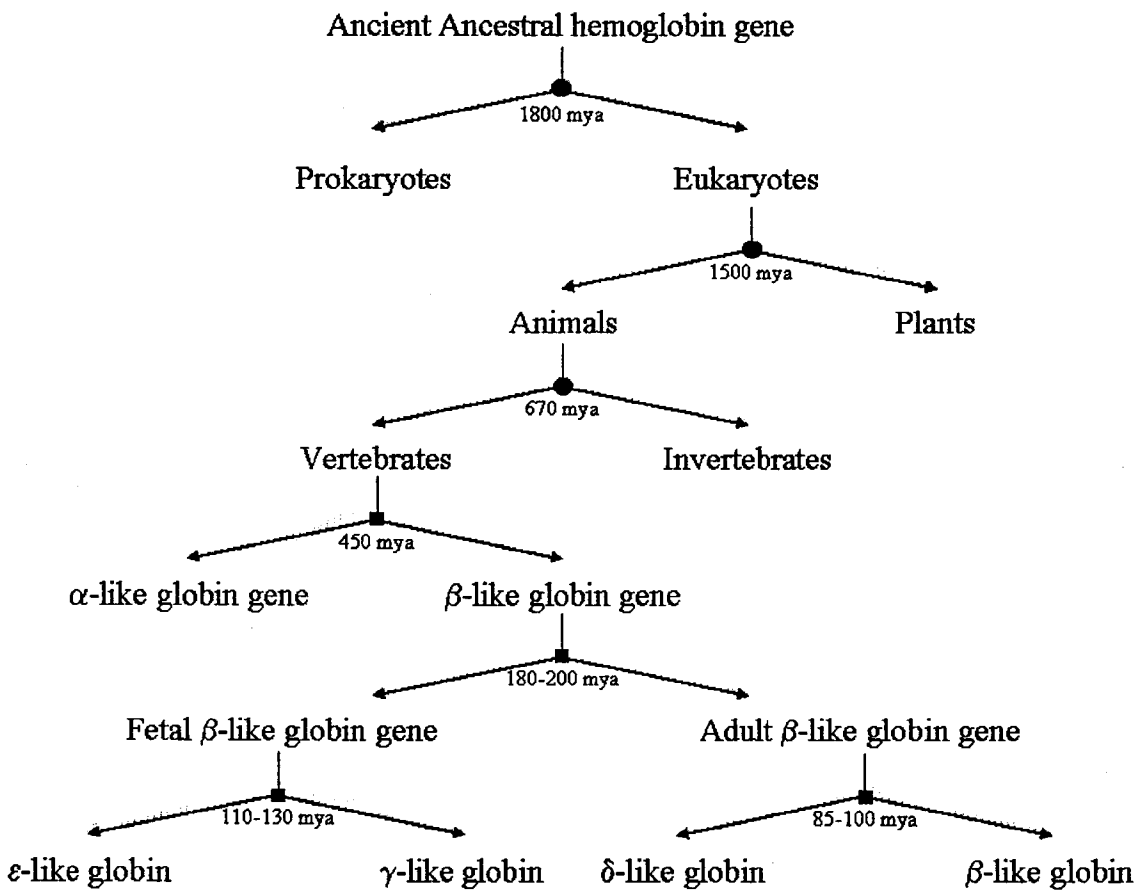


Figure 5 A diagram of LCR transgenic studies. Two constructs showed the importance of the LCR for β-globin gene activation. The first construct was a β-globin gene alone; its expression was very low compared to the endogenous β-globin gene expression. When the β-globin gene was linked downstream of the LCR in two different constructs, expression was at near endogenous levels. In addition, inclusion of the LCR in the transgenic construct resulted in expression in 100% of the mouse lines generated, while the β-globin gene alone allowed only spotty expression in mouse lines.

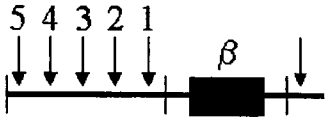
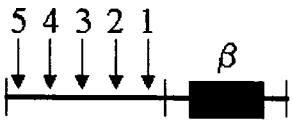
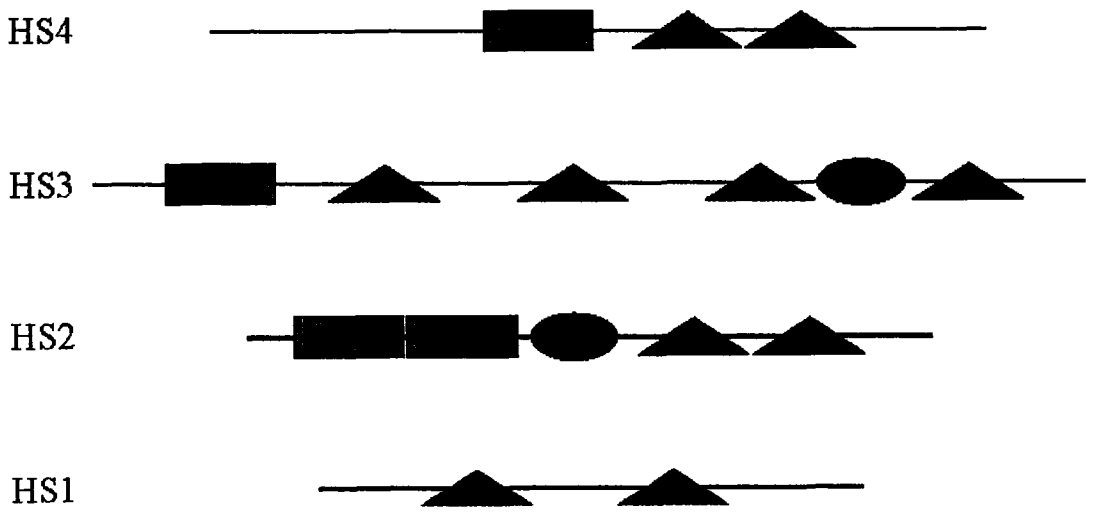
Construct	Group	Fraction of Expressors	Average % expression per gene copy
	Townes et al.	15/20	1.9
	Chada et al.	4/10	0.9
	Ryan et al.	7/23	0.3
	Grosveld et al.	7/7	100
	Ryan et al.	22/22	108.5

Figure 6. Conserved transcription factor binding sites in the HS of the β-globin LCR. HS4, HS3, HS2, and HS1 are represented by the black lines; conserved NF-E2-like, CACBP, and GATA sites are represented by purple boxes, red circles, and green triangles, respectively. HS4 has one NF-E2-like site and two GATA sites which are highly conserved. HS3 has one NF-E2-like site, one CACBP site, and four GATA sites that are highly conserved. HS2 has two NF-E2-like sites, one CACBP site, and two GATA sites which are conserved. Finally, HS1 has two GATA sites conserved between the five mammalian species.



CONSERVED NF-E2 BINDING SITES ARE NOT ESSENTIAL FOR HIGH-LEVEL, β-GLOBIN LOCUS CONTROL REGION ACTIVITY

By

ANDREA A. SVENDSEN, KEVIN M. PAWLIK, JINXIANG REN AND TIM M. TOWNES

Submitted to Nucleic Acid Research

Abstract

Conservation of NF-E2 binding sites in the β-globin Locus Control Region (LCR) of human, galago, rabbit, goat and mouse suggests that these sequences play an important functional role in globin gene expression. To define the role of these sequences in vivo, we replaced an 18 kb region of the wild-type LCR (HS1-4) in murine embryonic stem (ES) cells with an 18 kb LCR fragment (HS1-4) containing mutations of the conserved NF-E2 sites in HS4, HS3 and HS2. Mice were derived from these LCR knockin ES cells, and globin gene expression was analyzed. Surprisingly, β-globin gene expression is reduced by only 23% in these animals and mice that are homozygous for the mutations are not anemic. These results suggest that the highly conserved NF-E2 binding sites in the β-globin LCR play a significant, but nonessential, role in long-range enhancement of β-globin gene expression.

Introduction

The β-globin LCR was originally identified as a set of DNase I Hypersensitive sites (HSs) mapped from 6 to 22 kb upstream of the human ε-globin gene (1,2). Subsequently, the functional significance of these sequences was demonstrated in transgenic mice (3), and the role of the LCR in tissue (4-6) and developmental (7,8) specificity of globin gene expression was defined. Although a 22 kb fragment of DNA containing HS1-5 directs maximal globin gene expression, individual LCR HSs function independently to enhance β-globin gene expression in cultured cells and in transgenic mice (5,9,10). The sites appear to function additively and not synergistically (11). The contribution of individual transcription factor binding sites in HS2, HS3 or HS4 was determined by mutating these sequences in individual HSs and assaying the fragments in transgenic mice (12-16).

Hardison et al. (17) aligned the HSs of the human, galago, rabbit, goat, and mouse LCRs to identify regions of conservation. Within each HS, regions of high conservation are observed. These highly conserved regions correspond to known and putative transcription factor binding sites, such as CACCC, GATA-1, and NF-E2. The HS2 core (15) is the most highly conserved region in the LCR (17). Two NF-E2 sites are highly conserved in HS2; a single base pair difference is detected in the alignments (17). When these sites are deleted in an HS2/β-globin gene construct and tested in transgenic mice, β-globin gene expression is reduced by 20 fold (14,15,18). These results suggest that the NF-E2 sites play a critical role in HS2 enhancer activity. HS3 is not as highly conserved as HS2 but also contains blocks of conservation between the five species (17). In HS3, a single NF-E2 site is perfectly conserved (17). Deletion of this site in HS3 reduces reporter gene expression significantly in transgenic mice (19). In HS4, blocks of conservation are also observed including a single NF-E2 site (17). Mutation of this site in HS4 results in a loss of DNase I sensitivity but no loss in enhancer activity (16).

To understand the role of conserved NF-E2 sites in endogenous LCR activity, all four conserved NF-E2 sites were mutated in the endogenous mouse LCR. We demonstrate that conserved NF-E2 sites in the endogenous LCR play a significant but nonessential role in the expression of embryonic/fetal and adult β-globin genes.

Materials and Methods

Recombineering of the β-globin BAC

A 180 kb bacterial artificial chromosome (BAC) clone (BPRC, CHORI) containing the murine HbbD β-globin locus was maintained in the DY380 strain of *E.*

coli. Mutations were made following standard recombineering procedures (20-22). Briefly, BAC-containing DY380 cells were grown at 30 °C in 30 mL of Luria-Bertani (LB) broth to an optical density of 0.5 at an absorbance of 600 nm. The temperature was then increased to 42 °C for 15 min to induce expression of the bacteriophage λ-Red genes *exo*, *bet* and *gam*. After heat induction, cells were placed on ice for 25 minutes, washed three times in 10% glycerol and resuspended in a final volume of 200 μL of 10% glycerol. Fifty to two hundred nanograms of mutagenic oligonucleotides were electroporated into 50 μL of the cells at 1.8 kV, 200 Ω, and 25 μF (BioRad GenePulser II). After 1 hr of recovery at 30 °C, the cells were seeded into a 96 well BioBlock (Labcor) at 10,000 cells/well and 1,000 cells/well and incubated overnight at 30 °C. One microliter aliquots were assayed by PCR to identify pools containing mutated BAC clones. The positive pools were then diluted, cultured, and assayed. This process was repeated until 10 cell/well pools containing mutated BACs were obtained. Dilutions of these pools were then spread onto LB agar plates containing tetracycline (5 μg/mL) and chloramphenicol (12.5 μg/mL) and incubated at 30 °C overnight. The following oligos were used to make the mutations in HS4, HS3, and HS2, respectively: HS4, ATGTCAGAATGAGATGGCTTTCCCCACCCCTTTACTAGTCTCCCGGGCTCCCA CCCATGGCAAGGGGAGTGAGATCCTGCCAAG; HS3, TCTGTAAGTGTAAATTTTGGAGCACAGGTTTCTAAGACTTTCCCGGGGCAAAC CCTAGGCCTCCTAGGGACTGAGAGAGGCTGCTT; HS2, CAGCCTTGTGAGCCAGCATCAGGCTTGAGCACAGCAGTGCCCCGGGATGCCC CGGGATGCTGAGGCTTAGGGTGTGTGGCCAGATGTTTTCAGCTG.

The underlined portion of each oligo is the mutation of the NF-E2 site. All oligos were purchased from Integrated DNA Technologies (IDT) and purified by polyacrylamide gel electrophoresis (PAGE). All PCR primers used in this study were also purchased from IDT. For the identification of the mutations in the BAC, the following primers were used: HS4F1, ACCCCTTTACTAGTCTCCC; HS4R, GTTCCTGCCCTCTAAATCTGC; HS3F1, ACAGGTTTCTAAGACTTTCCC; HS3R, TTAGCAGGTGTTCTTGGAC; HS2F1, TTGAGCACAGCAGTGCCCC; HS2R, GTTATGTCACACAGCAAGGC.

Generation of ES cell targeting vector

A capture vector (pCV6) was designed to retrieve a 27.9 kb fragment of DNA containing LCR 5' HS5 through HS1 and the εy2 gene from the mouse β-globin BAC. The 5' and 3' capture homologies were produced by PCR and subcloned into plasmid pKP114 to construct pCV6 (see Supplemental Figure 3). In brief, the 395 bp 5' homology fragment was amplified from the mouse β-globin BAC using primers CV6-5' Fwd and CV6-5' Rev, and this fragment was subsequently digested with Sal I and Nsi I. The 430 bp 3' homology fragment was amplified using primers CV6-3' Fwd and CV6-3' Rev, and this fragment was then digested with Nsi I and Cla I. These two homologies were ligated into the Sal I- and Cla I-digested cloning vector pKP114 and transformed into *E. coli* (XL 10-Gold, Stratagene). PCR was performed using ExTaq polymerase (Takara) according to the manufacturer's protocols. The primer sequences were:

CV6-5' Fwd, CCACCGTCGACACGCGTGAGAATATGTCAGTACTTCCC;

CV6-5' Rev, CCACCATGCATCAGACATACTCCTCTACCC;

CV6-3' Fwd, CCACCATGCATAGATGCCATGGAAGGCTTCTAC;

CV6-3' Rev, CCACCATCGATTCGCGACCATGAAGCTTGCTGTGATAC. Restriction enzymes and T4 DNA ligase were used according to the manufacturer's protocols (Roche).

To recombineer the ES cell targeting vector, pCV6 was linearized with Nsi I and 50 to 400 ng of linearized plasmid was electroporated into DY380 cells containing the mutated BAC. After a 1 hr recovery at 30 °C, the cells were spread onto LB agar plates with ampicillin (50 μg/mL) for overnight culture at 30 °C. Single colonies from these plates were isolated and transferred to 3 mL LB broth with ampicillin for another overnight incubation at 30 °C. One microliter of culture was assayed in a 25 uL PCR reaction using the following primers to confirm the capture of the LCR: mBG52620F, GGGAAGCACAGGAAGACTCA; M13-20, ACGTTGTAAAACGACGGCCA; mBG25712R, CCCAAACACACAGCAGAACA; and ψM13R, CACAGGAAACAGCTATGACCA. Three positive clones were obtained and plasmid DNA was isolated using a standard alkaline lysis procedure (23). The plasmid DNA was then transformed into XL 10-Gold cells, and the cells were spread onto LB agar plates with ampicillin. After overnight incubation at 37 °C, single colonies were isolated, expanded in liquid culture, and assayed as above for LCR sequences. Positive cells were further analyzed by PCR to confirm the presence of the NF-E2 mutations. To ensure that the LCR was captured correctly, a series of restriction digests was performed. These digests confirmed that the entire 27.9 kb fragment of the LCR plus homology regions was captured.

Knockin mutations in ES cells

The targeting vector containing the LCR with NF-E2 site mutations was linearized with Nsi I and electroporated into HS1-4 KO ES cells (KMP and TMT, unpublished). HS1-4 KO ES cells are HM-1 (24) cells in which an 18 kb fragment of the LCR has been deleted and replaced with a mouse hypoxanthine phosphoribosyl transferase (HPRT) marker gene (25). These cells were maintained on $HPRT^+$ murine embryonic fibroblasts (MEFs) in media supplemented with 1.6 μg/mL hypoxanthine, 0.175 μg/mL aminopterin, and 4.8 μg/mL thymidine (HAT) prior to electroporation to confirm the presence of a functional HPRT gene. After electroporation, the cells were grown for 4 days in nonselective media to allow time for residual HPRT protein to decay. The cells were then replated at a low density onto gelatinized plates containing $HPRT^-$ MEFs and selected for the loss of HPRT with 10 μM 6-thioguanine (6-TG). Cells were cultured under selection for 14 days and then individual colonies were isolated.

Correctly modified HS1-4 KI ES cells were identified using PCR and DNA sequencing. Primers mHPRT(961.E1)F, ATGCCGACCCGCAGTCCCAG and mHPRT(61.E7)R, GGCCTGTATCCAACACTTCG were utilized to detect the loss of the HPRT marker gene. The previously described primers HS4F1, HS4R, HS3F1, HS3R, HS2F1 and HS2R were used to identify the mutations. Cell samples that were negative for HPRT and positive for the NF-E2 mutations by PCR were sequenced across the four HS sites. The following PCR primer pairs were used to generate the templates for sequencing: HS4F2, ATTTGCCCTCTCTTACATCT and HS4R; HS3F2, CCAGGACAGCCAGGACTAC and HS3R; HS2F2, CACTTCTTCATATTCTCTCTC

and HS2R; HS1F, CTCTCCCATAACCCATACCTC and HS1R, CTCACTCTCAAGCCTGTGTTAC.

Generation of the mice

C57BL/6 blastocysts were injected with HS1-4 KI ES cells and then transferred to pseudopregnant CD-1 females to generate chimeric mice. Male chimeras were bred to C57BL/6 females to obtain agouti pups whose paternal chromosomes are derived from the HS1-4 KI ES cells. The pups were genotyped using the aforementioned primers. Mice that were heterozygous for the mutations were mated to produce homozygotes. The NF-E2 site mutations were confirmed by sequencing as described above. In addition, the junctions between homology regions in the targeting vector and chromosomal DNA was sequenced to confirm correct gene targeting. The following PCR primer pairs were used to generate the templates for sequencing: mB24959F, TGGTAGAGGAGCAGGACAC; mB25957R, CTCATCAGTACCATCTCCTC; mB52700F, TACAGGCCCAGAGAGAGTC and mB53752R, TCCACATACTCTGTCTCAGC.

Allele specific RT-PCR

Blood RNA was extracted with Trizol LS (Invitrogen) according to the manufacturer's protocol. RNA was resuspended at the equivalent of $1x10^6$ cells/μL in diethylpyrocarbonate (DEPC)-treated H_2O.

Yolk sac RNA was prepared from 10.5 days post coitus (dpc) embryos. Individual yolk sacs were dissected from embryos and placed into 0.4 mL of Trizol (Invitrogen). Tissue was disaggregated by shearing through a 26 gauge needle and was

then incubated at room temperature for 15 min. RNA was subsequently isolated according to the manufacturer's protocol. RNA pellets from individual yolk sacs were resuspended in 8 μL of DEPC-treated H_2O.

Adult blood and yolk sac RNA samples were treated with DNase I (RNase-free; Invitrogen) in a 10 μL reaction for 30 min at 25 °C. From this reaction, 5 μL was used to generate cDNA using the SuperScript II First Strand Synthesis System for RT-PCR (Invitrogen).

Genomic DNA was obtained from adult mouse tail clips and 10.5 dpc embryos by overnight incubation at 55 °C in 500 μL of a solution containing 20 mM Tris, 20 mM EDTA, 100 mM NaCl, 1% SDS, and 500 μg/mL Proteinase K (PCR grade; Roche). DNA was extracted with phenol/chloroform (50:50), precipitated with ethanol, and resuspended in 350 μL or 100 μL of TE for tail or embryo DNA, respectively.

PCR of adult beta globins was performed using the following primers: mBG79746F, TGCTTCTGACATAGTTGTGTTGACT and mBG80043R, AAAGCTATCAAAGTACCGCTG. Two microliters of cDNA or genomic DNA was amplified for 32 cycles in a 20 μL PCR reaction. The mBG79746F primer was then end-labeled with γ^{32}P-ATP and purified on a Sephadex G-25 column (Boehringer Mannheim). A 20 μL primer extension reaction was performed using this labeled primer and 1 μL of the PCR reaction. To distinguish between the allele-specific products, 5 μL of the primer extension reaction was digested with Hae III for 6 hr at 37 °C. Two microliters of digested or undigested sample was electrophoresed through an 8% polyacrylamide gel. The gels were dried and bands were quantitated using a Molecular Dynamics Storm 860 PhosphorImager and ImageQuant software (GE Healthcare).

