PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The Regulation of Human Globin Gene Switching

Frank Grosveld, Mike Antoniou, Meera Berry, Ernie De Boer, Niall Dillon, James Ellis, Peter Fraser, Olivia Hanscombe, Jacky Hurst, Ali Imam, Mike Lindenbaum, Sjaak Philipsen, Sara Pruzina, John Strouboulis, Selina Raguz-Bolognesi and Dale Talbot

Phil. Trans. R. Soc. Lond. B 1993 **339**, 183-191
doi: 10.1098/rstb.1993.0015

The regulation of human globin gene switching

FRANK GROSVELD, MIKE ANTONIOU, MEERA BERRY, ERNIE DE BOER, NIALL DILLON, JAMES ELLIS, PETER FRASER, OLIVIA HANSCOMBE, JACKY HURST, ALI IMAM, MIKE LINDENBAUM, SJAAK PHILIPSEN, SARA PRUZINA, JOHN STROUBOULIS, SELINA RAGUZ-BOLOGNESI AND DALE TALBOT

Laboratory of Gene Structure & Expression, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, U.K.

SUMMARY

This paper describes the mechanism of regulation of the human β-globin on the basis of a number of natural mutations and experiments in transgenic mice. From these data we conclude that this multigene locus is regulated at a number of different levels involving specific interactions between the Locus Control Region (LCR) and the individual genes. Most important is the action of stage specific transcription factors acting on sequences immediately flanking the genes. In addition, specificity is obtained through specific interaction of the genes with the LCR and through competition of the genes for interaction with the LCR.

1. INTRODUCTION

The human β globin gene cluster spans a region of 70 kb on chromosome 11 and contains five developmentally regulated genes in the order 5′ ε,$\gamma_G\gamma_A$,δ,β 3′. In the early stages of human development, the embryonic yolk sac is the haematopoietic tissue and expresses the ε globin gene. This is followed by a switch to the γ globin genes in the foetal liver and the δ and β genes in adult bone marrow (for review, see Collins & Weissman (1984)). At all stages the output of these genes is balanced by that of the α gene locus. The important difference between the α globin genes and the β globin genes is the fact that they undergo a single switch and do not have an additional switch around the time of birth. At all developmental stages these genes are expressed at exceptionally high levels giving rise to 90% of the total soluble protein in circulating red blood cells.

The entire β-like gene locus has ben sequenced and a large number of structural defects, collectively known as the β-thalassaemias, have been documented in and around the β-globin gene (for review, see Collins & Weissman (1984); Poncz *et al.* (1988)). In a related condition, hereditary persistence of foetal haemoglobin (HPFH), γ-globin gene expression and hence HbF (foetal haemoglobin) production persist into adult life. These diseases are not only clinically important but they also provide natural models for the study of transcriptional regulation during development (see below).

2. THE β GLOBIN LCR

The entire locus is regulated by the locus control region (LCR) which first became apparent from the study of a Dutch thalassaemia (van der Ploeg *et al.* 1979) and later a Hispanic thalassaemia (Driscoll *et al.* 1989). The Dutch thalassaemia contained an intact β globin gene but had a deletion of the upstream part of the locus which removed the LCR and prevented activation of the β gene (Kioussis *et al.* 1983; Wright *et al.* 1984; Taramelli *et al.* 1986). The LCR is characterized by a set of developmentally stable, hypersensitive sites, 5′ HS1, 2, 3 and 4 and is distributed over approximately 15 kb located upstream of the ε globin gene (figure 1; Tuan *et al.* 1985; Grosveld *et al.* 1987; Forrester *et al.* 1987). The deletion of the LCR in these thalassaemias not only showed that it is required for the expression of the β-like globin genes but also that its presence affects chromatin structure over a distance of at least 150 kb (Forrester *et al.* 1990. However, despite the early data on the Dutch thalassaemia (Kioussis *et al.* 1983) and the hypersensitive sites (Tuan *et al.* 1985), the functional significance of the LCR only became clear when it was used to drive the expression of the human globin genes in transgenic mice.

When the regulation of the individual globin genes in the absence of the LCR was studied in transgenic mice, it showed that the γ and β globin genes were expressed in a developmental specific manner, albeit at low levels and dependent on the position of integration in the host genome (Magram *et al.* 1985; Chada *et al.* 1986; Kollias *et al.* 1986; Townes *et al.* 1985). When the β globin gene was studied in the context of the complete LCR region in transgenic mice a completely different pattern was observed, the gene was expressed at levels comparable to that of the endogenous mouse globin genes and it became independent of the position of integration in the mouse genome. In addition, the level of expression was

