The **HBS1L-MYB** intergenic interval associated with elevated HbF levels shows characteristics of a distal regulatory region in erythroid cells

Karin Wahlberg,1 Jie Jiang,1 Helen Rooks,1 Kiran Jawaid,1 Fumihiko Matsuda,2,3 Masao Yamaguchi,2,3 Mark Lathrop,3 *Swee Lay Thein,1,4 and *Steve Best1

1Division of Gene and Cell Based Therapy, King’s College London School of Medicine, London, United Kingdom; 2Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Yoshida, Japan; 3Centre National de Génotypage, Institut Génomique, Commissariat a l’Energie Atomique, Evry, France; and 4Department of Haematological Medicine, King’s College Hospital, London, United Kingdom

**HBS1L-MYB** intergenic polymorphism (HMIP) on chromosome 6q23 is associated with elevated fetal hemoglobin levels and has pleiotropic effects on several hematologic parameters. To investigate potential regulatory activity in the region, we have measured sensitivity of the sequences to DNase I cleavage that identified 3 tissue-specific DNase I hypersensitive sites in the core intergenic interval. Chromatin immunoprecipitation with microarray (ChIP-chip) analysis showed strong histone acetylation in a defined interval of 65 kb corresponding to the core **HBS1L-MYB** intergenic region in primary human erythroid cells but not in non–MYB-expressing HeLa cells. ChIP-chip analysis also identified several potential cis-regulatory elements as strong GATA-1 signals that coincided with the DNase I hypersensitive sites present in MYB-expressing erythroid cells. We suggest that **HMIP** contains regulatory sequences that could be important in hematopoiesis by controlling **MYB** expression. This study provides the functional link between genetic association of **HMIP** with control of fetal hemoglobin and other hematologic parameters. We also present a large-scale analysis of histone acetylation as well as RNA polymerase II and GATA-1 interactions on chromosome 6q, and α and β globin gene loci. The data suggest that GATA-1 regulates numerous genes of various functions on chromosome 6q. (Blood. 2009;114:1254-1262)

**Introduction**

Variable levels of fetal hemoglobin (HbF, α2γ2) persist into adulthood, and although they have no clinical consequences in otherwise healthy individuals, high HbF levels have a major impact on the principal β hemoglobin disorders—β thalassemia and sickle cell disease. Increased HbF production mitigates the severity of both diseases.5-7 The level of HbF in adults is inherited as a quantitative trait, and is largely genetically controlled with a heritability of 0.89.4

Three loci—**HBS1L-MYB** intergenic region on chromosome 6q23, BCL11A on chromosome 2p16, and the β globin cluster on chromosome 11—account for up to 50% of the variation in HbF levels in patients with sickle cell anemia or β thalassemia and in healthy European whites.1-3 The **HBS1L-MYB** intergenic region alone contributes approximately 20% of the overall trait variance in both diseases.1-3 The level of HbF in adults is inherited as a quantitative trait, and is largely genetically controlled with a heritability of 0.89.4

The panel of single nucleotide polymorphisms in the **HBS1L-MYB** region that account for the effects of the 6q locus6,8-10 reside in a nearly contiguous segment of 79 kb distributed in 3 linkage disequilibrium blocks, referred to as **HBS1L-MYB** intergenic polymorphism (HMIP) blocks 1, 2, and 3.9 Genetic variants that show the strongest effects are concentrated in 24 kb of HMIP block 2, located 33 kb upstream of **HBS1L** and 65 kb upstream of **MYB.**8 The mechanisms through which these variants operate to increase HbF are still not clear, but studies suggest that the biologic effects are likely to involve regulation of the flanking genes—**HBS1L** and **MYB**. **MYB** and **HBS1L** expression was significantly reduced in erythroid cultures of individuals with high HbF levels, whereas overexpression of **MYB** in K562 cells inhibited γ-globin expression supporting **MYB**’s role in HbF regulation.11 Further, **HBS1L** and **MYB** expression was positively correlated in erythroid progenitor cells, and **HBS1L** expression correlated with the genetic variants associated with HbF.8 Variability in HMIP block 2 was subsequently shown to have a pleiotropic effect on erythrocyte count and volume, and platelet and monocyte counts in healthy Europeans.12 These observations suggest that the **HBS1L-MYB** intergenic region is functionally active, containing distal regulatory sequences for the flanking genes—**HBS1L** and **MYB**. The function of **HBS1L** is unknown but it encodes a protein with apparent GTP-binding activity, involved in the regulation of a variety of critical cellular processes,13 and **MYB** encodes a transcription factor involved in oncogenesis and with an essential role in erythropoiesis.14-16

