Tissue-specific Histone Modification and Transcription Factor Binding in α Globin Gene Expression.

Short title: Epigenetic regulation of α globin genes

Marco De Gobbi, Eduardo Anguita, Jim Hughes, Jacqueline A Sloane-Stanley, Jacqueline A Sharpe, Christoph M Koch, Ian Dunham, Richard J Gibbons, William G Wood, Douglas R Higgs

1MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, Oxford University, Oxford, OX3 9DS, UK
2The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SA, UK
*Current address: Department of Hematology, Hospital Clínico San Carlos, Universidad Complutense, Madrid, Spain

Corresponding author:
Doug Higgs,
MRC Molecular Haematology Unit,
Weatherall Institute of Molecular Medicine,
Oxford University, Oxford, OX3 9DS, UK
doug.higgs@imm.ox.ac.uk
Phone: +44 1865 222393
Fax: +44 1865 222500
Abstract

To address the mechanism by which the human globin genes are activated during erythropoiesis, we have used a tiled microarray to analyse the pattern of transcription factor binding and associated histone modifications across the telomeric region of human chromosome 16 in primary erythroid and non-erythroid cells. This 220kb region includes the α globin genes and nine widely expressed genes flanking the α globin locus. This unbiased, comprehensive analysis of transcription factor binding and histone modifications (acetylation and methylation) described here not only identified all known cis-acting regulatory elements in the human α globin cluster but also demonstrated that there are no additional erythroid-specific regulatory elements in the 220kb region tested. In addition, the pattern of histone modification distinguished promoter elements from potential enhancer elements across this region. Finally, comparison of the human and mouse orthologous regions in a unique mouse model, with both regions co-expressed in the same animal, showed significant differences that may explain how these two clusters are regulated differently in vivo.
Introduction

Globin gene expression plays a central role in normal erythropoiesis and abnormalities of globin synthesis underlie the clinically important thalassaemia syndromes. A long-term goal in the treatment of thalassaemia has been to identify ways in which to manipulate gene expression to redress the imbalance in α and β globin chain synthesis. Progress in this area would be enhanced by understanding in detail how these genes are normally regulated. To address this, it is important to identify all the key cis-acting sequences controlling α and β globin expression, the transcription factors that bind them and the accompanying modifications of chromatin at different stages of differentiation.

To identify and characterise the cis-acting elements controlling α globin expression we have previously used a variety of approaches, including comparative genomics, a wide range of chromatin analyses, transgenic experiments, and the analysis of expression in inter-specific hybrids. Upstream of the four α-like globin promoters (ζ, αD, α2 and α1) lie four multi-species conserved sequences (MCSs) associated with erythroid-specific DNAse1 hypersensitive sites (HS), referred to, in human, as HS-48, HS-40, HS-33 and HS-10 (Figure 1A). The globin genes, the upstream MCSs and several widely expressed flanking genes all lie within a region of conserved synteny spanning 130kb of the telomeric region of human chromosome 16 (16p13.3) (Figure 1A). Functional analyses suggest that HS-40 is the most important of the MCSs.

Recently, to develop a relevant experimental system for studying human α globin expression and to evaluate the roles of individual elements we established a humanised mouse model in which the 117kb region containing the human α-like globin genes and all of their known regulatory elements (the α globin regulatory domain) completely replaces the corresponding, orthologous region of the mouse genome (Figure 1A). Although, in this model, the human α genes are
expressed in an appropriate tissue- and development stage-specific manner, their level of expression is suboptimal (40% of that of the endogenous genes). This could either be because one or more critical elements in the human cluster have been overlooked in previous analyses or because changes in the structure or recognition of sequences by the key transcription factors have altered during evolution such that binding or stability of mouse transcription factors on human sequences are suboptimal. These observations prompted us to re-assess the full extent of the human \( \alpha \) globin regulatory domain, and to compare, in detail, the binding of transcription factors and epigenetic modifications in the human and mouse \( \alpha \) globin clusters during erythropoiesis.