Phil. Trans. R. Soc. Lond. B (1993) **339,** 183–191
Printed in Great Britain

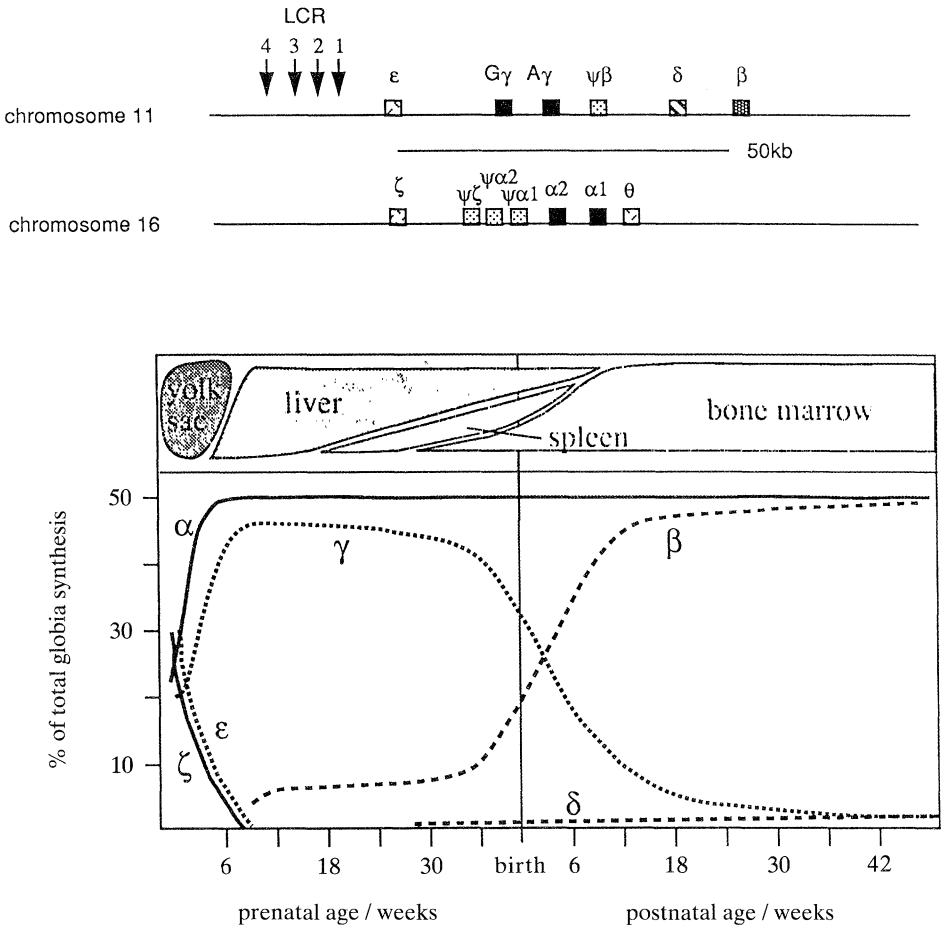


Figure 1. The human α and β loci and the developmental expression of the globin gene expression during human development. Speckled boxes indicate pseudogenes.

dependent on the number of copies of the introduced transgene (Grosveld *et al.* 1987).

3. PROPERTIES OF THE LCR

Position independence and copy number dependence can theoretically be explained by at least two independent mechanisms; either positive activation by the LCR is always achieved and can result in very high levels of expression which obscures small position effects that may still be present in the background, or (and) the region contains elements that insulate it from neighbouring regions, providing a locus border element (LBE) as have been described in Drosophila (Kellum & Schedl 1991). Matrix attachment sites (MAR; Gasser & Laemmli 1986; Jarman & Higgs 1988) or 'A' elements (Stief *et al.* 1989; Bonifer *et al.* 1990) could be LBEs and we initially speculated that these could be part of the LCR in addition to sequences activating sequences (Grosveld *et al.* 1987). However, preliminary experiments indicate that this is not the case and that such a border may be located further upstream (S. Pruzina & F. Grosveld, unpublished data). In addition, small fragments containing any one of the hypersensitive sites can also give rise to copy-number-dependent expression independent of the site of integration albeit at lower levels than the full LCR (Forrester *et al.* 1989; Ryan *et al.* 1989; Talbot *et al.* 1989; Fraser *et al.* 1990). This makes it highly unlikely that insulating sequences form the basis of the effect. It therefore appears that the position independence we observe is (at least in part) due to the fact that the LCR achieves activation of transcription in some dominant fashion perhaps by creating very stable interactions between the LCR and the genes (see below). Consequently positive position effects would only be present as part of the background and only become apparent in situations where the linked gene is suppressed (Dillon & Grosveld 1991; see below for discussion). Interestingly, position effects are not observed when low levels of expression are obtained by the use of inefficient promoters or part of the LCR or mutations in the LCR. This suggests that the stability of the interaction between the LCR and the promoter is not related to the rate of initiation of transcription, except when the promoter is suppressed (Dillon & Grosveld 1991). On the basis of this it may be expected that protein factors exist that are required for the stability of the LCR–promoter complex, in a manner which is separate from the trans-activating function of classical transcription factors. One candidate factor for such a function might be the γ globin

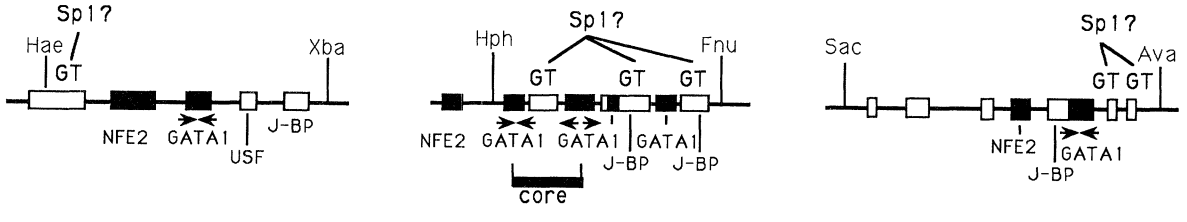
promoter factor that increases the competitive ability of the γ gene relative to the β gene in transfection assays (Jane *et al.* 1992; see below for discussion).

4. STRUCTURE OF THE LCR

The main activity of the LCR is associated with HS2, 3 and 4 (Forrester *et al.* 1989; Ryan *et al.* 1989; Tuan *et al.* 1989; Collis *et al.* 1990; Fraser *et al.* 1990; Fraser *et al.* 1992; Lowrey *et al.* 1992), in agreement with the deletion observed in a Hispanic γβ thalassaemia (Driscoll *et al.* 1989). Each of the sites contains elements capable of activating a linked transgene, independent of the site of integration. A number of erythroid specific and ubiquitously expressed protein binding sites have been mapped to these fragments. Several of specific binding sites are present in all three of the active HS regions (figure 2; Philipsen *et al.* 1990; Talbot *et al.* 1990; Pruzina *et al.* 1991). Very striking are closely spaced binding sites for the erythroid and megakaryocytic specific factor GATA1 (Martin & Orkin 1990; Romeo *et al.* 1990), which has been shown to be essential for erythroid development (Pevny *et al.* 1991). Deletion of GATA-1 binding sites in the promoter prevents erythroid specific induction of the β-globin gene (de Boer *et al.* 1988) and the protein has been shown to have transcriptional activation properties (Martin & Orkin 1990). GATA-1 binds to many erythroid specific genes including all of the globin genes, but the presence of GATA-1 binding sites per se is insufficient to give position independent expression. The immediate 5′ and 3′ flanking regions of the human β globin gene contain at least six GATA-1 binding sites (de Boer *et al.* 1988; Wall *et al.* 1988), but these do not confer integration site independent expression on the β-globin gene (Magram *et al.* 1985; Townes *et al.* 1985; Kollias *et al.* 1986). Our recent experiments with HS3 in transgenic mice indicate that the minimal combination required for position-independent expression is in actual fact not the presence of two closely spaced GATA1 sites, but two GATA1 sites immediately flanking a G rich motif on both sides (Philipsen *et al.* 1992; figure 2). This is in perfect agreement with the binding sites that appear to be occupied *in vivo* (Strauss & Orkin 1992). Interestingly these GATA1 binding sites are completely conserved between the human HS3 and the homologous sequence in the goat (Li *et al.* 1991), which supports our observation on the role of these sites. The G-rich motif in the human and goat HS3 efficiently bind the transcription factor Sp1 (Philipsen *et al.* 1992) *in vitro* suggesting that this factor may play an important role in erythroid specific transcriptional activation. Sp1 would be an attractive candidate because it has been shown to be able to loop DNA (Li *et al.* 1991; Mastrangelo *et al.* 1991; Su *et al.* 1991), a process thought to be central to gene activation in general (for review, see Ptashne (1988)) and to the interactions between the LCR and the globin genes in particular (Hanscombe *et al.* 1991). It should be noted however that neither the human or goat sequence contain the optimal Sp1 binding site (Letovsky & Dynan 1989) and although Sp1 can function through such sites (Li *et al.* 1991), it cannot be excluded that another as yet to be identified factor is important. It is interesting to note that two GATA sites are required for erythroid specific position independent activation which leaves the possibility that a combination of different GATA proteins may be required as part of the activating complex (Yamamoto *et al.* 1990).