PCR of εy2 was performed as above with the following primers: Ey-1R, AACCCTCATCAATGGCCTGTGG and Ey-2R, TCAGTGGTACTTGTGGGACAGC. These primers were previously described by Fiering et. al. (26). An end-labeled Ey-1R primer was used for the primer extension reaction. Xba I was used to distinguish between the allele-specific products and quantitation of bands was carried out as above.

PCR of βh1 was performed as above with the following primers: mbh1 182524F, GCACACTTGAGATCATCTCC and bh1-5R, TAACCCCCAAGCCCAAGGATG. The bh1-5R primer was previously described by Fiering et. al. (26). An end-labeled bh1-5R primer was used for the primer extension reaction. Xba I was used to distinguish between the allele-specific products and quantitation of bands was carried out as above.

Hematology

Red blood cell (RBC) counts were determined using a HEMAVET 1700 hematology analyzer (CDC Technologies, Inc.). Hemoglobin concentration was determined spectrophotometrically after conversion to cyanmethemoglobin with Drabkin's reagent (Sigma). Before determining the hemoglobin concentration, red cell membranes were pelleted at 14000 RPM for 5 min in an Eppendorf centrifuge. Pelleting the membranes inhibits artifactually high values caused by membrane-bound, denatured hemoglobin. Packed Cell Volume (PCV) was measured using a JorVet J503 micro-hematocrit centrifuge (Jorgensen Laboratories, Inc). Reticulocyte numbers were determined by flow cytometry (FACSCalibur flow cytometer; BD Biosciences) after staining with thiazole orange (Sigma).

Results

Recombineering and embryonic stem cell mutation

Figure 1 illustrates the conserved NF-E2 sites in HS4, HS3, and HS2 of the murine β-globin LCR. We mutated these four sites by recombineering a murine β-globin BAC (see Supplemental Fig. 1). Gap repair was used to capture the mutated LCR (18 kb) and flanking sequences into a plasmid to generate a 30 kb targeting vector for homologous recombination in ES cells. The targeting vector was electroporated into HS1-4 KO ES cells which contain an 18 kb deletion of the LCR and insertion of an HPRT marker gene (Fig. 2). ES cells were selected for loss of the HPRT gene by growth in 6-TG. Resistant colonies were analyzed by PCR for NF-E2 site mutations in HS4, HS3, and HS2. PCR was performed with primers that are specific for mutated HS4, HS3, and HS2 sequences (Fig. 3A-C). Figure 3D demonstrates that the HPRT gene was replaced with an 18 kb LCR fragment containing the mutated NF-E2 sites.

Generation of mice bearing NF-E2 site mutations

HS1-4 KI ES cells were injected into wild-type (WT) C57BL/6 mouse blastocycsts to produce chimeric males. The chimeric males were subsequently bred to WT C57BL/6 females. Agouti offspring were either LCR NF-E2 +/+, HbbD/HbbS or LCR NF-E2 +/- , HbbD/HbbS. Genotyping was performed using DNA isolated from tail clips with primers specific for each of the mutations (Fig. 4). Mice that were heterozygous for the LCR NF-E2 site mutations were mated to produce homozygous offspring (LCR NF-E2 -/-, HbbD/HbbD). HS sites and chromosomal junctions were

sequenced and compared to WT alleles. The results clearly demonstrate the presence of HS4, HS3, and HS2 NF-E2 site mutations (Fig. 5) and correct chromosomal junctions.

Down-regulation of adult β-globin gene expression in LCR NF-E2 +/- mice

Samples of blood were obtained from 5 WT and 6 LCR NF-E2 +/- mice and RT-PCR of blood RNA was performed with primers common to all four adult β-globin alleles (β^s, β^t, β^{maj}, β^{min}) (Fig. 6A). As a control, PCR of tail DNA from the same animals was performed with the same primers. After primer extension with a 5' end labeled primer, RT-PCR (Fig. 6B) and PCR (Fig. 6C) products were digested with Hae III and separated on 8% polyacrylimide gels to distinguish HbbD and HbbS sequences (HbbD, single Hae III site; HbbS, two Hae III sites) (see Fig. 6A). Labeled bands were quantitated using a phosphorimager and $\beta^{maj,min}/\beta^{s,t}$ values were calculated. These values were normalized to $\beta^{maj,min}/\beta^{s,t}$ values for tail DNA PCR products (Fig. 6C) and the results are illustrated in Figure 6D. Surprisingly, $\beta^{maj,min}$ gene expression in LCR NF-E2 +/- mice was reduced by only 23% compared to $\beta^{s,t}$. These data demonstrate that mutations of the highly conserved NF-E2 sites in HS4, HS3 and HS2 have a relatively mild, but significant, effect on endogenous β-globin gene expression in vivo.

Down-regulation of embryonic εy2-globin gene expression in LCR NF-E2 +/- embryos

Ten and one-half dpc yolk sacs were obtained from 4 WT and 4 LCR NF-E2 +/- mice and RT-PCR of yolk sac RNA was performed with primers specific for εy2-globin alleles ($\varepsilon y2^{HbbD}$, $\varepsilon y2^{HbbS}$) (Fig. 7A). As a control, PCR of DNA from the same embryos was performed with the same primers. After primer extension with a 5' end labeled

primer, RT-PCR (Fig. 7B) and PCR (Fig. 7C) products were digested with Xba I and separated on 8% polyacrylimide gels to distinguish HbbD and HbbS sequences (HbbD, no Xba I site; HbbS, one Xba I site) (see Fig. 7A). Labeled bands were quantitated using a phosphorimager and $\varepsilon y2^{HbbD}/\varepsilon y2^{HbbS}$ values were calculated. These values were normalized to $\varepsilon y2^{HbbD}/\varepsilon y2^{HbbS}$ values for embryo DNA PCR products (Fig. 7C) and the results are illustrated in Figure 7D. Consistent with the results described above for adult genes, $\varepsilon y2^{HbbD}$ gene expression in LCR NF-E2 +/- embryos was reduced by only 20% compared to $\varepsilon y2^{HbbS}$. These data demonstrate that mutations of the highly conserved NF-E2 sites in HS4, HS3 and HS2 have a relatively mild, but significant, effect on endogenous εy2-globin gene expression in vivo. Similar results were obtained for βh1-globin gene expression (see Supplemental Fig. 2)

Hematology of LCR NF-E2 -/- mice

To determine the effects of LCR NF-E2 mutations on hematological values in adult mice, red blood cell indices were obtained. The data are listed in Table 1. Red blood cell counts in LCR NF-E2 -/- mice were not significantly reduced compared to WT mice, and reticulocyte counts in LCR NF-E2 -/- and WT mice were not significantly different. Hemoglobin concentrations of LCR NF-E2 -/- mice were in the normal range, and blood smears of LCR NF-E2 +/- and LCR NF-E2 -/- mice are identical to blood smears of WT mice (Fig. 8). These results suggest that the 23% down-regulation of globin mRNA does not result in significant anemia. As previously described, relatively small decreases in β-globin mRNA do not result in significant decreases in hemoglobin

levels (27). Increased translation of β-globin mRNA compensates for the decrease in mRNA abundance.

Discussion

Within the β-globin LCR of five species, several NF-E2 sites are highly conserved. To understand the role of these NF-E2 sites in LCR activity, we mutated all four conserved sites in the endogenous mouse LCR. We demonstrate that these four NF-E2 sites play a significant role in long-distance enhancement of β-globin gene expression. We also show that mice homozygous for mutations of these sites are not anemic; therefore, these highly conserved sequences are not essential for hemoglobin synthesis in the normal range.

The mechanism by which the LCR activates transcription of β-globin genes is not completely understood. Looping (28-30) and linking (31,32) models have been proposed to explain the requirement of LCR sequences for high level globin gene expression at all developmental stages. Both models are supported by recent studies demonstrating that LCR sequences interact directly with downstream β-globin genes in erythroid cells (33,34). In the looping model, individual HSs form an LCR holocomplex through the cooperative binding of many factors to each HS. This LCR holocomplex interacts with proteins bound to downstream globin gene promoters/proximal enhancers to form an active transcription complex (35-37) that directs globin gene expression. DNA loops are created as a consequence of these long-distance interactions. In the linking model, the LCR holocomplex interacts with "transcription factories" (38-42) that are outside the β-globin locus chromosomal domain. DNA is then spooled into the transcription factory

until a promoter/proximal enhancer complex is encountered. The LCR/promoter/proximal enhancer interaction completes the transcription complex that activates gene expression. Interactions of LCR/promoter/proximal enhancer complexes with transcription factories are also accommodated in the looping model; however, in this model the LCR/promoter/proximal enhancer complex is formed prior to interaction with the transcription factory.

Multiple deletion studies have demonstrated that HSs in the LCR function in a highly cooperative manner. Deletion of individual HSs has a relatively minor effect on globin gene expression (11,26,43,44); however, deletion of the entire LCR drastically inhibits transcription of beta-like globin genes at all developmental stages (45-47). We anticipated that mutation of the highly conserved NF-E2 sites in multiple HSs would significantly inhibit cooperative interactions and mimic deletion of the entire LCR. Surprisingly, these mutations reduced globin gene expression by only 20%. These results suggest that there is significant redundancy involved in the formation of a functional holocomplex. If holocomplex-forming sequences are this redundant, why are specific LCR sequences highly conserved? Perhaps these sequences are conserved because LCR activity at or near 100% confers a selective advantage for animals in the wild, but significantly less activity is necessary for survival under laboratory conditions.

The 20% reduction of β-globin gene expression resulting from deletion of the conserved NF-E2 sites is consistent with both the looping and linking models. In the looping model, the absence of conserved NF-E2 sites reduces the rate or stability of LCR holocomplex formation and, consequently, alters formation of the LCR/promoter/proximal enhancer transcription complex. In the linking model, mutation

of conserved NF-E2 sites reduces the rate or stability of LCR interactions with transcription factories and, consequently, inhibits the spooling required to form an active transcription complex.

Comparison of LCR sequences of human, galago, rabbit, goat and mouse revealed many conserved regions in addition to NF-E2 sites (17). These regions and the proteins that bind to them may provide the large number of interaction surfaces that are necessary for full LCR activity. Future experiments involving the mutation of combinations of conserved sequences, including CACCC boxes in HS3 and HS2 and GATA-1 sites in HS4, 3, 2 and 1, may reveal the minimal sequences necessary for maximal LCR activity.

Acknowledgements

This work was supported by a grant to T.M.T. from the NHLBI.

References

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

2. Forrester, W.C., Thompson, C., Elder, J.T. and Groudine, M. (1986) A developmentally stable chromatin structure in the human beta-globin gene cluster. *Proc Natl Acad Sci U S A*, **83**, 1359-1363.

3. Grosveld, F., van Assendelft, G.B., Greaves, D.R. and Kollias, G. (1987) Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell*, **51**, 975-985.

4. Forrester, W.C., Novak, U., Gelinas, R. and Groudine, M. (1989) Molecular analysis of the human beta-globin locus activation region. *Proc Natl Acad Sci U S A*, **86**, 5439-5443.

5. Ryan, T.M., Behringer, R.R., Martin, N.C., Townes, T.M., Palmiter, R.D. and Brinster, R.L. (1989) A single erythroid-specific DNase I super-hypersensitive site activates high levels of human beta-globin gene expression in transgenic mice. *Genes Dev*, **3**, 314-323.

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

7. Enver, T., Raich, N., Ebens, A.J., Papayannopoulou, T., Costantini, F. and Stamatoyannopoulos, G. (1990) Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. *Nature*, **344**, 309-313.

8. Behringer, R.R., Ryan, T.M., Palmiter, R.D., Brinster, R.L. and Townes, T.M. (1990) Human gamma- to beta-globin gene switching in transgenic mice. *Genes Dev*, **4**, 380-389.

9. Fraser, P., Hurst, J., Collis, P. and Grosveld, F. (1990) DNaseI hypersensitive sites 1, 2 and 3 of the human beta-globin dominant control region direct position-independent expression. *Nucleic Acids Res*, **18**, 3503-3508.

10. Fraser, P., Pruzina, S., Antoniou, M. and Grosveld, F. (1993) Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression on the globin genes. *Genes Dev*, **7**, 106-113.

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

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

13. Philipsen, S., Pruzina, S. and Grosveld, F. (1993) The minimal requirements for activity in transgenic mice of hypersensitive site 3 of the beta globin locus control region. *Embo J*, **12**, 1077-1085.

14. Caterina, J.J., Ciavatta, D.J., Donze, D., Behringer, R.R. and Townes, T.M. (1994) Multiple elements in human beta-globin locus control region 5' HS 2 are involved in enhancer activity and position-independent, transgene expression. *Nucleic Acids Res*, **22**, 1006-1011.

15. Caterina, J.J., Ryan, T.M., Pawlik, K.M., Palmiter, R.D., Brinster, R.L., Behringer, R.R. and Townes, T.M. (1991) 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**, 1626-1630.

16. Stamatoyannopoulos, J.A., Goodwin, A., Joyce, T. and Lowrey, C.H. (1995) NF-E2 and GATA binding motifs are required for the formation of DNase I hypersensitive site 4 of the human beta-globin locus control region. *Embo J*, **14**, 106-116.

17. Hardison, R., Slightom, J.L., Gumucio, D.L., Goodman, M., Stojanovic, N. and Miller, W. (1997) Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. *Gene*, **205**, 73-94.

18. Moi, P. and Kan, Y.W. (1990) Synergistic enhancement of globin gene expression by activator protein-1-like proteins. *Proc Natl Acad Sci U S A*, **87**, 9000-9004.

19. Jackson, J.D., Petrykowska, H., Philipsen, S., Miller, W. and Hardison, R. (1996) Role of DNA sequences outside the cores of DNase hypersensitive sites (HSs) in functions of the beta-globin locus control region. Domain opening and synergism between HS2 and HS3. *J Biol Chem*, **271**, 11871-11878.

20. Swaminathan, S., Ellis, H.M., Waters, L.S., Yu, D., Lee, E.C., Court, D.L. and Sharan, S.K. (2001) Rapid engineering of bacterial artificial chromosomes using oligonucleotides. *Genesis*, **29**, 14-21.

21. Yu, D., Ellis, H.M., Lee, E.C., Jenkins, N.A., Copeland, N.G. and Court, D.L. (2000) An efficient recombination system for chromosome engineering in Escherichia coli. *Proc Natl Acad Sci U S A*, **97**, 5978-5983.

22. Zhou, D., Ren, J.X., Ryan, T.M., Higgins, N.P. and Townes, T.M. (2004) Rapid tagging of endogenous mouse genes by recombineering and ES cell complementation of tetraploid blastocysts. *Nucleic Acids Res*, **32**, e128.

23. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning: a laboratory manual*. 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

24. Selfridge, J., Pow, A.M., McWhir, J., Magin, T.M. and Melton, D.W. (1992) Gene targeting using a mouse HPRT minigene/HPRT-deficient embryonic stem cell system: inactivation of the mouse ERCC-1 gene. *Somat Cell Mol Genet*, **18**, 325-336.

25. Melton, D.W., Ketchen, A.M. and Selfridge, J. (1997) Stability of HPRT marker gene expression at different gene-targeted loci: observing and overcoming a position effect. *Nucleic Acids Res*, **25**, 3937-3943.

26. Fiering, S., Epner, E., Robinson, K., Zhuang, Y., Telling, A., Hu, M., Martin, D.I., Enver, T., Ley, T.J. and Groudine, M. (1995) Targeted deletion of 5'HS2 of the murine beta-globin LCR reveals that it is not essential for proper regulation of the beta-globin locus. *Genes Dev*, **9**, 2203-2213.

27. Cividalli, G., Natham, D.G. and Lodish, H.F. (1974) Translational control of hemoglobin synthesis in thalassemic bone marrow. *J Clin Invest*, **53**, 955-963.

28. Townes, T.M. and Behringer, R.R. (1990) Human globin locus activation region (LAR): role in temporal control. *Trends Genet*, **6**, 219-223.

29. Orkin, S.H. (1990) Globin gene regulation and switching: circa 1990. *Cell*, **63**, 665-672.

30. Engel, J.D. and Tanimoto, K. (2000) Looping, linking, and chromatin activity: new insights into beta-globin locus regulation. *Cell*, **100**, 499-502.

31. Bulger, M. and Groudine, M. (1999) Looping versus linking: toward a model for long-distance gene activation. *Genes Dev*, **13**, 2465-2477.

32. Li, Q., Barkess, G. and Qian, H. (2006) Chromatin looping and the probability of transcription. *Trends Genet*, **22**, 197-202.

33. Carter, D., Chakalova, L., Osborne, C.S., Dai, Y.F. and Fraser, P. (2002) Long-range chromatin regulatory interactions in vivo. *Nat Genet*, **32**, 623-626.

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

35. de Laat, W. and Grosveld, F. (2003) Spatial organization of gene expression: the active chromatin hub. *Chromosome Res*, **11**, 447-459.

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

37. Patrinos, G.P., de Krom, M., de Boer, E., Langeveld, A., Imam, A.M., Strouboulis, J., de Laat, W. and Grosveld, F.G. (2004) Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. *Genes Dev*, **18**, 1495-1509.

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

39. Kosak, S.T. and Groudine, M. (2004) Form follows function: The genomic organization of cellular differentiation. *Genes Dev*, **18**, 1371-1384.

40. Kosak, S.T. and Groudine, M. (2004) Gene order and dynamic domains. *Science*, **306**, 644-647.

41. Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W. *et al.* (2004) Active genes

dynamically colocalize to shared sites of ongoing transcription. *Nat Genet*, **36**, 1065-1071.

42. Ragoczy, T., Bender, M.A., Telling, A., Byron, R. and Groudine, M. (2006) The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation. *Genes Dev.*

43. Hug, B.A., Wesselschmidt, R.L., Fiering, S., Bender, M.A., Epner, E., Groudine, M. and Ley, T.J. (1996) Analysis of mice containing a targeted deletion of beta-globin locus control region 5' hypersensitive site 3. *Mol Cell Biol*, **16**, 2906-2912.

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

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

46. Schubeler, D., Groudine, M. and Bender, M.A. (2001) The murine beta-globin locus control region regulates the rate of transcription but not the hyperacetylation of histones at the active genes. *Proc Natl Acad Sci U S A*, **98**, 11432-11437.

47. Zhou, D., Pawlik, K.M., Ren, J., Sun, C.W. and Townes, T.M. (2006) Differential binding of EKLF to embryonic/fetal globin gene promoters during development. *J Biol Chem.*

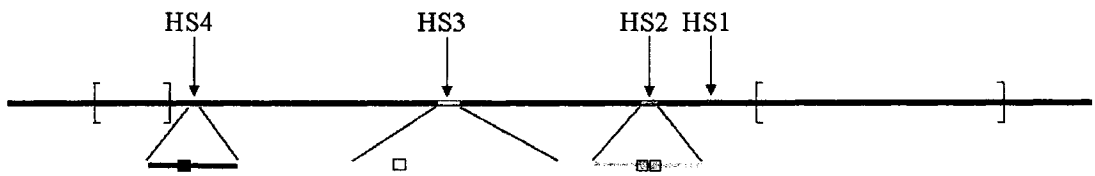
Table 1. Hematological values of WT and LCR NF-E2 -/- mice.

Mice	RBC $x10^6/\mu L$	Hb g/dL	PCV %	Retic %	RDW %	MCV fL	MCH pg
WT n=17	9.2±0.7	13.0±1.2	44.7±3.1	3.1 ± 0.8	18.2±0.8	48.7±1.8	14.2±0.7
NF-E2 -/- n=18	9.0±0.7 (0.546)	12.1±0.9 (0.016)	42.7±3.7* (0.094)	3.2 ± 0.8 (0.890)	20.4±1.2 (<0.0001)	47.4±2.9* (0.115)	13.4±0.6 (0.002)

RBC, red blood cell; Hb, hemoglobin; PCV, packed cell volume; Retic, reticulocytes; RDW, red cell distribution width; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin. Values represent the mean ± SD. P values in parentheses were calculated using a two-tailed unpaired Student's t-Test. *n= 17

Figure 1. Mouse β-globin LCR and mutagenic oligos. (A) Diagram of HS1-4 of the mouse β-globin LCR. The bracketed regions are 5' and 3' DNA fragments used for homologous recombination in ES cells. The core regions are highlighted in red, yellow, blue, and green for HS4, HS3, HS2, and HS1, respectively. Boxes in the enlarged regions of HS4, 3 and 2 indicate the relative position of conserved NF-E2 sites. (B) Sequence of conserved NF-E2 sites and surrounding regions. The conserved sites are indicated in bold, and the mutations are indicated below in red.

A.



B.