Using a tiled microarray, we used ChIP-chip technology (see “Materials and methods”), validated by ChIP-qPCR, to assess the binding of key erythroid transcription factors and characterise the associated histone modifications at defined stages of erythropoiesis. Applying this newly established ChIP-chip technology allowed us to obtain a comprehensive, unbiased analysis of the binding and activation of potential \textit{cis}-acting sequences in a contiguous 220kb region of chromosome 16, extending 50kb on either side of the previously defined \( \alpha \) globin regulatory domain. This demonstrated that, within this region, no sequences other than the known upstream regulatory elements and human \( \alpha \) globin promoters bound erythroid transcription factor complexes or became activated (as judged by relevant histone modifications) in erythroid cells. In basophilic erythroblasts, when the \( \alpha \) globin genes are fully active, a domain of histone H4 acetylation extending through 80kb became established. Thus, together, these findings re-define the limits of the human \( \alpha \) globin regulatory domain.

The humanised mouse model allowed, for the first time, a direct comparison of transcription factor binding and chromatin modification, in heterozygotes, at the human and mouse loci in nuclei with identical transcriptional and epigenetic programmes. Using this model, and by comparing ChIP data from primary
human and mouse erythroid cells,\textsuperscript{4} we identified significant differences in transcription factor binding and histone modification at the human and mouse clusters which may explain differences in their regulation.

Finally the use of the tiled microarray allowed us to compare histone modifications across the entire $\alpha$ globin regulatory domain with those encompassing the nine widely expressed flanking genes, whose epigenetic profile had not previously been studied. These epigenetic marks, in combination, distinguish promoter elements from other classes of regulatory element (e.g. enhancers).
Materials and Methods

Cells and cultures
K562 were grown in RPMI 1640 medium (Sigma), 50 U/ml penicillin G (GIBCO), 50 µg/ml streptomycin (GIBCO), 2 mM L-glutamine (GIBCO) supplemented with 10% (vol/vol) foetal calf serum (PAA Laboratories GmbH, Linz, Austria). Primary human erythroblasts were obtained from peripheral blood mononuclear cells (PBMCs) collected from blood donors and expanded in a two-phase system as previously described. Primary human T lymphocytes were obtained from PBMCs by culture for 3 days in RPMI/20% fetal calf serum in the presence of 2% phytohaemagglutinin M (GIBCO) and 20 U/ml interleukin-2. EBV-transformed lymphoblastoid cell lines were derived from normal subjects. Mature primary murine erythroblasts were obtained from phenylhydrazine-treated adult “humanized” mice, generated as described.

Antibodies
The antibodies used were: anti-diacetylated histone H3 (06-599), anti-tetra-acetylated histone H4 (06-866), anti mono-, di- and tri-methyl Lys4 histone H3 (07-436, 07-030, 07-473) from Upstate; RNA Pol II N20, RNA Pol II H224, NF-E2 C19, NF-E2 p18 C16, GATA2 H-116, GATA-1 C20, GATA-1 M20, E2A.E12 V18 from Santa Cruz Biotechnology; SCL, Ldb-1 and LMO2 (kindly provided by C. Porcher)

Chromatin immunoprecipitation (ChIP) assay
ChIPs were performed according to the Upstate ChIP protocol, as described previously with minor modifications. Cells (1.5 x 10⁷ per experiment) were fixed with 0.4% formaldehyde for 10 minutes at room temperature and chromatin was sonicated to a size of less than 500 bp. Immunoprecipitations were performed, after an overnight incubation with the appropriate antibody, with protein A agarose (Upstate), or with protein G agarose (Roche) when using GATA1 antibody. A sample containing no antibody was used as a negative control.
Immunoprecipitated DNA was analysed by quantitative real time PCR (qPCR) (ABI Prism 7000 Sequence Detection System). Primers and 5’FAM-3’TAMRA Taqman probes, for selected sequences of the human α globin locus (available on request), were designed by Primer Express software. All primers and probes were validated over a serial dilution of genomic DNA. For a given target sequence, the amount of product precipitated by a specific antibody was determined relative to the amount of non immunoprecipitated (input) DNA and these results were normalized to a control sequence in the 18S ribosomal RNA gene.

