SUPPLEMENTARY MATERIALS

SUMMARY

SUPPLEMENTARY NOTE 3
Empirical evaluation of genomic power 3
Genotyping in the replication cohort 3
In-depth analyses at known loci 4
β-globin gene cluster 4
HBS1L-MYB intergenic region 5
BCL11A 6
Gene-gene interaction analyses 6
$\beta^0$39 specific analysis 7
Clinical impact assessment 8
REFERENCES 10

SUPPLEMENTARY FIGURES

Supplementary Figure 1. Empirical evaluation of genomic power upon reference panel population. 13
Supplementary Figure 2. Manhattan and QQ plots for the assessed hemoglobin types. 14
Supplementary Figure 3. Regional plots of loci associated with HbA1. 16
Supplementary Figure 4. Regional plots of loci associated with HbA2. 18
Supplementary Figure 5. Regional plots of loci associated with HbF. 22

SUPPLEMENTARY TABLES

Supplementary Table 1. Characteristics of the studied cohort. 26
Supplementary Table 2. Association results for the $\beta^0$39 and –α 3.7 deletion type I. 27
Supplementary Table 3. Joint-association analysis results for loci with multiple associated variants. 28
Supplementary Table 4. Validation of top genome-wide findings. 29

Supplementary Table 5. Functional annotation of signals reported in Table 1. 30
Supplementary Table 6. Association results at SNPs previously associated with hemoglobin or red blood cell indices.....31
Supplementary Table 7. Genome-wide significant signals using 1000 Genomes phase III as reference panel.....37
Supplementary Table 8. Evidence of relationships between suggestive and genome-wide significant loci.....38
Supplementary Table 9. Markers associated with suggestive p-values to a single trait and demonstrating genome-wide significant pleiotropic effects considering between-trait combined p-values. 39
Supplementary Table 10. Summary of Extended Association for $\beta^0$39.....40
Supplementary Table 11. Effects of genome-wide significant loci in males and females.....41

SUPPLEMENTARY NOTE

Empirical evaluation of genomic power

To represent the additional power gained by the sequencing effort in the SardiNIA cohort, we empirically evaluated the power provided by imputation with different reference panels. When imputing the SardiNIA cohort using either the Sardinian reference panel or the 1000 Genomes reference panel phase III (version 5)[1], imputation quality metrics vary drastically for low frequencies. Consequently, the effective sample size available for discovery, defined by the number of samples multiplied by imputation accuracy metric, is reduced. We therefore calculated, for different frequency ranges, the power to detect a variant taking into account the resulting effective sample sizes, multiplying N for the average imputation accuracy of variants in each frequency range (as presented in **Supplementary Figure 1**). We selected SNPs that passed imputation accuracy filters in both imputation runs as described in Pistis et al., 2014[2], and separately analysed SNPs present in both panels and those specific to each panel. We consider the SardiNIA cohort roughly equivalent to a group of 3,000 unrelated individuals, and estimated that for a trait following a normal distribution this sample has 80% power to detect a variant explaining 1.3% of the variance at a $5x10^{-8}$ level of significance. If imputation accuracy were perfect for all variants, power to detect a variant with this effect size would be 80% for all frequency ranges. Taking into account the effect of imperfect genotype inference on effective sample size, power drops slightly below 80% only for rare variants (MAF <0.5%) when imputation is performed with the Sardinian panel, while a substantial decrease is notable even for variants with MAF <5% when 1000 Genomes data are used for imputation. For variants that are only imputable using the Sardinian panel, power is obviously 0% with the 1000 Genomes panel. For such variants, average imputation accuracy is high for variants with MAF >=0.5%, and estimated power remains over 60% (see **Supplementary Figure 1**). By contrast, for variants that are only imputable using the 1000 Genomes panel, power is below 60% even for more frequent alleles (MAF <5%).

Genotyping in the replication cohort

Replication tests were performed in the TwinsUK cohort[3]. Genotyping was performed using a combination of Illumina arrays (HumanHap300, HumanHap610Q, 1M-Duo and 1.2MDuo 1M). Normalized intensity data were pooled for each of the three arrays separately (with 1M-Duo and 1.2MDuo 1M pooled together). For each dataset, the Illuminus calling algorithm was used to assign genotypes in the pooled data. No calls were assigned if an individual's most likely genotype was called with less than a posterior probability threshold of 0.95. Validation of pooling was achieved via visual inspection of 100 randomly selected shared SNPs for overt batch effects. Finally, intensity cluster plots of significant SNPs were visually inspected for overdispersion biased no calling and/or erroneous genotype assignment. SNPs exhibiting any of these characteristics were discarded. Quality control filters were applied to each of the three datasets separately. Specifically, samples were excluded if 1) sample call rate <98%; 2) heterozygosity across all SNPs

≥2 standard deviations from the sample mean; 3) evidence of non-European ancestry as assessed by PCA comparison with HapMap3 populations; 4) observed pairwise IBD probabilities suggestive of sample identity errors; or 5) misclassified monozygotic and dizygotic twins based on IBD probabilities were corrected. SNPs were excluded if 1) Hardy-Weinberg p-value $<1x10^{-6}$, assessed in a set of unrelated samples; 2) MAF<1%, assessed in a set of unrelated samples; or 3) SNP call rate <97% (SNPs with MAF≥5%) or <99% (for 1%≤ MAF <5%). Alleles of all three datasets were aligned to HapMap2 or HapMap3 forward strand alleles. Data sets were then combined into a unique set.

Prior to merging, we performed pairwise comparison among the three datasets and further excluded SNPs and samples to avoid spurious genotyping effects, identified as follows: 1) concordance at duplicate samples <1%; 2) concordance at duplicate SNPs <1%; 3) visual inspection of QQ plots for logistic regression applied to all pairwise dataset comparisons; (iv) Hardy-Weinberg p-value $<1x10^{-6}$, assessed in a set of unrelated samples; or (v) observed pairwise IBD probabilities suggestive of sample identity errors. During merging, we selected the largest array for individuals typed with two different arrays. The merged dataset consists of 5,654 individuals (2,040 from the HumanHap300, 3,461 from the HumanHap610Q and 153 from the HumanHap1M and 1.M arrays) and up to 874,733 SNPs depending on the dataset (HumanHap300: 303,940, HumanHap610Q: 553,487, HumanHap1M and 1.M: 874,733).

In-depth analyses at known loci

β-globin gene cluster

Thalassemia and different hematological diseases are caused by genetic variation in this genomic region, which has been intensively studied in the past decades. However, our results suggest that some of the polymorphisms identified here are more functionally relevant than what already described. To evaluate whether our best association signal, rs67385638, was driven by any previously reported association, we corrected the association analysis using 40 previously correlated markers (r^2 <=0.80):

- 3 variants found within the (AT)x(T)y motif (within 5248828-5248852bp),
- 2 markers of the intergenic region containing the *HBD-HBG1* haplotypes (5258827-5259219bp),
- rs57966301 and rs2855039 from the pre-*HBG1* haplotype,
- the XmnI polymorphism (rs7482144), rs2855122 and rs2855121 from the pre-*HBG2* haplotype,
- 9 variants within the 3'HS, 5'HS1, 5'HS2, 5'HS3, 5'HS4 and 5'HS5 of the β-globin gene cluster (considering boundaries 5225972-5226553bp, 5297046-5297919bp, 5301801-5302172bp, 5305887-5306107bp, 5309436-5309702bp and 5312654-5312760bp respectively),
- and 21 variants from the ERV-9 5' LTR region (5314490-5316184bp).

Even after adding these 40 markers as covariates together with age, age^2, gender and β^0 mutations, rs67385638 was still the most associated marker ($p=3.10x10^{-07}$). Three previously correlated markers not used as covariates because of their r^2 >0.80 with rs67385638, showed

weaker evidence of association (rs2855039 with $p=1.6x10^{-05}$, rs2855121 with $p=6.5x10^{-05}$ and rs7482144 with $p=2.2x10^{-04}$). Noteworthy, rs67385638 is more strongly associated than the classically reported XmnI polymorphism.

In the same region, we identified an independently associated variant rs2855122. This variant along with rs2855121 and rs10128653, are known to define the pre-Ggamma haplotype, previously associated with variations of HbF levels in sickle cell and β-thalassemia patients[4]. However rs10128653 is monomorphic in the SardiNIA cohort. We hypothesized that our 2 genome-wide associated signals, rs67385638 and rs2855122, could better describe the variation of HbF levels in our cohort than the pre-Ggamma haplotypes. Therefore we tested, on 1,164 unrelated individuals, both the pre-Ggamma haplotype (here rs2855122 and rs2855121) and the haplotype formed by rs67385638 with rs2855122, estimating the contribution to variability of all possible allelic configurations at each haplotype (omnibus test in pLink)[5]. We observed that the haplotype we newly construct indeed correlates more strongly with HbF variations ($p=1.1x10^{-04}$ versus $7.1x10^{-04}$). This is in line with 1) the fact that rs2855122 was not nominally significant before conditioning on rs67385638 (p=0.1363), possibly indicating a combined action of these two variants, and 2) functional annotations of these 2 markers, showing evidence of chromatin interactions and RNA Polymerase II binding in K562 cell-lines between the regions containing these 2 markers (chr11:5274899..5283703-chr11:5287409..5292802,11 and chr11:5273135..5283072-chr11:5287409..5292550,13 from ENCODE/GIS-Ruan K562 Pol2 ChIA-PET Interactions)[6]. This further supports the notion that they jointly contribute to the transcription of gamma-globin and/or switching between fetal to adult globins.