HS4

ATGTCAGAATGAGATGGCTTTCCCCACCCCTTTACTAGTC**TGAGAT**CCTCCCACCCATGGCAAGGGGAGTGAGATCCTGCCAAG
TCTCCCGGGCTC

HS3

CTGTAAGTGTAAATTTTGGAGCACAGGTTTCTAAGACTT**TGACTCAG**CAAACCCTAGGCCTCCTAGGGACTGAGAGAGGCTGCTT
TTTCCCGGGGCA

HS2

GTGAGCCAGCATCAGGCTTGAGCACAGCAG*T***GCTGAGTCATGCTGAGTCAT**GCTGAGGCTTAGGGTGTGTGGCCAGATGTTTT
TGCCCCGGGATGCCCCGGGATG

Figure 2. Schematic of knockin strategy. HS1-4 KO ES cells are murine HM-1 ES cells in which an 18 kb fragment of the LCR has been deleted and replaced with an HPRT marker gene. The targeting vector is composed of homology regions totaling 10 kb (4 kb 5' and 6 kb 3') and an 18 kb fragment of the LCR containing mutated NF-E2 sites in HS4, HS3, and HS2. After homologous recombination, HS1-4 KI ES cells contained one mutant HbbD allele and one WT HbbD allele.

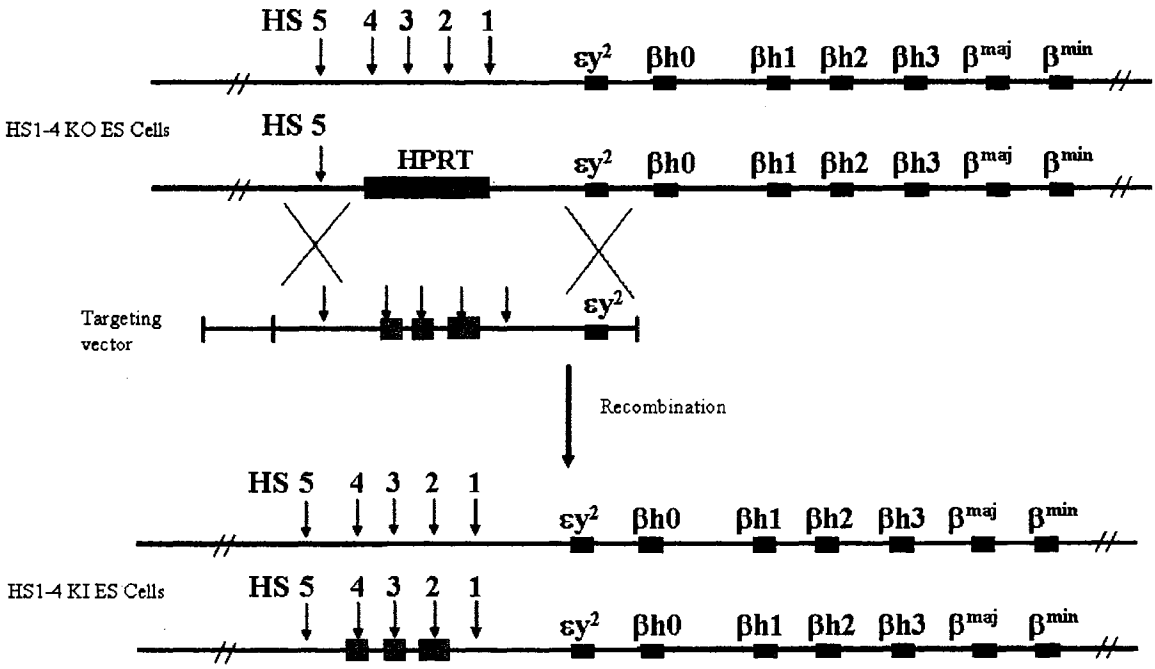


Figure 3. PCR genotyping of ES cells. The top panels of A, B and C demonstrate PCR products obtained with primers that are specific for the NF-E2 site mutations in HS1-4 KI ES cells. The lower panels are controls demonstrating intact core HS sites in HS1-4 KI cells. Panel D demonstrates the loss of HPRT in ES cells containing the knockin allele

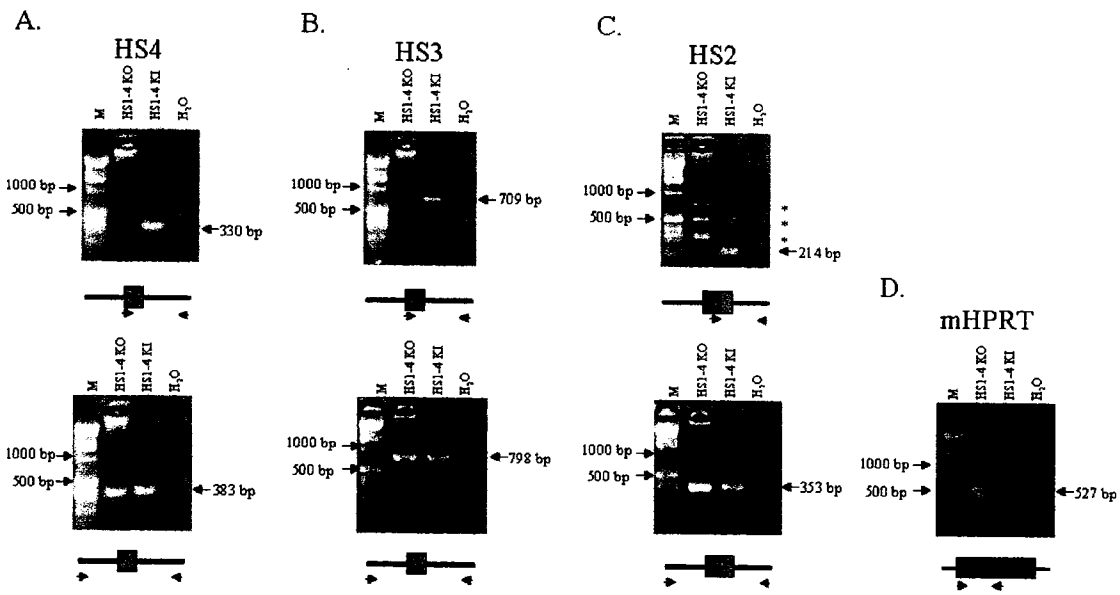


Figure 4. PCR genotyping of mice. M is a marker, +/+ is WT mouse tail DNA, +/- is mutant NF-E2 heterozygous mouse tail DNA, -/- is mutant NF-E2 homozygous mouse tail DNA and H_2O is a water (no DNA template) control. The top panel of A, B and C demonstrates the presence of NF-E2 site mutations in HS4, HS3 and HS2, respectively. The bottom panels of A-D demonstrate the presence of intact HS site cores in all mouse lines. Homozygous NF-E2 site mutant mice were confirmed by the presence of two HS3 HbbD alleles (Panel B, bottom). -/- represents an HbbD/HbbD mouse and +/- represents an HbbS/HbbD mouse.

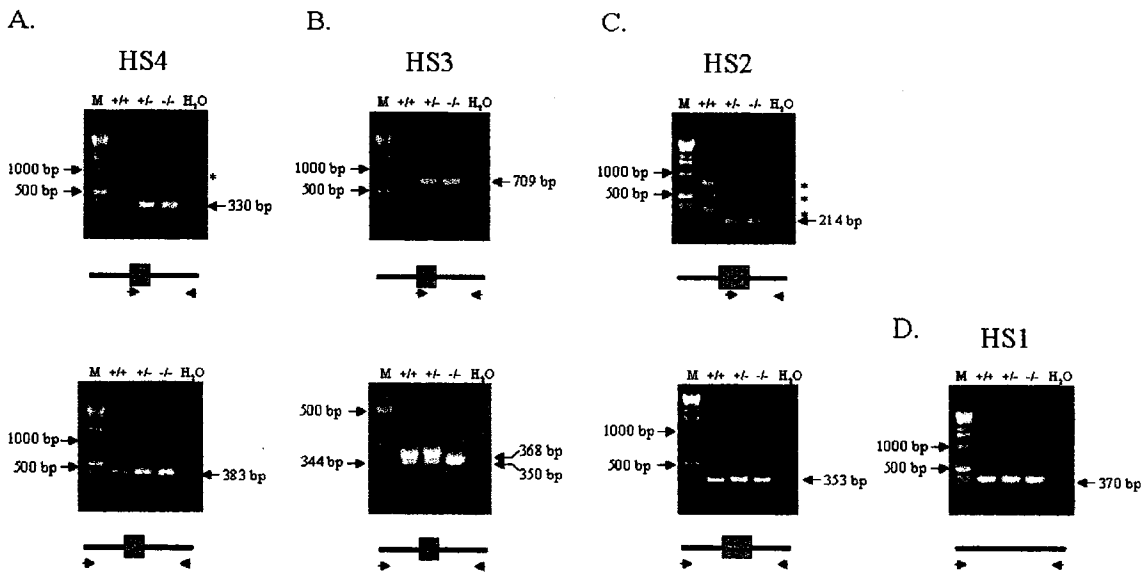


Figure 5. Sequence of HS4, HS3 and HS2 in homozygous NF-E2 site mutant mice. The top strand is sequence from homozygous NF-E2 -/- mice and the bottom strand is the reference BALB/c mouse sequence. Sequences in red are the mutations. The vertical line represents the junction of homology regions in the targeting vector and endogenous sequences. The sequence in italics identifies a portion of homology regions. (A) HS4 sequence (B) HS3 sequence (C) HS2 sequence (D) Junction of 5' homology and endogenous mouse chromosome 7 sequences (E) Junction of 3' homology and the endogenous mouse chromosome 7 sequences.

A. HS4

mt GCAGCAATGTCAGAATGAGATGGCTTTCCCCACCCCTTTACTAGTCTCCCGGG
wt GCAGCAATGTCAGAATGAGATGGCTTTCCCCACCCCTTTACTAGTCTGAGATC

mt CTCCCACCCATGGCAAGGGGAGTGAGATCCTGCCAAGACTCTGATAATA
wt CTCCCACCCATGGCAAGGGGAGTGAGATCCTGCCAAGACTCTGATAATA

B. HS3

mt AGGAGGTCTGTAAGTGTAAATTTTGGAGCACAGGTTTCTAAGACTTTCCCGGG
wt AGGAGGTCTGTAAGTGTAAATTTTGGAGCACAGGTTTCTAAGACTTTGACTCA

mt GCAAACCCTAGGCCTCCTAGGGACTGAGAGAGGCTGCTTTGGAAGATGT
wt GCAAACCCTAGGCCTCCTAGGGACTGAGAGAGGCTGCTTTGGAAGATGT

C. HS2

mt TGTGGGTGTGTTCAGCCTTGTGAGCCAGCATCAGGCTTGAGCACAGCAGTGC
wt TGTGGGTGTGTTCAGCCTTGTGAGCCAGCATCAGGCTTGAGCACAGCAGTGC

mt CCCGGGATGCCCCGGGATGCTGAGGCTTAGGGTGTGTGGCCAGATGTTTTCA
wt TGAGTCATGCTGAGTCATGCTGAGGCTTAGGGTGTGTGGCCAGATGTTTTCA

mt GCTGTGAGTGATCA
wt GCTGTGAGTGATCA

D. 5' boundary

mt CCTTCTCTTTTCCCCAAAATGTCTTCTGAGAATATGTCA|*GTACTTCCCTCTTTA*
wt CCTTCTCTTTTCCCCAAAATGTCTTCTGAGAATATGTCA|*GTACTTCCCTCTTTA*

mt *CAATACTTAAAGCTAAATCAGTAAAATATCACCTTGGATTTTGCAAATGCC*
wt *CAATACTTAAAGCTAAATCAGTAAAATATCACCTTGGATTTTGCAAATGCC*

E. 3' boundary

mt *GAGGTCTCATCGGCTTTATTTACCCCACCCATATTAGCAAATGTGTATCACAGCAA*|G
wt *GAGGTCTCATCGGCTTTATTTACCCCACCCATATTAGCAAATGTGTATCACAGCAA*|G

mt CTTCATGGTCTTTTGGCTCTTACAATCCTTTTCCCTTCTTCTG
wt CTTCATGGTCTTTTGGCTCTTACAATCCTTTTCCCTTCTTCTG

Figure 6. Allele specific RT-PCR of adult mouse blood. Blood samples were obtained from 5 WT and 6 LCR NF-E2 +/- mice. RT-PCR of blood RNA was performed with primers common to all four adult β-globin alleles (β^s, β^t, β^{maj}, β^{min}). (A) Diagram of expected bands from β-globin alleles. (B) RT-PCR of blood RNA. Undigested cDNA is 182 bp and cDNA digested with Hae III results in bands of 126 bp and 85 bp for HbbD and HbbS, respectively. (C) PCR of tail DNA. Undigested DNA is 298 bp and DNA digested with Hae III results in bands of 126 bp and 85 bp for HbbD and HbbS, respectively. (D) Histogram indicating the relative amounts of HbbD compared to HbbS. HbbD/HbbS for WT samples was set to 1. All values for mutant animals were normalized to WT. The NF-E2 mutations result in a 23% decrease in β^{maj}/β^{min}-globin gene expression. Raw values for WT DNA ranged from 1.06 to 1.09 with a mean of 1.07; mutant DNA ratios ranged from 1.07 to 1.10 with a mean of 1.09. The ratio of HbbD/HbbS for WT cDNA ranged from 0.92 to 1.01 with a mean of 0.97 and the ratio of HbbD/HbbS for NF-E2 mutant cDNA ranged from 0.72 to 0.82 with a mean of 0.75.

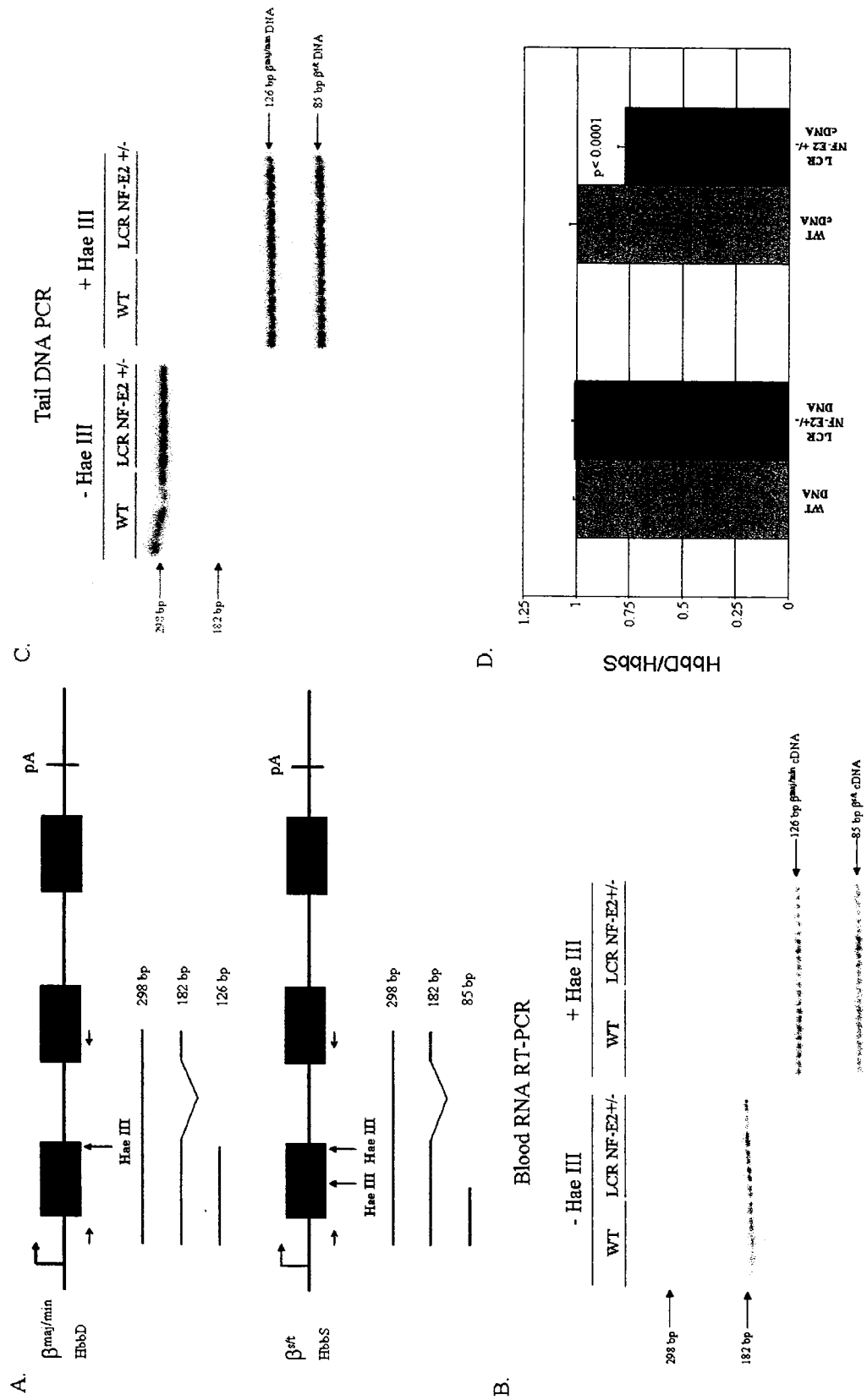


Figure 7. Allele specific RT-PCR of 10.5 dpc mouse yolk sacs. Four WT and 4 LCR NF-E2 +/- embryos were analyzed. RT-PCR was performed with primers common to HbbD and HbbS εy2 alleles. (A) Diagram of expected bands from εy2-globin alleles. (B) RT-PCR of 10.5 dpc yolk sac RNA. Undigested cDNA is 418 bp and cDNA digested with Xba I results in bands of 418 bp and 209 bp for HbbD and HbbS, respectively. (C) PCR of 10.5 dpc embryo DNA. Undigested DNA is 1,208 bp and 1,255bp for HddD and HbbS εy2 alleles, respectively. DNA digested with Xba I results in bands of 1,208 bp and 298 bp for HbbD and HbbS, respectively. (C) Histogram indicating the relative amounts of HbbD compared to HbbS. HbbD/HbbS for WT samples was set to 1. All values for mutant animals were normalized to WT. The NF-E2 mutations result in a 20% decrease in $\varepsilon y2^{HbbD}$-globin gene expression. Values for WT DNA ranged from 0.60 to 0.70 with a mean of 0.66; mutant DNA ratios ranged from 0.61 to 0.77 with a mean of 0.65. The ratio of HbbD/HbbS for WT cDNA ranged from 0.76 to 0.91 with a mean of 0.84, and the ratio of HbbD/HbbS for NF-E2 mutant cDNA ranged from 0.67 to 0.68 with a mean of 0.68.

