genes involved in the control of viral load. Indeed, these 3 GWASs identified the *HCP5* rs2395029 polymorphism, which is in linkage disequilibrium with HLA-B*57 and other immunity genes, such as *MICB*, *TNF*, *LTB*, *BAT1*, and *HLA-C*. An association with the HIV DNA level (reservoir) was also depicted by the PRIMO GWAS [2] for *SDC2* (chromosome 8), whose encoded protein is required for Tat internalization. Finally, the *ZNRD1* locus of chromosome 6 was associated with the control of disease progression but not of viral load in both the Euro-CHAVI (Center for HIV/AIDS Vaccine ImmunologyCenter for HIV/AIDS Vaccine Immunology) [3] and the nonprogressor Genomics of Resistance to Immunodeficiency Virus (GRIV) [4] GWASs.

To identify genetic loci predisposing a person to rapid progression of AIDS rather than disease control, we undertook a case-control GWAS involving a unique cohort of human immunodeficiency virus type 1 (HIV-1)–infected patients who experienced rapid progression, using Illumina HumanHap300 BeadChips. In a manner symmetric to the published nonprogressor GRIV GWAS [4], the use of this extreme phenotype should lead to an enrichment of our knowledge of the genetic factors involved in rapid disease progression. The power of this extreme design has indeed been demonstrated by previous candidate gene studies [5–7].

METHODS

The GRIV cohort. The GRIV cohort was established in France in 1995 to generate a large collection of DNA for genetic studies to identify host genes associated with either rapid progression or nonprogression to AIDS [5, 7]. Only white people of European descent living in France were eligible for enrollment to reduce confounding by population substructure. These criteria limit the influence of ethnic and environmental factors (all subjects live in a similar environment and are infected by B strains) and emphasize how the genetic makeup of each individual determines the various patterns of progression. Rapid progressors were included on the basis of the main clinical outcomes, CD4 T cell count and time to disease progression, and were defined as those who had 2 or more CD4 T cell counts below 300 cells/mm^3 within 3 years after the last seronegative test result. DNA was obtained from fresh peripheral blood mononuclear cells or from Epstein-Barr virus–transformed cell lines. The rapid progression group (n = 85) was composed of 73 men and 12 women aged 21–55 years (median, 32 years) at inclusion. At inclusion, the median CD4 T cell count was 230 cells/mm^3 (minimum and maximum values, 20 and 297 cells/mm^3). All patients provided written informed consent before enrollment in the GRIV GWAS.

The seropositive control population. To discriminate between positive signals corresponding to either an association with rapid progression or an association with HIV-1 infection, we needed a group of seropositive control subjects who were not rapid progressors. For that, we used 275 white French subjects who qualified as nonprogressors to AIDS (ie, those who had an asymptomatic HIV-1 infection for >8 years, no receipt of treatment, and a CD4 T cell count consistently remaining at >500 cells/mm^3). This control group was composed of 201 men and 74 women aged 19–62 years (median, 35 years) at inclusion. The median CD4 T cell count of this seropositive control population was 706 cells/mm^3 (minimum and maximum values, 501 and 2298 cells/mm^3).

The SU.VI.MAX seronegative control group. The SU.VI.MAX (Supplémentation en Vitamines et Minéraux Antioxydants) study was a randomized, double-blind, placebo-controlled, primary-prevention trial designed to test the efficacy of daily supplements of antioxidant vitamins and minerals at nutrition-level doses in reducing several major health problems in industrialized countries, especially the main causes of premature death, cancers and cardiovascular diseases. This cohort study was started in 1994 in France and included 12,735 subjects [8]. The control group genotyped in the present study comprised 1352 representative SU.VI.MAX participants, all of them white persons living in France who were HIV-1 seronegative. This control cohort was composed of 525 men and 827 women, with a mean age of 53.1 and 48.5 years, respectively.

The second seronegative control group. The D.E.S.I.R. (Data from an Epidemiological Study on Insulin Resistance Syndrome) program was a 9-year follow-up study designed to clarify the development of insulin resistance syndrome. Subjects were recruited from 1994 to 1996 from volunteers insured by the French social security system, which offers periodic health examinations free of charge [9]. This second control group comprised 697 participants who were both not obese and normoglycemic from the D.E.S.I.R. trial, all French and HIV-1 seronegative. It was composed of 281 men and 416 women aged 30–64 years.

Genotyping method. The GRIV cohort and the 2 seronegative groups were genotyped using Illumina Infinium II HumanHap300 BeadChips (Illumina). Genomic DNA (750 ng) was whole-genome amplified, fragmented, denatured, and hybridized on prepared HumanHap300 BeadChips for a minimum of 16 h at 48°C. Non–specifically hybridized fragments were removed by washing, and the remaining specifically hybridized DNA was fluorescently labeled by a single base extension reaction and detected using a BeadArray scanner (Illumina). Normalized bead-intensity data obtained for each sample were loaded into BeadStudio software (version 3.1; Illumina), which converted fluorescence intensities into single-nucleotide polymorphism (SNP) genotypes.

Quality control. Using the BeadStudio software, we analyzed the crude genotyping data, and SNPs were filtered according to the following parameters. First, samples with a call

Table 1. Best Results Obtained for the Comparison between Rapid Progressors and Control Subjects

SNP	Gene	Chr	Chr position	A1	A2	Allelic frequency (A1), %				OR (95% CI)	Fisher *P* Value, RP-$CTR_{SU.VI.MAX}$	FDR *q* value
						RP	$CTR_{SU.VI.MAX}$	$CTR_{D.E.S.I.R.}$	SCP			
rs4118325	Intergenic[a]	1	107379355	A	G	5.2	19.0	18.7	15.8	0.24 (0.12–0.46)	6.09×10^{-7}	0.17
rs1522232	*SOX5*	12	24285639	T	C	29.1	47.7	48.2	50.7	0.45 (0.32–0.63)	1.80×10^{-6}	0.20
rs1360517	Intergenic	9	12997129	A	G	16.3	5.9	7.0	5.8	3.09 (2.00–4.78)	3.27×10^{-6}	0.20
rs3108919	Intergenic	8	101910722	C	T	43.6	26.6	27.8	27.5	2.13 (1.56–2.91)	3.86×10^{-6}	0.20
rs10800098	*RXRG*	1	163675719	A	G	14.5	4.9	4.7	5.6	3.29 (2.08–5.20)	3.86×10^{-6}	0.20
rs10494056	Intergenic[b]	1	107349442	A	C	5.8	18.6	17.9	16.1	0.27 (0.14–0.52)	4.29×10^{-6}	0.20
rs12351740	Intergenic[c]	9	13010010	T	C	12.8	4.1	4.7	3.8	3.46 (2.12–5.62)	6.63×10^{-6}	0.25
rs1020064	*TGFBRAP1*[d]	2	105264172	T	G	9.3	23.2	25.6	25.2	0.34 (0.20–0.57)	7.04×10^{-6}	0.25

NOTE. *P* values were computed by the Fisher exact test in the allelic frequency mode and are presented with their corresponding allelic frequencies in the various populations (rapid progressors [RP], seronegative control subjects [$CTR_{SU.VI.MAX}$], and the seropositive control population [SCP]), chromosome (Chr) positions, odds ratios (ORs) with 95% confidence intervals (95% CIs), and false-discovery rate (FDR) *q* values. The frequencies in the D.E.S.I.R. second seronegative control group ($CTR_{D.E.S.I.R.}$) are also indicated and are similar to those in the SU.VI.MAX cohort. SNP, single-nucleotide polymorphism.

[a] This intergenic SNP is in linkage disequilibrium with *PRMT6* and intergenic SNPs.

[b] This intergenic SNP is in linkage disequilibrium with *PRMT6* and intergenic SNPs ($r^2 = 0.92$ with rs4118325).

[c] This intergenic SNP is in partial linkage disequilibrium with the intergenic SNP rs1360517 ($r^2 = 0.68$).

[d] This *TGFBRAP1* SNP is in linkage disequilibrium with an intergenic SNP.

rate (percentage of SNPs genotyped by sample) <95% in the Illumina clusters were deleted. Second, the SNPs with a call frequency (percentage of samples genotyped by SNP) <99% were reclustered. Third, after reclustering, samples with a call rate below 97% were deleted. The clustering step can create SNP genotyping errors, which can be prevented by following the Illumina procedure (see http://www.illumina.com/downloads/GTDataAnalysis_TechNote.pdf). This method evaluates the quality of the newly created clusters according to several criteria, which can be manually checked and corrected as necessary. In total, 1300 SNPs were excluded by this Illumina quality control procedure. Finally, after all the quality control steps were performed, the 15,731 SNPs with a call frequency <98% (>2% of missing data) were excluded. This procedure ensures reliable genotyping data with little missing data.

Hardy-Weinberg equilibrium analysis was performed for each SNP in each group by using an exact statistical test implemented in PLINK software ([10]; available at: http://pngu.mgh.harvard.edu/~purcell/plink/). Deviation from Hardy-Weinberg equilibrium in a group of patients suggests that the SNP has a biological effect, while deviation in the control group or all groups suggests a systematic error in genotyping. The 1475 SNPs that were not in the Hardy-Weinberg equilibrium in the SU.VI.MAX control group ($P < 1.0 \times 10^{-3}$) were rejected in this way.

In total, 235 SNPs with low minor allelic frequency (<1%) in the global population were also filtered.

Haplotype inference. Haplotype inference was obtained using the rapid and accurate Shape-IT algorithm [11].

Linkage disequilibrium. For each SNP exhibiting a significant association, we looked for the other SNPs in linkage disequilibrium ($r^2 \geq 0.8$) in the HapMap population of Western European ancestry (CEU, HapMap data Release 21a/phase II, January 2007, on NCBI B35 assembly, dbSNP125; available at http://www.hapmap.org) to identify the genes possibly concerned by the associations. A SNP was assigned to a gene if it was located in the gene or in the 2 kb flanking regions (potential regulatory sequence); otherwise, it was considered intergenic.

Statistical analysis. For each SNP, we performed a standard case-control analysis by using the Fisher exact test (with PLINK software) to compare allelic distributions between the rapid progressors and the control subjects.

To take into account the multiple tests while controlling for the risk of false discovery, we computed for each *P* value a false-discovery rate (FDR) under the form of a *q* value: the *q* value is an estimate of the proportion of false-positive signals below a threshold *P* value [12]. The FDR computation is more complex but more powerful than the standard Bonferroni corrections [13–15], because it allows the identification of more true-positive signals. For polyfactorial diseases in which several genes are at stake, it thus provides a more adapted outlook on the GWAS results than the "all or nothing" Bonferroni cutoff. We thus used an FDR approach called local base estimating, with a 25% threshold for our case-control study.

For all the SNPs meeting this statistical threshold (Table 1), the quality of genotyping was individually rechecked with the BeadStudio software. We also checked that the allelic frequen-

This figure is available in its entirety in the online version of the *Journal of Infectious Diseases.*

Figure 1. Number of independent ($r^2 < 0.5$) single-nucleotide polymorphisms (SNPs) meeting the false-discovery rate threshold of 25%.

This figure is available in its entirety in the online version of the *Journal of Infectious Diseases*.

Figure 2. Quantile-quantile plot for expected (*red*) versus observed (*black*) *P* values from the comparison of rapid progressors with control subjects.

cies in the seropositive control population were similar to those in the seronegative SU.VI.MAX control population for those SNPs of interest, confirming that the observed associations were indeed linked to rapid progression.