As for marker rs67385638, it is predicted to introduce a binding site for C-EBP-beta by ALGEN-PROMO[7,8] and a CP2 binding site by TFSEARCH[9] engines. It has already been reported that at the G-CRE site where the independent signal lies, CREB-1 interacts with transcription factors of the basic Leucine Zipper family, of which C-EBP-beta is a member[10]. CP2 is known to be involved in transcriptional activation of the human gamma-globin genes in erythroid cells as part of the stage selector protein complex that activates gamma and epsilon globin gene expression[11,12].

Finally, rs2855122 has also been implicated in gamma-globin expression induced by histone deacetylase inhibitors[10,13]. Because its C allele is associated with both a decrease of HbF and an opposite increase of HbA1 and HbA2, rs2855122 is a particularly strong candidate for a primary effect on the fetal to adult hemoglobin switch, and eventually as a predictor of inter-individual variation in drug response to histone deacetylase inhibitors.

HBS1L-MYB intergenic region

In our GWAS results, with SNPs imputed using the Sardinian sequencing data, the top signals at this locus coincide for HbA2 and HbF (locus 4 in **Supplementary Figure 4** and locus 2 in **Supplementary Figure 5**), at SNP rs9399137. When using the 1000 Genomes phase III[1] imputation panel, and therefore assessing both SNPs and indels, these signals showed the same statistical evidence for association of the well-known 3bp TAC deletion rs66650371 ($p=1.0x10^{-82}$ for HbF and $p= 2.0x10^{-20}$ for HbA2). We noticed that the previously reported most associated marker[14], rs7775698, overlaps with the deletion. On the sequenced samples, we investigated both SNPs and the deletion status, looking at raw read count and alignment, and compared genotypes with those

called using ExomeChip. Among the sequenced individuals, only a small fraction carried the SNP (MAF = 0.003) while a larger fraction carried the deletion (MAF = 0.163). Genotypes from the array instead indicate all samples bearing the deletion as carriers of SNP. Therefore we can predict that results previously reported for this SNP could be biased by the presence of the deletion.

Finally, we tested interaction terms of rs66650371 and rs11754265 with all markers in the intergenic region, and rs11754265 demonstrated highly significant interaction terms with different markers upstream or within *MYB* (rs62429816 with $p=3.2x10^{-12}$, rs9389278 with $p=2.7x10^{-11}$ and rs7744765 with $p=3.0x10^{-11}$). This element could support long range interaction evidences observed in previous studies[15,16], and the detection of an independent signal with stronger evidences of eQTL than the lead variant could reconcile previous findings of a region containing eQTLs for *HBS1L*[17] but with *MYB* as a modifier of HbF levels[18].

BCL11A

In the second intron of *BCL11A* gene, the top 3 signals are in strong LD and hardly distinguishable: rs4671393 with $p=2.596x10^{-130}$, rs766432 with $p=1.825x10^{-129}$ ($r^2=0.96$ with rs4671393) and rs1427407 with $p=5.597x10^{-129}$ ($r^2=0.92$ with rs4671393). Conditioning on each of these 3 markers results point to 2 independent signals in high LD ($r^2=0.94$): rs13019832 with $p=9.122x10^{-33}$ and rs7606173 with $p=6.092x10^{-33}$ respectively.

From all 2-markers combinations including one top and its independent signal the lowest non-nominally significant residual variance for the other variants of these 7 were obtained when conditioning on both rs4671393 and rs13019832 and when conditioning on both rs1427407 and rs7606173. The difference was beyond the limit of accurate phenotype determination of the present study, so that discriminating between the different possible combinations is difficult. The analysis of joint effects at this locus shows that the p-value for rs4671393 drops from $2.60x10^{-130}$ to $3.00x10^{-76}$ when taking into account the effect of rs13019832, so that complex synthetic associations may also be present (see **Supplementary Table 3**).

Gene-gene interaction analyses

Using the lmekin() function from the "Kinship" R package[19], we tested for interactions involving genome-wide significant variants as reported in **Table 1** plus $\beta^0 39$, –α 3.7 deletion type I (NG_000006.1:g.34164_37967del3804), age and gender. To declare significance we used a Bonferroni significance threshold of $1.67x10^{-04}$ (corresponding to 300 interactions). We found significant evidence for 16 interactions, mostly involving variants in the β and α globin genes clusters. Specifically:

- for HbA1, between $\beta^0 39$ and variants in the α-globin gene cluster (with –α 3.7 deletion type I at $p=3.6x10^{-20}$, with rs570013781 at $p=3.8x10^{-26}$, with rs141494605 at $p=1.3x10^{-28}$ and with chr16:391593 at $p=7.1x10^{-06}$) ;
- for HbA2 between $\beta^0 39$ and variants in the α-globin gene cluster (with –α 3.7 deletion type I at $p=2.3x10^{-15}$, with rs570013781 at $p=3.7x10^{-23}$, with rs141494605 at $p=5.2x10^{-26}$) and with chr16:391593 at $p=4.2x10^{-05}$), as well as between SNP rs7944544 in the *HBD-HBG* intergenic region and SNP rs141494605 in the α-globin gene cluster $p=3.9x10^{-05}$);

- for HbF between $\beta^0$39 and variants in the α-globin gene cluster (with rs570013781at $p=1.19x10^{-07}$, with rs141494605 at $p=5.6x10^{-06}$, and with rs67385638 at $p=7.55x10^{-06}$), as well as between the *HBS1L-MYB* region and SNPs in the *HBG2-HBE1* region (rs9399137 with both rs67385638 ($p=2.7x10^{-05}$) and rs2855122 ($p=1.4x10^{-04}$)).

Among interactions involving one SNP and either age or gender, only the interaction between $\beta^0$39 and gender passed the threshold for significance for HbA1 ($p=1.1x10^{-16}$) and HbA2 ($p=8.1x10^{-07}$). As for age and gender, 2 interactions were nominally significant between rs570013781 and gender ($p=6.3x10^{-03}$) for HbA1 and between rs13019832 and gender ($p=2.3x10^{-02}$) for HbF. For completeness, we also report all effects observed separately in males and females in **Supplementary Table 11**, for which no locus had significant heterogeneity after Bonferroni correction.

$\beta^0$39 specific analysis

In addition to the 3 hemoglobins, we tested 69 additional non-hematological phenotypes available from the SardiNIA study for association with $\beta^0$39. All traits were normalized using inverse normal transformation and age, age^2 and gender as covariates. The results were considered significant after Bonferroni correction for the 69 tests (critical alpha = $7.0x10^{-04}$), and are presented in **Supplementary Table 10**. Each of the tabulated traits was measured in 5,949 except for eosinophil counts (5,823 individuals).

The full list of additional tested traits (defined and measured as in Pilia et al., 2006)[20] is as follow:

Adiponectin (mg/mL)
Vitamin B12 (pg/ml)
Folate (ng/ml)
Homocysteine (umol/L)
High sensitivity C-reactive protein (ug/mL)
IL6
Leptin (pg/mL)
MCP1 (pg/mL)
VCAM (ng/ml)
Thyroid Antithyroglobulin Antibody
Anti-thyroglobulin
Basophils (%)
BMI
Diastolic diameter
Systolic diameter
Diastolic Blood Pressure
Platelets count
WBC (10^3/ul) White Blood Cells
Eosinophils (%)
Uric acid (mg/dL)
Alanine aminotransferase (ALT) (U/L)
Aspartate aminotransferase (U/L)
Bun (mg/dL)
Fractionated bilirubin (mg/dL)
Total bilirubin (mg/dL)
Total cholesterol (mg/dL)
Serum creatinine level (mg/dL)
Fibrinogen (mg/dL)
Gamma-glutamyltransferase (GGT) (U/L)
Serum glucose level (mg/dL)
HDL cholesterol (mg/dL)
Serum insulin level (U/mL)
C-reative protein (mg/dL)
Potassium (mEq/L)
PSA (microU/L)
Iron (microU/L)
Sodium (mEq/L)
Transferrin (mg/dL)
Triglycerides (mg/dL)
Erythrocyte sedimentation rate (mm/h)
Height
Anthropometric: Hip circumference (cm)
Serum insulin level (U/mL)
Low-density lipoprotein
Lymphocytes (%)
Mean blood pressure
Monocytes (%)
Neutrophils (%)
Pulse pressure
Normalized PWV
PWV: Pulse wave velocity
Supine Blood Pressure 1st systolic
Total IgE
Thyroid-stimulating hormone

Waist
Wall Lumen Ratio
Weight
Carotid Doppler:
CCA intima media thickness
End diastolic velocity
Integral time velocity
Systolic diastolic ratio
Peak systolic velocity
Acceleration time
Pulsatile Index
Vascular Mass
ECG:
QTc interval
Heart rate
PR

Clinical impact assessment

To assess the possible impact of the genome-wide significant variants on disease severity, we used a cohort of retrospectively studied β-thalassemia patients comprising 306 unrelated Sardinian individuals followed at the "Microcitemico" Hospital in Cagliari (153 females and 153 males). Of these patients, 45 had thalassemia intermedia (patients that never needed regular blood transfusion), and all were homozygous for the $\beta^0$39 mutation, which accounts for 96% of β-thalassemia alleles in Sardinia[21]. The age at which patients started regular transfusions was used as an indicator of hematological severity – ranging from 14 patients who began transfusion within the first 2 months of life to the 45 patients who never needed transfusions. Criteria to start transfusion were 1) hemoglobin level < 7 g/dl for more than 2 weeks in absence of infections, 2) moderate to severe spleen enlargement, and 3) initial skeletal changes and/or poor growth. Patients were considered as regularly transfused when undergoing more than 8 blood transfusions per year. The full cohort was genotyped using Affymetrix 6.0 microarray; quality of samples and genotypes were carefully assessed through sample and marker missing rates, deviation from Hardy-Weinberg equilibrium, and minor allele frequency (MAF), resulting in 602,036 autosomal markers available for imputation. As for the SardiNIA cohort, imputation was performed using Minimac[22] and the Sardinian whole-genome sequenced samples as reference panel, with genotypes pre-phased using MACH[23].