Initially, we investigated the regulatory potential of **HMIP** block 2 by measuring sensitivity of the sequences in this region to DNase I cleavage that identified multiple DNase I hypersensitive sites in the region in K562 cells. We then proceeded to a comprehensive analysis of the regulatory potential of a large region of chromosome 6q using chromatin immunoprecipitation (ChIP) and microarray (ChIP-chip) analysis on primary human erythroid progenitor cells. We identified strong signals of histones H3 and H4 acetylation in the **HBS1L-MYB** intergenic region (indicative of active chromatin) especially concentrated in block 2, in basophilic erythroblasts when the globin genes and **MYB** are fully active. Tissue specificity of the regulatory activity in the intergenic region...
was demonstrated by minimal histone acetylation in HeLa cells. ChIP-chip also demonstrated interactions of the erythroid-specific transcription factor GATA-1 with several sites in the HBBSIL-MYB interval. GATA-1 signals in coincidence with the DNase I hypersensitive sites in HMIP block 2 strongly suggest the presence of regulatory elements. Regulatory activity of the intergenic region was validated by presence of intergenic transcripts in erythroid precursor cells in a tiling microarray. We postulate that the regulatory elements distally control MYB expression, which in turn influences erythroid differentiation and, indirectly, the control of HbF levels.

The use of microarrays also allowed us to compare patterns of activity in the candidate interval with other regions encompassing widely expressed genes across 70 Mb of chromosome 6q. We further provide a large-scale analysis of GATA-1 occupancy in erythroid cells that includes the entire β and α globin gene clusters.

Methods

Cells and cell cultures

Cell lines were maintained in RPMI-1620 medium (Sigma-Aldrich) with the addition of 10% of fetal calf serum (FCS; PAA-laboratories), 2 mM l-glutamine (Sigma-Aldrich), 0.1 mg/mL streptomycin, and 18 units/mL penicillin (Sigma-Aldrich). Concentrations were kept at 0.5 to 1.0 × 10⁶ cells/mL. K562 cells were treated with 40 μM hemin (Sigma-Aldrich) for 24 hours to induce differentiation.

Primary human erythroid cells were cultured from peripheral blood using a 2-phase liquid system as previously described. Cytospins of erythroid progenitors from different days of culture were stained using a Giemsa staining set (Hema “Gurr”; VWR) according to the manufacturer’s protocol.

Flow cytometry of primary erythroid cells was performed with anti-human CD71 monoclonal antibodies (FITC conjugated, 555536; BD Biosciences) or anti-human glycophorin A (GPA, phycoerythrin [PE] conjugated, R7078; DAKO) as previously described.

DNase I hypersensitivity analyses

DNase I hypersensitivity analysis of HMIP-2 was performed on 2 biologic replicates of induced and uninduced K562 cells, and Jurkat cells. Nuclei were treated with 70 units of DNase I at 37°C for 3 minutes as determined by optimization experiments (supplemental methods and supplemental Figure 1, available on the Blood website; see the Supplemental Materials link at the top of the online article). Real-time quantitative polymerase chain reaction (PCR) was performed in triplicate on 20 ng DNA samples using SYBR Green PCR Mastermix (Applied Biosystems) and the ABI Prism 7900HT Sequence Detection system (Applied Biosystems).

The HMIP-2 region could be covered in 68 overlapping fragments of approximately 500 bp (PCR primer sequences available on request). Relative sensitivity to DNase I for each target was calculated by converting the DNA cross-linking efficiency in induced K562 cells to input DNA, and the results were normalized to a control sequence in exon 1 of the neurofilament gene (NEFM) representing an inactive gene (data not shown).

Microarray analysis of ChIP material

Two microarrays, both from Roche-NimbleGen, were used in these experiments as described in the supplemental methods. The first array encompassed 70 Mb (positions 93,424,310 to 165,905,673; hg18) on chromosome 6, including the 6q23 HbF locus. The second was a custom array and included the globin loci as controls. The microarray data have been deposited with Gene Expression Omnibus (GEO) under accession number GSE16541.