Statistical reliability of the rapid progression results. The GRIV rapid progression cohort is unique, and no other independent cohort was readily available for replication. To check in an independent fashion the statistical relevance of the associations with rapid progression, we decided to evaluate the odds of obtaining just as many associations with the same statistical method (ie, the Fisher exact test and FDR cutoff) by comparing the genotypes of 1000 control subgroups (randomly extracted from the D.E.S.I.R. control cohort and composed of 85 subjects each) with the genotypes of the entire SU.VI.MAX control group. For each simulation, we counted the number of independent ($r^2 < 0.5$) SNPs meeting the statistical threshold of an FDR of 25% (Figure 1).

Identification of population stratification. To correct for possible population stratification at the intercontinental level, case and control genotypes were analyzed using STRUCTURE software (version 2.2; see [16] and http://pritch.bsd.uchicago.edu/software.html). We selected a set of 328 SNPs informative for ancestral origin (*F* statistics fixation index, >0.2) based on the Perlegen data set and separated by 5 Mb to avoid linkage disequilibrium. We also included genotypes obtained from unrelated individuals representing the 3 populations studied by the HapMap project to better separate our rapid progressors and control individuals according to their continent of origin. While nearly all case and control subjects were within the range of the white individuals from HapMap, one individual was outside the white subjects cluster in the rapid progressors (decreasing the rapid progression group from 86 to 85 subjects).

To avoid spurious associations resulting from possible population stratification or genotyping errors, a quantile-quantile plot was also produced by plotting the ranked values of the test statistics against the approximated expected order statistic (Figure 2). We also computed the genomic inflation factor λ [17]. The result ($\lambda = 1.038$), along with the quantile-quantile plot, suggested little overall effect of stratification.

RESULTS

After the quality control tests, a total of 291,119 autosomal SNPs were tested for association with rapid progression. Figure 3 depicts the distribution of the *P* values along the chromosomes, and Table 1 presents the best signals (for an extended list, see Table 2). Eight associations with FDRs ≤25% were identified, corresponding to 6 independent ($r^2 < 0.5$) SNPs. Since no replication was readily available, we performed simulations on independent control populations to evaluate the reliability of these results (see Methods and Figure 1). Overall, less than 1% of these tests yielded 6 or more independent signals with a FDR ≤25% (mean ± standard deviation = 0.4 ± 1.29), suggesting that most of the rapid progression associations found in this study are likely to be true positives. This result underscores the statistical relevance of the rapid progression associations, even with a modest sample size.

The best rapid progression signals were observed for SNPs on chromosome 1 (rs4118325 [$P = 6.09 \times 10^{-7}$] and rs10494056 [$P = 4.29 \times 10^{-6}$]) (Figure 3), which are in linkage disequilibrium themselves ($r^2 = 0.92$) and are also in linkage disequilibrium with SNPs of the *PRMT6* gene. The rs4118325-A allele

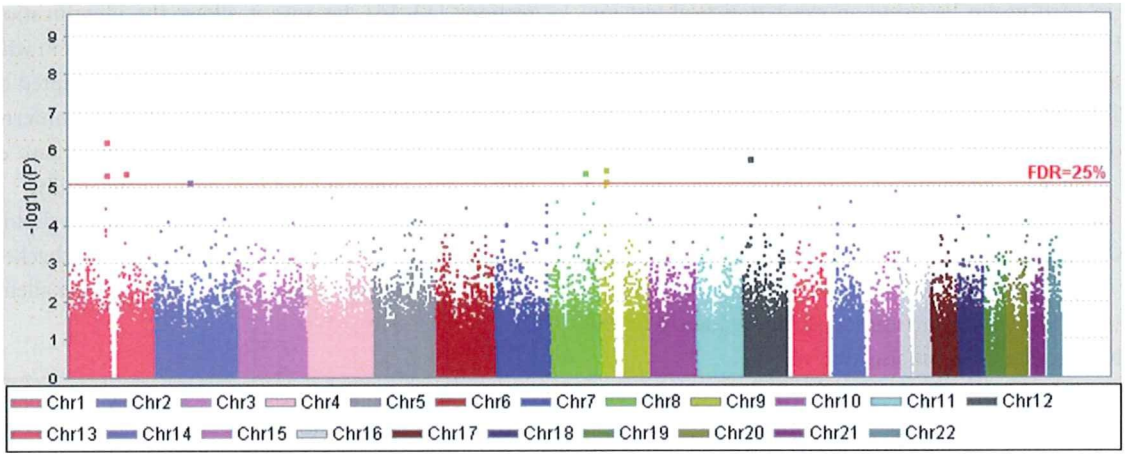

Figure 3. Distributions along the human autosomes of $-\log_{10}$(*P* values) obtained for the comparison of rapid progressors with control subjects. The red line marks the false-discovery rate threshold of 25%. Chr, chromosome.

Table 2. Fifty Best *P* Values Obtained for the Comparison between Rapid Progressors and Control Subjects

This table is available in its entirety in the online version of the *Journal of Infectious Diseases*.

and/or the alleles in linkage disequilibrium are associated with prevention of rapid progression, with an odds ratio (OR) of 0.24 [95% confidence interval [CI], 0.12–0.46] (Table 1). Another association was identified on chromosome 1 in *RXRG* gene ($P = 3.86 \times 10^{-6}$, OR, 3.29 [95% CI, 2.08–5.20]). Finally, SNPs modulating rapid progression were also found in *SOX5* ($P = 1.80 \times 10^{-6}$; OR, 0.45 [95% CI, 0.32–0.63]) and *TGFBRAP1* ($P = 7.04 \times 10^{-6}$; OR, 0.34 [95% CI, 0.20–0.57]). Two other independent SNPs, rs1360517 and rs3108919, met the FDR threshold of 25% (Table 1) but were not close to any known gene (distance, <20 kb).

To complete our analysis, we explored the influence of covariables—such as *CCR5*-Δ32 and *CCR5-P1* haplotypes [7], the HIV-1 infection mode (mucosal or parenteral), age at seroconversion, and sex—on all the associations presented in Table 1. None of them was found to affect the observed signals.

To deepen our genomic analysis, we also tried to combine our results with previously published data from the EuroCHAVI cohort, which assessed the viral load end point [3]. We combined their *P* values with ours using the classical Fisher method. Unfortunately, no combined *P* values met the FDR threshold of 25% (data not shown; best combined *P* value, 6.23×10^{-5}). The lack of significant common signals between the 2 studies may likely stem from the difference in inclusion criteria, in particular the use of viral load versus the use of

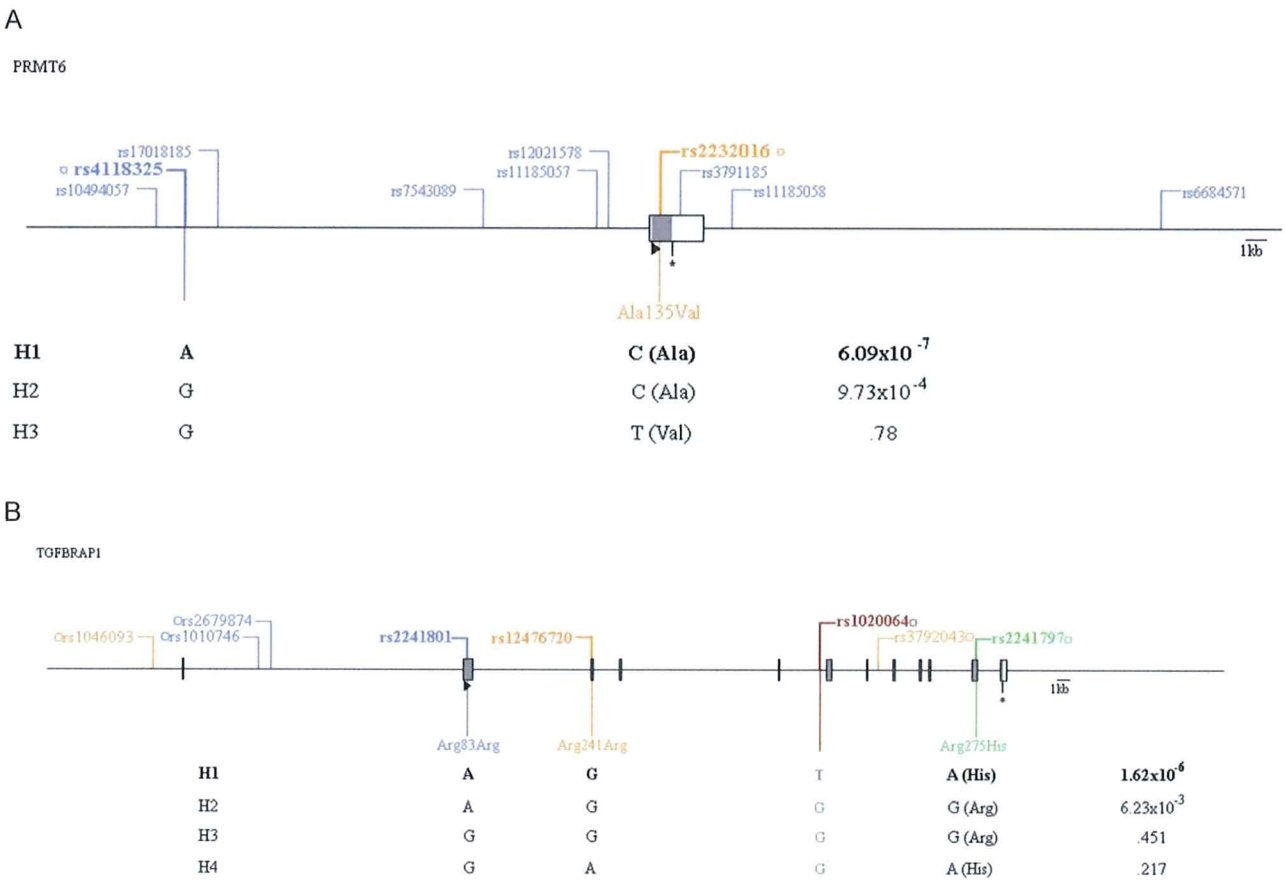

Figure 4. Haplotype maps for *PRMT6* (*A*) and *TGFBRAP1* (*B*). Exons and untranslated regions are symbolized by shaded and unshaded boxes, respectively. The positions of the ATG and stop codons are indicated by a triangle (▶) and by an asterisk (*), respectively. Single-nucleotide polymorphisms (SNPs) in high linkage disequilibrium ($r^2 > 0.9$) are represented with the same color, and the genotyped tagSNPs are marked with the symbol ¤. For greater clarity, only the exonic polymorphisms in linkage disequilibrium with the genotyped tagSNPs ($r^2 > 0.9$) are specified in panel B. The numbers in the right column correspond to the *P* values obtained when comparing the allelic frequency of each haplotype between rapid progressors and control subjects.

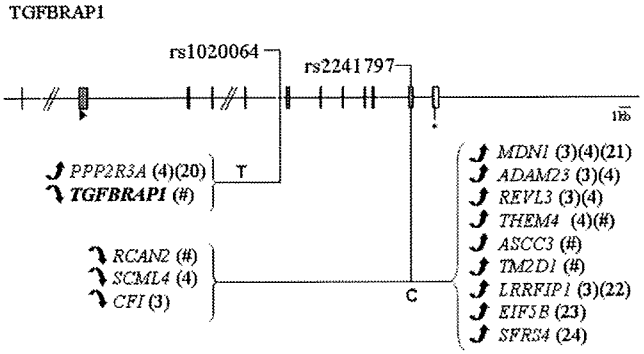

Figure 5. Correlation between some *TGFBRAP1* polymorphisms and differential gene expression according to the Genevar database [18] and the Dixon database [19]. Exons and untranslated regions are symbolized by shaded and unshaded boxes, respectively. The positions of the ATG and stop codons are indicated by a triangle (▶) and by an asterisk (*), respectively. The modulation of gene expression is indicated by the arrow direction: increased (↑) or decreased (↓) expression. The numbers correspond to the bibliographic references of previous works linking these genes to AIDS. The present genomewide association study of rapid progression is indicated by a pound sign (#).