Among the 23 unique variants associated with different hemoglobins in the general population, all samples were homozygous for the $\beta^0$39 mutation while for the α-globin genes we considered the – α 3.7 deletion type I as well as rs111033603 and rs41474145, determined using GAP-PCR or restriction enzyme digestion. Of the remaining 21 unique variants reported in **Table 1**, two (rs141006889 and rs148706947) were excluded because imputation RSQR was <0.50 and no acceptable surrogates were found. For two variants, rs5030868 (on the X chromosome, neither genotyped nor imputed) and rs11036338 (RSQR<0.50), we used the best available surrogates: for the first, rs4898455 (genotyped, r^2=0.6, p=5.9x10^{-08} for HbA1) and for the second, rs10837584 (imputed with RSQR=0.999, r^2=0.9, p=8.2x10^{-12} for HbA2). The available 19 SNPs, plus gender and α-globin gene defects forced as covariates, were considered in fitting a Cox proportional hazards model using a conditional backward procedure to iteratively remove the least significant variable at each step until all remaining variants were at least nominally significant.

This analysis identified 6 variants as significant predictors of the age at transfusion: marker chr16:391593 at the α-globin gene cluster (p=0.008, Hazard Ratio (HR) = 2.56); rs67385638 and rs12793110 at the β-globin gene cluster (p=0.005 with HR=3.01 and p=0.038 with HR=0.014 respectively); rs4671393 within the *BCL11A* gene (p=1.1x10^{-8}, HR=2.143); rs9399137 in the *HBS1L*-

MYB intergenic region ($p=1.8x10^{-04}$, HR=1.531); and rs59329875 in the *PLTP-PCIF1* intergenic region (p=0.036, HR=0.793). Among these variants, 3 (chr16:391593, rs67385638, and rs12793110) are new variants at loci already associated to β-thalassemia severity[24,25] and one revealed a new locus associated with β^0-thalassemia severity (rs59329875 at *PLTP-PCIF1;* whose association with HbA2 was replicated in the TwinsUK[3] cohort with $p=6.98x10^{-06}$).

REFERENCES

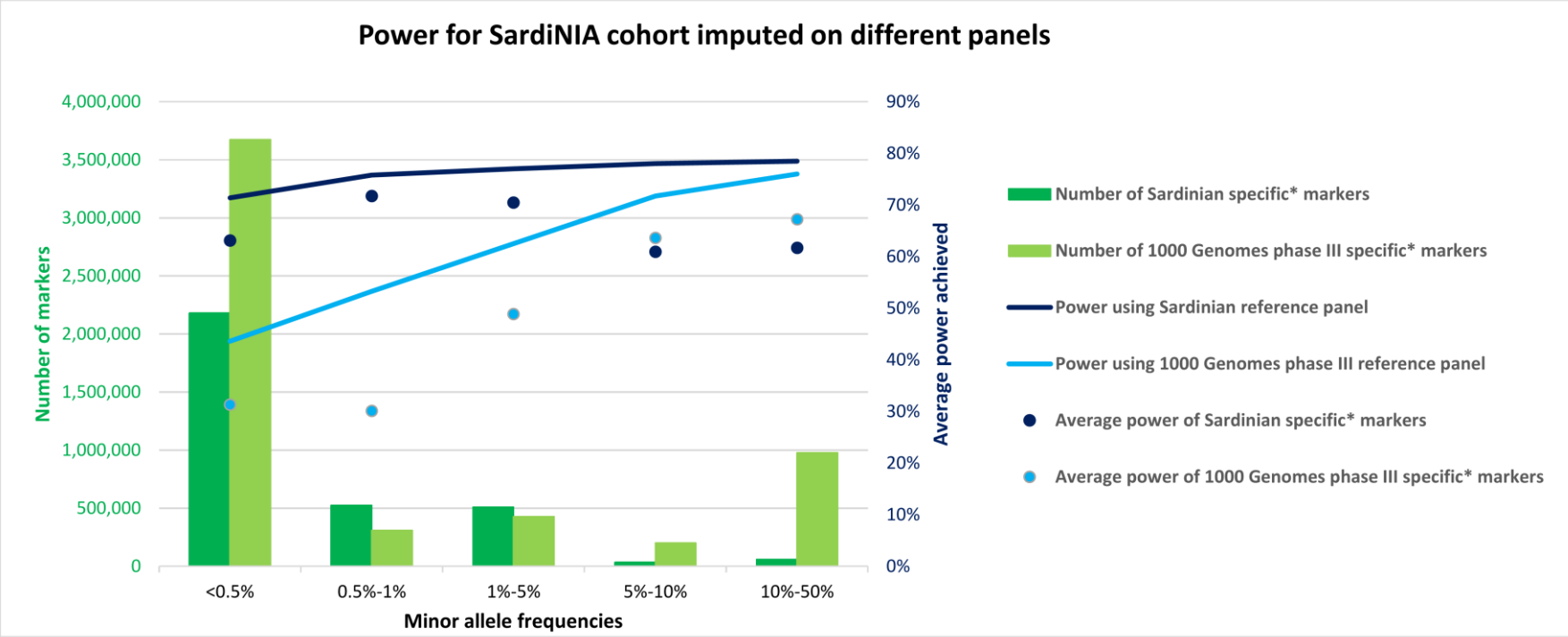
1. Consortium, T. 1000 G. P. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).
2. Pistis, G. *et al.* Rare variant genotype imputation with thousands of study-specific whole-genome sequences: implications for cost-effective study designs. *Eur. J. Hum. Genet. EJHG* (2014). doi:10.1038/ejhg.2014.216
3. Moayyeri, A., Hammond, C. J., Valdes, A. M. & Spector, T. D. Cohort Profile: TwinsUK and Healthy Ageing Twin Study. *Int. J. Epidemiol.* **42,** 76–85 (2013).
4. Pissard, S. & Beuzard, Y. A potential regulatory region for the expression of fetal hemoglobin in sickle cell disease. *Blood* **84,** 331–338 (1994).
5. Daly, M. J. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
6. Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148,** 84–98 (2012).
7. Messeguer, X. *et al.* PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinforma. Oxf. Engl.* **18,** 333–334 (2002).
8. Farré, D. *et al.* Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic Acids Res.* **31,** 3651–3653 (2003).
9. Akiyama, Y. Tfsearch: Searching transcription factor binding sites. (1998).
10. Liu, L. *et al.* Regulation of Gγ-Globin Gene by ATF2 and Its Associated Proteins through the cAMP-Response Element. *PLoS ONE* **8,** (2013).
11. Chae, J. H. & Kim, C. G. CP2 binding to the promoter is essential for the enhanced transcription of globin genes in erythroid cells. *Mol. Cells* **15,** 40–47 (2003).
12. Zhao, Q., Cumming, H., Cerruti, L., Cunningham, J. M. & Jane, S. M. Site-specific acetylation of the fetal globin activator NF-E4 prevents its ubiquitination and regulates its interaction with the histone deacetylase, HDAC1. *J. Biol. Chem.* **279,** 41477–41486 (2004).
13. Sangerman, J. *et al.* Mechanism for fetal hemoglobin induction by histone deacetylase inhibitors involves gamma-globin activation by CREB1 and ATF-2. *Blood* **108,** 3590–3599 (2006).
14. Menzel, S., Garner, C., Rooks, H., Spector, T. D. & Thein, S. L. HbA2 levels in normal adults are influenced by two distinct genetic mechanisms. *Br. J. Haematol.* **160,** 101–105 (2013).