A different result is obtained from the detailed functional analysis of HS2. Transient transfection experiments show that classical enhancer activity is associated with this site only (Tuan *et al.* 1989; Ney *et al.* 1990), and not with the others. Dissection of the HS2 showed that a number of proteins are bound to the core fragment (figure 2; Talbot *et al.* 1990). These included the erythroid specific factors NF-E2 (Mignotte *et al.* 1989*a*,*b*) and GATA1 and the ubiquitously expressed factors USF and J-BP (Ellis *et al.* 1993). The NF-E2 binding sites are responsible for the enhancer activity of HS2, but it does not appear to be necessary for position independent globin gene activation (Sorrentino *et al.* 1990; Talbot *et al.* 1990, 1991; Ney *et al.* 1990).

Combinations of deletions of the GATA1 and the ubiquitous factor binding sites of HS2 (USF and J-BP; Talbot & Grosveld 1991; Ellis *et al.* 1993), showed that it was impossible to find any particular binding site combination that was crucial for LCR activity (Ellis *et al.* 1993). However when using these minimal HS2 fragments it became apparent that LCR activity was only obtained when multiple copies of any of the transgene constructs had integrated. This suggests that LCR activity was only restored when these minimal sites could interact with each other, presumably by loop formation. This in turn could explain how the complete LCR might work in vivo, namely by the formation of a loop complex between the LCR elements to form a single larger complex that would



site 2 site 3 site 4

Figure 2. Summary of the protein binding sites HS2, 3 and 4 in the human β globin LCR. Black boxes indicate erythroid specific factor binding, open boxes the binding of ubiquitously expressed factors. The minimal functional core of HS3 is indicated by a thick black line (Philipsen *et al.* 1992).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

be capable of interacting with any of the genes. Such a mechanism would also explain why the genes compete with each other for interaction with the LCR, an observation which is difficult to reconcile with a mechanism involving different genes interacting with different elements of the LCR at the same time (see below).

5. DEVELOPMENTAL REGULATION OF THE β-LIKE GLOBIN GENES IN TRANSGENIC MICE

When the developmental regulation of the individual globin genes in the absence of the LCR was studied in transgenic mice, the ε gene was inactive (Shih *et al.* 1990), but the γ and β globin genes were expressed in a developmental specific manner, albeit at low levels and dependent on the position of integration in the host genome (Magram *et al.* 1985; Townes *et al.* 1985; Chada *et al.* 1986; Kollias *et al.* 1986). When the genes were studied in the context of the complete LCR region in transgenic mice, it was found that the ε gene was expressed at the embryonic stage only (P. Fraser, unpublished data) in agreement with the data published for the ε gene linked to part of the LCR (Shih *et al.* 1990; Raich *et al.* 1990). The γ globin gene like its murine structural homologue the βh1 gene, is expressed in the embryonic yolk sac. However in contrast to βh1 the γ gene is expressed in the early foetal liver and is only silenced after day 16 of development. It is not expressed in the adult (Dillon & Grosveld 1991). Although it has not been directly tested in transgenic mice an individual β gene linked to the LCR is expressed prematurely in differentiating embryonic stem cells, although not at maximal levels. Like the murine β gene, the human β gene is expressed at high levels in the foetal liver and adult (Blom *et al.* 1989; Enver *et al.* 1990; Behringer *et al.* 1990; Grosveld *et al.* 1987). This suggests that a large part but not all of the developmental regulation of the globin genes is specified by the regions immediately flanking the genes. However, it is also clear that the LCR is not developmentally neutral and that it influences the expression pattern of the genes (see below).

The simplest model on the basis of these data (see below) was the proposal that stage-specific suppressors are the very important regulators of the individual globin genes (figure 3; Dillon *et al.* 1991). According to this model the LCR can interact with each of the genes at any stage of development but it interacts preferentially with the ε gene in the yolk sac. When the interaction is negated by stage-specific suppressors in the foetal liver the LCR would interact preferentially with the γ globin genes which in turn are suppressed in the adult. At least two candidate binding sites for suppressor factors have been proposed, one in the upstream region of the promoter of the ε globin gene (Cao *et al.* 1989) and one at the distal CAAT box in the promoter of the γ globin genes (Superti Furga *et al.* 1988; Mantovani *et al.* 1989). However, this model does not explain how the later expressing γ and β genes are kept silent at the early stages of human erythroid development. Evidence from natural mutations and transgenic mouse experiments with combinations of genes suggest that competition of the genes for the LCR plays an important role in this process, in particular for the γ and β genes (see below).