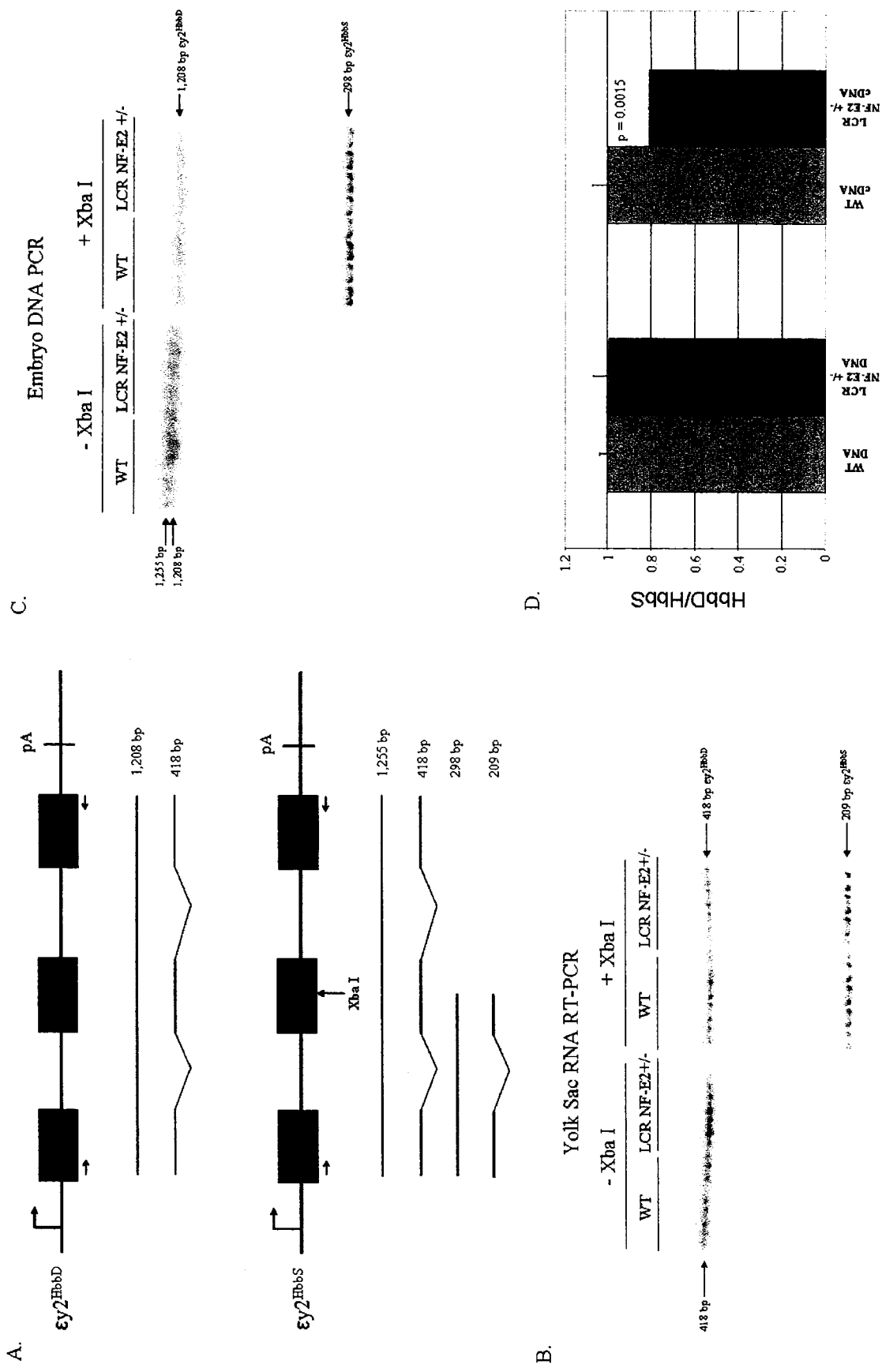
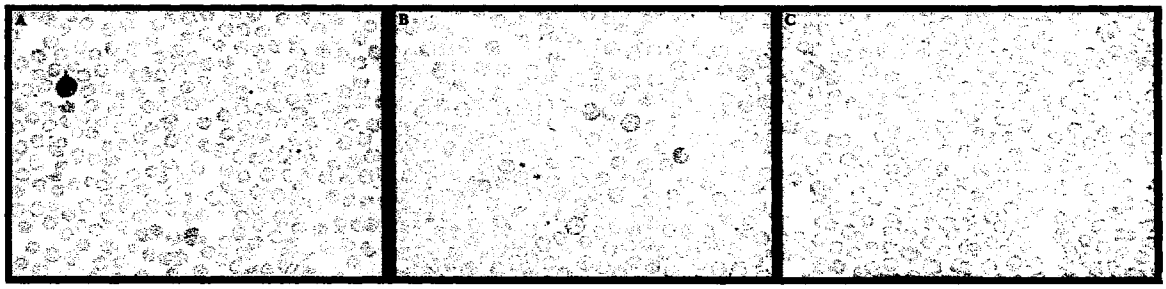
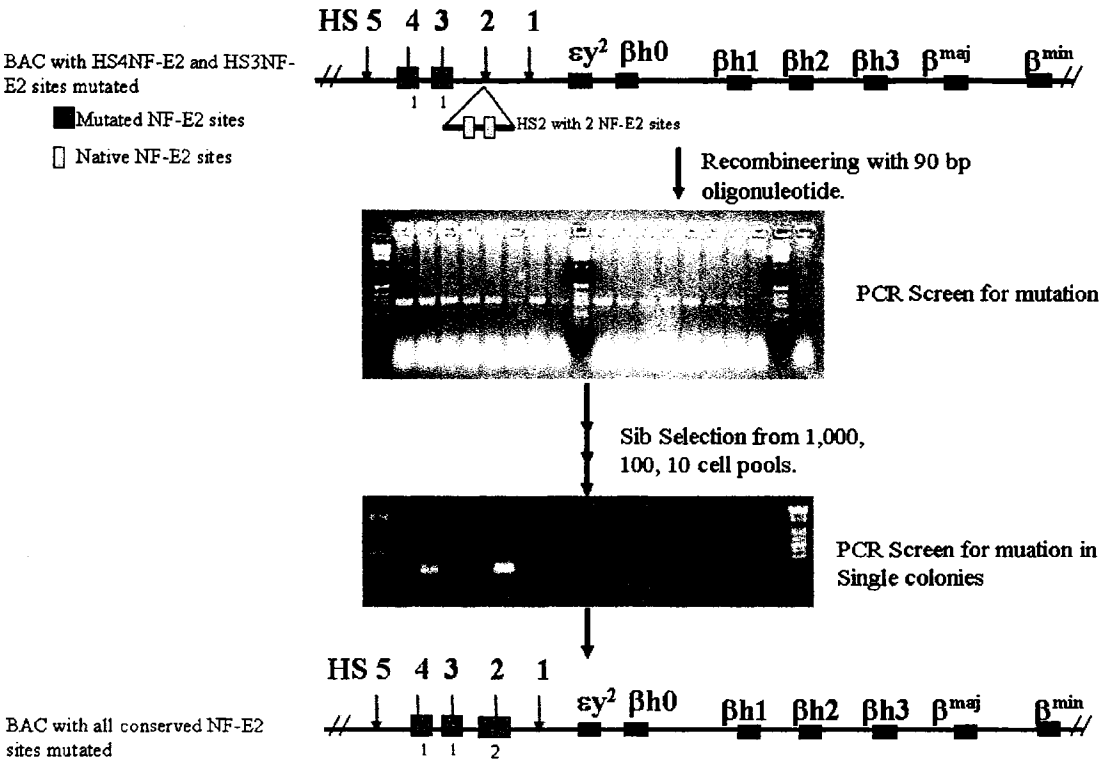


Figure 8. Blood smears of LCR NF-E2 +/+, +/- and -/- mice. No differences in the morphology of red blood cells were observed in these animals. (A) WT (B) LCR NF-E2 +/- (C) LCR NF-E2 -/-

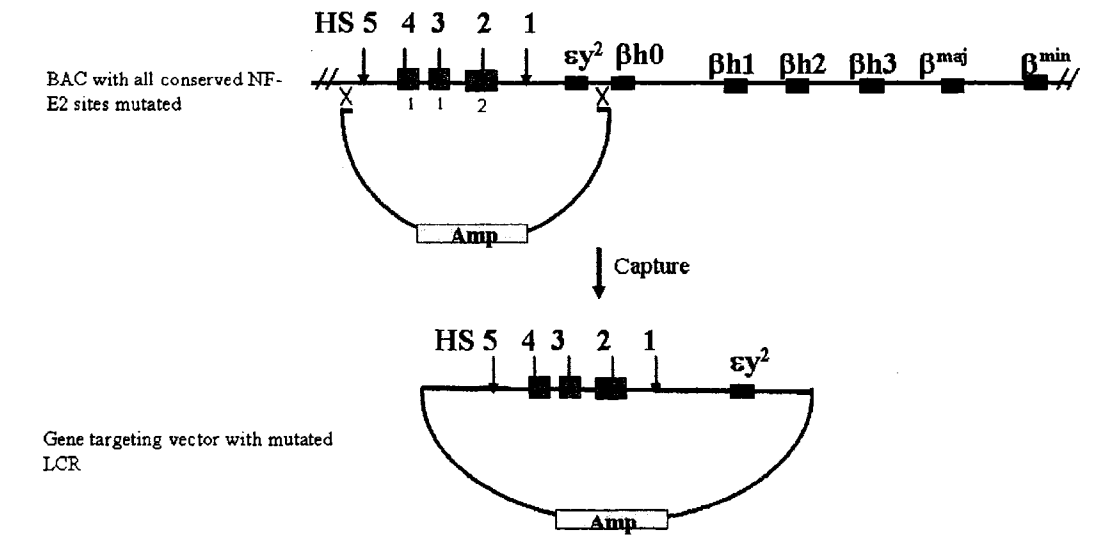


Supplemental Figure 1. BAC recombineering and capture of the mutant LCR fragment into the targeting vector. (A) A 90 bp mutagenic oligonucleotide was electroporated into DY380 cells. Serial dilutions were performed, and PCR was utilized to identify pools that were positive for the mutation. (B) After confirming NF-E2 mutations in HS4, HS3 and HS2, the LCR was captured by gap repair, which resulted in a targeting vector containing a total of 10 kb of homology (4 kb 5' and 6 kb 3') plus the entire mutated LCR.

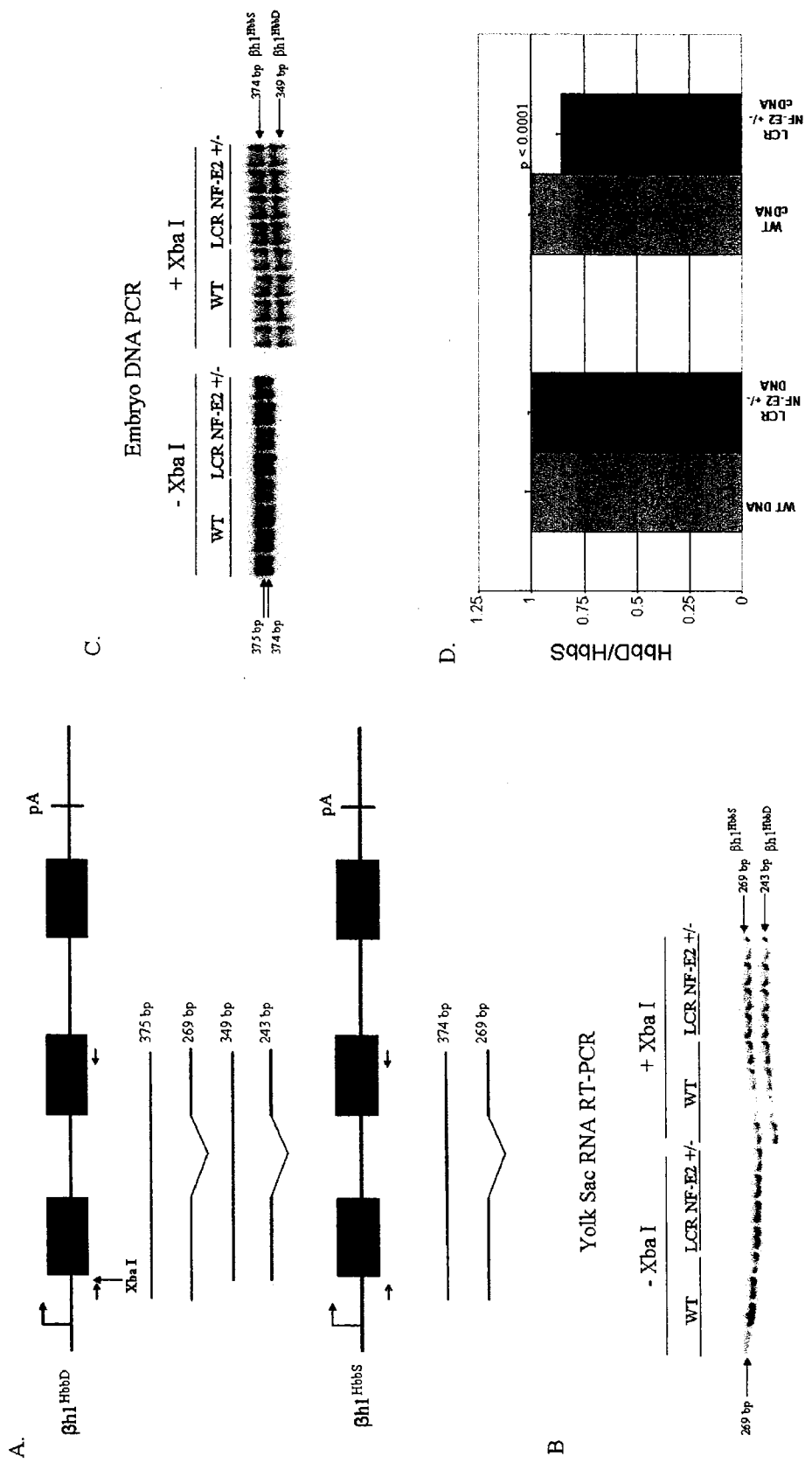
A.



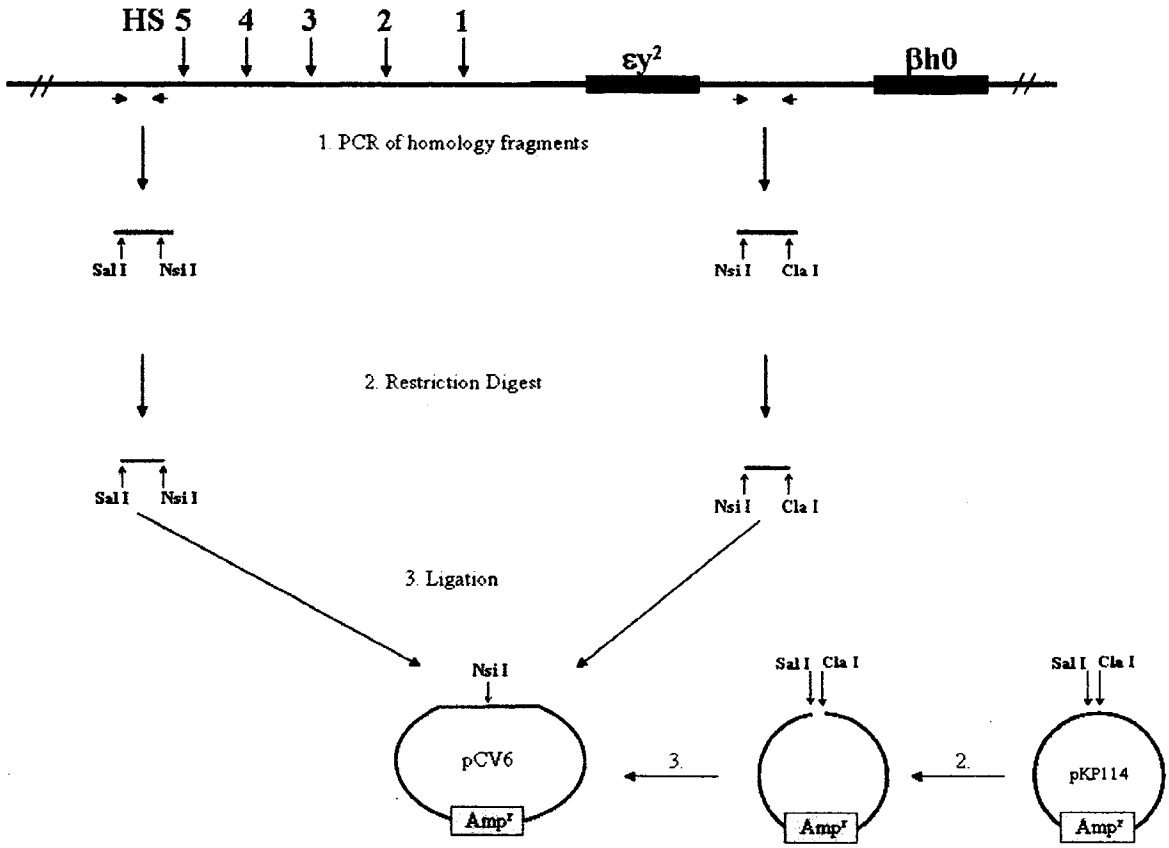
B.



Supplemental Figure 2. Allele specific RT-PCR of 10.5 dpc mouse yolk sacs. Four WT and 4 LCR NF-E2 +/- embryos were analyzed. RT-PCR was performed with primers common to HbbD and HbbS βh1 alleles. (A) Diagram of expected bands from βh1-globin alleles. (B) RT-PCR of 10.5 dpc yolk sac RNA. Undigested cDNA is 269 bp and cDNA digested with Xba I results in bands of 243 bp and 269 bp for HbbD and HbbS, respectively. (C) PCR of 10.5 dpc embryo DNA. Undigested DNA is 375 bp and 374 bp for HddD and HbbS βh1 alleles, respectively. DNA digested with Xba I results in bands of 349 bp and 374 bp for HbbD and HbbS, respectively. (C) Histogram indicating the relative amounts of HbbD compared to HbbS. HbbD/HbbS for WT samples was set to 1. All values for mutant animals were normalized to WT. The NF-E2 mutations result in a 15% decrease in $\beta h1^{HbbD}$-globin gene expression. Values for WT DNA ranged from 1.03 to 1.10 with a mean of 1.09; mutant DNA ratios ranged from 1.02 to 1.08 with a mean of 1.09. The ratio of HbbD/HbbS for WT cDNA ranged from 1.12 to 1.14 with a mean of 1.13, and the ratio of HbbD/HbbS for NF-E2 mutant cDNA ranged from 0.94 to 1.00 with a mean of 0.97.



Supplemental Figure 3. Generation of pCV6. A capture vector was designed to retrieve a 27.9 kb fragment of DNA containing LCR 5' HS5 through HS1 and the εy2 gene from the mouse β-globin BAC. The 5' and 3' capture homologies were produced by PCR and subcloned into plasmid pKP114 to construct the capture vector pCV6. In brief, the 395 bp 5' homology fragment was amplified from the mouse β-globin BAC using primers CV6-5' Fwd and CV6-5' Rev and this fragment was subsequently digested with Sal I and Nsi I; the 430 bp 3' homology fragment was amplified using primers CV6-3' Fwd and CV6-3' Rev and this fragment was then digested with Nsi I and Cla I. These two homologies were then ligated into the Sal I- and Cla I-digested cloning vector pKP114.



HIGHLY CONSERVED CACBP SITES IN HS3 AND HS2 ARE DISPENSABLE FOR HIGH-LEVEL β–GLOBIN LOCUS CONTROL REGION ACTIVITY

By

ANDREA A. SEVENDSEN, KEVIN M. PAWIK, JINXIANG REN AND TIM M. TOWNES

In preparation for Nucleic Acid Research

Abstract

Conservation of CAC Binding Protein (CACBP) sites in the β-globin Locus Control Region (LCR) between five mammalian species suggest that these sequences play an important functional role in globin gene expression. To define the role of these sequences in vivo, we replaced an 18 kb region of the wild-type LCR (HS1-4) in murine embryonic stem (ES) cells with an 18 kb LCR fragment (HS1-4) containing mutations of the conserved CACBP sites in HS3 and HS2. Mice were derived from these LCR knockin ES cells, and adult β-globin gene expression was analyzed. β-globin gene expression is reduced by only 10% in these animals and mice that are homozygous for the mutations are not anemic. These results suggest that the highly conserved CACBP binding sites in the β-globin LCR are dispensable in long-range enhancement of β-globin gene expression.

Introduction

Originally identified as a set of DNaseI Hypersensitive sites (HS) that were mapped from 6 to 22 kb upstream of the human ε-globin gene (1,2), the β-globin LCR was subsequently demonstrated to function as an enhancer in transgenic mice (3). The LCR was further demonstrated to function at all developmental stages of erythropoiesis (4,5) and specifically in erythroid cells (6-8). DNA fragments containing individual LCR HS sites have been demonstrated to function as enhancers of β-globin gene expression in cultured cells and in transgenic mice (7,9,10).

Hardison et al. (11) aligned the HS regions of the human, galago, rabbit, goat, and mouse LCRs and identifed regions of sequence conservation. Within each HS, highly

conserved regions corresponding to known and putative transcription factor binding sites, such as CACCC, GATA1, and NF-E2 were found. Recently, we demonstrated that the conserved NF-E2 sites have a significant but nonessential role in LCR enhancement of β-globin gene expression (manuscript submitted).

A single CACBP site is highly conserved in HS2 (11). When this site was deleted in an HS2/ β-globin gene construct and tested in transgenic mice, a modest reduction in transgene expression was observed compared to wild-type HS2/ β-globin transgene expression (12,13). HS3 also contains blocks of conservation between the five species, including a single CACBP site(11).

To understand the role of these CACBP sites in endogenous LCR activity, both were mutated in the endogenous mouse LCR. We demonstrate that these two CACBP sites, conserved between five species, have a nonessential role in the expression of adult β-globin genes.