Input and ChIP DNA were amplified using a whole genome amplification kit (WGA1; Sigma-Aldrich) applying a previously described protocol adapted for ChIP material. Amplified DNA was purified using a QIAquick PCR purification kit (QIAGEN) with buffer PBI substituted for buffer PB.

Arrays were hybridized and washed using Roche-NimbleGen kits according to the manufacturer’s protocol. Scanning was performed using a GenePix 4000B Scanner (Molecular Devices). Detailed protocol, including data extraction and analysis, is shown in supplemental methods.

Transcript Tiling Array

A customized Affymetrix GeneChip Tiling Array was designed to identify novel transcripts between positions 135,232,209 and 135,582,003 (NCBI build 36) that encompass the entire MYB and HBBSIL genes as well as the intergenic region: 116,858 bases containing human repetitive sequence were excluded from probe design. The remaining 141,937 bases were tiled by overlapping oligonucleotides. The starting position of each 25-mer oligo was shifted by 2 bases within HBSIL and MYB, and by one base within the intergenic region. Oligonucleotides were designed for both DNA strands. A total of 224,258 oligonucleotide probes were included in each array.

Total RNA was extracted from erythroid cells (liquid culture phase II, days 3 and 5) using TRI-Reagent (Sigma-Aldrich). Five micrograms was treated with the RiboMinus Transcriptome Isolation Kit (Invitrogen) to remove ribosomal RNA before cDNA synthesis according to the manufacturer’s instructions. Synthesis of cDNA, fragmentation, and end labeling was performed using the GeneChip Whole Transcript Double-Stranded Target Assay kit (Affymetrix) according to the manufacturer’s instructions. After hybridization for 16 hours, the arrays were washed and stained using a Fluidics Station 450 (Affymetrix). Scanning was performed by GeneChip Scanner 3000 7G (Affymetrix). Data analysis was performed using Affymetrix Tiling Analysis Software Version 1.1 and data were visualized using the Affymetrix Integrated Genome Browser. Because double-stranded cDNA was used, the transcript maps correspond to the signals from both strands.

Chromatin immunoprecipitation

Antibodies used for ChIP experiments included anti-diacylated histone H3 (K9 and 14; no. 06-599; Millipore), anti-tetra-acetylated histone H4 (K5, 8, 12, and 16; no. 06-866; Millipore), anti-RNA polymerase II (no. 05-623; Millipore), and anti-GATA-1 (no. sc-1234; Santa Cruz Biotechnology).

ChIP experiments for acetylated histones were performed using the EZ-ChIP protein G kit (no. 17-371; Millipore) according to the manufacturer’s protocol with minor modifications (Dr. David Garrick, Weatherall Institute of Molecular Medicine, Oxford, United Kingdom). ChIP for RNA polymerase II and GATA-1 was performed using the Magna ChIP protein G kit (no. 17-611; Millipore) according to the manufacturer’s protocol with minor modifications.

ChIP assays were performed on cultured primary human erythroid cells (phase II, day 10) and HeLa cells. Cells (5 × 10⁷ per experiment) were cross-linked in 10 mL growth medium with 1% formaldehyde (Sigma-Aldrich) for 10 minutes at room temperature, and the chromatin was sonicated (10 × 15 seconds, Sonic VibraCell at 40% efficiency) to a size of approximately 500 base pairs (bp; range 200-1000 bp). Immunoprecipitations were performed after an overnight incubation with 5 to 10 μg of the appropriate antibody, with protein G agarose beads, or with protein G magnetic beads. A sample containing no antibody was used as a negative control.

ChIP material was validated by SYBR Green quantitative PCR before microarray analyses using different positive control targets. Enrichment of a specific target sequence in ChIP material was calculated relative to input DNA, and the results were normalized to a control sequence in exon 1 of the neurofilament gene (NEFM) representing an inactive gene (data not shown).
Relative sensitivity to DNase I for each target was calculated from delta CT values between treated and untreated samples and normalized to a nonerythroid cell line. Two biologic repeats were performed in uninduced and hemin-induced K562 cells, and in Jurkat cells (T-cell leukemia), representing hypersensitivity profiles were studied in uninduced and hemin-induced K562 cells.

MYB expression increased 9.5-fold, that of HS3 control was not included for DNase I sensitivity analyses in Jurkat cells.