6. COMPETITION OF THE GLOBIN GENES IN TRANSGENIC MICE

Competition between the genes became apparent when combinations of genes were used in transgenic mouse experiments. The premature expression of the β gene when it was linked to the LCR and expressed in transgenic mice, could be abrogated by competition for the LCR with another globin gene, e.g. γ or α (Behringer *et al.* 1990; Enver *et al.* 1990; Hanscombe *et*

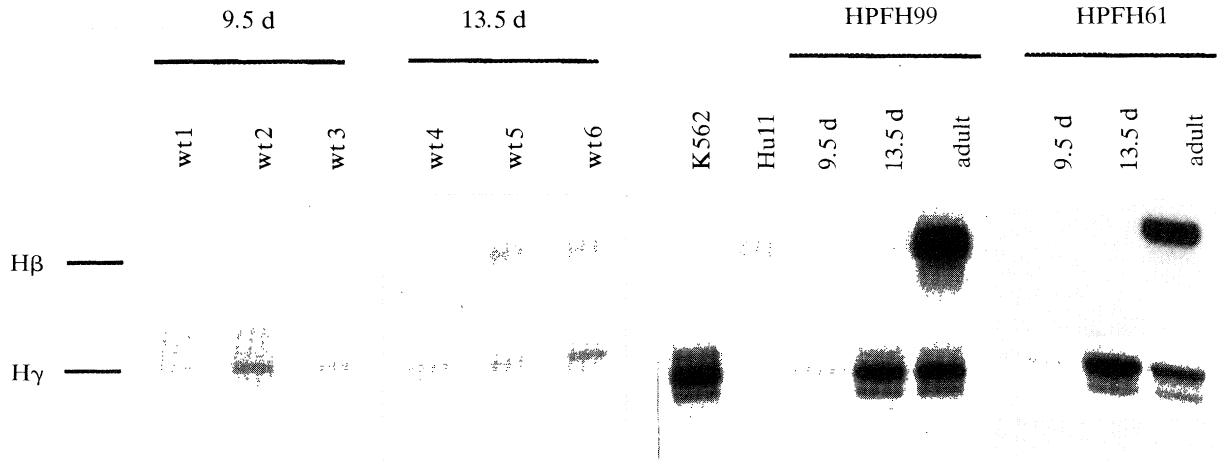


Figure 3. S1 nuclease mapping analysis of minilocus wild-type γβ and Greek HPFH γβ mice during development. RNA was prepared from transgenic 9.5 day yolk sacs, 13.5 day foetal livers and from blood of adult mice. The yolk sacs and foetal livers containing minilocus wtγβ were analysed directly after microinjection whereas yolk sac, foetal liver and adult blood containing the Greek HPFH γβ construct were obtained by breeding lines 61 and 99. Control RNAs were from K562 and Hu11 cells. The signals for human γ- and β-globin RNA are indicated on the left (Hγ, Hβ).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

al. 1991). This competition appears to operate in a polar fashion providing an advantage to a gene that is proximal to the LCR and a disadvantage to a gene that is distal to the LCR (Hanscombe *et al.* 1991). As mentioned above these results were subsequently confirmed by the expression of an HPFH (hereditary persistence of foetal haemoglobin) γ and β gene (Berry *et al.* 1992). One group of HPFH mutations is clustered around the distal CAAT box and appear to result in the loss of factor binding sites (Mantovani *et al.* 1989; Fucharoen *et al.* 1990; Gilman *et al.* 1988; Berry *et al.* 1992), suggesting that this region may contain a binding site for a negative regulator. Elevated γ expression in the Greek and Italian HPFHs is associated with a down regulation of the β gene on the same chromosome (Weatherall & Clegg 1981; Giglioni *et al.* 1984). These observations were confirmed by the expression of a γ globin gene which had a single point mutation engineered in the distal CAAT box to mimick the Greek HPFH. Expression of this gene in combination with a β gene in transgenic mice showed that the single point mutation is responsible for the increased expression of the γ gene in the adult and a decrease in expression of the linked β gene. This decrease is particularly strong in the foetal liver due to the increased competitive ability of the γ gene (figure 3; Berry *et al.* 1992).

Why would the expression of the upstream genes prevent expression of those located downstream but not vice versa? Existing competition models are based on the idea that the high level expression of the genes in the β globin locus is potentiated by direct interaction of each gene with the LCR. There is strong evidence that enhancers work through such interactions (Muller *et al.* 1989; Bickel & Pirotta 1990, and references therein). The polar competition which appears to operate in the locus would be explained if the LCR had an intrinsic preference for interaction with a more proximally located gene. A mechanism by which the LCR searches along the chromosome and interacts with the first gene it encounters would be energy requiring and would need very complex machinery. A second possibility would be the formation of chromatin structures which would bring gene and LCR together but these would require specific and developmentally regulated functions for the spacer regions in the locus. A comparison of the phenotypes of different deletions in the locus (reviewed by Poncz *et al.* (1989)), together with the transgenic mouse data, argues against such a role. An alternative and much simpler possibility is that the determining parameter is the relative frequency of contact between genes and LCR (Hanscombe *et al.* 1991). Where two genes each retain a significant capacity to form stable interactions with the LCR, a difference in their frequency of contact with the LCR would dramatically affect competition between them. During the foetal stage the interaction of the γ genes with the LCR would be much more frequent than that of the β gene and, owing to the action of stage specific factors, would also be stronger. The combination of these two parameters acting together would allow the γ genes to compete out β expression. In the adult stage, although the β gene would now have the stronger interaction with the LCR, its much lower frequency of contact would make it difficult for it to compete out a γ gene which retained a significant capability to form such interactions. It is possible that this effect could be achieved by a very strong β interaction but it seems unlikely that this would have evolved specifically to silence γ expression. The more likely alternative mechanism would be one of promoter mediated silencing of the early genes by stage specific factors and the transgenic mouse data indicate that this is the one which has in fact evolved (figure 4). In the case of an HPFH the γ gene would only partially be suppressed and hence retained a significant ability to interact with the LCR, resulting in γ espression and a lower level of β expression. Such a competition mechanism may have evolved to ensure that the switchover from γ to β globin expression is smooth and that the total output of the β gene locus is kept constant and in balance with the output from the α locus.