Materials and methods

Recombineering of the β-globin BAC

A 180 kb bacterial artificial chromosome (BAC) clone (BPRC, CHORI) containing the murine HbbD β-globin locus was maintained in the DY380 strain of *E. coli*. Mutations were made following standard recombineering procedures (14-16). Briefly, BAC-containing DY380 cells were grown at 30 °C in 30 mL of Luria-Bertani (LB) broth to an optical density of 0.5 at an absorbance of 600 nm. The temperature was then increased to 42 °C for 15 min to induce expression of the bacteriophage λ-Red genes *exo*, *bet* and *gam*. After heat induction, cells were placed on ice for 25 minutes, washed

three times in 10% glycerol and resuspended in a final volume of 200 μL of 10% glycerol. Fifty to two hundred nanograms of mutagenic oligonucleotides were electroporated into 50 μL of the cells at 1.8 kV, 200 Ω, and 25 μF (BioRad GenePulser II). After 1 hr of recovery at 30 °C, the cells were seeded into a 96 well BioBlock (Labcor) at 10,000 cells/well and 1,000 cells/well and incubated overnight at 30 °C. One microliter aliquots were assayed by PCR to identify pools containing mutated BAC clones. The positive pools were then diluted, cultured, and assayed. This process was repeated until 10 cell/well pools containing mutated BACs were obtained. Dilutions of these pools were then spread onto LB agar plates containing tetracycline (5 μg/mL) and chloramphenicol (12.5 μg/mL) and incubated at 30 °C overnight. The following oligos were used to make the mutations in HS3 and HS2, respectively:

HS3:TGCTGGCTCAGATAGATGACCATGTGTGGGTTTCAGCTCACCCGGGAGT

CTGTGACTCTGAGAAAAGATAGACACAAACATGAGAG

HS2:CACAGCAGTGCTGAGTCATGCTGAGTCATGCTGAGGCTTACCCGGGGTG

GCCAGATGTTTTCAGCTGTGAGTGATCAGTGCTATCT

The underlined portion of each oligo is the mutation of the CACBP site. All oligos were purchased from Integrated DNA Technologies (IDT) and purified by polyacrylamide gel electrophoresis (PAGE). All PCR primers used in this study were also purchased from IDT. For the identification of the mutations in the BAC, the following primers were used: HS3F3, TGTGGGTTTCAGCTCACCC; HS3R, TTAGCAGGTGTTCTTGGAC; HS2F3, AGTCATGCTGAGGCTTACCC; HS2R, GTTATGTCACACAGCAAGGC.

Generation of ES cell targeting vector

Previously, we (manuscript submitted) designed a plasmid, pCV6, to capture a 27.9 kb fragment of DNA containing LCR 5' HS5 through HS1 and the εy gene from the mouse β-globin BAC. To recombineer the ES cell targeting vector, pCV6 was linearized with Nsi I and 50 to 400 ng of linearized plasmid was electroporated into DY380 cells containing the mutated BAC. After a 1 hr recovery at 30 °C, the cells were spread onto LB agar plates with ampicillin (50 μg/mL) for overnight culture at 30 °C. Single colonies from these plates were isolated and transferred to 3 mL LB broth with ampicillin for another overnight incubation at 30 °C. One microliter of culture was assayed in a 25 uL PCR reaction using the following primers to confirm the capture of the LCR: mBG52620F, GGGAAGCACAGGAAGACTCA; M13-20, ACGTTGTAAAACGACGGCCA; mBG25712R, CCCAAACACACAGCAGAACA; and ψM13R, CACAGGAAACAGCTATGACCA. Multiple positive clones were obtained and plasmid DNA was isolated using a standard alkaline lysis procedure (17). The plasmid DNA was then transformed into XL 10-Gold cells, and the cells were spread onto LB agar plates with ampicillin. After overnight incubation at 37 °C, single colonies were isolated, expanded in liquid culture, and assayed as above for LCR sequences. Positive cells were further analyzed by PCR to confirm the presence of the CACBP mutations. To ensure that the LCR was captured correctly, a series of restriction digests was performed. These digests confirmed that the entire 27.9 kb fragment of the LCR plus homology regions was captured.

Knockin mutations in ES cells

The targeting vector containing the LCR with CACBP site mutations was linearized with Nsi I and electroporated into HS1-4 KO ES cells (KMP and TMT, unpublished). HS1-4 KO ES cells are HM-1 (18) cells in which an 18 kb fragment of the LCR has been deleted and replaced with a mouse hypoxanthine phosphoribosyl transferase (HPRT) marker gene (19). These cells were maintained on $HPRT^+$ murine embryonic fibroblasts (MEFs) in media supplemented with 1.6 μg/mL hypoxanthine, 0.175 μg/mL aminopterin, and 4.8 μg/mL thymidine (HAT) prior to electroporation to confirm the presence of a functional HPRT gene. After electroporation, the cells were grown for 4 days in nonselective media to allow time for residual HPRT protein to decay. The cells were then replated at a low density onto gelatinized plates containing $HPRT^-$ MEFs and selected for the loss of HPRT with 10 μM 6-thioguanine (6-TG). Cells were cultured under selection for 14 days and then individual colonies were isolated.

Correctly modified HS1-4 KI ES cells were identified using PCR and DNA sequencing. Primers mHPRT(961.E1)F, ATGCCGACCCGCAGTCCCAG and mHPRT(61.E7)R, GGCCTGTATCCAACACTTCG were utilized to detect the loss of the HPRT marker gene. The previously described primers HS3F3, HS3R, HS2F3 and HS2R were used to identify the mutations. Cell samples that were negative for HPRT and positive for the CACBP mutations by PCR were sequenced across the four HS sites. The following PCR primer pairs were used to generate the templates for sequencing: HS4F, ATTTGCCCTCTCTTACATCT and HS4R, GTTCCTGCCCTCTAAATCTGC; HS3F2, CCAGGACAGCCAGGACTAC and HS3R; HS2F2, CACTTCTTCATATTCTCTCTC

and HS2R; HS1F, CTCTCCCATAACCCATACCTC and HS1R, CTCACTCTCAAGCCTGTGTTAC.

Generation of the mice

C57BL/6 blastocysts were injected with HS1-4 KI ES cells and then transferred to pseudopregnant CD-1 females to generate chimeric mice. Male chimeras were bred to C57BL/6 females to obtain agouti pups whose paternal chromosomes are derived from the HS1-4 KI ES cells. The pups were genotyped using the aforementioned primers. Mice that were heterozygous for the mutations were mated to produce homozygotes. The NF-E2 site mutations were confirmed by sequencing as described above. In addition, the junctions between homology regions in the targeting vector and chromosomal DNA was sequenced to confirm correct gene targeting. The following PCR primer pairs were used to generate the templates for sequencing: mB24959F, TGGTAGAGGAGCAGGACAC; mB25957R, CTCATCAGTACCATCTCCTC; mB52700F, TACAGGCCCAGAGAGAGTC and mB53752R, TCCACATACTCTGTCTCAGC.

Allele specific RT-PCR

Blood RNA was extracted with Trizol LS (Invitrogene) according to the manufacturer's protocol. RNA was resuspended at the equivalent of $1x10^6$ cells/μL in diethylpyrocarbonate (DEPC)-treated H_2O.

Adult blood RNA samples were treated with DNase I (RNase-free; Invitrogen) in a 10 μL reaction for 30 min at 25 °C. From this reaction, 5 μL was used to generate cDNA using the SuperScript II First Strand Synthesis System for RT-PCR (Invitrogen).

Genomic DNA was obtained from adult mouse tail clips by overnight incubation at 55 °C in 500 µL of a solution containing 20 mM Tris, 20 mM EDTA, 100 mM NaCl, 1% SDS, and 500 µg/mL Proteinase K (PCR grade; Roche). DNA was extracted with phenol/chloroform (50:50), precipitated with ethanol, and resuspended in 350 µL.

PCR of adult beta globins was performed using the following primers: mBG79746F, TGCTTCTGACATAGTTGTGTTGACT and mBG80043R, AAAGCTATCAAAGTACCGCTG. Two microliters of cDNA or genomic DNA was amplified for 32 cycles in a 20 µL PCR reaction. The mBG79746F primer was then end-labeled with γ^{32}P-ATP and purified on a Sephadex G-25 column (Boehringer Mannheim). A 20 µL primer extension reaction was performed using this labeled primer and 1 µL of the PCR reaction. To distinguish between the allele-specific products, 5 µL of the primer extension reaction was digested with Hae III for 6 hr at 37 °C. Two microliters of digested or undigested sample was electrophoresed through an 8% polyacrylamide gel. The gels were dried and bands were quantitated using a Molecular Dynamics Storm 860 PhosphorImager and ImageQuant software (GE Healthcare).

Hematology

Red blood cell (RBC) counts were determined using a HEMAVET 1700 hematology analyzer (CDC Technologies, Inc.). Hemoglobin concentration was determined spectrophotometrically after conversion to cyanmethemoglobin with Drabkin's reagent (Sigma). Before determining the hemoglobin concentration, red cell membranes were pelleted at 14000 RPM for 5 min in an Eppendorf centrifuge. Pelleting the membranes inhibits artifactually high values caused by membrane-bound, denatured

hemoglobin. Packed Cell Volume (PCV) was measured using a JorVet J503 micro-hematocrit centrifuge (Jorgensen Laboratories, Inc). Reticulocyte numbers were determined by flow cytometry (FACSCalibur flow cytometer; BD Biosciences) after staining with thiazole orange (Sigma).

Results and Discussion

Recombineering and Embryonic Stem cell mutation

The conserved CACBP sites in HS3 and HS2 of the murine beta-globin LCR are illustrated in Figure 1. By recombineering a murine β-globin BAC, we mutated both of these sites. Capture, by gap repair, of the mutated LCR into pCV6 generated a 30 kb targeting vector for homologous recombination in ES cells. The targeting vector was electroporated into HS1-4KO ES cells, which contain an 18kb deletion of the LCR and insertion of a murine HPRT marker gene (Fig 2). The electroporated ES cells were cultured in media containing 6-TG to select for the loss of the HPRT marker due to correct targeting. Genomic DNA from surviving colonies was analyzed by PCR with primers specific for mutated HS3 and HS2 sequences (Fig. 3A-B). Figure 3C demonstrates that the HPRT gene was replaced with an 18kb LCR fragment containing the mutated NF-E2 sites.

Generation of Mice bearing CACBP site mutations

CACBPmtHS1-4 KI cells were injected into wild type C57BL/6 mouse blastocycsts to produce chimeric males. The chimeric males were subsequently bred to wild type C57BL/6 females. Agouti offspring were either LCR CACBP +/+,

HbbD/HbbS or the LCR CACBP +/- , HbbD/HbbS. Genotyping was performed on DNA isolated from tail clips using primers specific for each of the mutations (Fig 4). Mice that were heterozygous for the LCR CACBP site mutations were mated to produce homozygous offspring (LCR CACBP -/-, HbbD/HbbD). Each HS site and chromosomal junctions were sequenced and compared to wild-type alleles. The results clearly demonstrate the presence of HS3 and HS2 mutations (Fig 5) and correct chromosomal junctions.

Down-regulation of adult β-globin gene expression in LCR CACBP +/- mice

Samples of blood were obtained from 6 wild type and 6 LCR CACBP +/- mice and RT-PCR of blood RNA was performed with primers common to all four adult β-globin alleles (β^S, β^T, β^{maj}, β^{min}) (Fig. 6A). PCR of tail DNA from the same animals was performed with the same primers as a control (Data not shown). After primer extension with a 5' end labeled primer, RT-PCR (Fig. 6B) and PCR products were digested with HaeIII and separated on 8% polyacrylimide gels to distinguish HbbD and HbbS sequences (HbbD, single HaeIII site; HbbS, two HaeIII sites) (see Fig. 6A). Labeled bands were quantitated on a phosphorimager and $\beta^{maj,min}/\beta^{s,t}$ values were calculated. These values were normalized to $\beta^{maj,min}/\beta^{s,t}$ values for wild type PCR products and the results are illustrated in Figure 6C. Surprisingly, $\beta^{maj,min}$ gene expression in LCR CACBP +/- mice was reduced by only 10% compared to $\beta^{s,t}$. These data demonstrate that mutations of the highly conserved CACBP sites in HS3 and HS2 have a relatively mild, effect on endogenous β-globin gene expression in vivo.

Hematology of LCR NF-E2 -/- mice

To determine the effects of LCR CACBP mutations on hematological values in adult mice, red blood cell indices were obtained. The data are listed in Table 1. Red blood cell counts in LCR CACBP -/- mice were not significantly reduced compared to WT mice, and reticulocyte counts in LCR CACBP -/- and WT mice were not significantly different. Hemoglobin concentrations of LCR CACBP-/- mice were in the normal range. These results suggest that the 10% down-regulation of globin mRNA does not result in significant anemia.

Within the β-globin LCR HSs of human, mouse, galago, rabbit and goat, two CACBP sites are highly conserved. To understand the role of these CACBP sites in LCR activity, we mutated both sites in the endogenous mouse LCR. We demonstrate that these CACBP sites are dispensable in long-distance enhancement of beta-globin gene expression. We also show that mice homozygous for mutations of these sites are not anemic; therefore, these highly conserved sequences are not essential for hemoglobin synthesis in the normal range.

Recently Zhou et al. (20) demonstrated that Erythroid Kruppel-like Factor (EKLF), a CACBP, bound to HS4, HS3, HS2, and HS1 of the murine β-globin LCR. In addition to the two CACBP sites mutated in this report that are conserved between five species, 4 more CACBP sites are conserved between humans and mice. Whereas the position and sequence of the highly conserved sites in HS3 and HS2 are nearly the same in all five species, the position and sequence of the four additional sites are not as well conserved. Further understanding of the role of CACBP sites in long range enhancement

of globin gene expression by the β-globin LCR may be obtained by mutating these additional sites.

Acknowledgements

This work was supported by a grant to T.M.T. from the NHLBI.

References

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

2. Forrester, W.C., Thompson, C., Elder, J.T. and Groudine, M. (1986) A developmentally stable chromatin structure in the human beta-globin gene cluster. *Proc Natl Acad Sci U S A*, **83**, 1359-1363.

3. Grosveld, F., van Assendelft, G.B., Greaves, D.R. and Kollias, G. (1987) Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell*, **51**, 975-985.

4. Enver, T., Raich, N., Ebens, A.J., Papayannopoulou, T., Costantini, F. and Stamatoyannopoulos, G. (1990) Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. *Nature*, **344**, 309-313.

5. Behringer, R.R., Ryan, T.M., Palmiter, R.D., Brinster, R.L. and Townes, T.M. (1990) Human gamma- to beta-globin gene switching in transgenic mice. *Genes Dev*, **4**, 380-389.

6. Forrester, W.C., Novak, U., Gelinas, R. and Groudine, M. (1989) Molecular analysis of the human beta-globin locus activation region. *Proc Natl Acad Sci U S A*, **86**, 5439-5443.

7. Ryan, T.M., Behringer, R.R., Martin, N.C., Townes, T.M., Palmiter, R.D. and Brinster, R.L. (1989) A single erythroid-specific DNase I super-hypersensitive site activates high levels of human beta-globin gene expression in transgenic mice. *Genes Dev*, **3**, 314-323.

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

9. Fraser, P., Hurst, J., Collis, P. and Grosveld, F. (1990) DNaseI hypersensitive sites 1, 2 and 3 of the human beta-globin dominant control region direct position-independent expression. *Nucleic Acids Res*, **18**, 3503-3508.

10. Fraser, P., Pruzina, S., Antoniou, M. and Grosveld, F. (1993) Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression on the globin genes. *Genes Dev*, **7**, 106-113.

11. Hardison, R., Slightom, J.L., Gumucio, D.L., Goodman, M., Stojanovic, N. and Miller, W. (1997) Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. *Gene*, **205**, 73-94.

12. Caterina, J.J., Ryan, T.M., Pawlik, K.M., Palmiter, R.D., Brinster, R.L., Behringer, R.R. and Townes, T.M. (1991) 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**, 1626-1630.

13. Caterina, J.J., Ciavatta, D.J., Donze, D., Behringer, R.R. and Townes, T.M. (1994) Multiple elements in human beta-globin locus control region 5' HS 2 are involved in enhancer activity and position-independent, transgene expression. *Nucleic Acids Res*, **22**, 1006-1011.

14. Swaminathan, S., Ellis, H.M., Waters, L.S., Yu, D., Lee, E.C., Court, D.L. and Sharan, S.K. (2001) Rapid engineering of bacterial artificial chromosomes using oligonucleotides. *Genesis*, **29**, 14-21.

15. Yu, D., Ellis, H.M., Lee, E.C., Jenkins, N.A., Copeland, N.G. and Court, D.L. (2000) An efficient recombination system for chromosome engineering in Escherichia coli. *Proc Natl Acad Sci U S A*, **97**, 5978-5983.

16. Zhou, D., Ren, J.X., Ryan, T.M., Higgins, N.P. and Townes, T.M. (2004) Rapid tagging of endogenous mouse genes by recombineering and ES cell complementation of tetraploid blastocysts. *Nucleic Acids Res*, **32**, e128.

17. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning: a laboratory manual*. 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

18. Selfridge, J., Pow, A.M., McWhir, J., Magin, T.M. and Melton, D.W. (1992) Gene targeting using a mouse HPRT minigene/HPRT-deficient embryonic stem cell system: inactivation of the mouse ERCC-1 gene. *Somat Cell Mol Genet*, **18**, 325-336.

19. Melton, D.W., Ketchen, A.M. and Selfridge, J. (1997) Stability of HPRT marker gene expression at different gene-targeted loci: observing and overcoming a position effect. *Nucleic Acids Res*, **25**, 3937-3943.

20. Zhou, D., Pawlik, K.M., Ren, J., Sun, C.W. and Townes, T.M. (2006) Differential binding of EKLF to embryonic/fetal globin gene promoters during development. *J Biol Chem.*

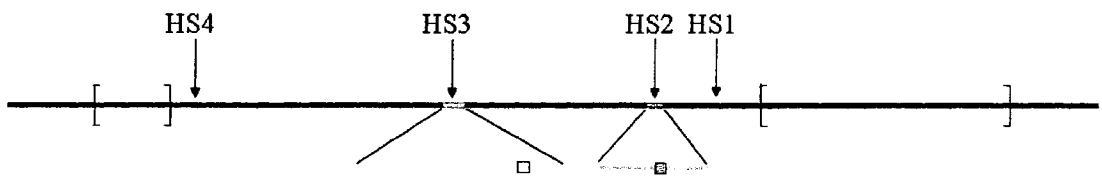
Table 1. Hematological values of WT and LCR CACBP -/- mice.

Mice	RBC M/μL	Hb g/dL	PCV %	RDW %	MCV fL	MCH pg
WT n=11	8.98±0.41	12.86±0.90	44.23±1.77	18.20±0.55	49.28±1.45	14.33±0.82
CACBP -/- n=6	8.52±0.43 (0.0489)	11.70±1.21 (0.0393)	43.33±2.36* (0.3885)	19.60±0.90 (0.0011)	50.86±1.62 (0.0583)	13.72±1.12 (0.2164)

RBC, red blood cell; Hb, hemoglobin; PCV, packed cell volume; Retic, reticulocytes; RDW, red cell distribution width; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin. Values represent the mean ± SD. P values in parentheses were calculated using a two-tailed unpaired Student's t-Test. *n= 5

Figure 1. Mouse β-globin LCR and mutagenic oligos. (A.) Diagram of HS1-4 of the mouse β-globin locus control region. The bracketed regions are 5' and 3' DNA fragments used for homologous recombination in ES cells. The core regions are highlighted in red, yellow, blue, and green for HS4, HS3, HS2, and HS1, respectively. Boxes in the enlarged regions of HS3 and HS2 indicate the relative position of conserved CACBP sites. (B) Sequence of conserved CACBP sites and surrounding regions. The conserved sites are indicated in **bold**, and the mutations are indicated below in red.

A.



B.

