CHAPTER 4
A SILENCER LOCATED 3'TO THE Aγ-GLOBIN GENE INFLUENCES TRANSCRIPTION OF THE HUMAN β-GLOBIN LOCUS AT THE EMBRYONIC YOLK SAC STAGE IN TRANSGENIC MICE

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

Summary

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

Introduction

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

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

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

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

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

Results

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

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

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

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

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

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

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

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

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

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

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

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

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

76

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

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

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

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

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

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

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

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

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

Discussion

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

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

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

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

Materials and Methods

DNA constructs

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

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

Homologous recombination in E. coli

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

Transgenic Mice

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

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

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

DNA FISH analysis

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

S1 nuclease protection assays

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

Primary transcript in situ hybridisation

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

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

Primer extensions

$^G\gamma$ and $^A\gamma$ transcripts were distinguished by the use of two $^G\gamma$- and $^A\gamma$-specific primers. Primers were end-labelled with T4 polynucleotide kinase (New England Biolabs) at 37^0C for 1 hour and unincorporated radionucleotides were removed by Sephadex G-25 chromatography. Specific activities were determined for both primers, as previously described (Lindenbaum and Grosveld, 1990) and in all experiments were found to be of very similar levels. Each labelled primer was phenol extracted, ethanol precipitated and re-dissolved in DEPC-treated TE (10mM Tris-HCl, 1mM EDTA pH 8.0) to a final concentration of 2ng/μl. Approximately 10μg of 10.5 dpc yolk sac or 12.5 dpc foetal liver RNA was annealed to 2ng of both $^G\gamma$ and $^A\gamma$ primers in a 33μl final reaction volume also containing 4μl of 5x Superscript buffer (Life Technologies) and 1μl of 0.1M DTT. Annealing of the primers to the RNA template was carried out by incubation at 65^0C for 10 minutes followed by incubation at 37^0C for 1 hour. The extension step was carried out by adding to each reaction 1μl of nucleotide mix containing 10mM each of dGTP, dATP, dCTP and 1.25mM of ddTTP, 0.5μl RNase inhibitor (RNAguard, Amersham-Pharmacia) and 0.5μl of Superscript reverse transcriptase (Life Technologies), followed by incubation at 42^0C for 1 hour. Samples were ethanol precipitated, resuspended in 5μl of 97.5% formamide loading dye, heat-denatured and resolved by electrophoresis on a 12% denaturing polyacrylamide sequencing gel.
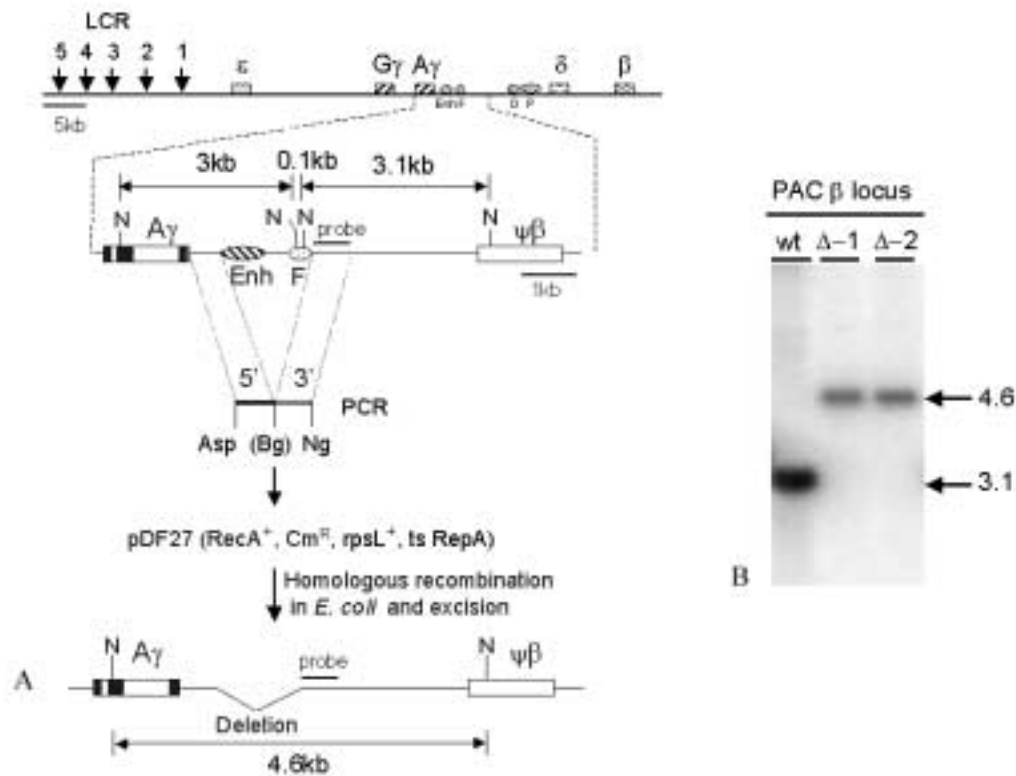
Figure 1: *Deletion of Enh and F from the human β-globin locus PAC by homologous recombination in E. coli—* (A) The 75 kb human β-globin locus is shown at the top with the region 5' to the δ-globin gene magnified above it to show the O and P elements (ovals). The region 3' of the ^{A}γ gene has also been magnified below the locus to show the Enh and F elements (ovals) (ψβ: pseudo-β-globin gene). The 5' and 3' regions of homology amplified for use in the targeting construct are indicated (dotted lines). These were cloned into the Asp 718/NgoM IV sites of the targeting vector pDF27 which carries a chloramphenicol resistance gene (Cm^R), a temperature sensitive replication initiation protein gene (ts RepA), the recA gene and a wild type copy of the rpsL gene required for *E. coli* DH10B