The model described above would explain the fact that the order of the genes in the β globin locus is largely conserved among mammals. However, it does not predict that the order of expression during development will follow that of the genes in the locus in all species. For example, a gene located proximal to

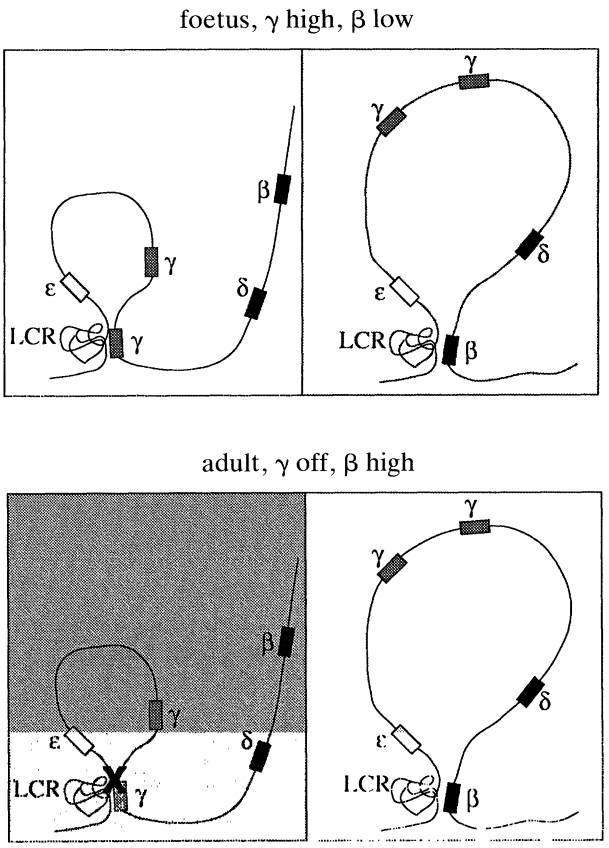


Figure 4. Model of globin gene switching in humans. The genes are indicated by small shaded boxes, high and low levels of expression are indicated by a white or shaded background of each panel. The LCR is symbolized by a large complex and the position of interaction between the LCR and genes are arbitrarily represented. The black cross (adult panel, γ gene) indicates suppression of the γ gene and destabilization of the LCR–γ interaction.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

a LCR could be subject to factor mediated repression early in development and then be activated at a later stage. This activation would bring competition into play and would result in a switch from a distal to a proximal gene.

Our experiments with a complete human β globin locus in transgenic mice indicate that such a mechanism may be operating in the early expression period in the mouse. The more distal γ globin genes are expressed and at a higher level than the more proximal ε gene in the yolk sac, which cannot be explained by a simple order–competition model only (Strouboulis *et al.* 1992). Preliminary experiments indicate that the ε and γ genes compete with each other (Fraser *et al.* 1993). This shows that the ε and γ genes behave differently in humans and transgenic mice and suggests that an important change in at least one transcription factor (an ε activator and/or γ suppressor) must have taken place in the process of foetal recruitement of the γ genes in humans (Strouboulis *et al.* 1992). Another possibility to change the patterns of expression of the genes would be to change the position of the LCR. This would also alter the parameters affecting competition. In the chicken β globin locus, for which competition was first proposed (Choi & Engel 1988), there is evidence that part of the LCR lies between the β and ε genes (Reitman *et al.* 1990).

7. DEVELOPMENTAL SPECIFICITY OF THE LCR

The fact that the LCR influences the developmental expression pattern of the genes by competition suggests that the LCR is unable to interact and stimulate transcription of multiple genes at the same time, even though the LCR contains several hypersensitive elements with LCR activity which are distributed over 15 kb of DNA. This could be explained if the individual HS elements would interact with each other (see above for HS2) to form a larger complex, which in turn would interact with the genes. Such a mechanism suggested that the individual HS sites may have a different developmental specificity and that the genes are competing for a specific element in the larger complex at a given developmental stage. We have therefore recently tested whether the individual hypersensitive regions, which make up the complete LCR, interact differently with the genes and whether each region has a different developmental specificity. To answer this question we introduced each hypersensitive region in combination with a set of human γ and β globin genes into transgenic mice and determined the expression pattern of the different human globin genes during development (figure 5; Fraser *et al.* (1993)). The results showed that the individual sites behave differently with the γ or β globin gene, which has a number of implications for the mechanism of LCR function and showed that the LCR is not developmentally neutral but that it influences the expression pattern of the genes. Most striking was the result that HS3 is the only part of the LCR that is capable of directing the expression of the γ globin gene

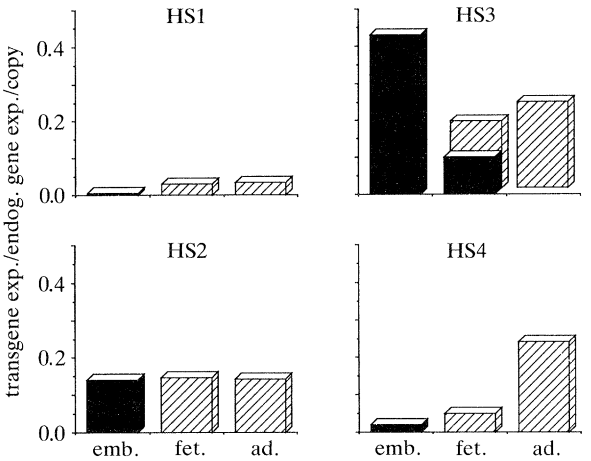


Figure 5. Transgene expression per copy as a percentage of the endogenous mouse genes. Transgene expression levels per copy were calculated by dividing the transgene signal (γ or β) by the endogenous gene signal (βH1 in the embryo; β major in the foetus and adult) and then dividing by 1/2 the transgene copy number (to account for the fact that the mouse has two copies of endogenous genes). Solid columns; γ gene expression. Striped columns; β gene expression.

in the foetal liver. This result suggests that this site is the part of the LCR for which the γ and β genes compete during the foetal stage of development. Since all of the other sites are capable of directing β expression during this period this result also indicates that the LCR elements may interact with each other to form a larger single complex (see also above), that interacts with the genes. Interestingly HS4 is the most active site in adult β expression, suggesting that it forms the most stable interaction with the β gene at that stage (Fraser *et al.* 1993). These combination of results indicate that a switch from HS3 to HS4 takes place at the foetal–adult switch and that the γ promoter is capable of forming stable complexes with HS3 but not with HS4 in the LCR complex. As the Greek HPFH gene is expressed effectively in adult mice and able to compete the β gene it suggests that the single point mutation leads to a change in the binding of a factor which positively changes the ability of γ to interact with HS4 in the adult.