15. Stadhouders, R. *et al.* Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development. *EMBO J.* **31,** 986–999 (2012).
16. Stadhouders, R. *et al.* HBS1L-MYB intergenic variants modulate fetal hemoglobin via long-range MYB enhancers. *J. Clin. Invest.* **124,** 1699–1710 (2014).
17. Zeller, T. *et al.* Genetics and Beyond – The Transcriptome of Human Monocytes and Disease Susceptibility. *PLoS ONE* **5,** e10693 (2010).
18. Jiang, J. *et al.* cMYB is involved in the regulation of fetal hemoglobin production in adults. *Blood* **108,** 1077–1083 (2006).
19. R Core Team. *R: A language and environment for statistical computing.* (R Foundation for Statistical Computing, Vienna, Austria, 2013).
20. Pilia, G. *et al.* Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet.* **2,** e132 (2006).
21. Rosatelli, M. C. *et al.* Molecular characterization of beta-thalassemia in the Sardinian population. *Am. J. Hum. Genet.* **50,** 422–426 (1992).
22. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44,** 955–959 (2012).
23. Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34,** 816–834 (2010).
24. Danjou, F. *et al.* Genetic modifiers of β-thalassemia and clinical severity as assessed by age at first transfusion. *Haematologica* **97,** 989–993 (2012).
25. Danjou, F. *et al.* A genetic score for the prediction of beta-thalassemia severity. *Haematologica* haematol.2014.113886 (2014). doi:10.3324/haematol.2014.113886
26. Tatusova, T., Ciufo, S., Fedorov, B., O'Neill, K. & Tolstoy, I. RefSeq microbial genomes database: new representation and annotation strategy. *Nucleic Acids Res.* **42,** D553–559 (2014).
27. Su, M. Y. *et al.* Identification of Biologically Relevant Enhancers in Human Erythroid Cells. *J. Biol. Chem.* **288,** 8433–8444 (2013).
28. Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26,** 2336–7 (2010).
29. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46,** 310–315 (2014).

30. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42,** D1001–D1006 (2014).
31. Raychaudhuri, S. *et al.* Identifying Relationships among Genomic Disease Regions: Predicting Genes at Pathogenic SNP Associations and Rare Deletions. *PLoS Genet* **5,** e1000534 (2009).
32. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25,** 25–29 (2000).
33. Kapushesky, M. *et al.* Gene Expression Atlas at the European Bioinformatics Institute. *Nucleic Acids Res.* **38,** D690–D698 (2010).

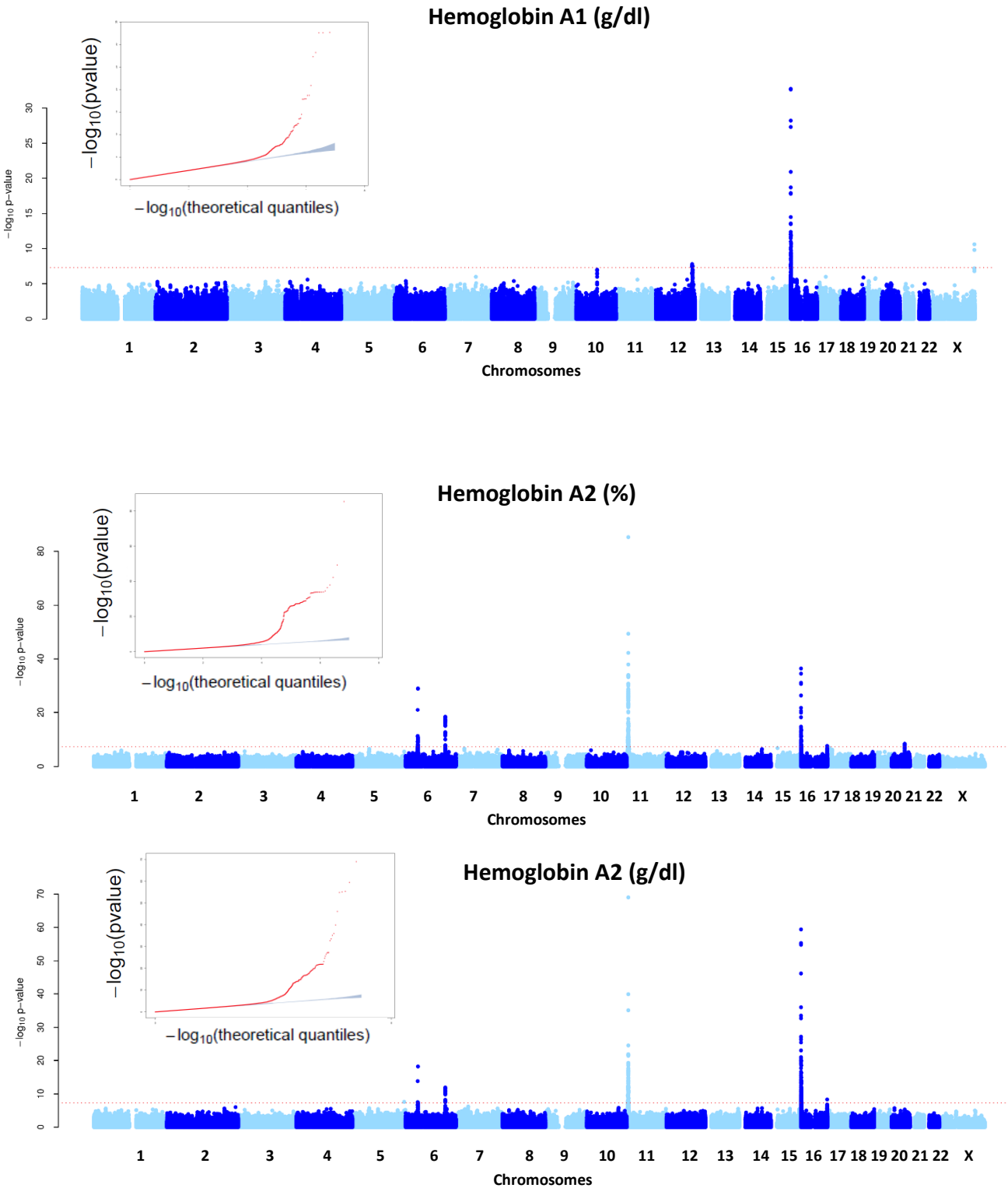
SUPPLEMENTARY FIGURES

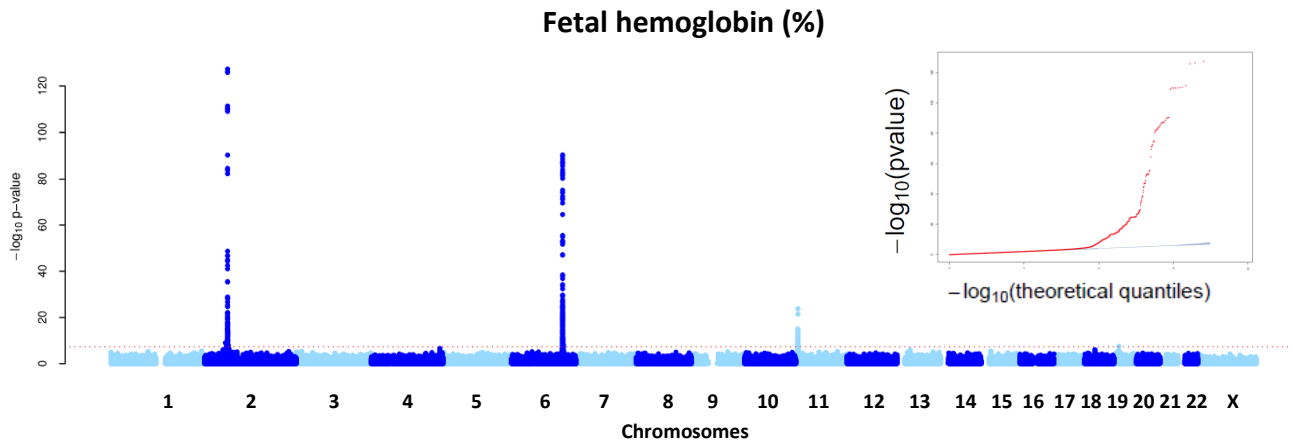
Supplementary Figure 1. Empirical evaluation of genomic power upon reference panel population.
The figure presents the power achieved for 1.3% of variance explained (equivalent to 80% power for 3,000 unrelated individuals and a MAF of 50%, at a p-value of $5x10^{-8}$), for five different ranges of allele frequencies. To calculate power we used the effective sample size as the product of sample size and average imputation quality (RSQR), so that the variation of power between populations is a direct consequence of the average RSQR for each range of allele frequencies.

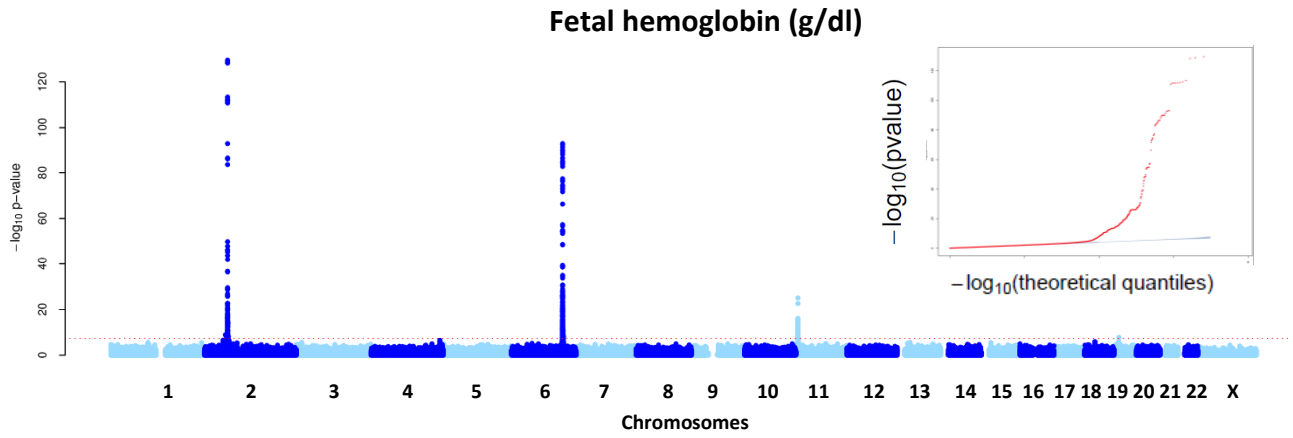


* markers specific to one panel refer to variants that could be inferred according to imputation quality filters as defined in Pistis et al., 2014[2].