CD4 T cell count, and also from the elimination of very rapid progressors in the Euro-CHAVI study, since such patients could not exhibit a sufficiently prolonged viral load set point [3].

In the past, we have often observed that haplotypes may be more informative than individual SNPs; this was notably shown for *CCR5* [7], *CXCR1* [6], and *HLA* [5]. We thus decided to explore the haplotype patterns for the genes exhibiting the best signals in our rapid progression GWAS, limiting the investigation to exonic and promoter SNPs (Figure 4). For *PRMT6*, we demonstrated that the effect of the *PRMT6* SNP rs4118325 could be tracking a haplotype (composed of rs4118325 and rs2232016 [Ala135Val]) through linkage disequilibrium (r^2 = 0.12, D' = 1) (Figure 4*A*). No differential messenger RNA (mRNA) expression could be significantly associated with either of these SNPs using the Genevar database [18] or the Dixon database [19].

For *TGFBRAP1*, the rs1020064 SNP was in high linkage disequilibrium (r^2 = 0.97) with a haplotype containing 3 exonic SNPs (rs2241801 [Arg83Arg], rs12476720 [Arg241Arg], and rs2241797 [Arg275His]) (Figure 4*B*). The SNP rs2241797 (Arg275His) appeared to be essential, since the signal disappeared when that SNP was removed from the haplotype ($P > 5.0 \times 10^{-2}$) and remained identical when either of the 2 synonymous SNPs were removed ($P = 1.62 \times 10^{-6}$). Interestingly, rs1020064 has been associated with differential expression of *TGFBRAP1* in the Genevar database and of *PPP2R3A* [20] in the Dixon database (Figure 5). More strikingly, in the Dixon database the nonsynonymous SNP rs2241797 was significantly associated with the differential expression of several proteins (Figure 5), among which 4 have been independently described to interact with HIV-1 (*MDN1* [21], *LRRFIP1* [22], *EIF5B* [23], and *SFRS4* [24]), and several have exhibited a positive association in an AIDS GWAS. Of note, no differential expression could be found for the 2 synonymous SNPs rs2241801 and rs12476720. For the *SOX5* rs1522232 and *RXRG* rs10800098 SNPs, no haplotype involving exonic or promoter SNPs and no differential mRNA expression could be found.

DISCUSSION

For the first time, the extreme HIV-1 rapid progression phenotype was specifically examined in a GWAS. The power of using the GRIV extreme design has been demonstrated by previous candidate gene studies [5–7, 25] and by the previous GWAS of nonprogressors [4]. Here we have identified 6 novel associations with rapid progression with ORs as high as 4, emphasizing the power of this extreme design in spite of a relatively modest sample size. As a comparison, there was only 1 independent ($r^2 < 0.5$) signal passing the FDR threshold of 25% in the GWAS of the nonprogressor GRIV cohort [4]. No replication was readily available, since this specific rapid progression design is rather unique in the world. The biological relevance of the rapid progression associations and their statistical validation through simulations with a second control group (<1% odds of finding as many independent signals) are nevertheless compelling. The use of an extreme population may provide a strong contrast and indeed help unravel new genetic factors, behaving as a magnifying glass [25, 26].

Four of the 6 SNPs associated with rapid progression to AIDS were clearly linked to a gene (distance, <2 kb). *SOX5* (OR, 0.45) encodes a transcription regulator notably expressed in lymph nodes and lymphoid tissues [27] and is also known to be involved in the transforming growth factor β (TGF-β)/SMAD chondrogenesis signaling pathway [28, 29]. There is no other experimental evidence linking this protein to the pathogenesis of HIV-1 infection. *RXRG* (OR, 3.29) encodes a retinoic acid nuclear receptor mediating the antiproliferative effects of retinoic acid and is known to repress HIV-1 transcription and replication [30, 31]. Several studies have also associated vitamin A deficiency with a bad prognosis in patients with AIDS, but the benefit of vitamin A supplementation in AIDS patients is still controversial [32, 33]. The genetic association for *PRMT6* (OR, 0.24) points toward a direct interaction between the host and the virus. Indeed, *PRMT6* codes for an arginine *N*-methyltransferase previously reported to methylate HIV-1 Tat and Rev, impairing their function [34, 35]. The product of *PRMT6* also methylates the high mobility group protein HMGA1, thereby modifying HMGA1 interactions with DNA [36], which could alter HIV-1 integration into the human genome [37]. The modulation of *PRMT6* transcription could

thus affect HIV-1 integration or replication and prevent rapid progression. Finally, an association with rapid progression was discovered for *TGFBRAP1* (OR, 0.34). *TGFBRAP1* is expressed in most lymphoid tissues and is involved in the TGF-β signaling pathway [38]. TGF-β is a pleiotropic immunosuppressive cytokine involved in immune homeostasis and in the differentiation of and balance between type 17 T helper (Th17) cells (T cells protecting the mucosal barrier integrity [39]) and regulatory T (T_{reg}) cells (T cells essential for immune suppression [40]). Interestingly, recent works have also supported a combined role for TGF-β and retinoic acid for T_{reg} cell differentiation [41, 42]. T_{reg} cells are rapidly induced after simian immunodeficiency virus (SIV) and HIV-1 infection [43, 44] and may have a deleterious effect during the chronic phase of infection [45, 46]. Moreover, an impairment in the balance between Th17 and T_{reg} cells during HIV-1 infection was recently described [40, 47, 48]. Finally, TGF-β can stimulate HIV-1 Tat transcription [49] and, reciprocally, be induced by Tat early during infection [50], supporting a key role for the intermingled effects of TGF-β and Tat in the early development of HIV-1 infection. Overall, our results support the central role played by the TGF-β pathway and the balance between Th17 and T_{reg} cells in AIDS progression. Interestingly, *PRMT6* and *TGFBRAP1* haplotypes involving promoter and exonic SNPs were also associated with disease progression; moreover, some of the haplotype SNPs could be associated with downstream differential mRNA expression (Figure 5).

In conclusion, as for all genetic studies, our results—including SNPs with higher *P* values, which were not discussed—will need to be confirmed by replication in other cohorts and by other investigations, such as fine gene mapping and biological experiments.

ANRS GENOMIC GROUP

The ANRS Genomic Group is composed of Prof Jean-François Delfraissy (Agence Nationale de Recherche sur le SIDA, Paris), Dr Laurence Meyer (Hôpital Kremin-Bicêtre, France), Prof Philippe Broët (Hôpital Kremlin-Bicêtre, France), Dr Cyril Dalmasso (Hôpital Kremlin-Bicêtre, France), Prof Patrice Debré (Hôpital La Salpétrière, Paris), Dr Ioannis Théodorou (Hôpital La Salpétrière, Paris), Prof Christine Rouzioux (Hôpital Necker, Paris), Cédric Coulonges (Conservatoire National des Arts et Métiers, Paris), Sigrid Le Clerc (Conservatoire National des Arts et Métiers, Paris), Sophie Limou (Conservatoire National des Arts et Métiers, Paris), and Prof Jean-François Zagury (Conservatoire National des Arts et Métiers, Paris).

Acknowledgments

We are grateful to all the patients and medical staff who have kindly collaborated with the GRIV project.

References

1. Kingsmore SF, Lindquist IE, Mudge J, Gessler DD, Beavis WD. Genome-wide association studies: progress and potential for drug discovery and development. Nat Rev Drug Discov **2008**; 7:221–30.
2. Dalmasso C, Carpentier W, Meyer L, et al. Distinct genetic loci control plasma HIV-RNA and cellular HIV-DNA levels in HIV-1 infection: the ANRS Genome Wide Association 01 study. PLoS ONE **2008**; 3:e3907.
3. Fellay J, Shianna KV, Ge D, et al. A whole-genome association study of major determinants for host control of HIV-1. Science **2007**; 317: 944–7.
4. Limou S, Le Clerc S, Coulonges C, et al. Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by *HLA* genes (ANRS Genomewide Association Study 02). J Infect Dis **2009**; 199:419–26.
5. Flores-Villanueva PO, Hendel H, Caillat-Zucman S, et al. Associations of MHC ancestral haplotypes with resistance/susceptibility to AIDS disease development. J Immunol **2003**; 170:1925–9.
6. Vasilescu A, Terashima Y, Enomoto M, et al. A haplotype of the human CXCR1 gene protective against rapid disease progression in HIV-1+ patients. Proc Natl Acad Sci U S A **2007**; 104:3354–9.
7. Winkler CA, Hendel H, Carrington M, et al. Dominant effects of CCR2-CCR5 haplotypes in HIV-1 disease progression. J Acquir Immune Defic Syndr **2004**; 37:1534–8.
8. Hercberg S, Galan P, Preziosi P, et al. Background and rationale behind the SU.VI.MAX Study, a prevention trial using nutritional doses of a combination of antioxidant vitamins and minerals to reduce cardiovascular diseases and cancers: SUpplementation en VItamines et Mineraux AntioXydants Study. Int J Vitam Nutr Res **1998**; 68:3–20.
9. Balkau B. An epidemiologic survey from a network of French Health Examination Centres, (D.E.S.I.R.): epidemiologic data on the insulin resistance syndrome. Rev Epidemiol Sante Publique **1996**; 44:373–5.
10. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet **2007**; 81:559–75.
11. Delaneau O, Coulonges C, Zagury JF. Shape-IT: new rapid and accurate algorithm for haplotype inference. BMC Bioinformatics **2008**; 9:540.
12. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc Ser B **1995**; 57:289–300.
13. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. Stati Med **1990**; 9:811–8.
14. Perneger TV. What's wrong with Bonferroni adjustments. BMJ **1998**; 316:1236–8.
15. Storey JD, Tibshirani R. Statistical significance for genomewide studies. PNAS **2003**; 100:9440–5.
16. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics **2000**; 155:945–59.
17. Devlin B, Roeder K. Genomic control for association studies. Biometrics **1999**; 55:997–1004.
18. Ge D, Zhang K, Need AC, et al. WGAViewer: software for genomic annotation of whole genome association studies. Genome Res **2008**; 18:640–3.
19. Dixon AL, Liang L, Moffatt MF, et al. A genome-wide association study of global gene expression. Nat Genet **2007**; 39:1202–7.
20. Ammosova T, Washington K, Debebe Z, Brady J, Nekhai S. Dephosphorylation of CDK9 by protein phosphatase 2A and protein phosphatase-1 in Tat-activated HIV-1 transcription. Retrovirology **2005**; 2:47.
21. Brass AL, Dykxhoorn DM, Benita Y, et al. Identification of host proteins required for HIV infection through a functional genomic screen. Science **2008**; 319:921–6.
22. Wilson SA, Brown EC, Kingsman AJ, Kingsman SM. TRIP: a novel double stranded RNA binding protein which interacts with the leucine rich repeat of flightless I. Nucleic Acids Res **1998**; 26:3460–7.
23. Wilson SA, Sieiro-Vazquez C, Edwards NJ, et al. Cloning and characterization of hIF2, a human homologue of bacterial translation initi-