In summary, the data suggest that the LCR may act through loop formation, possibly achieving activation through a looping between the LCR elements to form a larger single complex which interacts with the genes. The developmental expression pattern of the β globin locus is regulated at different levels. Most important is the action of stage specific regulatory factors on the sequences immediately flanking the genes and perhaps the LCR. In addition, the genes influence each other through a competitive mechanism with the LCR. It is clear that this mechanism involves a number of parameters which include the spatial distribution of the genes relative to the LCR and the stability of the interactions between individual promoters and each of the hypersensitive sites, although the relative contribution of each of these is as yet unknown.

This work was supported by the MRC (U.K.).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

REFERENCES

Behringer, R.R., Ryan, T.M., Palmiter, R.D., Brinster, R.L. & Townes, T.M. 1990 Human γ to β globin gene switching in transgenic mice. *Genes Dev.* **4**, 380–389.

Berry, M., Dillon, N. & Grosveld, F. 1992 A single point mutation is the cause of the Greek form of Hereditary Persistance of Foetal Haemoglobin. *Nature, Lond.* **358**, 499–502.

Blom van Assendelft, G., Hanscombe, O., Grosveld, F. & Greaves, D.R. 1989 The β-globin domain control region activates homologous and heterologous promoters in a tissue-specific manner. *Cell* **56**, 969–977.

Bickel, S. & Pirotta, V. 1990 Self association of the *Drosophila* zeste protein is responsible for transvection effects. *EMBO J.* **9**, 2959–2967.

Bonifer, C., Vidal, M., Grosveld, F. & Sippel, A.E. 1990 Tissue specific and position independent expression of the complete gene domain for chicken lysozyme in transgenic mice. *EMBO J.* **9**, 2843–2848.

Cao, S., Gutman, P.D., Dave, H.P.G. & Schechter, A.J. 1989 Identification of a transcriptional silencer in the 5′-flanking region of the human ε-globin gene. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5306–5309.

Chada, K., Magram, J. & Costantini, F. 1986 An embryonic pattern of expression of a human fetal globin gene in transgenic mice. *Nature, Lond.* **319**, 685–689.

Choi, O.-R. & Engel, J.D. 1988 Developmental regulation of human β globin gene switching. *Cell* **55**, 17–26.

Collins, F.S. & Weissman, S.M. 1984 The molecular genetics of human hemoglobin. *Prog. Acid Res. molec. Biol.* **31**, 315–462.

Collis, P., Antoniou, M. & Grosveld, F. 1990 Definition of the minimal requirements within the human beta-globin gene and the dominant control region for high level expression. *EMBO J.* **9**, 233–240.

deBoer, E., Antoniou, M., Mignotte, V., Wall, L. & Grosveld, F. 1988 The human beta-globin promoter; nuclear protein factors and erythroid specific induction of transcription. *EMBO J.* **7**, 4203–4212.

Dillon, N. & Grosveld, F. 1991 Human γ-globin genes silenced independently of other genes in the β-globin locus. *Nature, Lond.* **350**, 252–254.

Driscoll, C., Dobkin, C. & Alter, B. 1989 γδβ-Thalassemia due to a de novo mutation deleting the 5′ β-globin gene activation-region hypersensitive sites. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7470–7474.

Ellis, J., Talbot, D., Pruzina, S. & Grosveld, F. 1993 Synthetic 5′ HS2 construct function as a partially active locu control region. *EMBO J.* (In the press.)

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

Forrester, W., Takegawa, S., Papayannopoulou, T., Stamatoyannopoulos, G. & Groudine, M. 1987 Evidence for a locus activator region. *Nucl. Acids Res.* **15**, 10159–10177.

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

Forrester, W., Epner, E., Driscoll, C., Enver, T., Brice, M., Papayannopoulou, T. & Groudine, M. 1990 A deletion of the human β globin locus activation region causes a major alteration in chromatin structure and replication across the entire β globin locus. *Genes Dev.* **4**, 1637–1649.

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

Fraser, P., Pruzina, S. & Grosveld, F. 1993 Each hypersensitive site of the human beta globin locus control region confers a different developmental pattern of expression to the globin genes. *Genes Dev.* (In the press.)

Fucharoen, S., Shimiza, K. & Fukumaki M. 1990 A novel C-T transition within the distal CCAAT motif of the Gγ globin gene in the Japanese HPFH: implication of factor binding in elevated fetal globin expression. *Nucl. Acids Res.* **18**, 5245.

Giglioni, B., Casini, C., Mantovani, R., Merli, S., Comp, P., Ottolenghi, S., Saglio, G., Camaschella, C. & Mazza, U. 1984 A molecular study of a family with Greek hereditary persistence of fetal hemoglobin and β-thalassaemia. *EMBO J.* **11**, 2641–2645.

Gilman, J., Mishima, N., Wen, X., Stoming, T., Lobel, J. & Huisman, T. 1988 Distal CCAAT box deletion in the Aγ globin gene of two black adolescents with elevated fetal Aγ globin. *Nucl. Acids Res.* **18**, 10635–10642.

Gasser, S.M. & Laemmli, U.K. 1986 Cohabitation of scaffold binding regions with upstream/enhancer elements of three developmentally regulated genes of D. melanogaster. *Cell* **46**, 521–530.

Grosveld, F., Blom van Assendelft, G., Greaves, D. & Kollias, G. 1987 Position-independent high level expression of the human β-globin gene in transgenic mice. *Cell* **51**, 975–985.

Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N. & Grosveld, F. 1991 Importance of globin gene order for correct developmental expression. *Genes Dev.* **5**, 1387–1394.

Jane, S., Paul, P., Vanin, E., Gumucio, D. & Nienhuis, A. 1992 Identification of a stage selector element in the human γ-globin gene promoter that fosters preferential interaction with the 5′ HS2 enhancer when in competition with the β-promoter. *EMBO J.* (In the press.)

Jarman, A.P. & Higgs, D.R. 1988 Nuclear scaffold attachment sites in the human globin gene complexes. *EMBO J.* **7**, 3337–3344.

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

Kioussis, D., Vanin, E., deLange, T., Flavell, R.A. & Grosveld, F. 1983 β-globin gene inactivation by DNA translocation in γ-thalassaemia. *Nature, Lond.* **306**, 662–666.