Supplementary Figure 2. Manhattan and QQ plots for the assessed hemoglobin types.
The figure shows association results for all the hemoglobin types at QCed genotyped and imputed autosomal markers, and only genotyped markers on the X chromosome. Results are those obtained from imputation using the Sardinian sequencing data. The inner box shows the quantile-quantile plot for theoretical versus observed p-values. The gray shadow represents 95% confidence interval.



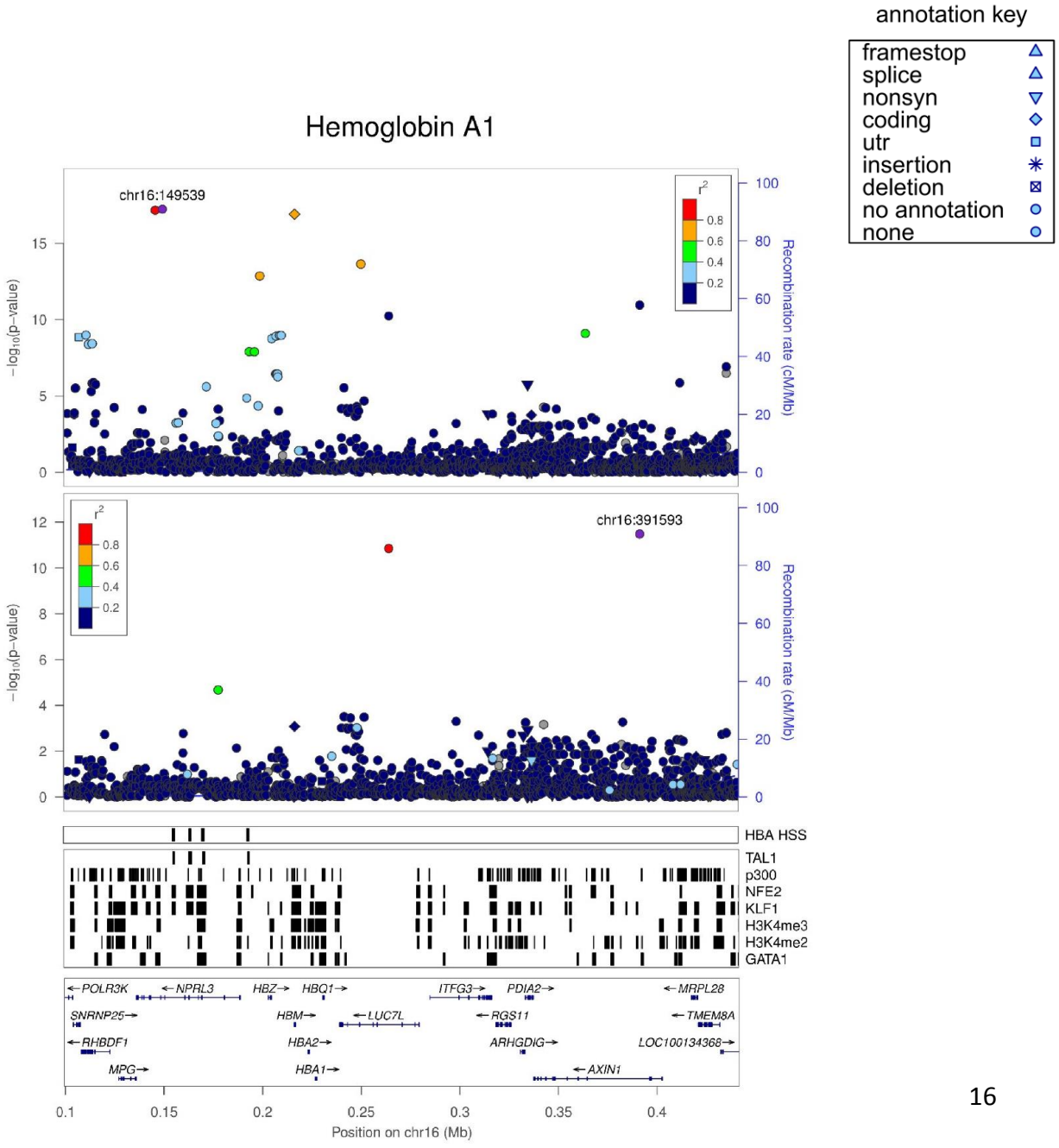

Fetal hemoglobin (%)
−log10 p−value
0
20
40
60
80
100
120
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Chromosomes
−log10(pvalue)
−log10(theoretical quantiles)


Fetal hemoglobin (g/dl)
−log10 p−value
0
20
40
60
80
100
120
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Chromosomes
−log10(pvalue)
−log10(theoretical quantiles)

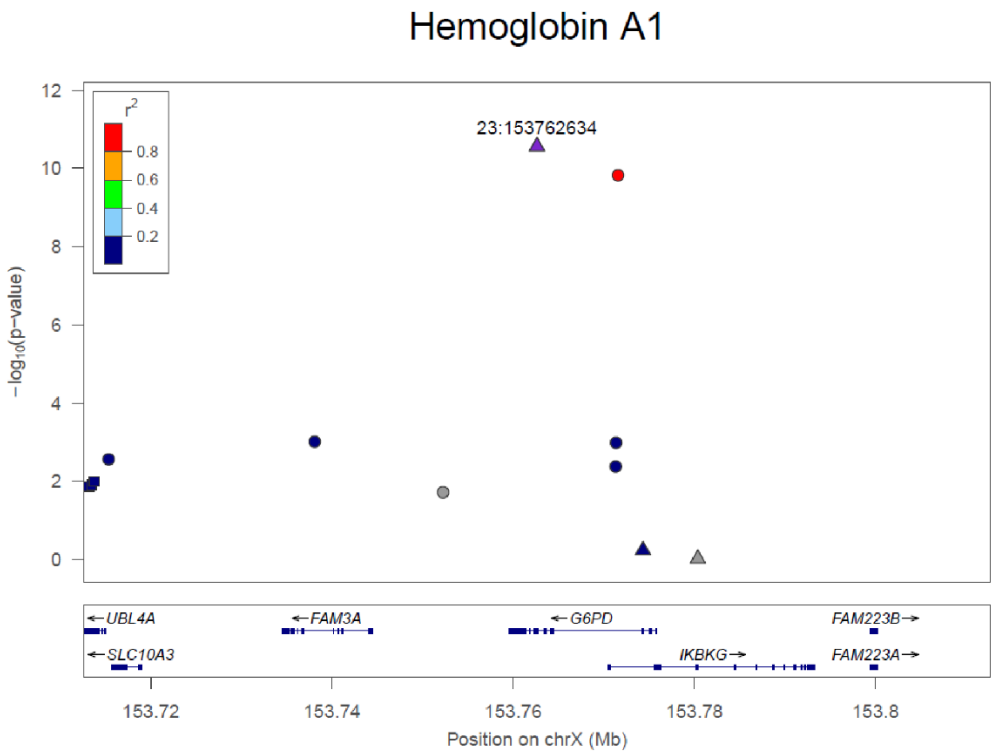
Supplementary Figure 3. Regional plots of loci associated with HbA1.

Regional association plots at the loci associated with HbA1, for imputation performed using the Sardinian sequencing-based reference panel. At each locus, we plotted association strength (Y axis showing the –log 10 p-value) versus the genomic positions (on the hg19/GRCh37 genomic build) around the most significant SNP, which is indicated with a purple dot. Other SNPs in the region are color-coded to reflect their LD with the top SNP as in the inset (taken from pairwise r^2 values calculated on Sardinians) or gray if LD was <0.01. Symbols reflect genomic functional annotation, as indicated in the inner box of the first figure. A blue line representing recombination rate in HapMap CEU populations overlie the plot. Symbols reflect genomic functional annotation, as indicated for Locus 1. A box at the bottom of each locus show location of genes and the position of exons, as well as the direction of transcription, according to RefSeq[26]. In the globin genes clusters two additional boxes report the classical DNAse hypersensitive sites and putative specific human erythroid cells enhancers (according to Su et al., 2014)[27], respectively. This plot was drawn using the standalone version of the LocusZoom package[28].