ChIP-chip analysis
The custom α globin tiling path microarray has been described previously.15 450 ng of ChIP DNA, obtained for each antibody from two independent IP reactions, and 450 ng of input DNA were labelled with Cy3-dCTP (Amersham) and Cy5-dCTP (Amersham), respectively, with Bioprime Labelling system (Invitrogen). Labelled DNA was purified through Amersham G50 columns. ChIP-Cy3 and input-Cy5 DNAs were combined, precipitated in ethanol with human Cot-1 DNA (Invitrogen) and yeast t-RNA (Invitrogen) and re-suspended in hybridization buffer (50% (v/v) formamide, 2xSSC, 10mM Tris (ph7.4), 5% (w/v) dextran sulfate, 0.1% (v/v) Tween 20). The arrays were hybridized with the hybridization solution containing the labelled DNAs in a HS 400 Pro Hybridization Station (Tecan) for 45 hours. The arrays were then scanned in a Scan Array Gx Plus scanner (Perkin Elmer) and the spot intensities quantified using Scan Array Express Version 3.0 (Perkin Elmer) with background subtraction. The ratio of the background corrected ChIP signal divided by the background corrected input signal (both globally normalized) were used for the analysis. The data were then plotted and visualized on a customized genome database (GBrowse) designed by the Computational Biology Research Group (Oxford University). Figures 1, 3 and 5 were generated by the GBrowse.
Results

To study chromatin modifications and transcription factor binding throughout the terminal region of chromosome 16, we analysed primary human erythroblasts together with the human erythroid cell line K562. Primary human T lymphocytes or EBV transformed lymphoblastoid cell lines were studied as examples of differentiated, non-erythroid cells. Erythroid progenitors were derived from peripheral blood mononuclear cells (PBMNs) cultured through a two-phase liquid culture system. It has been previously shown\textsuperscript{11,12} that circulating erythroid progenitors differentiate into CFU-E and terminally differentiate to erythrocytes after the addition of erythropoietin. Cell samples were collected on days 5, 8, 11 and 15 of the second phase of culture. Although heterogeneous cell populations are present at any particular stage, clear differences in maturation are distinguishable (Figure 2).

The pattern of histone acetylation across the terminal 220kb of human chromosome 16

Using a ChIP-chip approach, the pattern of histone H3 and H4 acetylation (H3ac and H4ac) was determined across a 220 kb, telomeric region of human chromosome 16, including the putative $\alpha$ globin regulatory domain. For each modification we performed two biological ChIP-chip replicates using primary human basophilic erythroblasts (obtained at day 8 of culture, Figure 2B) and T lymphocytes. These duplicate analyses provided very similar profiles (data available for comparison at: http://www.imm.ox.ac.uk/groups/mrc_molhaem/home_pages/Higgs).

By analysing chromatin from lymphocytes (Figure 1B) we detected relatively small peaks of H3 and H4 acetylation at the CpG islands associated with the promoters of the widely expressed genes that flank the $\alpha$ globin cluster. As previously noted,\textsuperscript{3} there were very low levels of H3ac and H4ac throughout the inactive $\alpha$ globin regulatory domain. In primary adult erythroblasts (Figure 1B),
H3ac was enriched at the CpG islands associated with the promoters of the expressed α globin genes (αD, α2 and α1), and to a lesser extent at the ζ and θ globin genes, which are also associated with CpG islands. Low levels of H3ac were also seen at all upstream MCS elements (HS-48, HS-40, HS-33 and HS-10). By contrast, H4ac appeared more uniformly enriched at both the upstream elements and the active globin gene promoters, in erythroid cells.

To validate these ChIP-chip data and to extend the analysis throughout erythroid differentiation, we carried out independent, duplicate ChIP-qPCR analyses in primary human erythroblasts at each stage of maturation, in T lymphocytes, and in the K562 cell line. The levels of H4ac and H3ac changed relatively little during erythroid differentiation (Supplemental Figure S1) but appeared to reach a maximum in basophilic erythroblasts (day 8, Figure 2B) and subsequently decrease. The lower levels of enrichment in the earliest stages of erythropoiesis could reflect the greater cellular heterogeneity in these populations. The pattern of acetylation observed in K562 (Supplemental Figure S1) was similar to that seen in primary erythroblasts, although an additional peak of acetylation was seen around the embryonic ζ globin gene, which is expressed in these cells. Moreover, in K562 the level of histone acetylation detected at HS-10 was greater than in primary erythroblasts, suggesting a potential role for this element in the regulation of embryonic gene expression.