Kollias, G., Wrighton, N., Hurst, J. & Grosveld, F. 1986 Regulated expression of human ^{A}γ-, β-, and hybrid γβ-globin genes in transgenic mice: manipulation of the developmental expression patterns. *Cell* **46**, 89–94.

Letovsky, J. & Dynan, W.S. 1989 Measurement of the binding of transcription factor Sp1 to a single GC box recognition sequence. *Nucl. Acids Res.* **17**, 2639–2653.

Li, Q., Zhou, B., Powers, P., Enver, T. & Stamatoyannopoulos, G. 1991 Beta globin locus activating regions: conservation of organization, structure and function. *Proc. natn. Acad. Sci. USA* **87**, 8207–8211.

Li, R., Knight, J., Jackson, S., Tjian, R. & Botchan, M. 1991 Direct interaction between Sp1 and the BPV enhancer E2 protein mediates synergistic activation of transcription. *Cell* **65**, 493–505.

Lowrey, H., Bodine, D. & Nienhuis, A. 1992 Mechanism of DNase I hypersensitivity site formation within the human globin locus control region. *Proc. natn. Acad. Sci. U.S.A.* **89**, 1143–1147.

Magram, J., Chada, K. & Costantini, F. 1985 Developmental regulation of a cloned adult β-globin gene in transgenic mice. *Nature, Lond.* **315**, 338–340.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Mantovani, R., Superti–Fuga, G., Gilman, J. & Ottolenghi, S. 1989 The deletion of the distal CCAAT box region of the Aγ globin gene in black HPFH abolishes the binding of the erythroid specific protein NFE3 and of the CCAAT displacement protein. *Nucl. Acids. Res.* **17**, 6681–6691.

Martin, D.I. & Orkin, S.H. 1990 Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. *Genes Dev.* **4**, 1886–1898.

Mastrangelo, I., Courey, A., Wall, J., Jackson, S. & Hough, V. 1991 DNA looping and Sp1 multimer links: A mechanism for transcriptional synergism and enhancement. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5670–5674.

Mignotte, V., Eleouet, J.F., Raich, N. & Romeo, P.H. 1989 Cis- and trans-acting elements involved in the regulation of the erythroid promoter of the human porphobilinogen deaminase gene. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6548–52.

Mignotte, V., Wall, L., deBoer, E., Grosveld, F. & Romeo, P.H. 1989 Two tissue-specific factors bind the erythroid promoter of the human porphobilinogen deaminase gene. *Nucl. Acids Res.* **17**, 37–54.

Muller, H., Sogo, J. & Schaffner, W. 1989 An enhancer stimulates transcription in trans when attached to the promoter via a protein bridge. *Cell* **58**, 767–777.

Ney, P.A., Sorrentino, B.P., McDonagh, K.T. & Nienhuis, A.W. 1990 Tandem AP-1-binding sites within the human beta-globin dominant control region function as an inducible enhancer in erythroid cells. *Genes Dev.* **4**, 993–1006.

Ney, P.A., Sorrentino, B.P., Lowrey, C.H. & Nienhuis, A.W. 1990 Inducibility of the HS II enhancer depends on binding of an erythroid specific nuclear protein. *Nucl. Acids Res.* **18**, 6011–6017.

Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.F., D'Agati, V., Orkin, S.H. & Costantini, F. 1991 Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature, Lond.* **349**, 257–260.

Philipsen, S., Talbot, D., Fraser, P. & Grosveld, F. 1990 The beta-globin dominant control region: hypersensitive site 2. *EMBO J.* **9**, 2159–67.

Philipsen, S., Pruzina, S. & Grosveld, F. 1992 The minimal requirements for activity in transgenic mice of hypersensitive site 3 of the β globin Locus Control Region. (Submitted.)

Poncz, M., Henthorn, P., Stoeckert, C. & Surrey, S. 1988 Globin gene expression in hereditary persistence of fetal hemoglobin and γβ thalassaemia. In *Oxford surveys on eukaryotic genes* (ed. N. McLean), vol. 5, pp. 163–203. Oxford University Press.

Pruzina, S., Hanscombe, O., Whyatt, D., Grosveld, F. & Philipsen, S. 1991 Hypersensitive site 4 of the human beta globin locus control region. *Nucl. Acids Res.* **19**, 1413–1419.

Ptashne, M. 1988 How eukaryotic transcriptional activators work. *Nature, Lond.* **335**, 683–689.

Raich, N., Enver, T., Nakamoto, B., Josephson, B., Papayannopoulou, T. & Stamatoyannopoulos, G. 1990 Autonomous developmental control of human embryonic globin switching in transgenic mice. *Science, Wash.* **250**, 1147–1149.

Reitman, M., Lee, E., Westphal, H. & Felsenfeld, G. 1990 Site independent expression of the chicken β^A globin gene in transgenic mice. *Nature, Lond.* **348**, 749–752.

Romeo, P.H., Prandini, M.H., Joulin, V., Mignotte, V., Prenant, M., Vainchenker, W., Marguerie, G. & Uzan, G. 1990 Megakaryocytic and erythrocytic lineages share specific transcription factors. *Nature, Lond.* **344**, 447–449.

Ryan, T.M., Behringer, R.R., Townes, T.M., Palmiter, R.D. & Brinster, R.L. 1989 High-level erythroid expression of human alpha-globin genes in transgenic mice. *Proc. natn. Acad. Sci. U.S.A.* **86**, 37–41.

Shih, D., Wall, R. & Shapiro, S. 1990 Developmentally regulated and erythroid-specific expression of the human embryonic β globin gene in transgenic mice. *Nucl. Acids Res.* **18**, 5465–5472.

Sorrentino, B., Ney, P., Bodine, D. & Nienhius, A.W. 1990 A 46 base pair enhancer sequence within the locus activating region is required for induced expression of the gamma-globin gene during erythroid differentiation. *Nucl. Acids Res.* **18**, 2721–2731.

Stief, A., Winter, D.M., Stratling, W.H. & Sippel, A.E. 1989 A nuclear DNA attachment element mediates elevated and position-independent gene activity. *Nature, Lond.* **341**, 343–345.

Strauss, E. & Orkin, S. 1992 In vivo protein-DNA interactions at hypersensitive site 3 of the human beta globin locus control region. *Proc. natn. Acad. Sci. U.S.A.* **89**, 5809–5813.