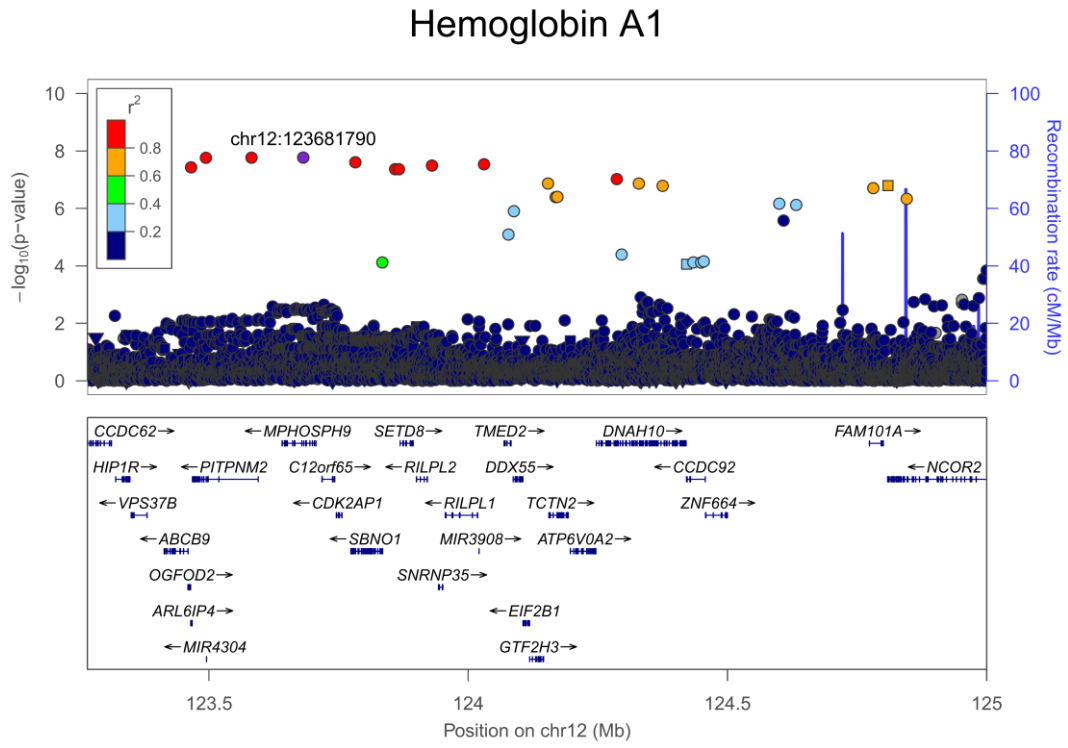
Locus 1



Locus 2

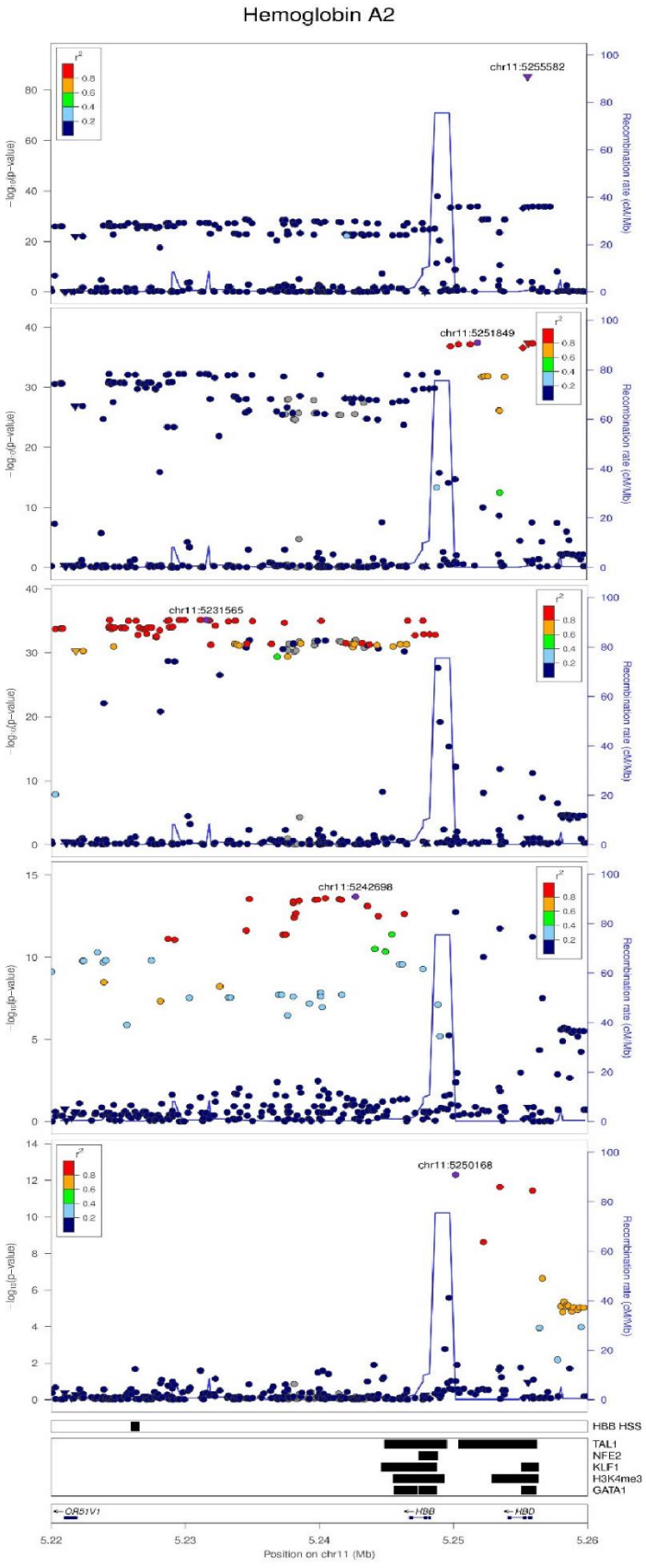


Locus 3

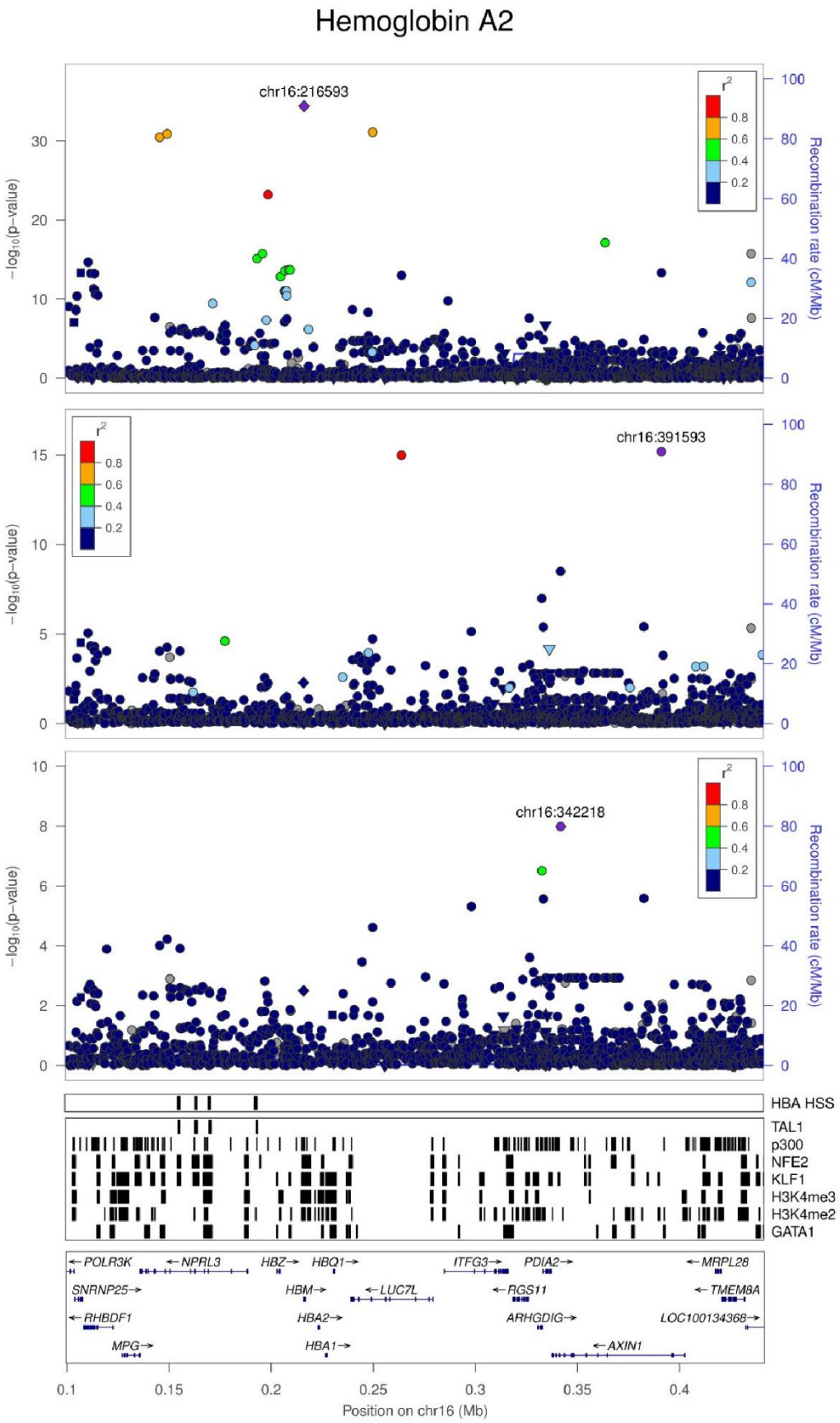


Supplementary Figure 4. Regional plots of loci associated with HbA2.
Regional association plots at the loci associated with HbA2, for imputation performed using the Sardinian reference panel. For a description of the plot and specific features see **Supplementary Figure 3** legend.