In uninduced K562 cells, several sites within HMIP-2 region account for the majority of the genetic association with HbF in the 6q QTL region. We performed DNase I hypersensitivity analysis along the entire 24 kb of HMIP-2 region in K562 cells as a screen for genetic regulatory elements. K562 cells treated with hemin were induced to differentiate, resulting in strong up-regulation of globin gene expression and simultaneous down-regulation of MYB expression. After 24 hours of hemin treatment (at 40 μM), β globin expression increased 9.5-fold, that of γ globin, 5-fold, whereas MYB expression decreased by 7-fold (data not shown). DNase I hypersensitivity profiles were studied in uninduced and hemin-induced K562 cells, and in Jurkat cells (T-cell leukemia), representing a nonerythroid cell line. Two biologic repeats were performed for each cell type. Analysis of both HS2 and HS3 (positive controls) and NEFM (negative control) were included for K562 cells and HS2 and NEFM, for Jurkat cells.

In uninduced K562 cells, several sites within HMIP-2 showed sensitivity to DNase I above background levels, indicating an open chromatin structure (Figure 1A). When cells were induced to differentiate, the region showed a general increase in DNase I sensitivity; 3 sites referred to here as HS1, HS2, and HS3, in particular, showed a marked increase in sensitivity compared with background levels (Figure 1A). DNase I sensitivity also increased for HS2 and HS3 controls, consistent with the induction in globin gene expression. HMHS1, HMHS2, and HMHS3 also showed stronger sensitivity to DNase I than the βHS3 control, thereby reaching a threshold level for hypersensitivity (Figure 1A). There was no difference in DNase I sensitivity for NEFM in uninduced and induced K562 cells.

As expected, HS2 was not sensitive to DNase I in Jurkat cells, a nonerythroid cell line (Figure 1B). In the HMIP block 2 region, Jurkat cells show similar background levels and generally, a similar DNase I sensitivity profile to induced K562 cells, but with much less sensitivity at HMHS1, HMHS2, and HS3, indicating that DNase I sensitivity at these sites is tissue specific (Figure 1B). The strongest sensitivity to DNase I in Jurkat cells coincided with the putative promoter region of the alternative HS1 exon (exon 1a), which showed a low degree of sensitivity in K562 cells (both uninduced and induced). This is consistent with expression of the HS1-1a transcript in Jurkat cells but not in K562 cells. To validate the promoter prediction at HS1-1a, we examined its functional property in a reporter assay; the region showed activity in Jurkat cells but not in K562 cells (supplemental Figure 2). The HS1-1a promoter therefore served as a positive internal control for DNase I hypersensitivity at an active regulatory element within the HMIP block 2 region.

Characterization of erythroid and nonerythroid cells for use in chromatin immunoprecipitation

The identification of DNase I hypersensitive sites in HMIP block 2 in hemin-induced K562 cells suggested that the intergenic region contained regulatory elements active in erythroid lineages and encouraged further functional analysis of the interval. We assessed the chromatin activity and transcription factor binding throughout the HS1-1a and flanking regions of chromosome 6q, and tissue-specific activity of these profiles. Primary human erythroid progenitors cultured in a 2-phase liquid system were analyzed to
select optimum time of harvest when HBS1L, MYB, and the globin genes were fully expressed. HeLa cells served as examples of a MYB-negative cell line.

Basophilic erythroblasts from phase II, day 10, were chosen for ChIP analysis (supplemental methods and supplemental Figure 3). At this stage, expression profiles indicated that the cells were in a state of high transcriptional activity; the candidate genes (HBS1L and MYB) as well as GATA1 and the globin genes were expressed at high levels (data not shown).

To investigate tissue specificity of activity in the intergenic region and the relation between activity and candidate gene expression, we compared histone acetylation patterns in the intergenic region between erythroid precursor cells and a cell line lacking HBS1L and MYB expression. Eight cell lines, which included Jurkat, HL60 (promyelocytic leukemia), U937 (monocytic leukemia), HeLa (cervical cancer), HEK293 (kidney), HuH-7 (liver carcinoma), U2OS (osteosarcoma), and HKC-8 (renal epithelial), were screened, together with K562 cells and primary erythroid cells for HBS1L and MYB expression using TaqMan reverse transcription PCR. HBS1L was highly expressed in all lines analyzed (real-time PCR C_T values of 23-26), which indicates a housekeeping function (supplemental Figure 4A). In contrast, MYB expression varied dramatically; it was highly expressed in all hematopoietic cell–related lines (C_T values of 24-26) but minimally expressed in other lines (C_T values of 31-35; supplemental Figure 4B). HeLa, which showed relatively low HBS1L expression (10% of expression in erythroid precursors) and insignificant MYB expression (0.01% of expression observed in erythroid precursors), was chosen to represent a MYB-negative cell line in ChIP experiments.