Stroubuolis, J., Dillon, N. & Grosveld, F. 1992 Developmental regulation of a complete 70 kb human β-globin locus in transgenic mice. *Genes Dev.* (In the press.)

Su, W., Jackson, S., Tjian, R. & Echols, H. 1991 DNA looping between sites for transcriptional activation: self association of DNA-bound Sp1. *Genes Dev.* **5**, 820–826.

Superti-Furga, G., Berberis, A., Schaffner, G. & Busslinger, M. 1988 The −117 mutation in Greek HPFH affects the binding of three nuclear factors to the CCAAT region of the γ-globin gene. *EMBO J.* **7**, 3089–3107.

Talbot, D., Collis, P., Antoniou, M., Vidal, M., Grosveld, F. & Greaves, D.R. 1989 A dominant control region from the human beta-globin locus conferring integration site-independent gene expression. *Nature, Lond.* **338**, 352–355.

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

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

Taramelli, R., Kioussis, D., Vanin, E., Bartram, K., Groffen, J., Hurst, J. & Grosveld, F.G. 1986 Gamma delta beta-thalassaemias 1 and 2 are the result of a 100 kbp deletion in the human beta-globin cluster. *Nucl. Acids Res.* **14**, 7017–7029.

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

Tuan, D., Solomon, W., Li, Q. & London, I. 1985 The "β-like globin" gene domain in human erythroid cells. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6384–6388.

Tuan, D., Solomon, W., London, I. & Lee, D. 1989 An erythroid specific, developmental stage independent enhancer far upstream of the human β-like globin genes. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2554–2558.

Van der Ploeg, L., Konings, A., Oort, M., Roos, D., Bernini, L. & Flavell, R. 1992 γβ thalassaemia: deletion of the γ and δ genes influences β globin expression in man. *Cell* **38**, 265–273.

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

Weatherall, D.J. & Clegg, J.B. 1981 *The thalassaemia syndromes*. Boston: Blackwell Scientific Publications.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Wright, S., deBoer, E., Rosenthal, A., Flavell, R.A. & Grosveld, F.G. 1984 DNA sequences required for regulated expression of the β-globin genes in murine erythroleukaemia cells. *Phil. Trans. R. Soc. Lond.* B **307**, 271–282.

Yamamoto, M., Ko, L.J., Leonard, M.W., Beug, H., Orkin, S.H. & Engel, J.D. 1990 Activity and tissue-specific expression of the transcription factor NF-E1 multigene family. *Genes Dev.* **4**, 1650–1662.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

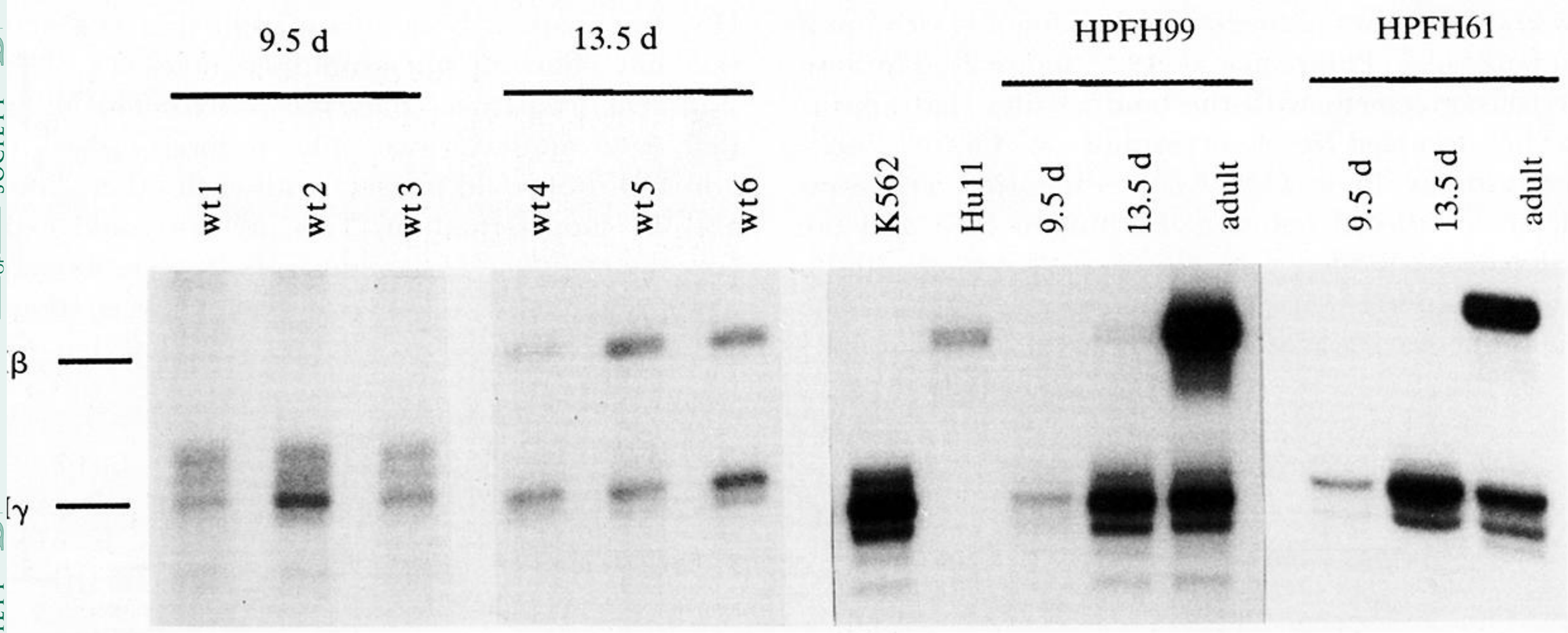


Figure 3. S1 nuclease mapping analysis of minilocus wild-type γβ and Greek HPFH γβ mice during development. RNA was prepared from transgenic 9.5 day yolk sacs, 13.5 day foetal livers and from blood of adult mice. The yolk sacs and foetal livers containing minilocus wtγβ were analysed directly after microinjection whereas yolk sac, foetal liver and adult blood containing the Greek HPFH γβ construct were obtained by breeding lines 61 and 99. Control RNAs were from K562 and Hu11 cells. The signals for human γ- and β-globin RNA are indicated on the left (Hγ, Hβ).