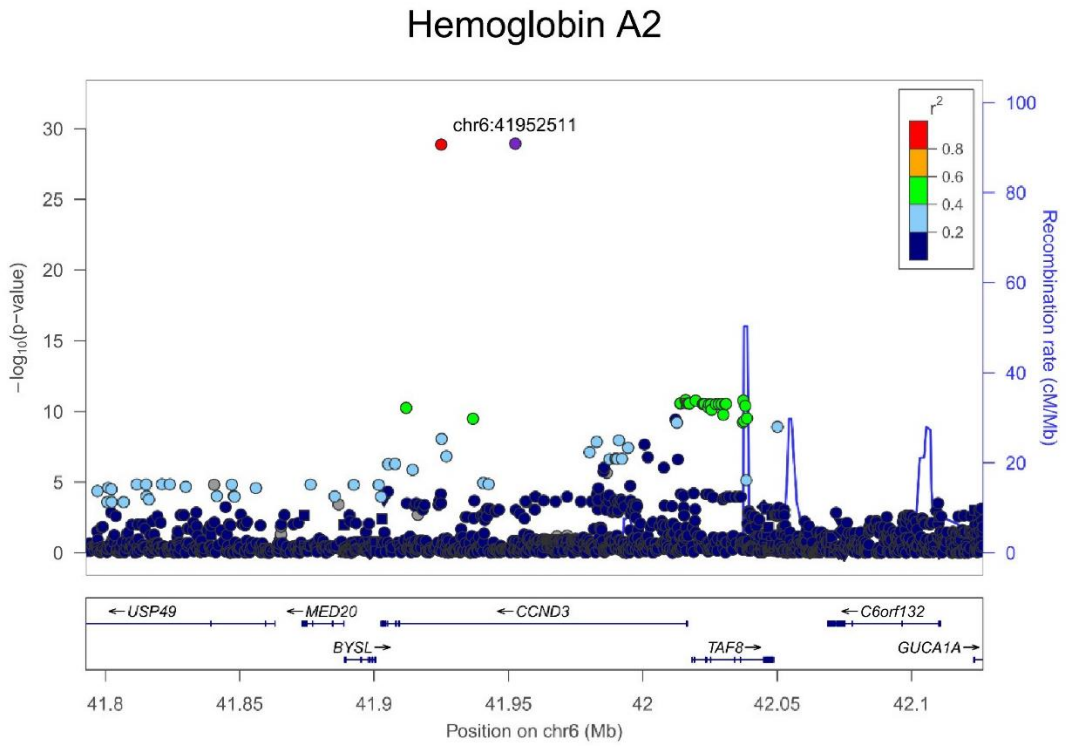
Locus 1



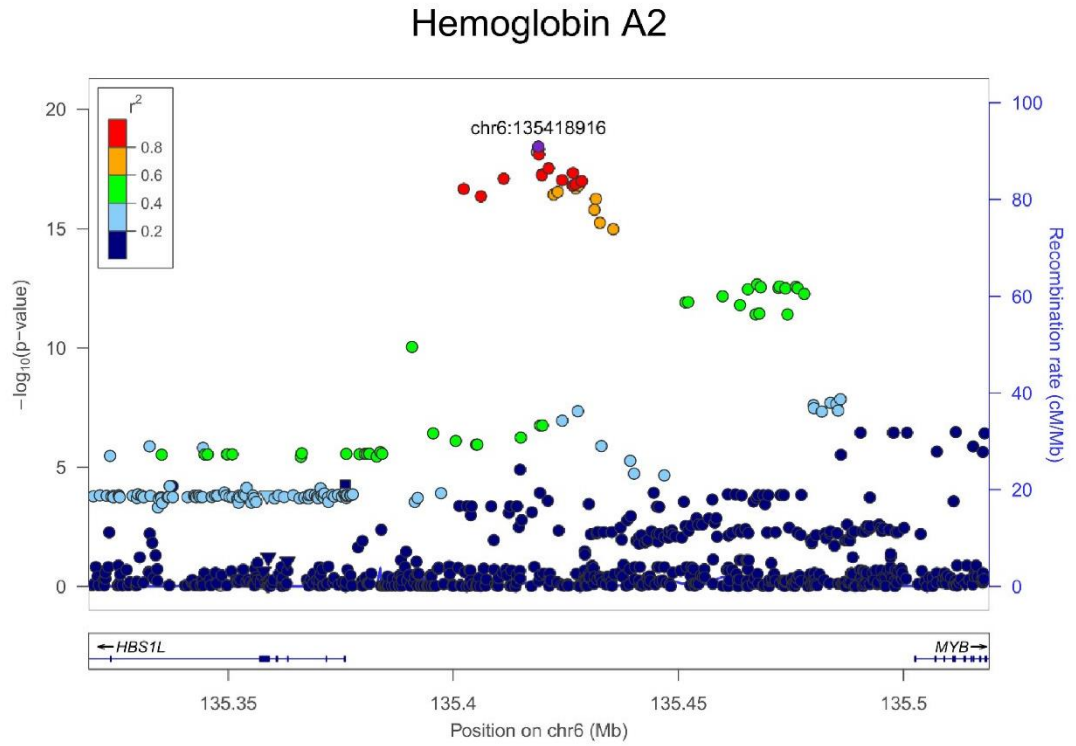
Locus 2


Hemoglobin A2
chr16:216593
chr16:391593
chr16:342218
−log10(p−value)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
HBA HSS
TAL1
p300
NFE2
KLF1
H3K4me3
H3K4me2
GATA1
POLR3K
NPRL3
HBZ
HBQ1
ITFG3
PDIA2
MRPL28
SNRNP25
HBM
LUC7L
RGS11
TMEM8A
RHBDF1
HBA2
ARHGDIG
LOC100134368
MPG
HBA1
AXIN1
Position on chr16 (Mb)