In summary, the domain of histone acetylation that appears in human erythroid cells extends from coordinates 92000 to 172000 and peaks of acetylation were seen only at previously characterised erythroid MCS elements and promoters. This comprehensive and extensive analysis thus provides important confirmation of the previously defined limits of the human α globin regulatory domain and established that there are no additional elements that may have been overlooked in the 220kb region studied.
The pattern of transcription factor and RNA Polymerase II binding across the terminal 220kb of human chromosome 16

We next determined whether the activated promoters and MCS elements in the human \( \alpha \) globin cluster bound the same or similar repertoires of erythroid transcription factors \textit{in vivo} as previously described for the mouse \( \alpha \) globin cluster. We therefore performed ChIP-chip experiments for GATA1, SCL and NF-E2 in day 8 basophilic erythroblasts (Figure 3) and ChIP-qPCR analysis for GATA2. No GATA2 binding was detected (data not shown), reflecting the fact that our analysis focussed on the later steps of erythroid differentiation when GATA1 expression has largely replaced GATA2. As in the mouse cluster, the four upstream MCS elements (HS-48, HS-40, HS-33 and HS–10) all bind GATA1, SCL (Figure 3) and the entire pentameric erythroid complex (data not shown) which includes GATA1, SCL, E2A, Lim-only 2 (LMO2) and LIM domain binding protein 1 (Ldb1). This complex is thought to play an important role in the positive regulation of erythroid-specific genes.\textsuperscript{16-18} In addition, enrichment of both p45 and p18 NF-E2 subunits were observed at HS-40, and to a lesser extent, at HS-48. However, unlike the mouse \( \alpha \) globin locus, no GATA1 binding was detected at the human \( \alpha \) globin promoters. Furthermore, consistent with the lack of any DNase1 HS (DRH and MDG unpublished data) and lack of histone modification (Figure 1B), no transcription factor binding (GATA1, SCL and NF-E2) was found in the orthologous region corresponding to the mouse HS-12 element (hoHS-12). This ~250bp region of the human genome (coordinates Human build HG18 chr16:125635-125878) aligned maximally with the region corresponding to mouse HS-12, as defined by a multi-species aligment,\textsuperscript{2} and is 53% homologous to the mouse region. However, unlike other MCS elements it contains no highly conserved transcription factor binding sites and, in particular, no consensus GATA1 binding site. No further erythroid transcription factor binding was identified across the 220kb region analysed, confirming that no additional erythroid-specific regulatory elements were present in this region.
We also examined chromatin structure in the heterozygote humanised mouse model. In this model, the human and mouse \( \alpha \) globin loci lie in the same nucleus and share an identical murine transcription factor environment. ChIP-qPCR analysis was performed in mature erythroid cells (Ter119+) isolated from the spleen of a phenylhydrazine-treated mouse. Using H4ac and H3K4me1 (see “Analysis of histone methylation”) antibodies, it was shown that unlike hoHS-12, which had only a basal level of histone acetylation and methylation (as judged by comparison with a non-regulatory sequence, MPG in Figure 4), HS-12 in mouse was modified as active chromatin. Consistent with this observation, analysis of transcription factor binding showed that whereas GATA1 and NF-E2 bound the HS-12 element in the mouse chromosome, no binding was seen at the corresponding region on the humanised chromosome (Figure 4). This shows that, even in the mouse background, there is no “cryptic” binding site at hoHS-12. We then tested the effect of the mouse transcription factor environment on GATA1 recruitment at human \( \alpha \) globin promoters. As seen in primary human erythroblasts (Supplemental Figure S2), no GATA1 binding was detected at the human \( \alpha \) globin promoters in this model (Figure 4).

Next, we mapped RNA Polymerase II (PolII) binding throughout the terminal region of human chromosome 16 (Figure 3). ChIP-chip analysis using anti-PolII antibody showed the binding of PolII to the \( \alpha \) globin promoters and to the regulatory elements HS-48 and HS-40, consistent with previous data from the mouse locus demonstrating the recruitment of PolII at these upstream elements during terminal erythroid differentiation. Although no binding was seen at HS-33 and HS-10 in the ChIP on chip analysis, these elements do bind PolII as judged by qPCR (Supplemental Figure S2).