ation factor 2, and its interaction with HIV-1 matrix. Biochem J **1999**; 342:97–103.
24. Exline CM, Feng Z, Stoltzfus CM. Negative and positive mRNA splicing elements act competitively to regulate human immunodeficiency virus type 1 vif gene expression. J Virol **2008**; 82:3921–31.
25. Huber C, Pons O, Hendel H, et al. Genomic studies in AIDS: problems and answers—development of a statistical model integrating both longitudinal cohort studies and transversal observations of extreme cases. Biomed Pharmacother **2003**; 57:25–33.
26. Froguel P, Blakemore AI. The power of the extreme in elucidating obesity. N Engl J Med **2008**; 359:891–3.
27. Su AI, Cooke MP, Ching KA, et al. Large-scale analysis of the human and mouse transcriptomes. Proc Natl Acad Sci U S A **2002**; 99:4465–70.
28. Furumatsu T, Tsuda M, Taniguchi N, Tajima Y, Asahara H. Smad3 induces chondrogenesis through the activation of SOX9 via CREB-binding protein/p300 recruitment. J Biol Chem **2005**; 280:8343–50.
29. Ikeda T, Kawaguchi H, Kamekura S, et al. Distinct roles of Sox5, Sox6, and Sox9 in different stages of chondrogenic differentiation. J Bone Miner Metab **2005**; 23:337–40.
30. Kiefer HL, Hanley TM, Marcello JE, Karthik AG, Viglianti GA. Retinoic acid inhibition of chromatin remodeling at the human immunodeficiency virus type 1 promoter: uncoupling of histone acetylation and chromatin remodeling. J Biol Chem **2004**; 279:43604–13.
31. Maeda Y, Yamaguchi T, Hijikata Y, et al. All-trans retinoic acid attacks reverse transcriptase resulting in inhibition of HIV-1 replication. Hematology **2007**; 12:263–6.
32. Austin J, Singhal N, Voigt R, et al. A community randomized controlled clinical trial of mixed carotenoids and micronutrient supplementation of patients with acquired immunodeficiency syndrome. Eur J Clin Nutr **2006**; 60:1266–76.
33. Mehta S, Fawzi W. Effects of vitamins, including vitamin A, on HIV/AIDS patients. Vitam Horm **2007**; 75:355–83.
34. Invernizzi CF, Xie B, Richard S, Wainberg MA. PRMT6 diminishes HIV-1 Rev binding to and export of viral RNA. Retrovirology **2006**; 3:93.
35. Xie B, Invernizzi CF, Richard S, Wainberg MA. Arginine methylation of the human immunodeficiency virus type 1 Tat protein by PRMT6 negatively affects Tat interactions with both cyclin T1 and the Tat transactivation region. J Virol **2007**; 81:4226–34.
36. Sgarra R, Lee J, Tessari MA, et al. The AT-hook of the chromatin architectural transcription factor high mobility group A1a is arginine-methylated by protein arginine methyltransferase 6. J Biol Chem **2006**; 281: 3764–72.
37. Li L, Yoder K, Hansen MS, Olvera J, Miller MD, Bushman FD. Retroviral cDNA integration: stimulation by HMG I family proteins. J Virol **2000**; 74:10965–74.
38. Charng MJ, Zhang D, Kinnunen P, Schneider MD. A novel protein distinguishes between quiescent and activated forms of the type I transforming growth factor beta receptor. J Biol Chem **1998**; 273:9365–8.
39. Manel N, Unutmaz D, Littman DR. The differentiation of human T_H-17 cells requires transforming growth factor-beta and induction of the nuclear receptor RORgammat. Nat Immunol **2008**; 9:641–9.
40. de St Groth BF, Landay AL. Regulatory T cells in HIV infection: pathogenic or protective participants in the immune response? AIDS **2008**; 22:671–83.
41. Benson MJ, Pino-Lagos K, Rosemblatt M, Noelle RJ. All-trans retinoic acid mediates enhanced T reg cell growth, differentiation, and gut homing in the face of high levels of co-stimulation. J Exp Med **2007**; 204:1765–74.
42. Mucida D, Park Y, Kim G, et al. Reciprocal TH17 and regulatory T cell differentiation mediated by retinoic acid. Science **2007**; 317:256–60.
43. Estes JD, Wietgrefe S, Schacker T, et al. Simian immunodeficiency virus–induced lymphatic tissue fibrosis is mediated by transforming growth factor beta 1-positive regulatory T cells and begins in early infection. J Infect Dis **2007**; 195:551–61.
44. Kekow J, Wachsman W, McCutchan JA, Cronin M, Carson DA, Lotz M. Transforming growth factor beta and noncytopathic mechanisms of immunodeficiency in human immunodeficiency virus infection. Proc Natl Acad Sci U S A **1990**; 87:8321–5.
45. Kared H, Lelievre JD, Donkova-Petrini V, et al. HIV-specific regulatory T cells are associated with higher CD4 cell counts in primary infection. AIDS **2008**; 22:2451–60.
46. Weiss L, Donkova-Petrini V, Caccavelli L, Balbo M, Carbonneil C, Levy Y. Human immunodeficiency virus-driven expansion of CD4+CD25+ regulatory T cells, which suppress HIV-specific CD4 T-cell responses in HIV-infected patients. Blood **2004**; 104:3249–56.
47. Brenchley JM, Paiardini M, Knox KS, et al. Differential Th17 CD4 T-cell depletion in pathogenic and nonpathogenic lentiviral infections. Blood **2008**; 112:2826–35.
48. Favre D, Lederer S, Kanwar B, et al. Critical loss of the balance between Th17 and T regulatory cell populations in pathogenic SIV infection. PLoS Pathog **2009**; 5:e1000295.
49. Li JM, Shen X, Hu PP, Wang XF. Transforming growth factor beta stimulates the human immunodeficiency virus 1 enhancer and requires NF-kappaB activity. Mol Cell Biol **1998**; 18:110–21.
50. Zocchi MR, Contini P, Alfano M, Poggi A. Pertussis toxin (PTX) B subunit and the nontoxic PTX mutant PT9K/129G inhibit Tat-induced TGF-beta production by NK cells and TGF-beta–mediated NK cell apoptosis. J Immunol **2005**; 174:6054–61.

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 Best[1]

[1]Division of Gene and Cell Based Therapy, King's College London School of Medicine, London, United Kingdom; [2]Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Yoshida, Japan; [3]Centre National de Génotypage, Institut Génomique, Commissariat a l'Energie Atomique, Evry, France; and [4]Department 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, $\alpha_2\gamma_2$) 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.[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]

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.[5-7] The *HBS1L-MYB* intergenic region alone contributes approximately 20% of the overall trait variance in healthy European whites,[5,8] and 3% to 7% of the trait variance in African-American and Brazilian patients with sickle cell anemia.[6]

The panel of single nucleotide polymorphisms in the *HBS1L-MYB* region that account for the effects of the 6q locus[6,8-10] 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.[8] 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

Submitted March 12, 2009; accepted June 7, 2009. Prepublished online as *Blood* First Edition paper, June 15, 2009; DOI 10.1182/blood-2009-03-210146.

*S.L.T. and S.B. contributed equally to this work.

The online version of this article contains a data supplement.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

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 *HBS1L-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 × 10^6 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.[11,17] 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.[11]

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)[18,19] 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 delta C_T (difference in C_T values between treated and untreated DNAs) to a linear scale and plotted as a function of primer position. Values were normalized to the negative control *NEFM* (supplemental methods and supplemental Figure 1) to account for differences in treatment conditions.

Chromatin immunoprecipitation

Antibodies used for ChIP experiments included anti-diacetylated 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 × 10^7 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).

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.[20]

Input and ChIP DNA were amplified using a whole genome amplification kit (WGA1; Sigma-Aldrich) applying a previously described protocol adapted for ChIP material.[21] 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[22] between positions 135 323 209 and 135 582 003 (NCBI build 36[23]) that encompass the entire *MYB* and *HBS1L* 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 25mer oligo was shifted by 2 bases within *HBS1L* and *MYB*, and by one base within the intergenic region. Oligonucleotides were designed for both DNA strands. A total of 224 258 oligonucleotide probes was included in each array.

Total RNA was extracted from erythroid cells (liquid culture phase II, days 3 and 5)[11,17] 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.

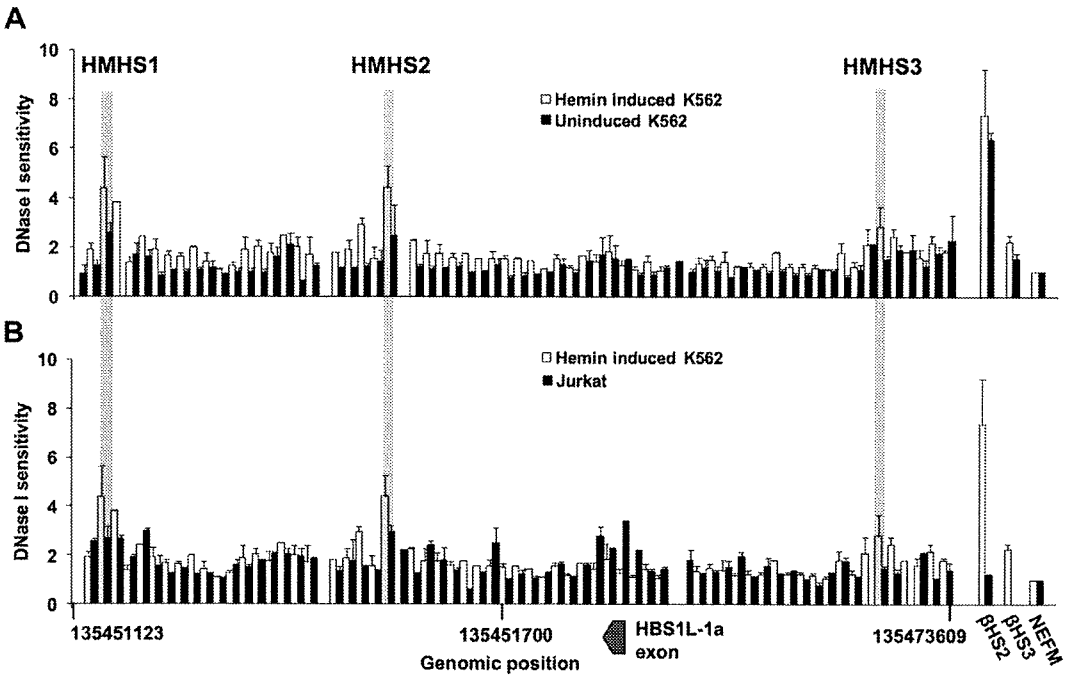

Figure 1. DNase I digestion profiles of the *HBS1L-MYB* intergenic region in *HMIP* block 2. DNase I sensitivity was analyzed by real-time PCR using 68 primer pairs spanning the 24-kb *HMIP-2* region. Control primers targeting the β-globin LCR HS2 and HS3 (βHS2 and βHS3) and the *NEFM* gene were also included in experiments. Relative sensitivity to DNase I for each target was calculated from delta C_T values between treated and untreated samples and normalized to *NEFM*. DNase I sensitivity was plotted as a function of primer position. Error bars represent differences between 2 biologic repeats. (A) DNase I sensitivity in *HMIP-2* compared between hemin-induced and uninduced K562 cells. Three DNase I hypersensitive sites (HMHS1, HMHS2, and HMHS3; gray bars) were identified in hemin-induced K562 cells. (B) DNase I sensitivity in *HMIP-2* compared between hemin-induced K562 cells and Jurkat cells. The 3 hypersensitive sites in hemin-induced K562 cells showed no or less sensitivity to DNase I in Jurkat cells. Hypersensitivity was instead observed in Jurkat cells at the site of the *HBS1L-1a* exon, which showed no hypersensitivity in K562 cells (induced and uninduced). The βHS3 control was not included for DNase I sensitivity analyses in Jurkat cells.