HS3

ATGCTGGCTCAGATAGATGACCATGTGTGGGTTTCAGCTC**AGGGTGGAG**TCTTGTGACTCTGAGAAAAGATAGACACAAACATGAG
TCACCCGGGACT

HS2

CACAGCAGTGCTGAGTCATGCTGAGTCATGCTGAGGCTT**AGGGTGTGT**GGCCAGATGTTTTCAGCTGTGAGTGATCAGTGCTATCT
TTACCCGGGGTC

Figure 2. Schematic of knockin strategy. HS1-4 KO ES cells (KMP and TMT, unpublished) are murine HM1 ES cells in which an 18 kb fragment of the LCR has been deleted and replaced with an HPRT marker gene. The targeting vector is composed of homology regions totaling 10 kb (4 kb 5' and 6 kb 3') and an 18 kb fragment of the LCR containing mutated CACBP sites in HS3, and HS2. After homologous recombination, HS1-4 KI ES cells contained one mutant HbbD allele and one wild-type HbbD allele

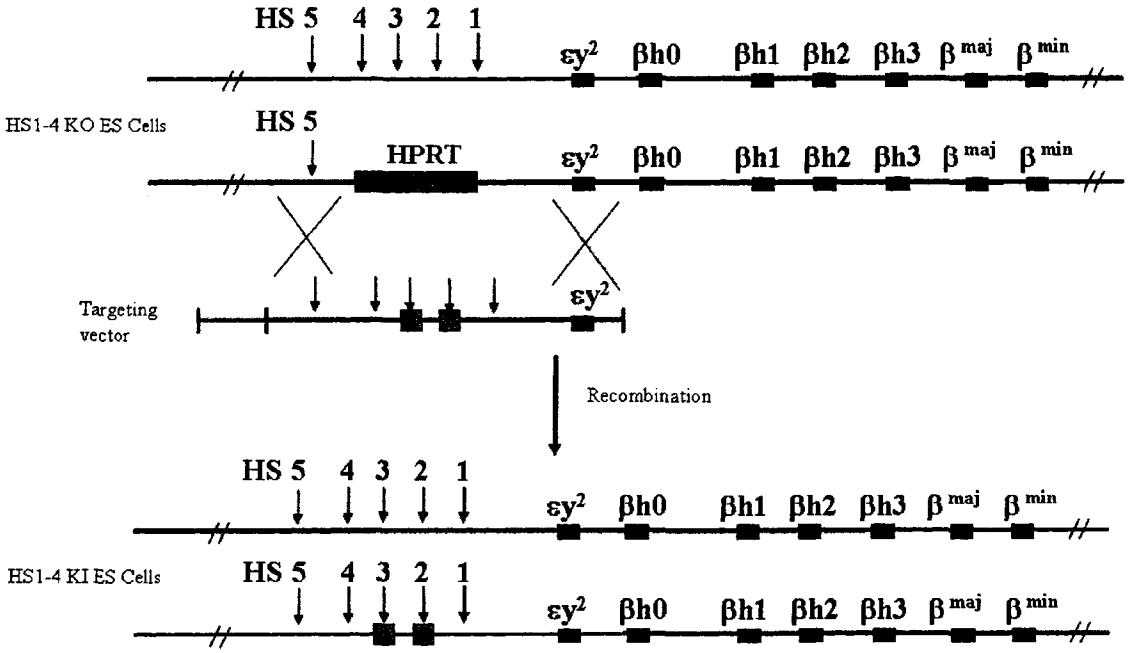


Figure 3. PCR genotyping of ES cells. The top panels of A and B demonstrate PCR products obtained with primers that are specific for the CACBP site mutations in HS1-4 KI ES cells. The lower panels are controls demonstrating intact core HS sites in HS1-4 KI cells. Panel D demonstrates the loss of HPRT in ES cells containing the knockin allele.

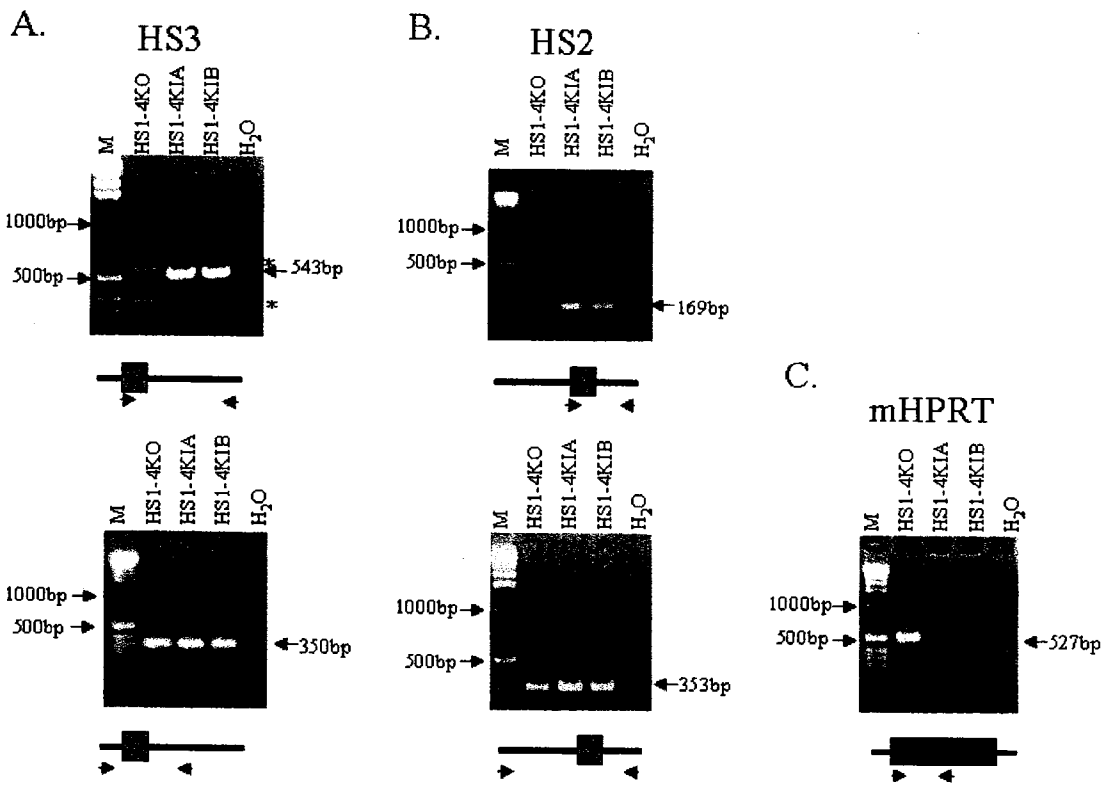


Figure 4. PCR genotyping of mice. M is a marker, +/+ is wild-type mouse tail DNA, +/- is mutant CACBP heterozygous mouse tail DNA, -/- is mutant CACBP homozygous mouse tail DNA and H_2O is a water control. The top panel of A and B demonstrates the presence of CACBP site mutations in HS3 and HS2, respectively. The bottom panels of A-D demonstrate the presence of intact HS site cores in all mouse lines. Homozygous CACBP site mutant mice were confirmed by the presence of two HS3 HbbD alleles (Panel B, bottom). -/- represents an HbbD/HbbD mouse and +/- represents an HbbS/HbbD mouse

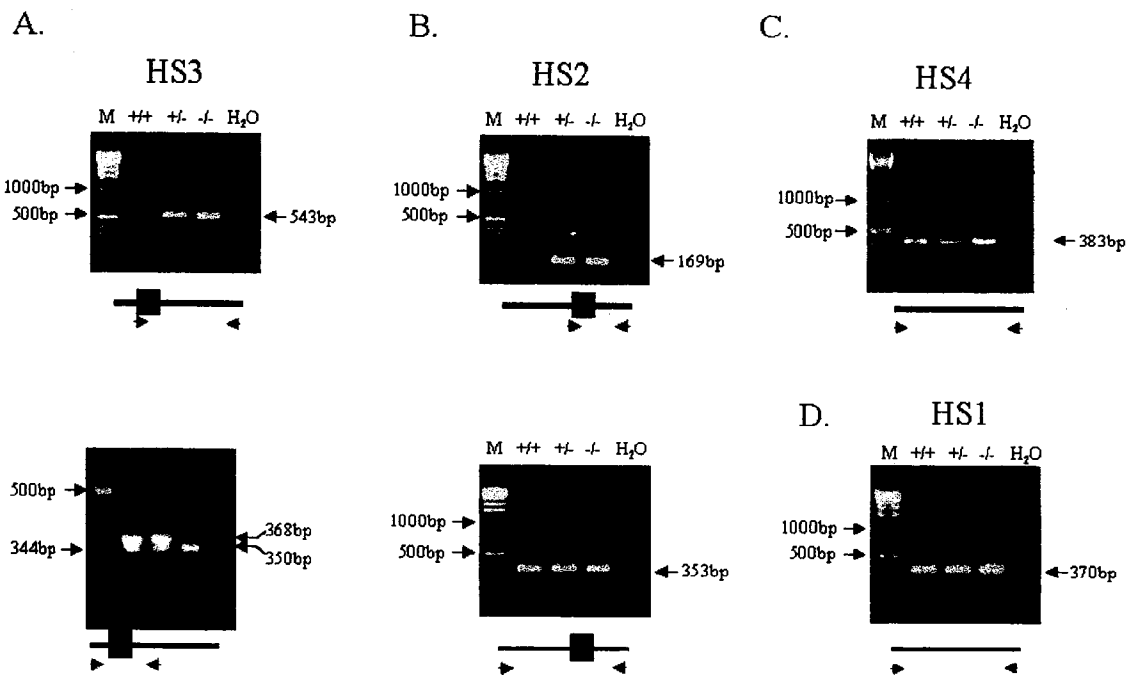


Figure 5. Sequence of HS3 and HS2 in homozygous CACBP site mutant mice. The top strand is sequence from homozygous CACBP site -/- mice and the bottom strand is the reference BalbC mouse sequence. Sequences in RED are the mutations. The vertical line represents the junction of homology regions in the targeting vector and endogenous sequences. The sequence in italics identifies a portion of homology regions. (A) HS3 sequence (B) HS2 sequence (C) Junction of 5' homology and endogenous mouse chromosome 8 sequences (D) Junction of 3' homology and the endogenous mouse chromosome 8 sequences.

A. HS3

mt AGGAGGTCTGTAAGTGTAAATTTTGGAGCACAGGTTTCTAAGACTTTCCCGGG
wt AGGAGGTCTGTAAGTGTAAATTTTGGAGCACAGGTTTCTAAGACTTTGACTCA

mt GCAAACCCTAGGCCTCCTAGGGACTGAGAGAGGCTGCTTTGGAAGATGT
wt GCAAACCCTAGGCCTCCTAGGGACTGAGAGAGGCTGCTTTGGAAGATGT

B. HS2

mt TGTGGGTGTGTTCAGCCTTGTGAGCCAGCATCAGGCTTGAGCACAGCAGTGC
wt TGTGGGTGTGTTCAGCCTTGTGAGCCAGCATCAGGCTTGAGCACAGCAGTGC

mt CCCGGGATGCCCCGGGATGCTGAGGCTTAGGGTGTGTGGCCAGATGTTTTCA
wt TGAGTCATGCTGAGTCATGCTGAGGCTTAGGGTGTGTGGCCAGATGTTTTCA

mt GCTGTGAGTGATCA
wt GCTGTGAGTGATCA

C. 5' boundary

mt CCTTCTCTTTTCCCCAAAATGTCTTCTGAGAATATGTCA|*GTACTTCCCTCTTTA*
wt CCTTCTCTTTTCCCCAAAATGTCTTCTGAGAATATGTCA|*GTACTTCCCTCTTTA*

mt *CAATACTTAAAGCTAAATCAGTAAAATATCACCTTGGATTTTGCAAATGCC*
wt *CAATACTTAAAGCTAAATCAGTAAAATATCACCTTGGATTTTGCAAATGCC*

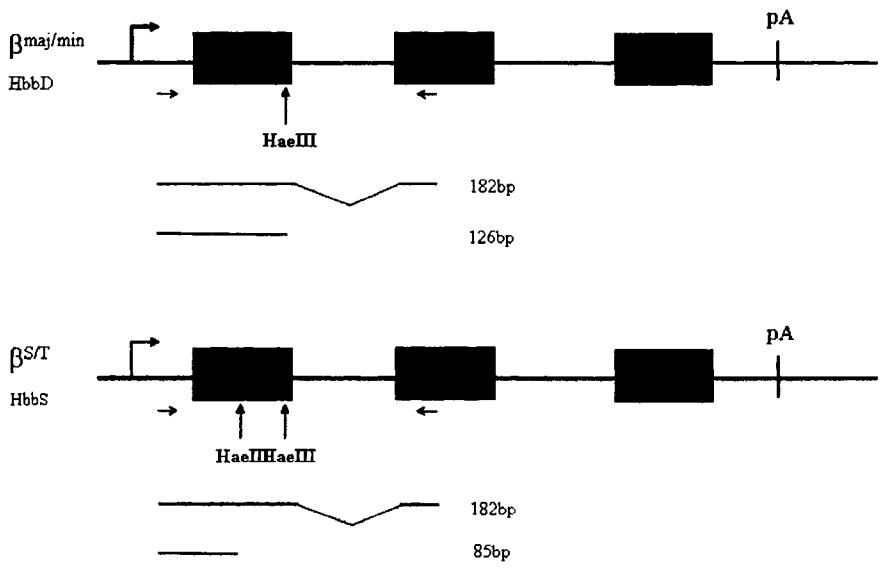
D. 3' boundary

mt *GAGGTCTCATCGGCTTTATTTACCCCACCCATATTAGCAAATGTGTATCACAGCAA*|G
wt *GAGGTCTCATCGGCTTTATTTACCCCACCCATATTAGCAAATGTGTATCACAGCAA*|G

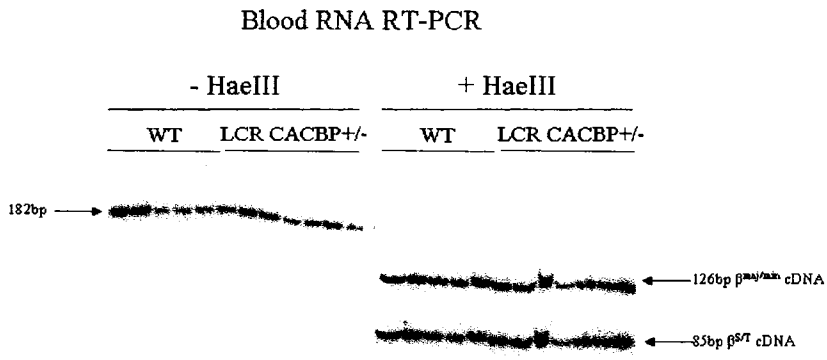
mt CTTCATGGTCTTTTGGCTCTTACAATCCTTTTCCCTTCTTCTG
wt CTTCATGGTCTTTTGGCTCTTACAATCCTTTTCCCTTCTTCTG

Figure 6. Allele specific RT-PCT of adult mouse blood. Blood samples were obtained from 5 wild type and 6 LCR CACBP +/- mice. RT-PCR of blood RNA was performed with primers common to all four adult β-globin alleles (β^S, β^T, β^{maj}, β^{min}) (A) Diagram of expected bands from β-globin alleles (B) RT-PCR of blood RNA. Undigested cDNA is 182bp and cDNA digested with HaeIII results in bands of 126bp and 85bp for HbbD and HbbS, respectively. (C) Histogram indicating the relative amounts of HbbD compared to HbbS. HbbD/HbbS for wild-type samples were set to 1. All values for mutant animals were normalized to wild-type. The CACBP mutations result in a 10% decrease in β^{maj}/β^{min} -globin gene expression. The ratio of HbbD/ HbbS for wild type cDNA ranged from 0.946 to 1.006 with a mean of 0.984 and the ratio of HbbD/HbbS for CACBP mutant cDNA ranged from 0.836 to 0.929 with a mean of 0.878.

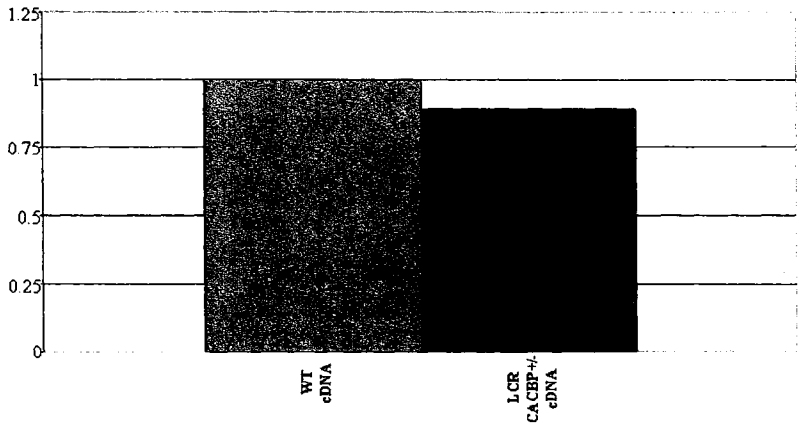
A.



B.



C.



CONCLUSIONS

Hemoglobin is a tetrameric protein composed of two β-globin polypeptides, two α-globin polypeptides and four heme groups. It is one of the most abundant proteins in blood. The function and regulation of hemoglobin are very well studied. In the previous chapters I presented the work conducted to further characterize the enhancement of β-globin gene expression by the LCR. The β-globin gene locus of mice is composed of four functional genes and three pseudogenes. The genes are located on chromosome 7 in the following order 5'-εy2–βh0–βh1–βh2–βh3–β1–β2-3'. These genes are expressed in a regulated developmental pattern with βh1 and εy2 being expressed during early embryonic development and β1 and β2 expression during late fetal development and adult life. The developmental switch in polypeptides that compose hemoglobin allows hemoglobin to act as a sensitive oxygen delivery system in embryonic, fetal, and adult life.

Upstream of the εy2 gene there are a series of DNAseI hypersensitive sites (HS). Multiple studies have also demonstrated the necessity of these HS sites for high levels of β-globin gene expression, and so have they have collectively been named a locus control region (LCR). When a LCR sequence is linked to a transgenic gene it confers tissue-specific and physiological levels of expression of the linked genes in a position-independent and copy number dependent manner (34-37). In addition to the β-globin

gene locus, the tyrosinase gene locus (30,31), the Th2 cytokine locus (32), and the T-cell receptor α/δ locus (33) are just a few of the over 38 known loci that contain LCRs.

The β-globin locus is highly conserved throughout the mammalian order. Alignments of the HSs of human, galago, mouse, rabbit, and goat LCRs reveals regions of high sequence conservation that correspond with transcription factor binding sites. To further understand the role of the LCR and these conserved sites in β-globin gene activation, I set out to create mutations of these sites and replaced the endogenous mouse LCR with a mutant LCR.

A Brief Discussion on the Methods Used

Hardison et al. aligned the HS sites of human, galago, rabbit, goat, and mouse LCRs to identify regions of conservation.(68) Within each HS, blocks with very high conservation were observed. These highly conserved regions corresponded with known putative transcription factor binding sites, such as CACCC, GATA1, and NF-E2-like binding sites. I set out to understand the role of the conserved NF-E2 and CACBP sites in endogenous LCR activity.

Mutation of the conserved sites were made in a bacterial artificial chromosome (BAC) which contains mouse sequence upstream of HS5 of the β-globin locus and extends well downstream of the β2 gene. This BAC was electroporated into the *E. coli* strain DY380 developed by Lee *et. al.* (84). This strain has an inducible recombination system that allowed the mutation of the sites by introduction of an 80-100 bp oligonuleotide. These mutagenic oligonuleotieds are composed of a 40 bp 5' homology, the desired mutation, and a 40bp 3' homology region. Using PCR screens and sib

selection, single colonies were identified as positive for the mutation of the conserved sites. After multiple rounds of recombineering a BAC with all the conserved NF-E2 or CACBP sites mutated was obtained.

Gap-repair was utilized to capture a targeting vector for electroporation in to ES cells. This targeting vector was composed of a 5' homology region, the mutated LCR, and a 3' homology region. This vector was linerized and electroporated into HS1-4 KO ES cell which had an HPRT marker gene in place of the LCR. The ES cells positive for homologous recombination with the mutant LCR were identified by selection with 6-TG. The ES cells with the mutations of the conserved NF-E2 or CACBP sites were then injected into C57BL/6 blastocysts and then transferred to pseudopregnant CD-1 females to generate chimeric mice. The chimeric males generated where bred to C57BL/6 females to generate the NF-E2 or CACBP mutant heterozygous mice. These mice were then analyzed to determine the role of highly conserved NF-E2 and CACBP sites in LCR activation of the downstream β-globin genes.

Traditional subcloning would have been extremely complex and very cumbersome if utilized to make the desired mutation. One reason is the mutations of the conserved sites were 6 bp in size and spanned 18kb of DNA sequence. Traditional subcloning would have involved a detailed search for unique restriction sites at each site, a task that is very difficult when covering such a large area. Using recombination to generate the mutations was simple in comparison as oligonucleotides can be designed quickly with the desired changes and short homology regions and so changes can be made in any sequence. In addition, small changes such as I made are easily detected with PCR, and serial screens allowed for the rapid identification of single colonies

positive for each mutation. Also by generating the mutations in a BAC I was able to avoid the recombination that can occur when larger pieces of DNA are inserted and maintained in traditional vectors. In fact after the capture of the 28 kb necessary of targeting in ES cells into pCV6 it was necessary to grow the cultures under conditions which prevented the cells from entering the stationary phase where this recombination is most likely to occur.