Using ChIP-qPCR, we characterised the temporal association of transcription factors and PolII within the locus during erythroblast maturation (Supplemental Figure S2). The maximal enrichment of GATA1, SCL and NF-E2 occurred in the intermediate stages of maturation (basophilic and polychromatic erythroblasts).
At the later stages of erythropoiesis, GATA1 and SCL were no longer binding, whereas NF-E2 was still detectable at very low levels. An identical pattern of GATA1 and SCL complex (Ldb1, LMO2 and E2A) binding was found in K562 cells (Supplemental Figure S3). Finally, PolII was shown to be bound at the α globin promoter in human erythroid cells from the first recognised stages of erythropoiesis and had a maximum recruitment after 8-11 days of second phase culture when a low enrichment was found at all the remote conserved elements (HS-48, HS-40, HS-33 and HS-10).

**Analysis of histone methylation**

Recent evidence suggests that *cis*-acting elements may not only be identified but may also be classified on the basis of the associated histone modifications, giving some insight into the role of individual elements. In fact, ChIP-chip analysis of 1% of human genome\(^{19}\) showed that whereas most promoters of active genes are enriched in H3K4me3, enhancers appear to be marked by H3K4me1. Therefore, in addition to analysing the patterns of histone acetylation, we also investigated the patterns of H3K4 mono-, di-, and tri-methylation across the 220kb region of human chromosome 16 (Figure 5).

In erythroid cells, the H3K4me1 modification was found at the upstream MCS elements and equally enriched at the promoters of the globin genes. No significant enrichment of H3K4me1 was seen at the upstream MCSs in non-erythroid cells, whereas it was slightly enriched at all the promoters (globin and non-globin). H3K4me2 modified histones were clearly present at all promoters but not at upstream regulatory elements in non-erythroid cells. In erythroid cells H3K4me2 was readily detectable at the upstream elements and was further enriched at the globin promoters. H3K4me3 was enriched at the promoters of non-globin genes in non-erythroid cells and became, in comparison with the surrounding promoters, significantly enriched at the globin promoters only in erythroid cells. The level of H3K4 methylation changed relatively little during
erythroid differentiation (data not shown). Therefore, consistent with a global analysis of these modifications,\(^9\) H3K4me1 appears to be equally enriched at enhancer and promoter regions whereas H3K4me3 is more highly enriched at active promoters than enhancers in cells where these genes are expressed.
Discussion

Superficially, the transcriptional and epigenetic programs controlling basic biological processes (e.g., erythropoiesis) appear very similar when comparing one mammalian species with another and the proteins involved are often conserved. When considering hemopoiesis, comparisons are also frequently extended and drawn between mammals, amphibians and fish. However, there is increasing evidence that the detailed mechanisms by which each species achieves fully regulated expression of a given gene in the context of its own transcriptional/epigenetic program may be quite different. In this case, simple extrapolation of mechanisms established in one species might be misleading when considering another.