Results

HBS1L-MYB intergenic region is sensitive to DNase I in erythroid cells

Variants in block 2 of the *HBS1L-MYB* intergenic 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* 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.[24,25] 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 βHS2 and βHS3 (positive controls) and *NEFM* (negative control) were included for K562 cells and βHS2 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 *HBS1L-MYB* (HM) HS1, HS2, and HS3, in particular, showed a marked increase in sensitivity compared with background levels (Figure 1A). DNase I sensitivity also increased for βHS2 and βHS3 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, βHS2 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 HMHS3, 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 *HBS1L* exon (exon 1a), which showed a low degree of sensitivity in K562 cells (both uninduced and induced). This is consistent with expression of the *HBS1L-1a* transcript in Jurkat cells but not in K562 cells.[8] To validate the promoter prediction at *HBS1L-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 *HBS1L* exon *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 *HBS1L-MYB* 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 over the entire 70-Mb region in coincidence with RNAP II signals at active genes. The gene-free *HBS1L-MYB* interval showed strong signals for all antibodies in erythroid cells, indicating a high level of activity in this region.

Publicly available expression data from the UCSC Genome Browser (http://genome.ucsc.edu)[27] revealed that a majority of the genes that showed strong signals for histone acetylation, RNAP II, and GATA-1 were hematopoietic specific, but a few ubiquitous genes were also included. The genes represented a wide range of functions including transcription factors, adhesion receptors, and signaling proteins, reflecting a broad target repertoire for GATA-1 as a transcription factor (supplemental Table 1).

The *HBS1L-MYB* intergenic region is highly active in erythroid precursors

ChIP-chip data of GATA-1, AcH3, AcH4, and RNAP II 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 RNAP II signal in the 6q23 locus in erythroid precursors were consistent with previous expression analysis of genes in the locus[11] (Figure 2). The *PDE7B* and *ALDH8A1* genes are not expressed in erythroid precursors and consistently showed no signal of acetylation or RNAP II interaction. *AHI1* 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 RNAP II signal and histone acetylation around the promoter regions as well as coding regions. The RNAP II antibody used detects both nonphosphorylated inactive RNAP II 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, RNAP II 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 RNAP II signals were also

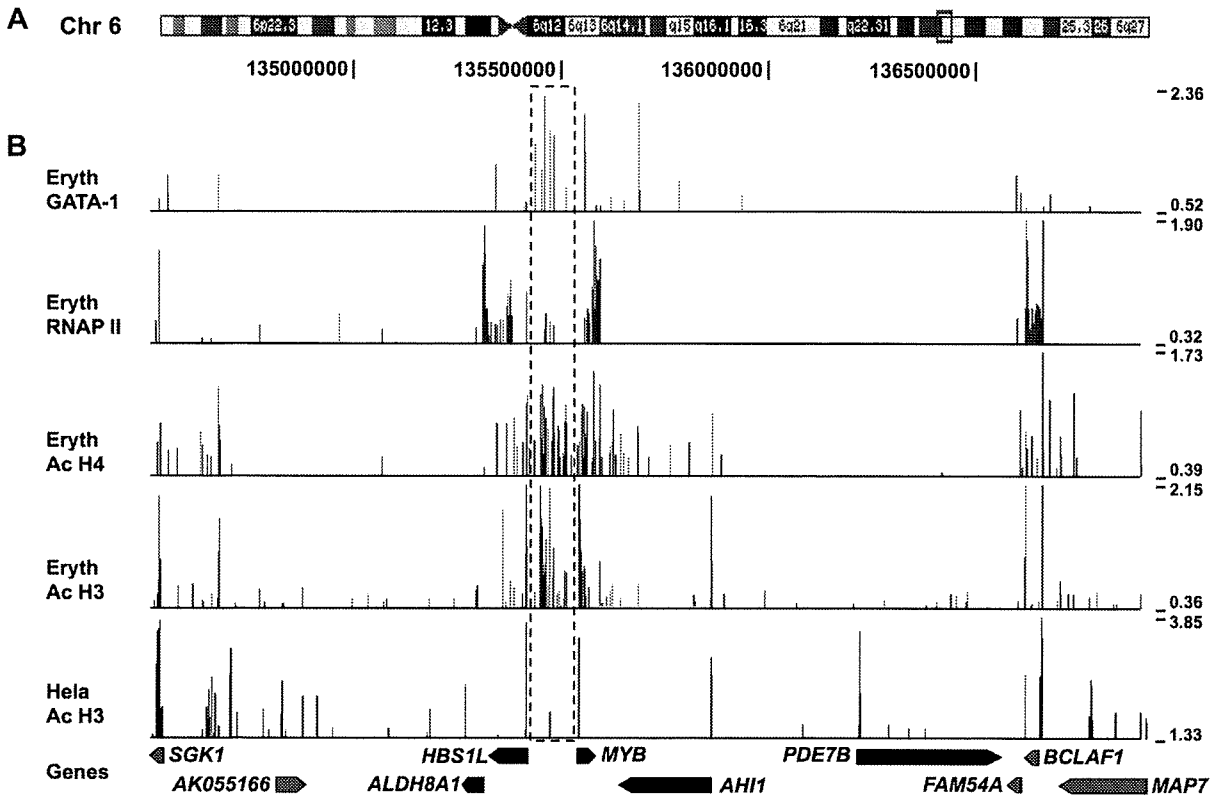

Figure 2. ChIP-chip data for the 6q23 locus in erythroid precursor and HeLa cells. (A) The box on chromosome 6 represent a 2.5-Mb region (position chr6: 134 503 000-136 928 000), which is covered in this figure and includes the 6q23 locus. (B) Results from ChIP-chip experiments for GATA-1, RNAP II, AcH4, 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 Browser[27] using autoscale settings.

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.

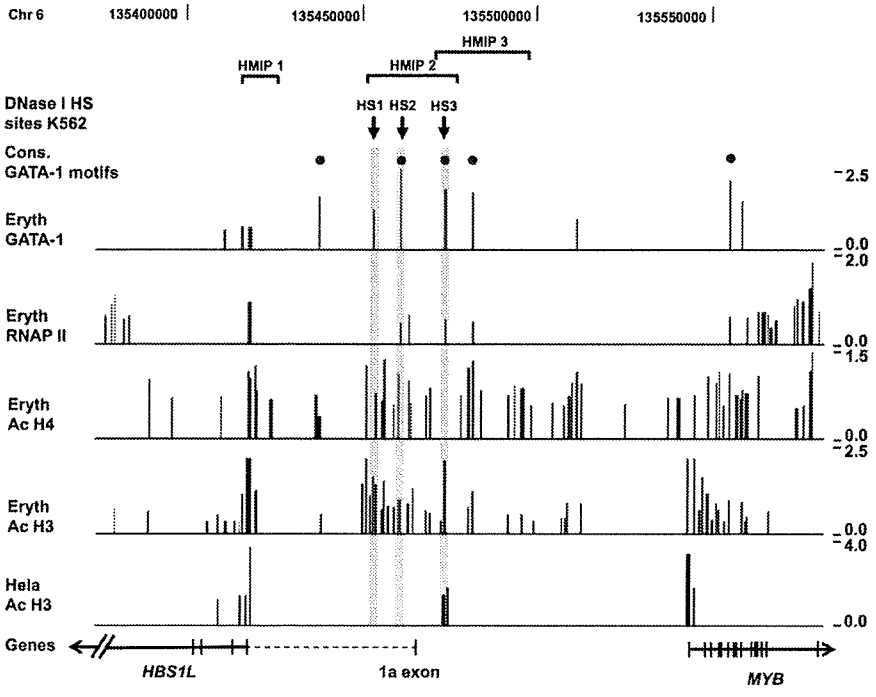

Figure 3. ChIP-chip data for the *HBS1L-MYB* intergenic region. Results from ChIP-chip experiments for GATA-1, RNAP II, AcH4, and AcH3 in erythroid precursors and AcH3 in HeLa cells. The figure covers a 210-kb region (position chr6: 135 373 000-135 582 000) and includes a part of the *HBS1L* gene, the 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.

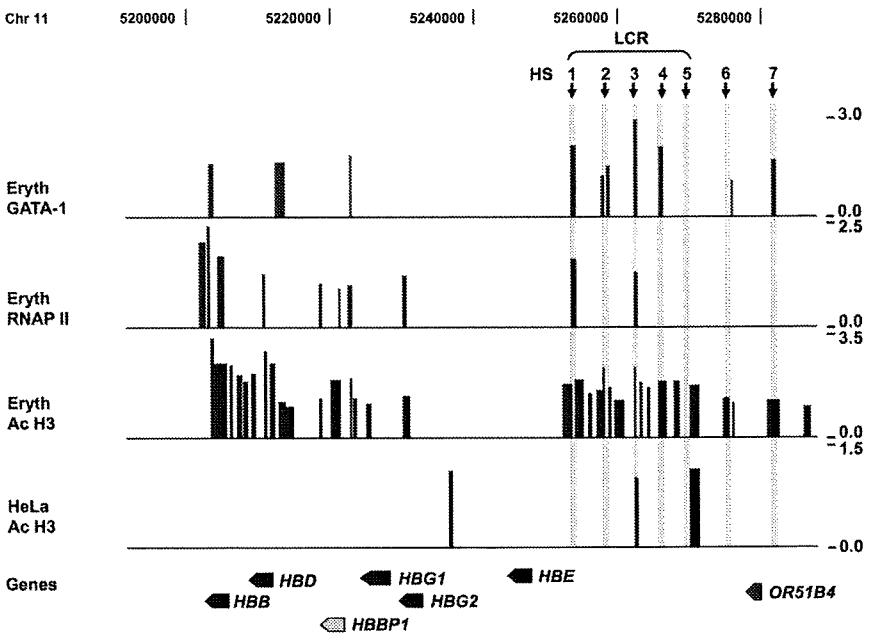

Figure 4. ChIP-chip data for the β globin locus in erythroid precursor and HeLa cells. Results from ChIP-chip experiments for GATA-1, RNAP II, and AcH3 in erythroid precursors and AcH3 in HeLa cells analyzed on the custom designed array. The figure covers a 120-kb region (position chr11: 5 180 000-5 302 000) that includes the entire β globin locus. The globin genes are indicated below and include β globin *(HBB)*, δ globin *(HBD)*, γ globin *(HBG1* and *HBG2)* and ε globin *(HBE)*, and the βψ pseudogene *(HBBP1)*. The region also includes the olfactory gene *(OR51B4)*. Hypersensitive sites in the β-globin LCR (HS1-HS5) and the upstream HS6 and HS7 are marked with ↓.

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 *HBS1L-MYB* intergenic interval that would facilitate the evaluation of the region upstream of *HBS1L* and *MYB* as a potential distal regulatory element. Interestingly, strong similarities in patterns of histone acetylation as well as GATA-1 and RNAP II interactions were observed between the *HMIP* block 2 and 3 regions and the α and β globin control regions.