sensitivity to high doses of streptomycin (Imam *et al*., 2000). Homologous recombination followed by excision of the targeting construct results in the deletion of the Enh and F elements 3' of the ^{A}γ gene in a subset of the modified PAC clones (Imam *et al*., 2000). As a result, two Nco I sites are deleted giving rise to a new 4.6 kb fragment from the original 3.0 kb and 3.1 kb bands. (B) Southern blot hybridisation of Nco I-digested PAC DNA obtained from the original unmodified human β locus PAC (wt lane) and two modified PAC clones bearing the 3' ^{A}γ deletion (lanes Δ-1 and Δ-2). The probe (shown in panel A) detects a 3.1 kb fragment in the unmodified PAC and a 4.6 kb fragment in the modified PAC resulting from the deletion of two Nco I sites (panel A). Abbreviations: N-Nco I, Asp-Asp 718, Bg-Bgl II, Ng-NgoM IV.
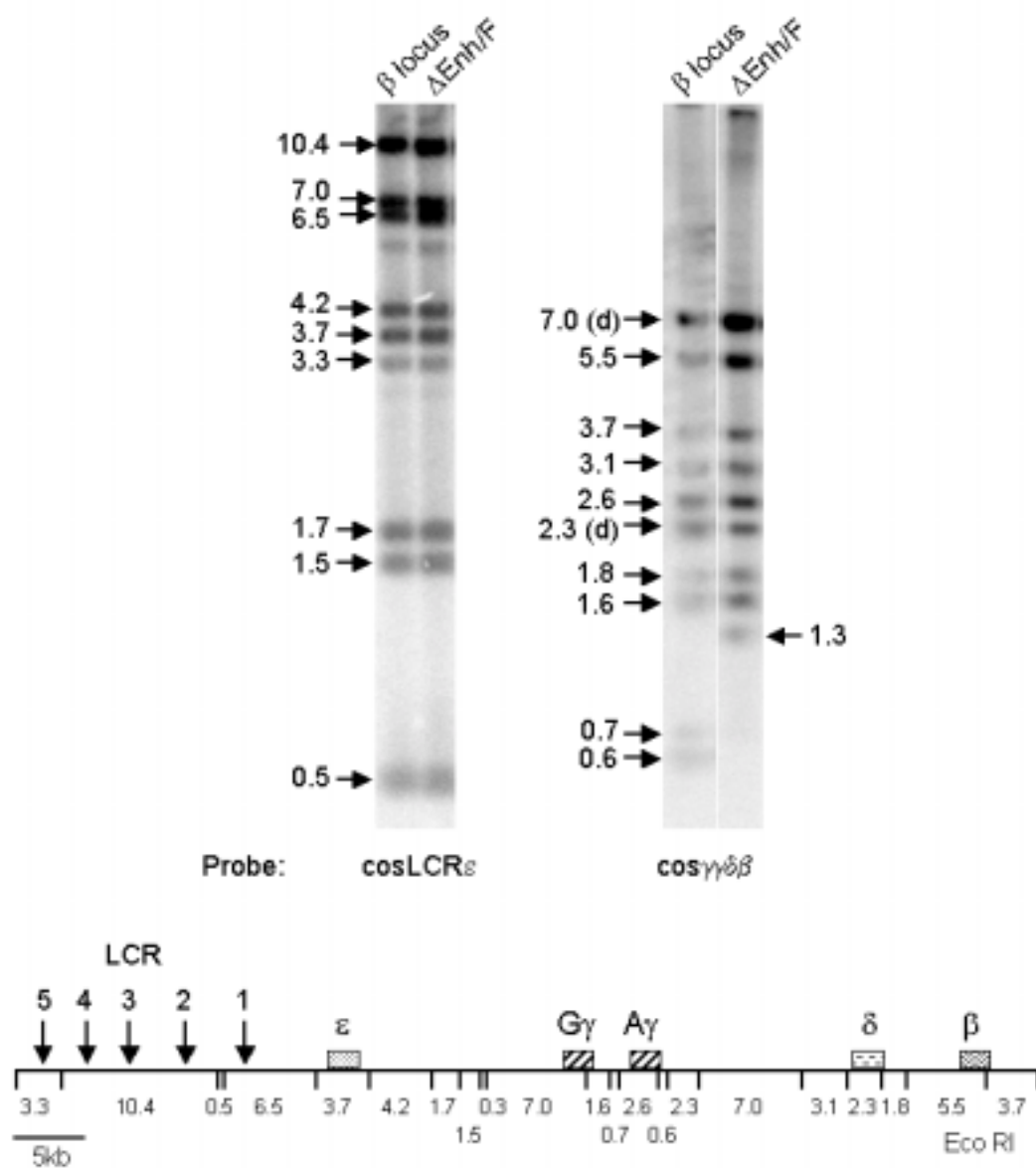
Figure 2*: Mapping the integrity of the 75 kb human β-globin locus in the unmodified (wt lane) and Enh/F-deleted (ΔEnh/F) globin locus PACs*—Eco RI digested PAC DNA was probed with cosmids LCRε (left panel) and γγδβ (right panel). The modified PAC clone DNA that was used for transgenesis is shown. The Eco RI map of the locus is shown below the blots. Deletion of the Enh and F elements results in a 2.3 kb Eco RI fragment that is part of a doublet (d) in the unmodified PAC, becoming a new 1.3 kb band.
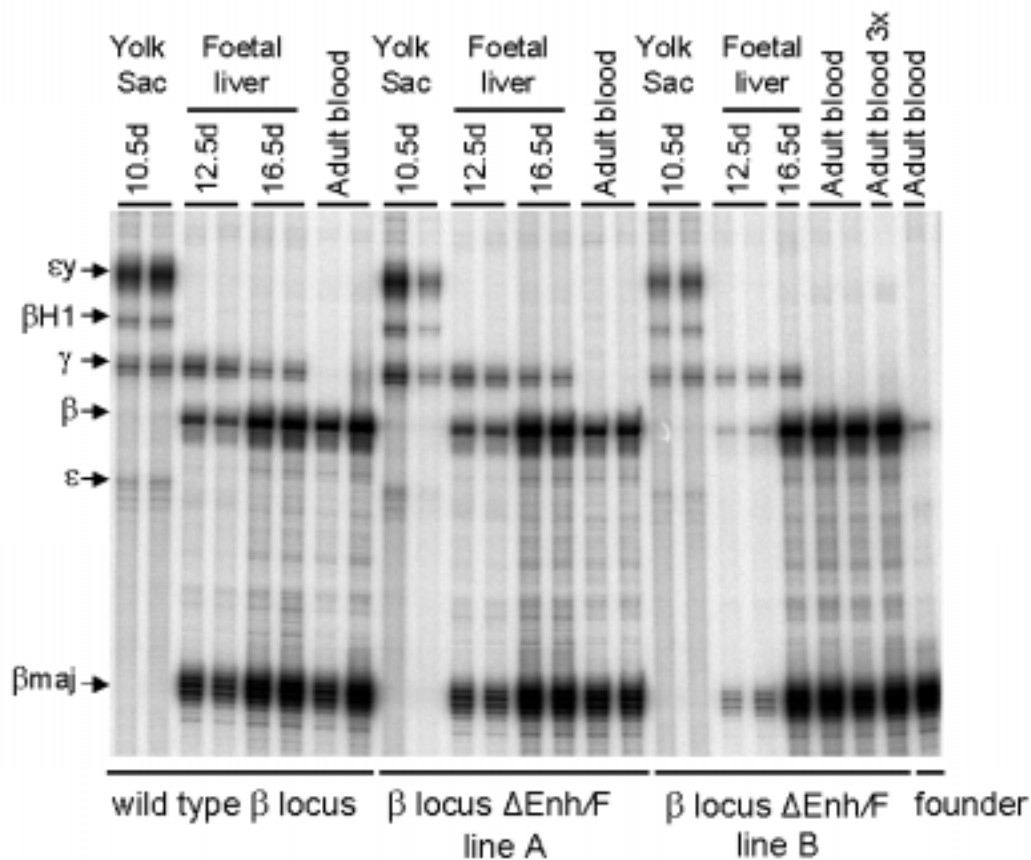
Figure 3: *S1 nuclease protection analysis of expression of human versus mouse β-like globin genes in embryonic yolk sac, foetal liver and adult blood*—Duplicate RNA samples obtained from embryos (littermates) and adult blood from two mouse lines transgenic for a single copy of the Enh and F-deleted β locus PAC (ΔEnh/F line A and B) were analysed (only one embryo was analysed for the 16.5d time point in line ΔEnh/F-B). As control, one line transgenic for a single copy of the unmodified (wild type) human β globin locus PAC was analysed. Adult blood RNA from a transgenic founder that did not transmit the deleted PAC transgene, was also included in the analysis (last lane). The protected fragments are indicated to the left of the panel. The relative specific activities of probes used in this experiment are given in the legend for Table 1.
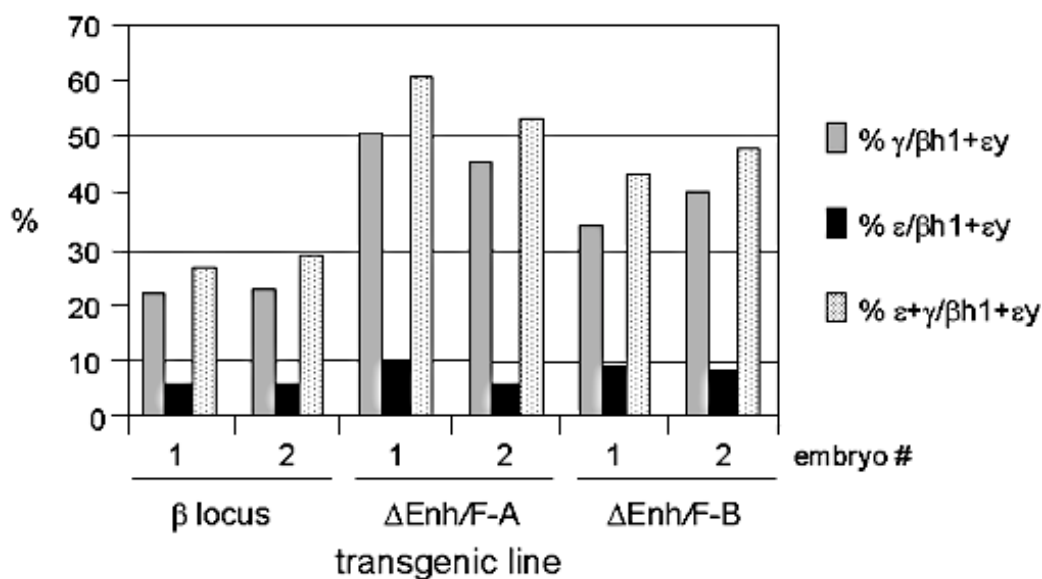
Figure 4: *Bar chart representation of human ε- and γ-globin expression levels at 10.5 day yolk sac, as a percentage of mouse εy-and βh1-expression*—The values are those presented in Table 1 and were calculated from the S1 nuclease protection assay in Figure 3.