Here we have confronted this general issue by asking if the human and mouse $\alpha$ globin clusters are regulated in the same or different ways. To fully address this point we first needed to be sure that we have identified all of the sequences required for fully regulated expression of the human $\alpha$ cluster in the context of its own transcriptional/epigenetic program. Here we used a newly developed tiling array to comprehensively analyse the patterns of transcription factor binding and histone modifications across a region of 220kb of human chromosome 16 (extending 50kb either side of the currently defined domain). Based on this study we can conclude that there appear to be no additional erythroid cis-elements in this region although we cannot formally rule out even more distant regulatory elements. Therefore, it seems most likely that the previously reported, sub-optimal expression of the human $\alpha$ gene cluster (containing all known regulatory elements) in the mouse results from altered interactions, deriving from species-specific differences between the mouse transcription factors/co-factors and the human cis-acting elements. In this case, fully equivalent expression of the wild type, human $\alpha$ globin cluster in a mouse transcription/epigenetic environment may not be achievable.
Clearly evolution has ensured that each cluster is expressed optimally within its own environment. But has this process “tinkered” with the same set of orthologous cis-elements, have new elements been recruited, and have others become redundant? Evidence from other systems would suggest that all of these processes may be in operation.\textsuperscript{22} Using sequence comparisons between erythroid cis-elements (marked by erythroid-specific Dnasel HSs) in the human and mouse β globin clusters there are clearly five conserved orthologous sequences (excluding hoHS-12 Table 1). However, these conserved regulatory elements appear to play different roles in each cluster. For example, removal of HS-40 alone from the human β globin cluster leads to a severe reduction (<5%) in PolII recruitment and β gene expression, whereas removal of the orthologous element (HS-26) in mouse causes only a moderate (~50%) decrease in expression.\textsuperscript{23} It seems that another element (or elements) must normally contribute significant levels of β globin expression in the mouse. A comparison of transcription factor binding in orthologous, conserved non-coding sequences in mouse and human is shown in Table 1. This highlights two significant differences. First, whereas the mouse β globin promoter includes a consensus GATA sequence that is bound by GATA1 (and its co-factor ZBP, Vernimmen unpublished data) \textit{in vivo}, the human β globin promoter does not. In fact, the β globin promoter, when compared in multiple, diverse mammalian species does not contain an evolutionarily conserved GATA site. Second, hoHS-12 is not marked by Dnasel HS and contains no conserved transcription factor binding sites corresponding to those at the mouse cis-element HS-12 (Figure 1A). In mouse erythroid cells, this element (like HS-40 and HS-48 in human) binds both the pentameric erythroid complex (GATA-1, SCL, E2A, LMO2 and Ldb-1) and NF-E2 and becomes activated (as judged by histone acetylation) during erythropoiesis. No corresponding transcription factor binding or chromatin activation was seen in the human, even in a mouse erythroid cell background (Figure 4). These observations show that the human and mouse clusters achieve fully regulated expression in somewhat different ways. Not only do the orthologous sequences play different roles in these two species but an additional,
species-specific element (HS-12) appears to have been recruited in the mouse cluster. Nevertheless, further studies will be required to define the precise role of this new element in vivo.

Knowing which transcription factors bind each conserved sequence may help determine the role each sequence plays in regulating α globin expression. In human, natural deletions associated with α thalassaemia (which all delete HS-48 and HS-40) show that HS-33 and HS-10 alone are not able to drive α globin expression whereas either HS-40, with or without HS-48, is essential. This might be related to the fact that HS-33 and HS-10 bind only the GATA1/SCL complex whereas HS-48 and HS-40, bind both this complex and NF-E2 (Table 1). Previous experiments on the β globin LCR elements have suggested that GATA-1 may be required to recruit PolII to such elements, but PolII is then transferred to the promoter in an NF-E2 dependent manner. Thus when HS-40 and HS-48 are deleted, PolII might be recruited to the remaining upstream elements (HS-33 and HS-10) but not transferred to the α globin promoters. Although the mechanism of transfer is unknown it seems most likely to involve physical interaction between the upstream elements and the promoters. It is interesting to note that the additional recruited element in mouse (HS-12), which, together with HS-26, may subsume the role of HS-40 in human, also binds NF-E2.

The different patterns of histone modification found at the upstream elements and the α globin promoters may add to the previously proposed model describing how the upstream elements enhance α globin transcription. Here, and elsewhere, we have shown that acetylation of histone H4 is equally strong at the upstream elements and the promoters in erythroid cells. This could be explained by the recruitment of histone acetylases by specific and general DNA binding transcription factors at all cis-acting elements. By contrast, acetylation of histone H3 is much less prominent at the upstream elements than at the promoters. Interestingly, monomethylation of histone H3K4 (catalysed by SET7/9 histone
methyltransferase\textsuperscript{28,29}) is equally enriched at the upstream elements and the promoters in erythroid cells. H3K4me2 (catalysed by Set1-like family of histone methyltransferases) is more prominent at the promoters, and H3K4me3 (also catalysed by Set1-like complexes\textsuperscript{30-32}) is still further enriched at the promoters compared with the upstream elements. It has previously been shown that patterns of histone H3K4 methylation may be influenced by the state of the associated PolII.\textsuperscript{33,34} In yeast, Set1 is recruited to chromatin only when the C-terminal domain of PolII is phosphorylated at serine 5 (i.e. when activated for initiation of transcription).\textsuperscript{35} In human, MLL1, homologous to yeast Set1, co-localizes with PolII to the 5’ end of actively transcribed genes.\textsuperscript{36} This suggests that the PolII recruited at the upstream elements might be inactive (and possibly not even engaged with DNA) and therefore the associated histones are predominantly modified with H3K4me1. By contrast, the PolII at the promoters is engaged and activated and in a state to recruit Set1 like proteins; thus, the associated histones are modified with H3K4me2 and H3K4me3. This implies that PolII may be initially recruited to the upstream elements in an inactive form and then transferred to the promoters where activation takes place.