In the β globin locus, 2 isolated domains of activity (AcH3 signals) were clearly discernable and these were concentrated around the β LCR and a region covering the β and δ globin genes *(HBB* and *HBD)* and the area around the βψ pseudogene *(HBBP1*; Figure 4). These 2 domains of activity were separated by an inactive domain comprising the ε globin *(HBE)* 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.[29] HeLa cells showed little signal for H3 acetylation in the β globin locus. Some signal was detected in the LCR at HS3 and HS5, and upstream 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.[30] 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 previous observations,[26] 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 ζ 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 *HBS1L*, 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 *HBS1L* 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 *HMIP-2*. Several GATA-1 binding sites were also identified in the *HBS1L-MYB* intergenic interval; within *HMIP-2*, all the GATA-1 signals coincided with the 3 DNase I hypersensitive sites identified in induced K562 cells, providing strong support for the sites being active regulatory elements in erythroid cells. Interestingly some of the GATA-1 sites also coincided with weak RNAP II binding. Weak

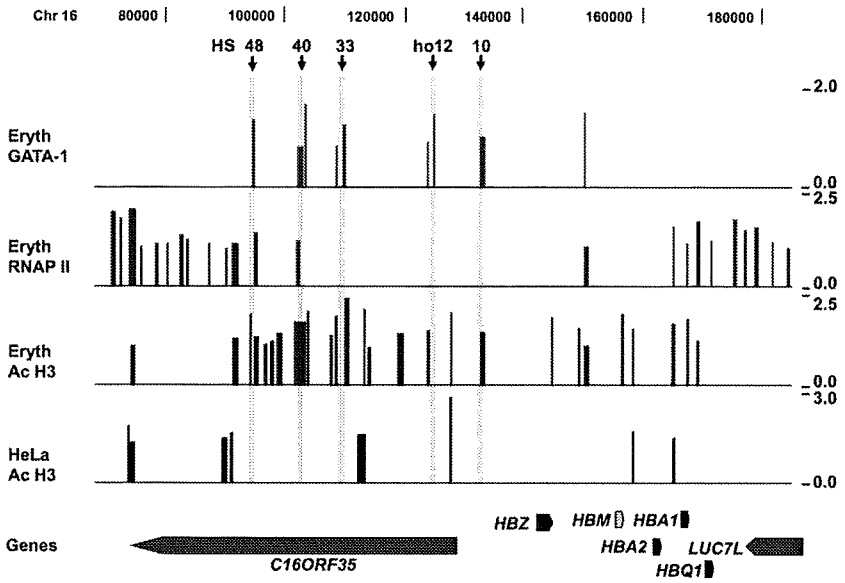

Figure 5. ChIP-chip data for the α globin locus in erythroid precursor and HeLa cells. Results from ChIP-chip experiments for GATA-1, RNAP II, and AcH3 in erythroid precursors and AcH3 in HeLa cells analyzed on the custom-designed array. The figure covers a 120-kb region (position chr16: 70 000-187 000) that includes the α-globin locus. The globin genes are indicated below and include ζ globin *(HBZ)*, α globins *(HBA1* and *HBA2)*, and θ globin *(HBQ1)*, and the *ψα2* pseudogene *(HBM)*. Other genes include *C16ORF35* and *LUC7L*. Hypersensitive sites in the human α globin locus (HS48, HS40, HS33, and HS10) and the human orthologous mouse HS12 (ho12) are marked 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.[31] 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.[32]

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,

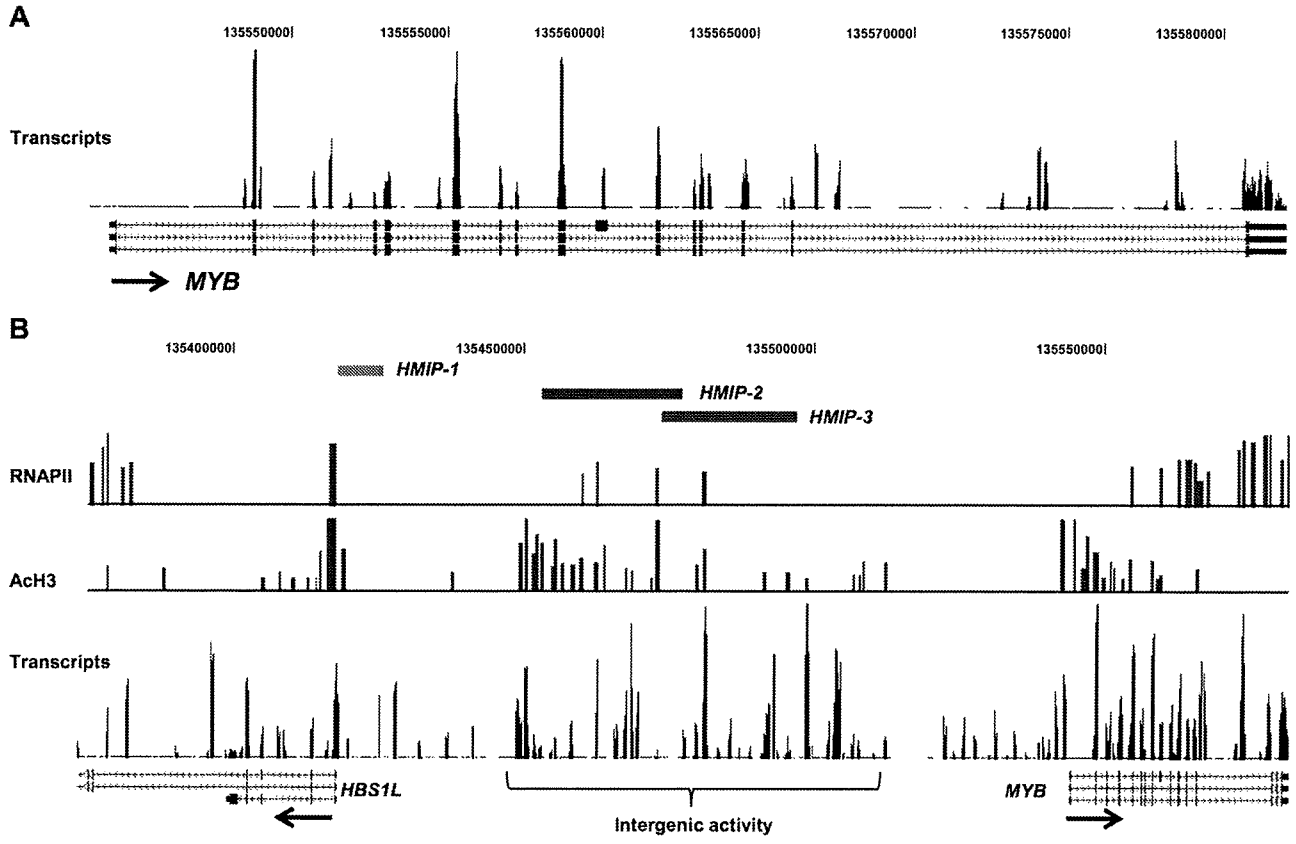

Figure 6. Transcript tiling array data from erythroid precursor cells. Results from 1 representative experiment are presented. (A) Transcript signal from the *MYB* genomic region shows very clear activity in exons and demonstrates that the array design and methodology are robust, given that *MYB* is highly expressed in the cells used. Occasional strong intronic signals were not coincident with spliced or unspliced ESTs and may therefore represent uncharacterized control elements or ncRNA. (B) Transcript data including *HBS1L*, *MYB*, and the intergenic region. The HbF-associated *HMIP* blocks 1 and 2 cover a region of high transcriptional activity. Again, there is no evidence of either spliced or unspliced ESTs. Given their magnitude in relation to the *MYB* signals and their complete consistency between different cultures tested, the intergenic signals are likely to represent important regulatory sequences.

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.[33] 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 *HBS1L-MYB* region contains a distal regulatory locus. Except for *HBS1L*-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[34]), 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.[29,35,36] The patterns of histone acetylation, RNAP II 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.[37] 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.[38] 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.[37] 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.[39]

Previously, we showed that *MYB* is a quantitative trait gene, with variable expression in healthy adults.[11] 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.[11] 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 thrombocytosis.[39-44] 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.[45] 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.[45] We propose that the *HBS1L-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 *HBS1L-MYB* interval on chromosome 6q have been shown to be highly associated, not only with HbF levels, but also with the control of 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 *HBS1L* and *MYB* genes, located within *HMIP-2* and *-3* blocks as identified in genetic association studies.

Delineation of the key variants in this *HBS1L-MYB* control region may lead to an improved understanding of *MYB* control and dysregulation[46] that underlies many of the leukemias and cancers, and may also provide targets for therapeutic activation of HbF[47] 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