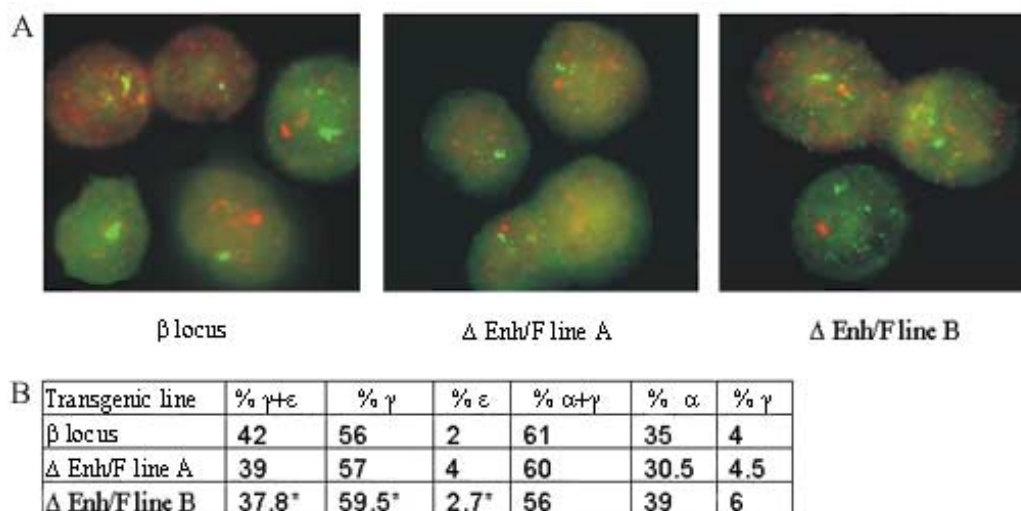| Transgenic line	% γ+ε	% γ	% ε	% α+γ	% α	% γ
β locus	42	56	2	61	35	4
Δ Enh/F line A	39	57	4	60	30.5	4.5
Δ Enh/F line B	37.8*	59.5*	2.7*	56	39	6