In conclusion, this study has demonstrated how the use of a genomic microarray to study in-vivo chromatin modifications may help explain some aspects of gene regulation at a specific locus. In addition, it has highlighted some differences in the mechanism of human and mouse \(\alpha\) globin regulation which impose important caveats when using the mouse as a model to study human gene regulation.
Acknowledgments
This work was supported by the Medical Research Council (UK) and NHGRI grant U01HG003168 to I.D. and the Wellcome Trust. M.DG. is PhD student in Pharmacology and Clinical and Experimental Therapy at Turin University, Italy. We would like to thank the Computational Biology Research Group, Oxford University, Gayle Clelland and Sarah Wilcox for the technical assistance, Cordelia Langford, Peter Ellis and the staff of the welcome Trust Sanger Institute Microarray Facility for array printing, Catherine Porcher for SCL complex antibodies and Doug Vernimmen for advice and comments on the manuscript.

Authorship contribution
M.DG., E.A., I.D., R.J.G., W.G.W., D.R.H designed research; M.DG., J.A.S-S., J.A.S performed experiments; M.DG, E.A., J.H., C.M.K. analysed data; M.DG., R.J.G., W.G.W and D.R H. wrote the manuscript.

Conflict-of-interest disclosure
The authors declare no competing financial interest
References


16. Wadman IS, Osada H, Grutz GG, et al. The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which include TAL1, E47, GATA-1, and Ldb1/NL1 proteins. EMBO J. 1997;16:3145-3157.


thalassemia clarifies the importance of the major human alpha globin regulatory element. Blood. 2006;107:3811-3812


35. Ng HH, Robert F, Young RA, Struhl K. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol Cell. 2003;11:709–719

Figure legends.

Figure 1. ChIP-chip analysis of histone acetylation across 220kb of chromosome 16p13.3

(A) Overview of the human $\alpha$ globin locus. The positions of erythroid-specific DHSs (eDHSs) associated with the murine MCS elements are shown at the top. The genes located in the telomeric region of chromosome 16p are shown. The red shadow box represents the putative $\alpha$ globin regulatory domain exchanged between human and mouse to produce the humanised mouse model. The region of conserved synteny is represented as a horizontal bar. Below this, all CpG islands, DHSs, eDHS and multi-species conserved regulatory sequences (MCS-R1-4) are shown. The human eDHS corresponding to MCS-R are labelled.

(B) H4ac and H3ac in cells collected on day 8 of the second phase of culture (Day 8 Ery) and in T lymphocytes (T Ly) are shown. The y axis represents enrichment of ChIP DNA over input DNA from one experiment. A very similar pattern of enrichment was seen for each biological replicate. All replicates are available at the website: http://www.imm.ox.ac.uk/groups/mrc_molhaem/home_pages/Higgs). The x axis represents the $\alpha$ globin locus. Grey columns run from CpG islands through the array data. CpG islands not represented on the array are marked with an asterisk. V marks CpG islands which are not associated with promoters. Dashed vertical lines run from mouse eDHS through the array data.
Figure 2. Cytospin preparations representing morphology of erythroid cells at different stages of the culture.