The ES cell targeting stratagem I used allowed me to insert an 18 kb region of the LCR back into its endogenous space with no extraneous sequences remaining. The deletion of the WT 18 kb region of interest before I introduced my mutated LCR was necessary, other wise the probability of recombination between the HS was great and it would have been very difficult to insert all the mutation into the ES cells. Fortuitously, Dr. Pawlik of our lab had previously replaced the region of interest with an HPRT marker gene. This marker gene has positive and negative selections available, and so no other marker was necessary for the replacement of it with the mutated LCRs. Though an extra marker may have increased the selection's power resulting in fewer colonies, it would have been essential to remove that marker as it has been shown that the presence of a marker gene in the LCR can influence the expression of the downstream globin genes (45). This removal of the marker would require a second step and the cloning efficiency of the cells may have been impaired. Instead, the ES cell targeting I used allowed single step generation of cells with the mutated LCR desired to create chimeric mice.

NF-E2 Site Mutations

Chapter 2 describes the mutation of the conserved NF-E2 sites in HS4, HS3, and HS2 and the effect these mutations have on downstream β-globin gene expression. The mutation of these sites demonstrated that they have a significant but non-essential role in β-globin gene activation. The four conserved NF-E2 sites were mutated to *SmaI* site by recombinerring of a BAC that contained the mouse β-globin locus. The mutated LCR was captured into a targeting vector and linearized vector was electroporated into ES cells which contained an 18 kb section of the LCR replaced with an HPRT marker gene. ES cells positive for the mutation were identified by PCR screens. These ES cells were injected into C57Bl/6 blastocysts and then implanted into pseudopregnant CD1 females. Chimeras generated by this method were bred to produce agouti offspring, of which 50% should be LCR NF-E2 +/+ and 50% should be LCR NF-E2 -/+. All agouti offspring had one hbbD and one hbbS β-globin locus allele. Allele-specific reverse transcription-polymerase chain reaction (RT-PCR) of adult β-globin genes revealed a 23% decrease in expression from the mutant locus compared to the wild-type locus. Analysis of yolk sac gene expression was consistent with the adult, a 20% decrease and a 16% decrease in εy2-globin and βh1-globin gene expression, respectively. These data clearly demonstrates that the conserved NF-E2 sites have a significant role in LCR activation of the downstream β-globin genes. In addition the similar results from adult and fetal tissues suggest that the sites are not essential from proper developmental switching of the globin genes. This further suggests that the protein(s) bound to the conserved NF-E2 sites are constitutive components of the LCR holocomplex. Mice homozygous for the mutation of

the NF-E2 sites are non-anemic, and so these sites are nonessential for high levels of β-globin.

CACBP Site Mutations

Interestingly, mutation of the conserved CACBP sites in HS3 and HS2 revealed that these sites are dispensable for high-level globin gene expression. Chapter 3 covered the mutation and analysis of these sites in the endogenous setting. The two conserved CACBP sites were mutated to *SmaI* site by recombinerring of a BAC that contained the mouse β-globin locus. The mutated LCR was captured into a targeting vector and linearized vector was electroporated into ES cells which contained an 18 kb section of the LCR replaced with an HPRT marker gene. ES cells positive for the mutation were identified by PCR screens. These ES cells were injected into C57Bl/6 blastocysts and then implanted into pseudopregnant CD1 females. Chimeras generated by this method were bred to produce agouti offspring, of which 50% should be LCR CACBP +/+ and 50% should be LCR CACBP -/+. All agouti offspring had one hbbD and one hbbS β-globin locus allele. Allele-specific reverse transcription-polymerase chain reaction (RT-PCR) of adult β-globin genes revealed a 10% decrease in expression from the mutant locus compared to the wild-type locus. Also mice homozygous for these mutations are non-anemic.

Additive Effects in the LCR

In combination the results from chapter 2 and 3 suggest an interesting model of LCR holocomplex activity. The activity of the LCR may not result from just the additive

nature of the HSs (43) but also additively for the transcription factor sites. This additive nature of the conserved sites is suggested by the fact that mutation of 4 conserved sites has a greater effect then mutation of 2 conserved sites. One example of an additive effect in enchancer activity is the binding of brachyury proteins to the enhancer of the homeobox gene *otp*. In *Drosophila* it has been shown that multiple conserved brachyury protein binding site act additively to enhance the expression of the *otp* gene (88). Christensen TH, and Kedes L (89) characterized the specific DNA regulatory elements of the human cardiac troponin C gene (cTnC) muscle-specific enhancer in myogenic cells. They concluded from their observations that the cTnC enhancer functions through the combined interactions of at least five transcription factor binding sites (89). Their data suggests that all of these sites contribute to enhancer function in an additive way but that none are absolutely required for enhancer activity (89). In addition, it is also implied that the expression level of cTnC in myogenic tissues lacking one of the factors would not be wholly suppressed (89). The cTnC gene enhancer has many parallels to the β-globin gene LCR. Both are necessary for high levels of expression of a linked gene and are composed of multiple classes of transcription factor binding sites. These parallels support the idea that the transcription factor binding sites of the LCR act additively with no one factor being essential for LCR activity.

To confirm the additive nature of the transcription factor binding sites a series of mutation throughout the HSs needs to be preformed. The mutations should vary from mutation of 2 NF-E2, or CACBP, or GATA sites to mutation of all the conserved sites in the HSs of the β-globin LCR (Figure 1). The series suggested would determine if the

effect is sites specific or if the activation of globin gene expression is solely dependent on the number of binding sites regardless of identity.

Redundancy in the LCR

Recently Zhou et al. (90)demonstrated that Erythroid Kruppel-like Factor (EKLF), a CACBP, bound to HS4, HS3, HS2, and HS1 of the murine β-globin LCR. Interestingly, the alignments of the HSs only revealed two highly conserved CACBP sites, one in HS3 and one in HS2. Examination of the mouse sequences revealed 4 more CACBP sites are partially conserved between humans and mice. Whereas the position and sequence of the highly conserved sites in HS3 and HS2 are nearly the same in all five species, the position and sequence of the four additional sites are not well conserved. These other CACBP sites, in the light of the apparently additive nature of the transcription factor binding site, suggest that the HSs have some redundancy build in. The system extra sites which are not highly conserved helps to ensure high levels of β-globin gene expression in the event of a mutation that disrupts binding at the presumably preferred conserved sites in HS3 and HS2. Further understanding of the role of CACBP sites in long range enhancement of globin gene expression by the β-globin LCR may be obtained by mutating these additional sites.

Holocomplex Formation

I hypothesized that mutations of highly conserved sites would mimic the deletion of the LCR in mice. In fact the mutation of these sites resulted in a relatively mild decrease in globin gene expression. This suggest the possibility that the protein/DNA

complex that forms on the LCR leading to gene activation does not form in a completely DNA dependent manner. If the complex formation is completely DNA dependent the proteins that bind to these conserved sites would first need to bind to the DNA and then they could interact to form the LCR holocomplex (Figure 2a). If this was the method of the complex formation then the deletion of the conserved sites would be expected to have a larger effect as one could assume that the absence of a set of proteins would prevent the formation of the complex. I show that with the NF-E2 and CACBP site mutations that the absence of theses site does not prevent high level globin gene expression and so appears to not significantly impair the formation of the LCR holocomplex. This suggest that small protein complexes for each HS form in the absence of DNA in the nucleus and then these complexes bind to the DNA through the interaction with multiple conserved sites in each HSs (figure 2b). After each HS complex has bound DNA they then interact further with each other to form the final LCR holocomplex leading to activation of the globin genes. In this model disruption of a few binding sites would weaken the HS complex's association constant and so would lead to slower binding of the complex to each HS, a less stable binding, and finally a higher rate of disassociation. This effect would be modulated by the number of sites present for complex binding. In the wild-type these multiple DNA interaction surfaces created by the small HS complexes would create a very stable DNA/protein complex at each HS and so allow the formation of a stable LCR holocomplex. This stability is very desirable for high level of gene expression.

Future Studies of holocomplex formation

I propose performing chromatin immunoprecipitation (ChIP) to determine which model of holocomplex formation is likely. EKLF is a CAC binding protein that is

essential for adult β-globin gene expression (81,82). In addition, using a HA tagged form of EKLF, it was recently demonstrated to bind HS 1-4 during all developmental stages in erythroid cells of mice (90). The Townes' lab also demonstrated that EKLF is bound to the promoter of the adult β-globin genes is the absences of the LCR in mice (90). This strongly suggested that binding of ELKF to the promoter and to the LCR are to different events and so EKLF should bind to the LCR in the absence of the β-globin promoters. To test if complex formation is DNA independent I must first confirm that EKLF binds to the LCR in the absence of β^1 and β^2. To confirm this I propose breeding the HA-EKLF mice to β-thalassemia mice in which β^1 and β^2 have been deleted and performing ChIP for the HSs using anti-HA antibody for the immunoprecipitation. If the assumption that EKLF binds to the LCR independent to the binding of EKLF to promoter is correct, binding of EKLF to HS1-4 should be detectable. If EKLF binds to the LCR in the absence of adult β-globin, I propose to perform a similar ChIP in mice which have both HA-EKLF and mutations of all the CACBP sites in the LCR. If the formation of the holocomplex is DNA dependent it would be expected that EKLF would be excluded from the region and no binding of EKLF to the HSs would be detectable. On the other hand if my model of DNA independent complex formation is correct I expect ChIP with anti-HA antibodies would still pull down LCR sequences. Similar ChIP experiments using antibodies to other holocomplexe factors, such as NF-E2 and GATA1, would confirm this model.

LCR Mutations are Possible β-Thalassemia Severity Modifiers

β-Thalassemia is a heterogeneous group of naturally occurring, inherited mutations. It's characterized by abnormal globin gene function resulting in the total absence or reduction of β-globin chain synthesis in human erythroid cells β-Thalassemia mainly affects people of Mediterranean, African and Asian ancestry. The severity of β-thalassemia varies greatly person to person and my data suggest a possible mechanism for the variability. The mutations in the globin gene which result in β-thalassemia are group into two classes: β^0-thalassemias, mutations that result in the total absence of β-chain synthesis; β^+-thalassemias, mutations that result in a marked decrease in β-chain synthesis.

There are three different groups of mutations that result in β^0-thalassemia. First is single base substitution that creates termination or nonsense codons in the adult β-globin gene (91). Secondly, frame shift mutation due to base deletions or insertions, such as the deletion of A in codon 6, which creates a new termination codon downstream of the deletion (92,93). Mutations in intervening sequences that prevent normal splicing are a third cause of β^0-thalassemias (94).

There are two major groups of mutations that result in β^+-thalassemia. The first group of mutations encompasses single base mutations in the globin gene that result in abnormal splicing (95). Mutations in the promoter region of the globin gene are the second cause of β^+-thalassemia. It is interesting to note that one of these mutations is found in the EKLF binding sites of the globin promoter (94,96).

Individuals which are homozygous for β-thalassemia experience a severe anemia and are often transfusion dependent from an early age. The mutations I generated in the

LCR in many ways mimic mild β^+-thalassemia. Mutation of the conserved sites resulted in a decrease in β-chain synthesis but did not result in anemic mice. This was surprising as Hispanic thalassemia is caused by deletion of a large region of DNA which spans the LCR. I suggest that single base mutation in the conserved protein binding sites may act as modifiers of β-thalassemia in humans. Heterozygous β^+-thalassemic individuals are usally asymptomatic, with a mild anemia through a wide range of hemoglobin levels (97). Although, moderately severe anemia (β-thalassemia intermedia) has been described occasionally in heterogeneous β^+-thalassemic individual (98-101). One mechanism by which the LCR may act as a modifier of thalassemia is if a mutation in the LCR, such as the mutation of the CACBP sites in HS3 and HS2, is linked to a mutation in the globin gene which impairs globin chain synthesis, a greater decrease in globin levels would occur, resulting in a more severe anemia. A second mechanism would be mutation in the LCR of a normal globin gene on one allele and a thalassemic mutation in the globin gene on the second allele. This would create an environment in which there would be less production of normal globin messenger RNA which could not be compensated for by the second allele of globin as its β-chain synthesis is impaired. This combination of mutations could potentially shift a mild β-thalassemia to an imtermidia form. In the second case if the mutation of the globin gene is of the β^0 type any impairment of the normal alleles globin expression could be catastrophic and create a severe anemia for an individual.

Figure 1. A proposed series of mutations of conserved sites in the HSs of the β-globin LCR. The top line represents the wild-type LCR with the conserved NF-E2, CACBP and GATA sites as purple boxes, green triangles, and red circles, respectively. The following lines are proposed mutations in the LCR determine if the sites act in an additive manner. Mutation of a site is represented by a black shape. Mutation of two NF-E2 or two GATA sites should mimic the effect of mutation of two CACBP sites if the number of sites is the critical factor in determining the activity of the LCR. The addition al mutations of a larger number of sites in different combinations will also help clarify the additive nature of the LCR.

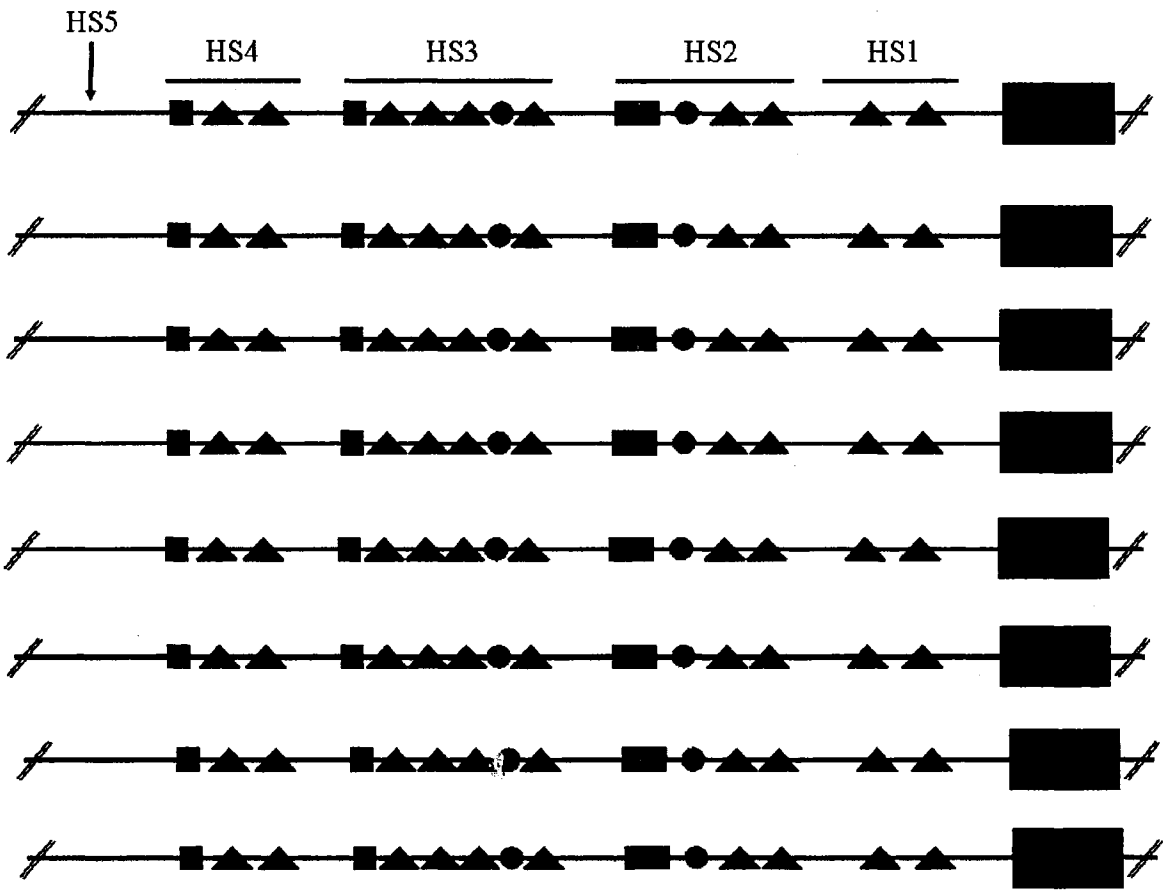
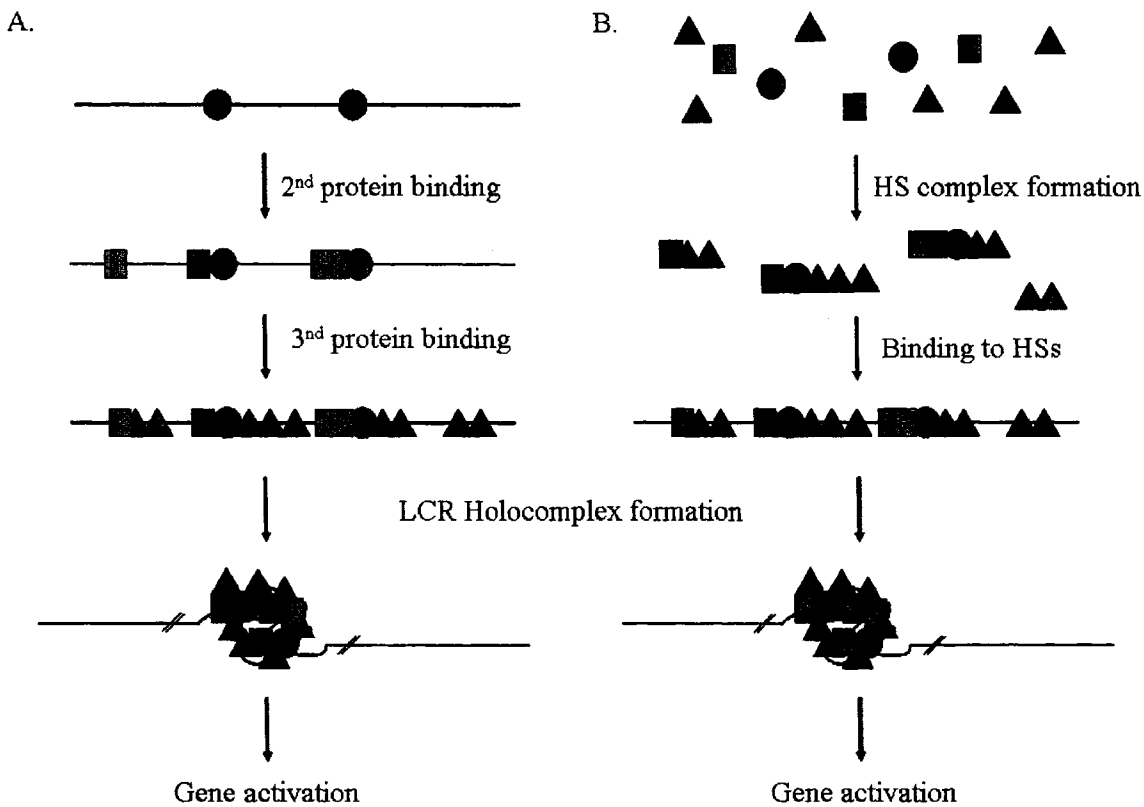


Figure 2. Models of LCR holocomplex formation. A. DNA Dependent Model. B. DNA Independent Model. Circles, boxes, and triangles represent proteins that bind to the CACBP, NF-E2, and GATA sites, respectively. In the DNA dependent model each protein binds to the LCR individually and then interacts on the DNA to form the LCR holocomplex. The DNA independent model suggests that small HS complexes form off the DNA and that these complexes then bind to each HS leading to the eventual formation of a LCR holocomplex. In both models gene activation occurs after LCR holocomplex formation.



GERENRAL LIST OF REFERENCES

1. Lavoisier, A. (1777) Memoir on Combustion in General. *Histoire de l'academie royale des Sciences*, 592-600.
2. Perutz, M., Rossman, MG, Cullis, AF et al. (1960) Structure of haemoglobin: A three- dimensional Fourier synthesis at 5.5(A)angstrom resolution obtained by x-ray analysis. *Nature*, **185**, 416-422.
3. Efremov, G.D., Huisman, T.H., Stanulovic, M., Zurovec, M., Duma, H., Wilson, J.B. and Jeremic, V. (1974) Haemoglobin M Saskatoon and haemoglobin M Hyde Park in two Yugoslavian families. *Scand J Haematol*, **13**, 48-60.
4. Ingram, V.M. (1959) Abnormal human haemoglobins. III. The chemical difference between normal and sickle cell haemoglobins. *Biochim Biophys Acta*, **36**, 402-411.
5. Bunn, F.H.a.F., Bernard G. (1986) *Hemoglobin: Molecular, Genetic and Clinical Aspects*. W.B Saunders Company, Philadelphia, Pa.
6. Dzierzak, E. and Medvinsky, A. (1995) Mouse embryonic hematopoiesis. *Trends Genet*, **11**, 359-366.
7. Medvinsky, A. and Dzierzak, E. (1996) Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell*, **86**, 897-906.
8. Medvinsky, A.L. and Dzierzak, E.A. (1998) Development of the definitive hematopoietic hierarchy in the mouse. *Dev Comp Immunol*, **22**, 289-301.