1. Platt OS, Brambilla DJ, Rosse WF, et al. Mortality in sickle cell disease: life expectancy and risk factors for early death. *N Engl J Med.* 1994;330(23): 1639-1644.
2. Ho PJ, Hall GW, Luo LY, Weatherall DJ, Thein SL. Beta thalassemia intermedia: is it possible to consistently predict phenotype from genotype? *Br J Haematol.* 1998;100(1):70-78.
3. Weatherall DJ, Clegg JB. The Thalassaemia Syndromes. 4th Ed. Oxford, United Kingdom: Blackwell Science; 2001.
4. Garner C, Tatu T, Reittie JE, et al. Genetic influences on F cells and other hematologic variables: a twin heritability study. *Blood.* 2000;95(1):342-346.
5. Menzel S, Garner C, Gut I, et al. A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. *Nat Genet.* 2007;39(10):1197-1199.
6. Lettre G, Sankaran VG, Bezerra MA, et al. DNA polymorphisms at the BCL11A, *HBS1L-MYB,* and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. *Proc Natl Acad Sci U S A.* 2008;105(33):11869-11874.
7. Sedgewick AE, Timofeev N, Sebastiani P, et al. BCL11A is a major HbF quantitative trait locus in three different populations with beta-hemoglobinopathies. *Blood Cells Mol Dis.* 2008;41(3):255-258.
8. Thein SL, Menzel S, Peng X, et al. Intergenic 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):11346-11351.
9. Uda M, Galanello R, Sanna S, et al. Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. *Proc Natl Acad Sci U S A.* 2008;105(3):1620-1625.
10. So CC, Song YQ, Tsang ST, et al. The HBS1L-MYB intergenic region on chromosome 6q23 is a quantitative trait locus controlling fetal haemoglobin level in carriers of beta-thalassaemia. *J Med Genet.* 2008;45(11):745-751.
11. Jiang J, Best S, Menzel S, et al. cMYB is involved in the regulation of fetal hemoglobin production in adults. *Blood.* 2006;108(3):1077-1083.
12. Menzel S, Jiang J, Silver N, et al. The HBS1L-MYB intergenic region on chromosome 6q23.3 influences erythrocyte, platelet, and monocyte counts in humans. *Blood.* 2007;110(10):3624-3626.
13. Wallrapp C, Verrier S-B, Zhouravleva G, et al. The product of the mammalian orthologue of the Saccharomyces cerevisiae HBS1 gene is phylogenetically related to eukaryotic release factor 3 (eRF3) but does not carry eRF3-like activity. *FEBS Lett.* 1998;440(3):387-392.
14. Oh IH, Reddy EP. The myb gene family in cell growth, differentiation and apoptosis. *Oncogene.* 1999;18(19):3017-3033.
15. Cantor AB, Orkin SH. Transcriptional regulation of erythropoiesis: an affair involving multiple partners. *Oncogene.* 2002;21(21):3368-3376.
16. Vegiopoulos A, Garcia P, Emambokus N, Frampton J. Coordination of erythropoiesis by the transcription factor c-Myb. *Blood.* 2006;107(12): 4703-4710.
17. Fibach E, Manor D, Oppenheim A, Rachmilewitz EA. Proliferation and maturation of human erythroid progenitors in liquid culture. *Blood.* 1989; 73(1):100-103.
18. McArthur M, Gerum S, Stamatoyannopoulos G. Quantification of DNaseI-sensitivity by real-time PCR: quantitative analysis of DNaseI-hypersensitivity of the mouse beta-globin LCR. *J Mol Biol.* 2001;313(1):27-34.
19. Dorschner MO, Hawrylycz M, Humbert R, et al. High-throughput localization of functional elements by quantitative chromatin profiling. *Nat Methods.* 2004;1(3):219-225.
20. National Center for Biotechnology Information. Gene Expression Omnibus (GEO). http://www.ncbi.nlm.nih.gov/geo. Accessed June 2, 2009.
21. O'Geen H, Nicolet CM, Blahnik K, Green R, Farnham PJ. Comparison of sample preparation methods for ChIP-chip assays. *Biotechniques.* 2006;41(5):577-580.
22. Kapranov P, Drenkow J, Cheng J, et al. Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res.* 2005;15(17):987-997.
23. National Center for Biotechnology Information. Build 36. http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=10090. Accessed March 14, 2007.
24. Eisbruch A, Blick M, Evinger-Hodges MJ, et al. Effect of differentiation-inducing agents on oncogene expression in a chronic myelogenous leukemia cell line. *Cancer.* 1988;62(6):1171-1178.
25. Villeval JL, Pelicci PG, Tabilio A, et al. Erythroid properties of K562 cells: effect of hemin, butyrate and TPA induction. *Exp Cell Res.* 1983;146(2): 428-435.
26. De Gobbi M, Anguita E, Hughes J, et al. Tissue-specific histone modification and transcription factor binding in {alpha} globin gene expression. *Blood.* 2007;110(13):4503-4510.
27. University of California Santa Cruz. UCSC Genome Browser. http://genome.ucsc.edu. Accessed May 15, 2008.
28. Lian Z, Karpikov A, Lian J, et al. A genomic analysis of RNA polymerase II modification and chromatin architecture related to 3′ end RNA polyadenylation. *Genome Res.* 2008;18(8):1224-1237.
29. Miles J, Mitchell JA, Chakalova L, et al. Intergenic transcription, cell-cycle and the developmentally regulated epigenetic profile of the human beta-globin locus. *PLoS ONE.* (htttp://www.ncbi.nlm.nih.gov/entrez/query.fcgi?) 2007;2(7):e630.
30. Behringer RR, Hammer RE, Brinster RL, Palmiter RD, Townes TM. Two 3′ sequences direct adult erythroid-specific expression of human β-globin genes in transgenic mice. *Proc Natl Acad Sci U S A.* 1987;84(20):7056-7060.
31. Heintzman ND, Stuart RK, Hon G, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet.* 2007;39(3):311-318.
32. Vakoc CR, Letting DL, Gheldof N, et al. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol Cell.* 2005;17(3):453-462.
33. Layon ME, Ackley CJ, West RJ, Lowrey CH. Expression of GATA-1 in a non-hematopoietic cell line induces beta-globin locus control region chromatin structure remodeling and an erythroid pattern of gene expression. *J Mol Biol.* 2007;366(3): 737-744.
34. National Center for Biotechnology Information. http://www.ncbi.nlm.nih.gov. Accessed February 20, 2009.
35. Ashe HL, Monks J, Wijgerde M, Fraser P, Proudfoot NJ. Intergenic transcription and transinduction of the human β-globin locus. *Genes Dev.* 1997;11(19):2494-2509.
36. Gribnau J, Diderich K, Pruzina S, Calzolari R, Fraser P. Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus. *Mol Cell.* 2000;5(2): 377-386.
37. Hanlon L, Barr NI, Blyth K, et al. Long-range effects of retroviral insertion on c-myb: overexpression may be obscured by silencing during tumor growth in vitro. *J Virol.* 2003;77(2):1059-1068.
38. Haviernik P, Festin SM, Opavsky R, et al. Linkage on chromosome 10 of several murine retroviral integration loci associated with leukaemia. *J Gen Virol.* 2002;83(4):819-827.
39. Mukai HY, Motohashi H, Ohneda O, Suzuki N, Nagano M, Yamamoto M. Transgene insertion in proximity to thec-myb gene disrupts erythroid-megakaryocytic lineage bifurcation. *Mol Cell Biol.* 2006;26(21):7953-7965.
40. Mucenski ML, McLain K, Kier AB, et al. A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis. *Cell.* 1991;65(4): 677-689.
41. Emambokus N, Vegiopoulos A, Harman B, Jenkinson E, Anderson G, Frampton J. Progression through key stages of haemopoiesis is dependent on distinct threshold levels of c-Myb. *EMBO J.* 2003;22(17):4478-4488.
42. Carpinelli MR, Hilton DJ, Metcalf D, et al. Suppressor screen in Mpl−/− mice: c-Myb mutation causes supraphysiological production of platelets in the absence of thrombopoietin signaling. *Proc Natl Acad Sci U S A.* 2004;101(17):6553-6558.
43. Sandberg ML, Sutton SE, Pletcher MT, et al. c-Myb and p300 regulate hematopoietic stem cell proliferation and differentiation. *Dev Cell.* 2005; 8(2):153-166.
44. Kasper LH, Boussouar F, Ney PA, et al. A transcription-factor-binding surface of coactivator p300 is required for haematopoiesis. *Nature.* 2002;419(6908):738-743.
45. Lu J, Guo S, Ebert BL, et al. MicroRNA-mediated control of cell fate in megakaryocyte-erythrocyte progenitors. *Dev Cell.* 2008;14(6):843-853.
46. García P, Frampton J. Hematopoietic lineage commitment: miRNAs add specificity to a widely expressed transcription factor. *Dev Cell.* 2008; 14(6):815-816.
47. Perrine SP. Hemoglobin F: new targets, new path. *Blood.* 2006;108(3):783-784.

Mod Rheumatol (2009) 19:401–406
DOI 10.1007/s10165-009-0173-1

ORIGINAL ARTICLE

Lack of association between tyrosine kinase 2 (*TYK2*) gene polymorphisms and susceptibility to SLE in a Japanese population

Chieko Kyogoku · Akio Morinobu · Kunihiro Nishimura · Daisuke Sugiyama · Hiroshi Hashimoto · Yoshiaki Tokano · Tsuneyo Mimori · Chikashi Terao · Fumihiko Matsuda · Takayoshi Kuno · Shunichi Kumagai

Received: 24 February 2009 / Accepted: 31 March 2009 / Published online: 8 May 2009

Abstract Tyrosine kinase 2 (*TYK2*) is a type I interferon (IFN) signaling pathway gene and was previously reported to be a risk factor for systemic lupus erythematosus (SLE) in Caucasian populations. In order to test for its genetic association with SLE in a Japanese population, *TYK2* single nucleotide polymorphisms (SNPs), rs2304256, rs12720270 and rs280519, were genotyped. A case–control association study was performed in a total of 411 Japanese SLE patients and 467 healthy controls. Linkage disequilibrium (LD) among *TYK2* SNPs was examined. According to the data from 94 healthy controls, non-synonymous rs2304256 resulting in Val → Phe substitution was revealed to be in a LD with rs12720270 and rs280519. Therefore, we further genotyped rs2304256 as a tag SNP in the full sample sets. As a result, no differences in genotype distribution and allelic frequencies of rs2304256 were found between SLE patients and healthy controls. In conclusion, *TYK2* is not a genetic risk factor for SLE in a Japanese population. Our result suggests that there is an ethnic difference in the susceptibility genes for SLE.

Keywords Systemic lupus erythematosus (SLE) · Tyrosine kinase 2 (*TYK2*) · Single nucleotide polymorphism (SNP) · Type I interferon (IFN) signaling pathway · Association study

C. Kyogoku (✉) · A. Morinobu · S. Kumagai
Department of Clinical Pathology and Immunology,
Kobe University Graduate School of Medicine,
7-5-1 Kusunoki-cho, Chuo-ku, Kobe, Hyogo 650-0017, Japan
e-mail: k-chieko@umin.ac.jp

K. Nishimura · D. Sugiyama
Department of Evidence-based Laboratory Medicine,
Kobe University Graduate School of Medicine, Kobe, Japan

H. Hashimoto · Y. Tokano
Department of Internal Medicine and Rheumatology,
Juntendo University School of Medicine, Tokyo, Japan

T. Mimori · C. Terao
Department of Rheumatology and Clinical Immunology,
Kyoto University Graduate School of Medicine, Kyoto, Japan

F. Matsuda
Center for Genomic Medicine,
Kyoto University Graduate School of Medicine, Kyoto, Japan

T. Kuno
Division of Molecular Pharmacology and Pharmacogenomics,
Department of Biochemistry and Molecular Biology, Kobe
University Graduate School of Medicine, Kobe, Japan

Introduction

Systemic lupus erythematosus (SLE) (OMIM 152700) is a systemic inflammatory autoimmune disease characterized by production of autoantibodies against nuclear antigens and deposition of immune complex (IC) in various organs. The pathogenesis of SLE is complex and involves a combination of multiple genetic and environmental factors [1]. Among these, the role of type I interferon (IFN) in SLE is suggested by the clinical finding of elevated serum levels of type I IFN in the patients, as well as by recent reports showing the genetic contributions of type I IFN signaling pathway genes to SLE [2]. The latter includes the reports of IFN-regulatory factor 5 (*IRF5*) [3–6], tyrosine kinase 2 (*TYK2*) [3], and signal transducer and activator of transcription 4 (*STAT4*) [7].

TYK2 belongs to the Janus kinase (JAK) family [8] and is located on chromosome 19p13.2, which previous linkage analysis suggests is a susceptibility region for SLE [9]. TYK2 is a cytoplasmic protein whose recruitment to IFN-α

receptor (IFNAR) is followed by the binding of type I IFN, which initiates a JAK-STAT signaling cascade leading to transcription of a second wave of type I IFN and IFN-inducible genes [10].

Sigurdsson et al. recently reported that two single nucleotide polymorphisms (SNPs) in *TYK2*, rs2304256 and rs12720356, are significantly associated with SLE in Scandinavian and Finnish populations [3]. These SNPs respectively cause a Val → Phe substitution at position 362 in the JH4 region and a Ile → Ser substitution at position 684 in the JH2 pseudo-kinase region. Following that study, Graham et al. reported a significant association of individual *TYK2* SNPs and their haplotypes with susceptibility to SLE in UK families [11]. A significant core association was observed in/between rs12720270 and rs280519, which respectively tag the unique under-transmitted haplotype 2 and the over-transmitted haplotype 1. These SNPs are intronic, though they are situated in the vicinity of exon-intron boundaries. Graham et al. also mentioned that rs2304256, whose association with SLE was suggested by Sigurdsson et al., is in a strong linkage disequilibrium (LD) with rs12720270, while rs12720356 lies outside the core association region they detected.

Our aim in this study was to assess the association between *TYK2* SNPs and SLE in a Japanese population by comparing the genotypes of SNPs detected in Caucasian populations (rs12720270, rs2304256 and rs280519), in Japanese patients and healthy controls. This was the first attempt to demonstrate an association study of *TYK2* with SLE in Asian population. We also examined the LD between *TYK2* SNPs that were genotyped in Japanese healthy controls.