The percentage of erythroid cells at each stage of differentiation were determined microscopically following May-Grünwald-Giemsa (MGG)-stained cytospins. Images were captured with an Olympus BX40 microscope, 40x/0.95 numerical aperture lens, fitted with Olympus Camedia C-3040ZOOM camera, and processed with Adobe Photoshop (Adobe Systems, Mountain View, CA). (A) Proerythroblasts (40%, black arrow) are present on day 5, although lymphocytes (white arrow) are the most abundant cell type. (B) Basophilic normoblasts (50%, arrow) are discernible on day 8. (C) Day 11 preparation demonstrates polychromatic erythroblasts (65%, arrow). (D) Day 15 culture shows late erythroblasts with pycnotic and shrunken nuclei (95%, arrow).
Figure 3. ChIP-chip analysis of transcription factors and PolII across 220 kb of chromosome 16p13.3.

Binding profile of GATA1 in primary erythroblasts from a homozygote “humanized” mouse (HM Ery), and of SCL, NF-E2 p45, NF-E2 p18 and PolII in day 8 human erythroblasts (Day 8 Ery) are shown. A schematic representation of the α globin locus is shown at the top annotated as in Figure1. Note that the binding profile of GATA1, is derived from “humanized” mouse, since antibodies recognizing human GATA1 do not perform well in ChIP assay.
Figure 4. ChIP-qPCR analysis in humanized mouse erythroid cells.
H3K4me1, H4ac, GATA1 and NF-E2 p45 ChIP-qPCR analysis in Ter119+ cells isolated from heterozygote humanised mouse. The y axis represent average enrichment over input DNA from independent ChIPs, normalized to a mouse GAPDH control sequence. Error bars correspond to $1 \pm SD$ from at least two independent ChIPs. Human amplicons (white columns) and mouse amplicons (grey columns) are indicated on the x axis. MPG, intronic sequence in MPG gene, $\alpha$ prom, $\alpha$ globin promoter sequence.
Figure 5. ChIP-chip analysis of H3K4 methylation across 220kb of chromosome 16p13.3.

H3K4me1, -me2 and -me3 in cells collected on day 8 of the second phase of culture (Day 8 Ery) and in EBV transformed lymphoblastoid cell lines (EBV-Ly) are shown. The y axis represents enrichment of ChIP DNA over input DNA from one experiment. A very similar pattern of enrichment was seen for each biological replicate. All replicates are available at the website: http://www.imm.ox.ac.uk/groups/mrc_molhaem/home_pages/Higgs). A schematic representation of the α globin locus is shown at the top and annotated as in Figure 1.
Table 1. Comparison of histone modifications and transcription factor binding at the human and mouse α globin loci.

<table>
<thead>
<tr>
<th></th>
<th>MCS-R1</th>
<th>MCS-R2</th>
<th>MCS-R3</th>
<th>Human</th>
<th>Mouse</th>
<th>Human</th>
<th>Mouse</th>
<th>Human</th>
<th>Mouse</th>
<th>Human</th>
<th>Mouse</th>
<th>Human</th>
<th>Mouse</th>
<th>Human</th>
<th>Mouse</th>
<th>Human</th>
<th>Mouse</th>
<th>Human</th>
<th>Mouse</th>
<th>Human</th>
<th>Mouse</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HS-48</td>
<td>HS-31</td>
<td>HS-40</td>
<td>HS-26</td>
<td>HS-33</td>
<td>HS-21</td>
<td>hoHS-12</td>
<td>HS-12</td>
<td>HS-10</td>
<td>HS-8</td>
<td>α globin promoter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H4ac</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>+++</td>
<td>-</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H3ac</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H3K4me1</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H3K4me2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H3K4me3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GATA1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCL</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NF-E2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pol2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table summarizes ChIP results from this study and from Anguita et al\(^4\) at erythroid-specific DHSs and α globin gene promoters. Differences are shown in grey shadow boxes. The multi-species conserved elements are annotated (as previously\(^2\)) as MCS-R1-4 (regulatory elements) and MCS-P9 (α globin promoter).
Figure 1
Figure 2
Figure 3
Figure 4
Tissue-specific histone modification and transcription factor binding in α globin gene expression

Marco De Gobbi, Eduardo Anguita, Jim Hughes, Jacqueline A Sloane-Stanley, Jacqueline A Sharpe, Christoph M Koch, Ian Dunham, Richard J Gibbons, William G Wood and Douglas R Higgs