Overview of histone acetylation, GATA-1, and RNA polymerase II interactions across chromosome 6q

On viewing the ChIP-chip data from erythroid precursors across 70 Mb of chromosome 6q represented on the array, it was evident that all antibodies showed similar patterns of activity. The strongest peaks were found in gene-rich areas, whereas large intergenic sequences lacked signal (supplemental Figure 5). Well-defined areas of high levels of AcH3 were seen at transcriptional start sites (TSSs) with a less defined pattern for AcH4 as previously described.26 Interestingly, abundant GATA-1 signals were found in a defined 65-kb interval encompassing the HBS1L/MYB region (Figure 3). Within this interval, the strongest peaks in the core intergenic region. In fact, the GATA-1 signal in the HBS1L–MYB intergenic region represented the most significant GATA-1 peaks in the entire 70-Mb region covered on the 6q array.

The HBS1L–MYB intergenic region is highly active in erythroid precursors

ChIP-chip data of GATA-1, AcH3, AcH4, and RNAPII in erythroid precursors and AcH3 in HeLa cells was analyzed across a 2.5-Mb region of the 6q23 HbF locus, encompassing the 5 protein coding genes (including the HBS1L–MYB intergenic region), and flanking sequences (Figure 2). Biologic ChIP-chip replicates using primary human erythroblasts (phase II day 10 of culture) from 2 individuals provided very similar profiles.

Histone acetylation patterns and RNAPII signal in the 6q23 locus in erythroid precursors were consistent with previous expression analysis of genes in the locus11 (Figure 2). The PDE7B and ALDH3A1 genes are not expressed in erythroid precursors and consistently showed no signal of acetylation or RNAPII interaction. AH11 that is expressed at low levels showed some histone acetylation signal at the promoter regions. In contrast, the highly expressed HBS1L and MYB genes were associated with strong RNAPII signal and histone acetylation around the promoter regions as well as coding regions. The RNAPII antibody used detects both nonphosphorylated inactive RNAPII as well as phosphorylated actively elongating forms. Interestingly, although MYB is the most highly expressed gene in the region, we see no RNAPII interaction at the immediate promoter-proximal 5’ end, but instead, high levels in the body of the gene and a large accumulation toward the 3’ end and beyond (Figures 2–3). We speculate that the lack of RNAPII at the promoter is a result of rapid transcription leaving no inactive RNAPII stalling at the initiation complex. The phenomenon of RNAPII accumulation at the 3’ end of actively transcribed genes has previously been observed and is likely to reflect pausing and dephosphorylation of the polymerase before release.28 With nonexpressed genes, RNAPII was often seen as a strong signal but only at the 5’ preinitiation complex. Strikingly, the AcH3 and AcH4 signals were equally strong in the HBS1L–MYB intergenic region as around the HBS1L and MYB promoters and exons, indicating that this region is highly active in erythroid precursors. In the context of the whole of 6q region analyzed, there was little evidence of such high intergenic activity elsewhere.

In contrast to the strong signals observed in erythroid cells, HeLa cells showed minimal histone acetylation, consistent with the low/absent candidate gene expression in these cells.

GATA-1 signal in the 6q23 locus was concentrated around the MYB and HBS1L area, with the strongest peaks in the core intergenic region. In fact, the GATA-1 signal in the HBS1L–MYB intergenic region represented the most significant GATA-1 peaks in the entire 70-Mb region covered on the 6q array.