9. Palis, J., Robertson, S., Kennedy, M., Wall, C. and Keller, G. (1999) Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. *Development*, **126**, 5073-5084.

10. Kovach, J.S., Marks, P.A., Russell, E.S. and Epler, H. (1967) Erythroid cell development in fetal mice: ultrastructural characteristics and hemoglobin synthesis. *J Mol Biol*, **25**, 131-142.

11. Whitelaw, E., Lamb, P., Hogben, P. and Proudfoot, N.J. (1989) The globin switch at the level of mRNA in the developing mouse. *Prog Clin Biol Res*, **316A**, 323-333.

12. Wong, P.M., Chung, S.W., White, J.S., Reicheld, S.M., Patterson, M., Clarke, B.J. and Chui, D.H. (1983) Adult hemoglobins are synthesized in murine fetal hepatic erythropoietic cells. *Blood*, **62**, 1280-1288.

13. Popp, R.A., D'Surney, S.J. and Wawrzyniak, C.J. (1987) Changes in expression of murine alpha- and beta-globin genes during development. *Prog Clin Biol Res*, **251**, 81-89.

14. Konkel, D.A., Maizel, J.V., Jr. and Leder, P. (1979) The evolution and sequence comparison of two recently diverged mouse chromosomal beta--globin genes. *Cell*, **18**, 865-873.

15. Jahn, C.L., Hutchison, C.A., 3rd, Phillips, S.J., Weaver, S., Haigwood, N.L., Voliva, C.F. and Edgell, M.H. (1980) DNA sequence organization of the beta-globin complex in the BALB/c mouse. *Cell*, **21**, 159-168.

16. Phillips, S.J., Hardies, S.C., Jahn, C.L., Edgell, M.H. and Hutchison, C.A., 3rd. (1984) The complete nucleotide sequence of a beta-globin-like structure, beta h2, from the [Hbb]d mouse BALB/c. *J Biol Chem*, **259**, 7947-7954.

17. Farace, M.G., Brown, B.A., Raschella, G., Alexander, J., Gambari, R., Fantoni, A., Hardies, S.C., Hutchison, C.A., 3rd and Edgell, M.H. (1984) The mouse beta h1 gene codes for the z chain of embryonic hemoglobin. *J Biol Chem*, **259**, 7123-7128.

18. Hansen, J.N., Konkel, D.A. and Leder, P. (1982) The sequence of a mouse embryonic beta-globin gene. Evolution of the gene and its signal region. *J Biol Chem*, **257**, 1048-1052.

19. Weaver, S., Comer, M.B., Jahn, C.L., Hutchison, C.A., 3rd and Edgell, M.H. (1981) The adult beta-globin genes of the "single" type mouse C57BL. *Cell*, **24**, 403-411.

20. Hardison, R. (1998) Hemoglobins from bacteria to man: evolution of different patterns of gene expression. *J Exp Biol*, **201**, 1099-1117.

21. Steinberg, M.H. and Adams, J.G., 3rd. (1991) Hemoglobin A2: origin, evolution, and aftermath. *Blood*, **78**, 2165-2177.

22. Hardison, R. and Miller, W. (1993) Use of long sequence alignments to study the evolution and regulation of mammalian globin gene clusters. *Mol Biol Evol*, **10**, 73-102.

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

24. Forrester, W.C., Thompson, C., Elder, J.T. and Groudine, M. (1986) A developmentally stable chromatin structure in the human beta-globin gene cluster. *Proc Natl Acad Sci U S A*, **83**, 1359-1363.

25. Moon, A.M. and Ley, T.J. (1990) Conservation of the primary structure, organization, and function of the human and mouse beta-globin locus-activating regions. *Proc Natl Acad Sci U S A*, **87**, 7693-7697.

26. Hug, B.A., Moon, A.M. and Ley, T.J. (1992) Structure and function of the murine beta-globin locus control region 5' HS-3. *Nucleic Acids Res*, **20**, 5771-5778.

27. Van der Ploeg, L.H., Konings, A., Oort, M., Roos, D., Bernini, L. and Flavell, R.A. (1980) gamma-beta-Thalassaemia studies showing that deletion of the gamma- and delta-genes influences beta-globin gene expression in man. *Nature*, **283**, 637-642.

28. Curtin, P., Pirastu, M., Kan, Y.W., Gobert-Jones, J.A., Stephens, A.D. and Lehmann, H. (1985) A distant gene deletion affects beta-globin gene function in an atypical gamma delta beta-thalassemia. *J Clin Invest*, **76**, 1554-1558.

29. Driscoll, M.C., Dobkin, C.S. and Alter, B.P. (1989) Gamma delta beta-thalassemia due to a de novo mutation deleting the 5' beta-globin gene activation-region hypersensitive sites. *Proc Natl Acad Sci U S A*, **86**, 7470-7474.

30. Montoliu, L., Umland, T. and Schutz, G. (1996) A locus control region at -12 kb of the tyrosinase gene. *Embo J*, **15**, 6026-6034.

31. Regales, L., Giraldo, P., Garcia-Diaz, A., Lavado, A. and Montoliu, L. (2003) Identification and functional validation of a 5' upstream regulatory sequence in

the human tyrosinase gene homologous to the locus control region of the mouse tyrosinase gene. *Pigment Cell Res*, **16**, 685-692.

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

33. Diaz, P., Cado, D. and Winoto, A. (1994) A locus control region in the T cell receptor alpha/delta locus. *Immunity*, **1**, 207-217.

34. Grosveld, F., van Assendelft, G.B., Greaves, D.R. and Kollias, G. (1987) Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell*, **51**, 975-985.

35. Ganss, R., Montoliu, L., Monaghan, A.P. and Schutz, G. (1994) A cell-specific enhancer far upstream of the mouse tyrosinase gene confers high level and copy number-related expression in transgenic mice. *Embo J*, **13**, 3083-3093.

36. Giraldo, P., Martinez, A., Regales, L., Lavado, A., Garcia-Diaz, A., Alonso, A., Busturia, A. and Montoliu, L. (2003) Functional dissection of the mouse tyrosinase locus control region identifies a new putative boundary activity. *Nucleic Acids Res*, **31**, 6290-6305.

37. Fields, P.E., Lee, G.R., Kim, S.T., Bartsevich, V.V. and Flavell, R.A. (2004) Th2-specific chromatin remodeling and enhancer activity in the Th2 cytokine locus control region. *Immunity*, **21**, 865-876.

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

39. Chada, K., Magram, J., Raphael, K., Radice, G., Lacy, E. and Costantini, F. (1985) Specific expression of a foreign beta-globin gene in erythroid cells of transgenic mice. *Nature*, **314**, 377-380.

40. Ryan, T.M., Behringer, R.R., Martin, N.C., Townes, T.M., Palmiter, R.D. and Brinster, R.L. (1989) A single erythroid-specific DNase I super-hypersensitive site activates high levels of human beta-globin gene expression in transgenic mice. *Genes Dev*, **3**, 314-323.

41. Fraser, P., Hurst, J., Collis, P. and Grosveld, F. (1990) DNaseI hypersensitive sites 1, 2 and 3 of the human beta-globin dominant control region direct position-independent expression. *Nucleic Acids Res*, **18**, 3503-3508.

42. Fraser, P., Pruzina, S., Antoniou, M. and Grosveld, F. (1993) Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression on the globin genes. *Genes Dev*, **7**, 106-113.

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

44. Fiering, S., Epner, E., Robinson, K., Zhuang, Y., Telling, A., Hu, M., Martin, D.I., Enver, T., Ley, T.J. and Groudine, M. (1995) Targeted deletion of 5'HS2 of the murine beta-globin LCR reveals that it is not essential for proper regulation of the beta-globin locus. *Genes Dev*, **9**, 2203-2213.

45. Ley, T.J., Hug, B., Fiering, S., Epner, E., Bender, M.A. and Groudine, M. (1998) Reduced beta-globin gene expression in adult mice containing deletions of locus control region 5' HS-2 or 5' HS-3. *Ann N Y Acad Sci*, **850**, 45-53.

46. Hu, X., Bulger, M., Bender, M.A., Fields, J., Groudine, M. and Fiering, S. (2006) Deletion of the core region of 5' HS2 of the mouse beta-globin locus control region reveals a distinct effect in comparison with human beta-globin transgenes. *Blood*, **107**, 821-826.

47. Jia, C.P., Huang, S.Z., Yan, J.B., Xiao, Y.P., Ren, Z.R. and Zeng, Y.T. (2003) Effects of human locus control region elements HS2 and HS3 on human beta-globin gene expression in transgenic mouse. *Blood Cells Mol Dis*, **31**, 360-369.

48. Hug, B.A., Wesselschmidt, R.L., Fiering, S., Bender, M.A., Epner, E., Groudine, M. and Ley, T.J. (1996) Analysis of mice containing a targeted deletion of beta-globin locus control region 5' hypersensitive site 3. *Mol Cell Biol*, **16**, 2906-2912.

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

50. Farrell, C.M., West, A.G. and Felsenfeld, G. (2002) Conserved CTCF insulator elements flank the mouse and human beta-globin loci. *Mol Cell Biol*, **22**, 3820-3831.

51. Tanimoto, K., Sugiura, A., Omori, A., Felsenfeld, G., Engel, J.D. and Fukamizu, A. (2003) Human beta-globin locus control region HS5 contains CTCF- and developmental stage-dependent enhancer-blocking activity in erythroid cells. *Mol Cell Biol*, **23**, 8946-8952.

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

53. Townes, T.M. and Behringer, R.R. (1990) Human globin locus activation region (LAR): role in temporal control. *Trends Genet*, **6**, 219-223.

54. Orkin, S.H. (1990) Globin gene regulation and switching: circa 1990. *Cell*, **63**, 665-672.

55. Engel, J.D. and Tanimoto, K. (2000) Looping, linking, and chromatin activity: new insights into beta-globin locus regulation. *Cell*, **100**, 499-502.

56. Bulger, M. and Groudine, M. (1999) Looping versus linking: toward a model for long-distance gene activation. *Genes Dev*, **13**, 2465-2477.

57. Li, Q., Barkess, G. and Qian, H. (2006) Chromatin looping and the probability of transcription. *Trends Genet*, **22**, 197-202.

58. de Laat, W. and Grosveld, F. (2003) Spatial organization of gene expression: the active chromatin hub. *Chromosome Res*, **11**, 447-459.

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

60. Patrinos, G.P., de Krom, M., de Boer, E., Langeveld, A., Imam, A.M., Strouboulis, J., de Laat, W. and Grosveld, F.G. (2004) Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. *Genes Dev*, **18**, 1495-1509.

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

62. Kosak, S.T. and Groudine, M. (2004) Form follows function: The genomic organization of cellular differentiation. *Genes Dev*, **18**, 1371-1384.

63. Kosak, S.T. and Groudine, M. (2004) Gene order and dynamic domains. *Science*, **306**, 644-647.

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

65. Ragoczy, T., Bender, M.A., Telling, A., Byron, R. and Groudine, M. (2006) The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation. *Genes Dev.*

66. Carter, D., Chakalova, L., Osborne, C.S., Dai, Y.F. and Fraser, P. (2002) Long-range chromatin regulatory interactions in vivo. *Nat Genet*, **32**, 623-626.

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

68. Hardison, R., Slightom, J.L., Gumucio, D.L., Goodman, M., Stojanovic, N. and Miller, W. (1997) Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. *Gene*, **205**, 73-94.

69. Caterina, J.J., Ryan, T.M., Pawlik, K.M., Palmiter, R.D., Brinster, R.L., Behringer, R.R. and Townes, T.M. (1991) 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**, 1626-1630.

70. Moi, P. and Kan, Y.W. (1990) Synergistic enhancement of globin gene expression by activator protein-1-like proteins. *Proc Natl Acad Sci U S A*, **87**, 9000-9004.

71. Caterina, J.J., Ciavatta, D.J., Donze, D., Behringer, R.R. and Townes, T.M. (1994) Multiple elements in human beta-globin locus control region 5' HS 2 are involved in enhancer activity and position-independent, transgene expression. *Nucleic Acids Res*, **22**, 1006-1011.

72. Jackson, J.D., Petrykowska, H., Philipsen, S., Miller, W. and Hardison, R. (1996) Role of DNA sequences outside the cores of DNase hypersensitive sites (HSs) in functions of the beta-globin locus control region. Domain opening and synergism between HS2 and HS3. *J Biol Chem*, **271**, 11871-11878.

73. Stamatoyannopoulos, J.A., Goodwin, A., Joyce, T. and Lowrey, C.H. (1995) NF-E2 and GATA binding motifs are required for the formation of DNase I hypersensitive site 4 of the human beta-globin locus control region. *Embo J*, **14**, 106-116.

74. Ney, P.A., Sorrentino, B.P., Lowrey, C.H. and Nienhuis, A.W. (1990) Inducibility of the HS II enhancer depends on binding of an erythroid specific nuclear protein. *Nucleic Acids Res*, **18**, 6011-6017.

75. Caterina, J.J., Donze, D., Sun, C.W., Ciavatta, D.J. and Townes, T.M. (1994) Cloning and functional characterization of LCR-F1: a bZIP transcription factor

that activates erythroid-specific, human globin gene expression. *Nucleic Acids Res*, **22**, 2383-2391.

76. Farmer, S.C., Sun, C.W., Winnier, G.E., Hogan, B.L. and Townes, T.M. (1997) The bZIP transcription factor LCR-F1 is essential for mesoderm formation in mouse development. *Genes Dev*, **11**, 786-798.

77. Miller, I.J. and Bieker, J.J. (1993) A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. *Mol Cell Biol*, **13**, 2776-2786.

78. Feng, W.C., Southwood, C.M. and Bieker, J.J. (1994) Analyses of beta-thalassemia mutant DNA interactions with erythroid Kruppel-like factor (EKLF), an erythroid cell-specific transcription factor. *J Biol Chem*, **269**, 1493-1500.

79. Orkin, S.H., Antonarakis, S.E. and Kazazian, H.H., Jr. (1984) Base substitution at position -88 in a beta-thalassemic globin gene. Further evidence for the role of distal promoter element ACACCC. *J Biol Chem*, **259**, 8679-8681.

80. Kulozik, A.E., Bellan-Koch, A., Bail, S., Kohne, E. and Kleihauer, E. (1991) Thalassemia intermedia: moderate reduction of beta globin gene transcriptional activity by a novel mutation of the proximal CACCC promoter element. *Blood*, **77**, 2054-2058.

81. Nuez, B., Michalovich, D., Bygrave, A., Ploemacher, R. and Grosveld, F. (1995) Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature*, **375**, 316-318.

82. Perkins, A.C., Sharpe, A.H. and Orkin, S.H. (1995) Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature*, **375**, 318-322.

83. Yu, D., Ellis, H.M., Lee, E.C., Jenkins, N.A., Copeland, N.G. and Court, D.L. (2000) An efficient recombination system for chromosome engineering in Escherichia coli. *Proc Natl Acad Sci U S A*, **97**, 5978-5983.

84. Lee, E.C., Yu, D., Martinez de Velasco, J., Tessarollo, L., Swing, D.A., Court, D.L., Jenkins, N.A. and Copeland, N.G. (2001) A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics*, **73**, 56-65.

85. Swaminathan, S., Ellis, H.M., Waters, L.S., Yu, D., Lee, E.C., Court, D.L. and Sharan, S.K. (2001) Rapid engineering of bacterial artificial chromosomes using oligonucleotides. *Genesis*, **29**, 14-21.

86. Nobelstiftelsen. (1964) *Chemistry, 1942-1962*. Published for the Nobel Foundation by Elsevier Pub. Co., Amsterdam; New York.

87. Ryan, T.M. (1990), University of Alabama at Birmingham, Birmingham.

88. Kusch, T., Storck, T., Walldorf, U. and Reuter, R. (2002) Brachyury proteins regulate target genes through modular binding sites in a cooperative fashion. *Genes Dev*, **16**, 518-529.

89. Christensen, T.H. and Kedes, L. (1999) The myogenic regulatory circuit that controls cardiac/slow twitch troponin C gene transcription in skeletal muscle involves E-box, MEF-2, and MEF-3 motifs. *Gene Expr*, **8**, 247-261.

90. Zhou, D., Pawlik, K.M., Ren, J., Sun, C.W. and Townes, T.M. (2006) Differential binding of EKLF to embryonic/fetal globin gene promoters during development. *J Biol Chem.*

91. Chang, J.C. and Kan, Y.W. (1979) beta 0 thalassemia, a nonsense mutation in man. *Proc Natl Acad Sci U S A*, **76**, 2886-2889.

92. Kazazian, H.H., Jr., Orkin, S.H., Boehm, C.D., Sexton, J.P. and Antonarakis, S.E. (1983) beta-Thalassemia due to a deletion of the nucleotide which is substituted in the beta S-globin gene. *Am J Hum Genet*, **35**, 1028-1033.

93. Chang, J.C., Alberti, A. and Kan, Y.W. (1983) A beta-thalassemia lesion abolishes the same Mst II site as the sickle mutation. *Nucleic Acids Res*, **11**, 7789-7794.

94. Orkin, S.H., Kazazian, H.H., Jr., Antonarakis, S.E., Goff, S.C., Boehm, C.D., Sexton, J.P., Waber, P.G. and Giardina, P.J. (1982) Linkage of beta-thalassaemia mutations and beta-globin gene polymorphisms with DNA polymorphisms in human beta-globin gene cluster. *Nature*, **296**, 627-631.

95. Westaway, D. and Williamson, R. (1981) An intron nucleotide sequence variant in a cloned beta +-thalassaemia globin gene. *Nucleic Acids Res*, **9**, 1777-1788.

96. Treisman, R., Orkin, S.H. and Maniatis, T. (1983) Specific transcription and RNA splicing defects in five cloned beta-thalassaemia genes. *Nature*, **302**, 591-596.

97. Weatherall, D.J., and Clegg, J.B. (1981) *The Thalassemia Syndromes*. 3rd ed. Blackwell Scientific Publications, Oxford.

98. Aksoy, M. (1970) Thalassaemia intermedia: a genetic study in 11 patients. *J Med Genet*, **7**, 47-51.

99. Stamatoyannopoulos, G., Woodson, R., Papayannopoulou, T., Heywood, D. and Kurachi, S. (1974) Inclusion-body beta-thalassemia trait. A form of beta thalassemia producing clinical manifestations in simple heterozygotes. *N Engl J Med*, **290**, 939-943.

100. Friedman, S., Ozsoylu, S., Luddy, R. and Schwartz, E. (1976) Heterozygous beta thalassaemia of unusual severity. *Br J Haematol*, **32**, 65-77.

101. Adams, J.G., 3rd, Boxer, L.A., Baehner, R.L., Forget, B.G., Tsistrakis, G.A. and Steinberg, M.H. (1979) Hemoglobin Indianapolis (beta 112[G14] arginine). An unstable beta-chain variant producing the phenotype of severe beta-thalassemia. *J Clin Invest*, **63**, 931-938.

APPENDIX A

IACUC APPROVAL LETTER

THE UNIVERSITY OF ALABAMA AT BIRMINGHAM

Office of the Provost

NOTICE OF APPROVAL

DATE: January 9, 2006

TO: Tim M. Townes, Ph.D.
KAUL-537 0024
FAX: 934-2889

FROM: Judith A. Kapp, Ph.D., Chair
Institutional Animal Care and Use Committee

SUBJECT: Title: Genetic Modifiers of Sickle Cell Disease
Sponsor: NIH
Animal Project Number: 060105655

On January 9, 2006, the University of Alabama at Birmingham Institutional Animal Care and Use Committee (IACUC) reviewed the animal use proposed in the above referenced application. It approved the use of the following species and numbers of animals:

Species	Use Category	Number in Category
Mice	A	5157
Mice	B	268

Animal use is scheduled for review one year from January 2006. Approval from the IACUC must be obtained before implementing any changes or modifications in the approved animal use.

Please keep this record for your files, and forward the attached letter to the appropriate granting agency.

Refer to Animal Protocol Number (APN) 060105655 when ordering animals or in any correspondence with the IACUC or Animal Resources Program (ARP) offices regarding this study. If you have concerns or questions regarding this notice, please call the IACUC office at 934-7692.

Institutional Animal Care and Use Committee
B10 Volker Hall
1717 7th Avenue South
205.934.7692 • Fax 205.934.1188
iacuc@uab.edu
www.uab.edu/iacuc

The University of
Alabama at Birmingham
Mailing Address:
VH B10
1530 3RD AVE S
BIRMINGHAM AL 35294-0019