Patients and methods

Study subjects

Among a total of 411 SLE patients, the first set of 69 cases was recruited from Kobe University Hospital, and a second set of 342 cases was recruited from Juntendo University Hospital ($n = 125$) and Kyoto University Hospital ($n = 217$). Patients from Kobe University Hospital are 4 males and 65 females with average age 40.1 ± 13.4 years, from Juntendo University Hospital are 13 males and 112 females with average age 39.9 ± 14.6 years, and from Kyoto University Hospital are 12 males and 205 females with average age of 40.8 ± 12.9 years. A total of 467 healthy individuals, consisting of a first set of 94 controls (58 males and 36 females, average age 39.4 ± 8.7 years) and a second set of 373 controls (162 males and 211 females, average age 42.9 ± 16.9 years), were healthy Japanese volunteers, mainly living or working in the Kobe area. All patients and controls were unrelated Japanese. Patients were diagnosed and classified for the presence or absence of clinical characteristics, and all fulfilled the American College of Rheumatology criteria for SLE [12]. Written informed consents were obtained from all donors. The study was approved by the ethics committee at Kobe University and proceeded in agreement with the 1964 Helsinki declaration.

Genomic DNA isolation

Peripheral blood mononuclear cells (PBMCs) were obtained from heparinized blood by Ficoll-Hypaque gradient centrifugation. Genomic DNA was extracted from the PBMCs using a QIAamp blood kit (QIAGEN, Hilden, Germany).

Genotyping

SNPs were genotyped using a 7900HT Fast Real-Time PCR System with TaqMan® SNP Genotyping Assays and TaqMan® Genotyping Master Mix (Applied Biosystems, Foster City, CA). We used invented assays for rs12720270 (Assay ID C-1931075-10) and s2304256 (Assay ID C-25473911-10), and a custom-made assay for rs280519 (forward primer, 5′-CCACATCCCACCCAGAGG-3′; reverse primer, 5′-TTGCCTGCCCCAACCA-3′; probes, VIC-CCCCCTATC[A]TCGTACC and FAM-CCCCTATC [G]TCGTACC; square brackets indicate the location of the SNPs). The genotypes of patients and controls, regarding these three SNPs, were assessed by a standard analysis of TaqMan end-point fluorescent data, in which clear identification of three tight clusters permitted assignment of allele-1 (AA), allele-2 (BB) and heterozygous (AB) genotypes for each SNP.

Sequencing

To confirm the accuracy of the SNP genotyping assays, we randomly selected 12 individuals (6 patients and 6 controls) for each SNP and undertook direct sequencing of the SNP and its adjacent area using a BigDye® Terminator v3.1 Cycle Sequencing Kit and an ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA). The primers used were: for rs12720270, GAAGGTCCTCC ATCTGCTTA (forward) and TGCTGACACAGTGCTC TTTC (reverse); for rs2304256, AAATAGCCGTCCACC AGCGA (forward) and GGTCAGGATTCTTCTCTCTG (reverse); and for rs280519, GTTAGCAGCTGATCT CCCAG (forward) and TCCGAAAGTTCCCCATTGAG (reverse).

Statistical analysis

The association between SNPs and disease susceptibility for SLE was assessed using the χ^2 test with 2 × 2 and 2 × 3 contingency tables with Yates' correction. Values of $P < 0.05$ were considered significant.

Calculation of LD (D' and r^2) using SNP genotyping data from 94 healthy Japanese was performed using Haploview v3.32 (http://www.broad.mit.edu/mpg/haploview/).

Results

Association of *TYK2* SNPs with SLE in the Japanese

Data from the International HapMap Project (http://www.hapmap.org/) [13] indicate that among the four *TYK2* SNPs identified in the studies of Caucasian populations [3, 11], rs12720356 is absent or quite rare in Asian populations. These data showed that rs12720356 is non-polymorphic in 45 JPTs (Japanese) and 45 HCBs (Han Chinese), whereas it exists uniquely in Caucasian populations with a ~10% minor allele frequency (MAF). For that reason, we decided to genotype rs12720270, rs2304256 and rs280519, but not rs12720356, in 69 Japanese SLE patients and 94 healthy controls using TaqMan® SNP Genotyping Assays. In addition, we sequenced the region of each SNP in 12 randomly selected individuals (6 patients and 6 controls) and confirmed that the genotype data obtained by the two methods matched.

We found no significant difference in the genotype or allele frequencies for any of the SNPs examined (Table 1). However, the frequencies of the homozygous T allele for rs12720270 (SLE: 18.8%, control: 10.6%), homozygous A allele for rs2304256 (SLE: 18.8%, control: 11.7%) and homozygous G allele for rs280519 (SLE: 26.5%, control: 22.3%) tended to be higher in the SLE patients than the controls. The MAFs of rs12720270 T (Japanese 37.2% vs. Caucasians 17.0%) and rs2304256 A (38.3% vs. 25.0%) were 1.5–2.2 times higher in Japanese than Caucasians (Table 1) [11]. On the other hand, the MAF of rs280519 was quite similar in Japanese and Caucasians (~50%; Table 1) [11]. All the genotyping data were consistent with the Hardy-Weinberg equilibrium.

Table 1 Genotype and allele frequencies of *TYK2* rs12720270, rs2304256 and rs280519 in 69 Japanese SLE patients and 94 healthy controls

SNP	Genotype/allele	SLE n=69	%	Control n=94	%	P*	Allele frequency in Caucasians*** (%)	LD**** (Control n=94)
rs12720270	C/C	25	36.2	34	36.2	NS**		
	C/T	31	44.9	50	53.2			
	T/T	13	18.8	10	10.6			
								D'=1, r^2=0.956
	C allele	81	58.7	118	62.8		83.0	
	T allele	57	41.3	70	37.2		17.0	
rs2304256	C/C	25	36.2	33	35.1	NS		
	C/A	31	44.9	50	53.2			
	A/A	13	18.8	11	11.7			
	C allele	81	58.7	116	61.7		75.0	
	A allele	57	41.3	72	38.3		25.0	
								D'=1, r^2=0.662
rs280519	A/A	17	25.0	24	25.5	NS		
	A/G	33	48.5	49	52.1			
	G/G	18	26.5	21	22.3			
	A allele	67	49.3	97	51.6		50.0	
	G allele	69	50.7	91	48.4		50.0	

Association of *TYK2* SNPs with susceptibility to SLE was calculated using the χ^2 test with 2 × 2 and 2 × 3 contingency tables $P < 0.05$ was considered significant. Calculation of LD (D' and r^2) between two SNPs were carried out in 94 controls using Haploview v3.32

*P value, ** not significant, *** data are from Graham et al. [11], **** linkage disequilibrium

Linkage disequilibrium between *TYK2* SNPs

We examined the LD between the SNPs using genotyping data from the 94 healthy controls. As shown in Table 1, nearly absolute LD was observed between rs12720270 and rs2304256 ($D' = 1$, $r^2 = 0.956$); accordingly they were located only 108 bp apart each other. The rs12720270 T allele and rs2304256 A allele, which were both increased in the SLE patients, were almost completely linked (called "perfect proxy set") with a ~1% minor haplotype (data not shown). There was also strong LD between rs2304256 and rs280519 ($D' = 1$, $r^2 = 0.662$), which were located 2,719 bp apart each other. Again the SLE-increased A allele of rs2304256 was linked to the SLE-increased G allele of rs280519 with a ~10% minor haplotype (data not shown). These data suggested that rs12720270, rs2304256 and rs280519 are on the same haplotype block.

Association of rs2304256 with SLE in the Japanese

Because we did not find any statistical significance but did see a tendency for certain alleles to be increased in SLE, we further genotyped a second set of 342 SLE patients and 373 healthy controls. This time, one of the three SNPs, rs2304256, was chosen as the tag SNP and genotyped in the additional samples, as the other two SNPs were in tight LD with the tested one (Table 1). Table 2 shows the genotyping results from the second sample set and the combined sample set. Data from the second sample set were similar to those from the first sample set. However, when combined to give samples of 411 SLE patients and 467 controls, association of *TYK2* rs2304256 with SLE did not reach statistical significance [$P = 0.1$ when comparing (C/C + C/A) vs. A/A; SLE: CC 38.9%, CA 43.8%, AA 17.3%, control: CC 37.5%, CA 49.5%, AA 13.1%]. Again the genotyping data were consistent with the Hardy-Weinberg equilibrium.

Table 2 Association of *TYK2* rs2304256 in 411 Japanese SLE patients and 467 healthy controls

SNP	Genotype/ allele	SLE		Control		*P**
		n	%	*n*	%	
rs2304256		$n = 342$		$n = 373$		
Second set	C/C	135	39.5	142	38.1	NS**
	C/A	149	43.6	181	48.5	
	A/A	58	17.0	50	13.4	
	C allele	419	61.3	465	62.3	
	A allele	265	38.7	281	37.7	
Combined		$n = 411$		$n = 467$		
	C/C	160	38.9	175	37.5	NS***
	C/A	180	43.8	231	49.5	
	A/A	71	17.3	61	13.1	
	C allele	500	60.8	581	62.2	
	A allele	322	39.2	353	37.8	

Association of *TYK2* SNP with susceptibility to SLE was calculated using the χ^2 test with 2 × 2 and 2 × 3 contingency tables. $P < 0.05$ was considered significant

*P value, ** not significant, *** $P = 0.1$ when comparing (C/C + C/A) vs. A/A

Association of clinical characteristics with rs2304256

We next examined the genotype distribution of *TYK2* rs2304256 in SLE patients exhibiting different clinical features (lupus nephritis, CNS disease, serositis) and expressing various anti-nuclear antibodies [anti-ds DNA, anti-Sm, anti-RNP, anti-SS-A/Ro, anti-SS-B/La, aCL (anti-cardiolipin) and rheumatoid factor] (data not shown). We observed no correlation between any clinical features and the genotype distribution for rs2304256.

Discussion

Since the time SLE patients were first reported to display elevated serum type I IFN levels, extensive examination has revealed the crucial role played by type I IFN in the pathogenesis of SLE [2]. Moreover, recent studies of type I IFN signaling pathway genes have revealed an association of *TYK2, IRF5* and *STAT4* with susceptibility to SLE in Caucasian populations [3, 7]. Among these, *IRF5* was subsequently reported also to be a risk factor in Japanese [6] and Korean [14] populations, and *STAT4* in Japanese population [15], though *TYK2* has not been reported in any other populations than Caucasian. Here we tested for the first time an association of *TYK2* to SLE in an Asian population, suggesting *TYK2* is not a risk factor for SLE in a Japanese population.

The inconsistent results between Japanese and Caucasian populations could be understood due to the different genetic backgrounds of the study populations. In fact, it is noteworthy that the MAFs for *TYK2* SNPs in Japanese are different from those of Caucasians (Table 1) [11]. There are many controversial reports on SLE-associated genes in the different ethnic groups. For example, the protein tyrosine phosphatase non-receptor type 22 (*PTPN22*) gene is a widely convinced susceptibility gene for SLE in the Caucasian population [16], but not for Japanese SLE. The main reason for this is that the allele frequency of *PTPN22* R620 W polymorphism is exclusively small in a Japanese compared to in a Caucasian population [17]. Therefore, we think our result does not deny the hypothesis that *TYK2* may be a Caucasian-specific risk factor for SLE.