The HMIP-2 and -3 regions show characteristics of a distal regulatory region in erythroid precursors

A closer view of the HBS1L–MYB intergenic region revealed that the histone H3 and H4 acetylation in erythroid precursor cells was found in a defined 65-kb interval encompassing the HMIP-2 and -3 regions (Figure 3). Within this interval, the strongest AcH3 and GATA-1 signals were concentrated in the HMIP-2 region. The intergenic region included 7 peaks of GATA-1 signal, 3 of which were within HMIP-2 and 1 was in HMIP-3. The GATA-1 signals in the HMIP-2 region all coincided with DNase I hypersensitive sites identified in induced K562 cells as a further indication of these sites being functional regulatory elements. In addition to a GATA-1 signal approximately 7 kb 5’ of MYB, strong signals were seen in intron 5 of MYB (just upstream of exon 6) and in intron 8 (just upstream of exon 9). No GATA-1 signal, however, was detected at the immediate MYB promoter. GATA-1 signal was also observed at the HBS1L promoter region. There was a prominent coincidence of GATA-1 signal on the array with conserved GATA-1 motifs (human/mouse/rat alignment available from the UCSC Genome Browser; http://genome.ucsc.edu),27 which support a functional relevance of these sites. Some weak RNAPII signals were also
observed in the HMIP-2 and -3 regions and in coincidence with GATA-1 signal.

In contrast to the strong signals observed in erythroid precursors, histone H3 acetylation was absent in the intergenic region in HeLa cells with the exception of a restricted but significant peak in the HMIP-2 region, coincident with HMHS3 in K562 cells and a strong GATA-1 signal in erythroid precursors. This peak is in the vicinity of the putative promoter region of HBS1L exon 1a.

Figure 2. ChIP-chip data for the 6q23 locus in erythroid precursor and HeLa cells. (A) The box on chromosome 6 represent a 2.5-Mb region (position chr6: 134 503 000-136 928 000), which is covered in this figure and includes the 6q23 locus. (B) Results from ChIP-chip experiments for GATA-1, RNAP II, AcH4, and AcH3 in erythroid precursors and AcH3 in HeLa cells. Genes in the 6q23 locus are shown below as black arrows and other genes in gray. The HBS1L-MYB intergenic region is indicated by the boxed area marked with dashed lines to highlight the strong histone acetylation and GATA-1 binding signals in erythroid precursor cells. Data are displayed in the UCSC Genome Browser27 using autoscale settings.

Figure 3. ChIP-chip data for the HBS1L-MYB intergenic region. Results from ChIP-chip experiments for GATA-1, RNAP II, AcH4, and AcH3 in erythroid precursors and AcH3 in HeLa cells. The figure covers a 210-kb region (position chr6: 135 373 000-135 582 000) and includes a part of the HBS1L gene, the intergenic region, and the MYB gene. The HMIP block 1, 2, and 3 are shown above indicated by the horizontal brackets. The figure also includes conserved GATA-1 motifs (●) and DNase I hypersensitive sites (△) as identified in hemin-induced K562 cells.
By including the well-characterized α and β globin loci on our custom-designed array, we introduced positive controls for histone acetylation and GATA-1 binding in erythroid cells to evaluate the quality of our ChIP material and data analysis. In addition, the custom array allowed us to compare patterns of histone acetylation and transcription factor binding between the distal regulatory regions of the globin loci and the upstream regulatory domain in erythroid precursors but not in HeLa cells (Figure 5). Consistent with previous observations, we observed GATA-1 signals at HS48, HS40, HS33, and HS10 upstream of the α globin genes. GATA-1 signal was also found at the human orthologous region corresponding to the mouse HS12 (hoHS12) and at a site in-between the β globin (HBZ) and the ψα2 pseudogene (HBM). RNAP II signals were observed at the α globin promoters and the upstream elements HS-48 and HS-40, again consistent with previous observations.

Given the unusually high and concentrated levels of H3 acetylation and evidence of RNAPII binding in the intergenic region, we decided to investigate the intergenic region for evidence of transcription using a high-resolution tiling array (supplemental methods). Primary human erythroid precursor cells (phase II, days 3 and 5 liquid culture) from 2 individuals were analyzed. All samples gave essentially identical results. As expected, there is strong transcriptional activity at the exons of MYB and HBSIL, with relatively little in the introns. However, very strong and well-defined transcriptional activity was identified in the intergenic locus spanning HMIP-2 and -3. In several areas, the signal intensity is even greater than from the MYB exons (Figure 6).

Discussion

Here, we provide a first characterization of the intergenic sequences upstream of the HBSIL and MYB genes, strongly supporting the hypothesis of a regulatory region being located in this interval. The conclusion is supported by parallel analysis of histone acetylation, GATA-1, and RNAP II interaction patterns across the erythroid-specific α and β globin loci. Chromatin immunoprecipitation showed significant histone acetylation in the intergenic region in a restricted interval that encompasses HMIP-2 and -3 linkage disequilibrium blocks as defined from genetic analysis. The H3 acetylation was particularly well defined and concentrated in the intergenic region.