Figure 5: *Primary transcript in situ hybridisation in 10.5d mouse embryonic blood*—(A) Representative fields of cells hybridised with mouse α-globin (green signal) and human γ-globin (red signal) intron-specific probes. (B) Quantitation of ε- and γ-globin primary transcripts (columns 2-4) and γ-globin and mouse α-globin primary transcripts (columns 5-7) by *in situ* hybridisation of 10.5d embryonic blood cells. At least 100 cells were counted in each case, with the exception of (*) where 36 cells were counted.
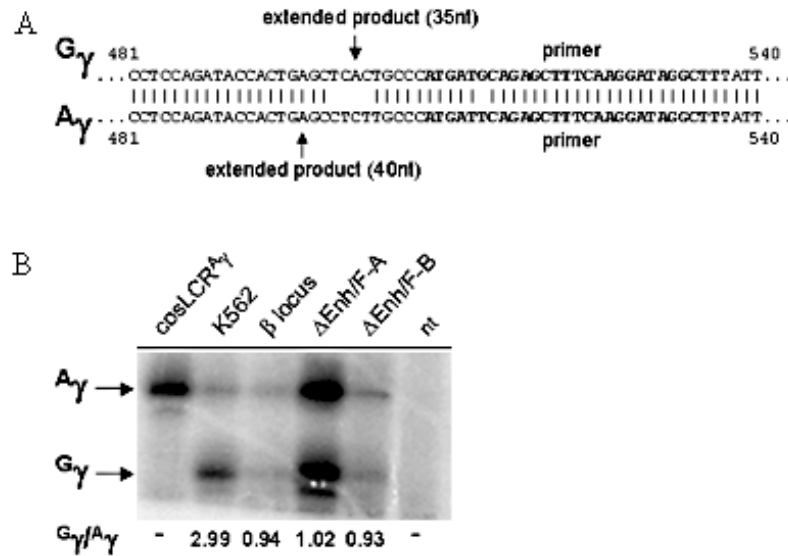
Figure 6: *Analysis of $^G\gamma$-versus $^A\gamma$-globin expression by primer extension—* (A) alignment of sequences from the 3' of the $^G\gamma$-and $^A\gamma$-cDNAs (nucleotides 481-540) which shows the 5nt region of non-homology between the two sequences. The sequences of the $^G\gamma$- and $^A\gamma$-specific primers used in the assay are shown in bold. The use of dideoxy-TTP in the extension mix results in 35nt and 40nt extended products from the $^G\gamma$ and $^A\gamma$ mRNAs, respectively (arrows). (B) resolution of $^G\gamma$ and $^A\gamma$ primer extension products on a 10% polyacrylamide gel. 10.5d yolk sac RNA was analysed from the control single copy unmodified β locus PAC line (β locus lane) and the two lines carrying the Enh/F-deleted locus PAC (ΔEnh/F-A and -B lanes). Controls include a 12.5 day foetal liver RNA sample from an embryo transgenic for a single copy of an LCR $^A\gamma$ construct (cosLCR $^A\gamma$ lane), RNA from K562 cells that express γ-globin and a 10.5 day yolk sac RNA from a non- transgenic embryo (nt lane).

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

A.

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

B.

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