In the β globin locus, 2 isolated domains of activity (AcH3 signals) were clearly discernable and were concentrated around the β LCR and a region covering the β and δ globin genes (HBB and HBD) and the area around the ψβ pseudogene (HBBP1; Figure 4). These 2 domains of activity were separated by an inactive domain comprising the ε globin (HE) gene and the region up to the γ globin genes (HBG1 and 2). This pattern was similar to previously published data of H3 acetylation in human erythroid precursors. HeLa cells showed little signal for H3 acetylation in the β globin locus. Some signal was detected in the LCR at HS5 and upsteam of the γ-globin genes. Acetylation at HS5 in HeLa cells is consistent with this site being ubiquitously active.

Strong GATA-1 signal was detected in the β globin LCR at HS1, HS2, HS3, and HS4, and at the upstream hypersensitive sites 6 and 7. In addition, GATA-1 signals were observed upstream of the HBD and HBBP1 as well as in HBB. A closer view revealed that the GATA-1 signal in HBB coincided with exon 3 (Figure 4), which has previously been shown to contain an enhancer element involved in the control of β globin expression. RNAP II binding in the β globin locus coincided with GATA-1 signals at HS1 and HS3 and was also associated with HBG1 and the adult globin genes. The strongest RNAP II signal was observed around HBB, consistent with high expression of β globin in erythroid precursors at the time of harvest.

The α globin locus showed strong H3 acetylation around the α globin genes and the upstream regulatory domain in erythroid precursors but not in HeLa cells (Figure 5). Consistent with
RNAP-II signals have previously been observed from ChIP-chip experiments at the site of enhancers and could be a marker of physical interactions with active promoters. Such enhancer elements could affect distal transcriptional control via long-range physical interaction as supported by the observations of GATA-1 involvement in looping formation within the β-globin locus. GATA-1, together with FOG-1, functions as an anchor in the formation of chromatin looping, and is required for physical interactions between the β-LCR and β-globin promoter. We show that the pattern of H3 acetylation in the HBS1L-MYB region differs between erythroid precursors and HeLa cells. HeLa cells that do not express hematopoietic transcription factors,
including MYB, showed substantially less H3 acetylation in the intergenic region, suggesting a link between activity in the region and MYB expression. We identified a restricted peak of H3 acetylation in the intergenic region in HeLa cells that coincided with GATA-1 signal in erythroid precursors and HMHS3 in induced K562 cells. It is possible that this peak represents a basal level of chromatin acetylation and could conceivably be the crucial activation core for the locus in erythroid cells. Binding of specific transcription factors including GATA-1 to this site could lead to the activation of the regulatory region with consequent induction of expression of MYB and other genes in the domain.

Our study has also provided, for the first time, a large-scale analysis of GATA-1 occupancy in human erythroid cells. The abundant GATA-1 signals over the entire 70-Mb interval of chromosome 6q suggest that GATA-1 has an important regulatory role in erythroid cells. The powerful influence of GATA-1 on erythroid commitment and development was recently suggested in overexpression studies of GATA-1 that resulted in the transformation of HeLa cells to a more erythroid phenotype, including formation of the β globin LCR and expression of globin mRNA. Our findings are supportive of GATA-1 having a role as a general transcription factor in erythroid cells in regulating the transcription of ubiquitous as well as erythroid-specific genes.

The identification of high levels of intergenic transcription provides further evidence that the HBSIL-MYB region contains a distal regulatory locus. Except for HBSIL-exon 1a, there is no evidence of ESTs, spliced or unspliced, or GeneScan gene predictions in the region (NCBI database; http://www.ncbi.nlm.nih.gov), suggesting there are no undiscovered genes in the interval. Several studies on locus control regions have suggested that intergenic transcription may be involved in chromatin decondensation and looping, which is fundamental to gene activation. Alternatively, it may represent a “tracking” mechanism that enables a transcription complex to move along the locus until a transcriptionally competent promoter is encountered. The patterns of histone acetylation, RNAPII binding, and GATA-1 interactions, in coincidence with the multiple DNase I hypersensitive sites and the intergenic transcripts in this defined interval, are highly similar to previously well-characterized erythroid control regions.

The regulatory potential of the region upstream of the MYB region and its influence on MYB expression has previously been emphasized. In murine models, Myb has been shown to be a key target for transcriptional activation by long-range upstream and downstream retroviral insertion. Integration of proviruses in a region 25- to 90-kb upstream of Myb in mice is associated with tumorigenesis, suggesting a functional importance of these sequences. Further, it has been observed that increased expression of the flanking genes occurred only in the presence of Myb overexpression. The observations suggest the possibility that regulation of Myb may affect a wider chromatin domain surrounding the gene. Alternatively, there may be common transcription factors or a common cis-regulatory element(s) that controls the expression of Myb and another gene(s) in its vicinity. Further evidence supporting the regulatory potential of this region comes from a serendipitous insertion of a transgene in this intergenic region, 77-kb upstream of the mouse Myb gene, that resulted in reduced Myb expression and markedly decreased megakaryocyte/erythrocyte lineage-restricted progenitors of the homozygous mutant mice.

Previously, we showed that MYB is a quantitative trait gene, with variable expression in healthy adults. Our previous studies also showed that human erythroid precursor cells from individuals with higher HbF and higher F-cell levels have lower MYB expression that was also associated with lower erythrocyte count but higher erythrocyte volume, and higher platelet count. Further, mouse models in which Myb activity was reduced, due to either mutation or integration of a transgene near the Myb locus, displayed anemia and thrombocytopenia. It is clear that MYB, a transcription factor that is also involved in oncogenesis, has multiple essential roles throughout the different stages of erythropoiesis.

What is not clear, however, are the regulatory sequences controlling MYB expression. Recent studies show that MYB is a major target of the microRNA 150 (miR-150), and that one pathway of MYB regulation is through the 2 conserved miR-150 binding sites in the 3′ UTR of MYB mRNA. miR-150 repression of MYB in CD34+ human bone marrow cells not only supported MYB’s key role in erythroid and megakaryocytic differentiation, but also suggested that modulations of its level are critical to its role. We propose that the HBSIL-MYB region upstream of MYB contains distal regulatory elements that form a key part of the overall control of MYB expression. The intergenic variants may account for some of the cis-control of the intrinsic quantitative variation in MYB expression. Genetic variants in the HBSIL-MYB interval on chromosome 6q have been shown to be highly associated, not only with HbF levels, but also with other hematologic parameters. Taken together, our data have provided a functional basis for this association and strongly support the hypothesis of a regulatory locus upstream of the HBSIL and MYB genes, located within HMIP-2 and -3 blocks as identified in genetic association studies.

Delineation of the key variants in this HBSIL-MYB control region may lead to an improved understanding of MYB control and dysregulation that underlies many of the leukemias and cancers, and may also provide targets for therapeutic activation of HbF in the treatment of sickle cell disease and β thalassemia.

Acknowledgments

We thank Claire Steward for help in preparation of the manuscript; Drs Marco de Gobbi, Jim Hughes and David Garrick for their help with ChIP-chip experiments, and Dr Mike Antoniou, Dr Stephan Menzel, Professors Doug Higgs and Bill Wood for helpful discussions.

This work was supported by a grant from the Medical Research Council, United Kingdom (MRC G0000111 and ID51640) to S.L.T. and an MRC training studentship to K.W. The research at the Center for Genomic Medicine, Kyoto University, is partly supported by the Ministry of Education, Culture, Sports, Science and Technology of Japan. We also thank the London University Central Research Fund (CRF) and British Society for Hematology for support (S.B.).

Authorship

Contribution: K.W. performed research, analyzed data, and wrote the paper; J.J., H.R., K.J., F.M., and M.Y. performed research; M.L. contributed to data analysis and writing of the paper; S.L.T. codirected research and wrote the paper; and S.B. codirected research, analyzed data, and wrote the paper.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Swee Lay Thein, King’s College London School of Medicine, James Black Centre, 125 Coldharbour Lane, London SE5 9NU, United Kingdom; e-mail: sl.thein@kcl.ac.uk.
References

7. Sedgwick AE, Timofeev N, Sebastiani P, et al. The HBS1L-MYB intergenic region on chromosome 6q23 is a major quantitative trait locus on chromosome 6q23 variants of HBS1L-MYB are responsible for a major quantitative trait locus on chromosome 6q23 influencing fetal hemoglobin levels in adults. Proc Natl Acad Sci U S A. 2007;104(23):7577-7580.