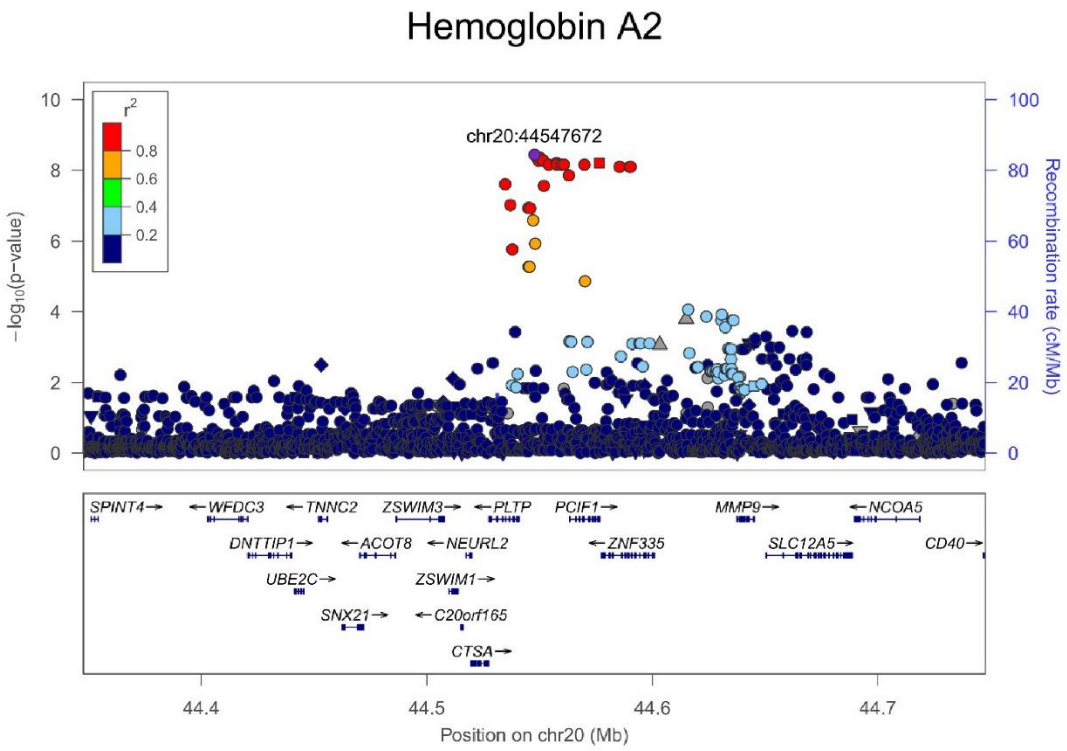
Locus 3



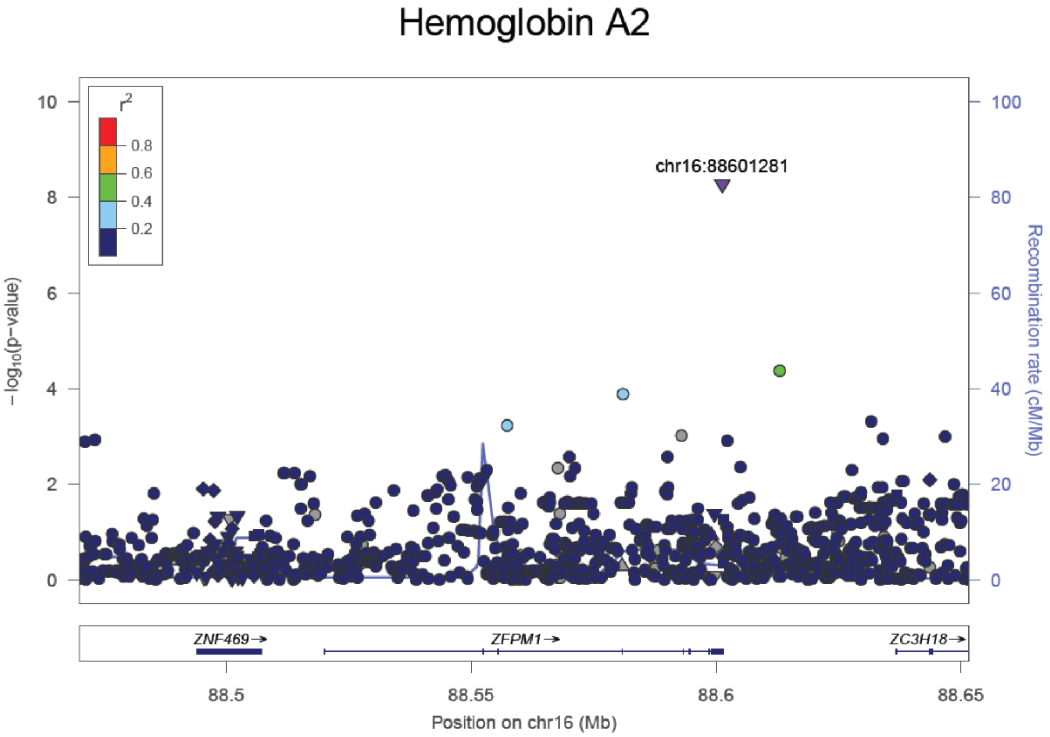
Locus 4



Locus 5

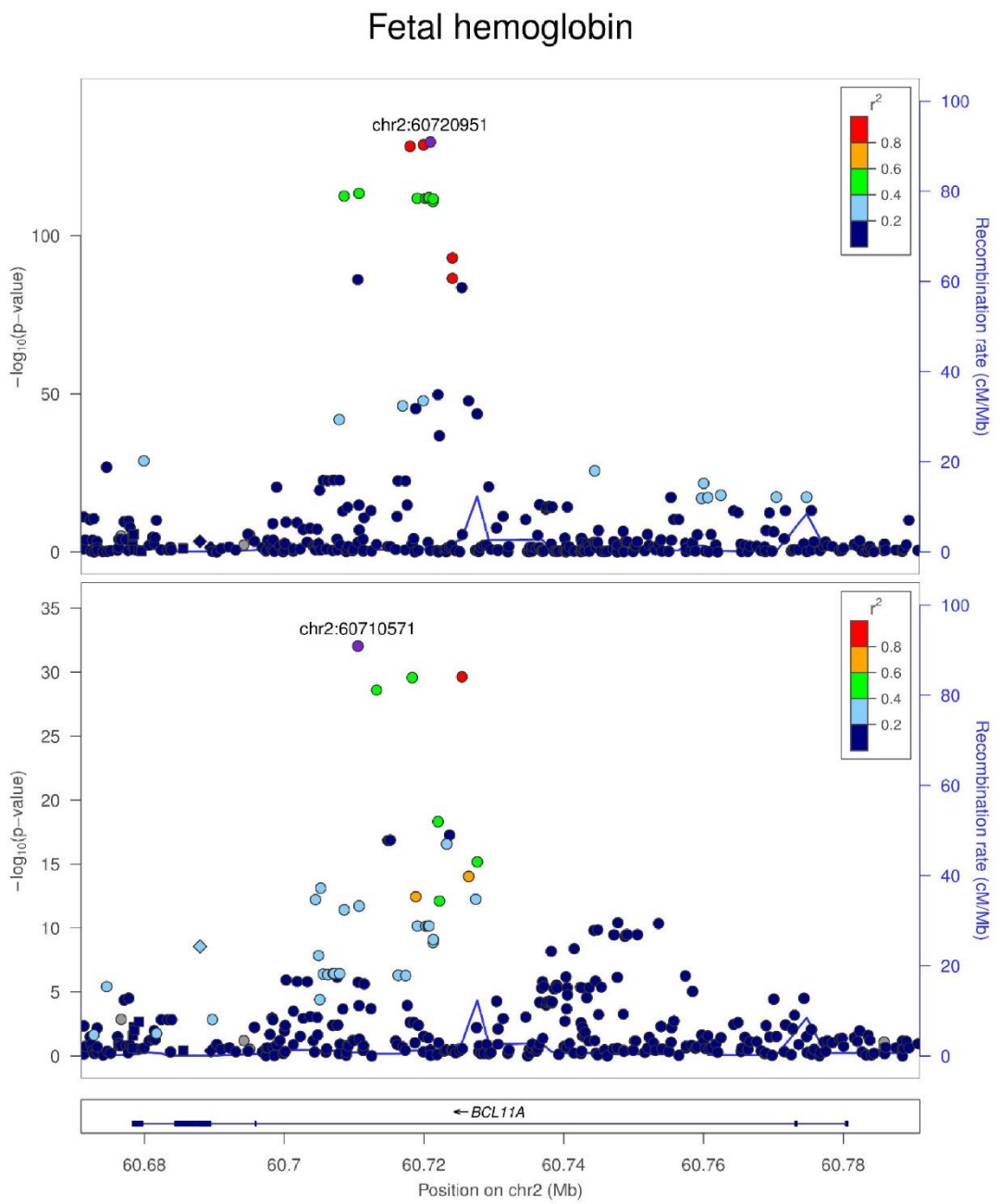


Locus 6

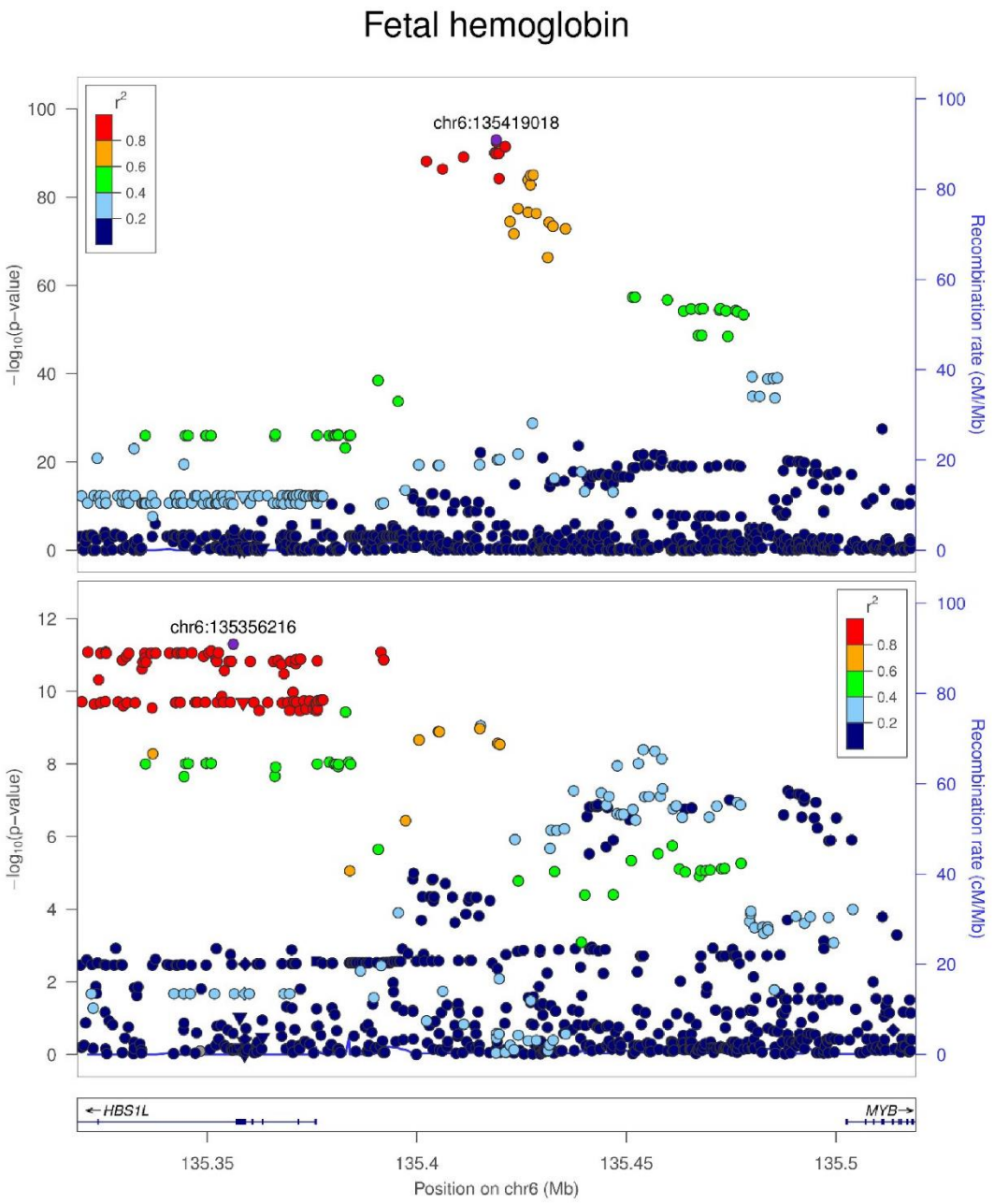


Supplementary Figure 5. Regional plots of loci associated with HbF.
Regional association plots at the loci associated with HbF, for imputation performed using the Sardinian reference panel. For a description of the plot and specific features see **Supplementary Figure 3** legend.

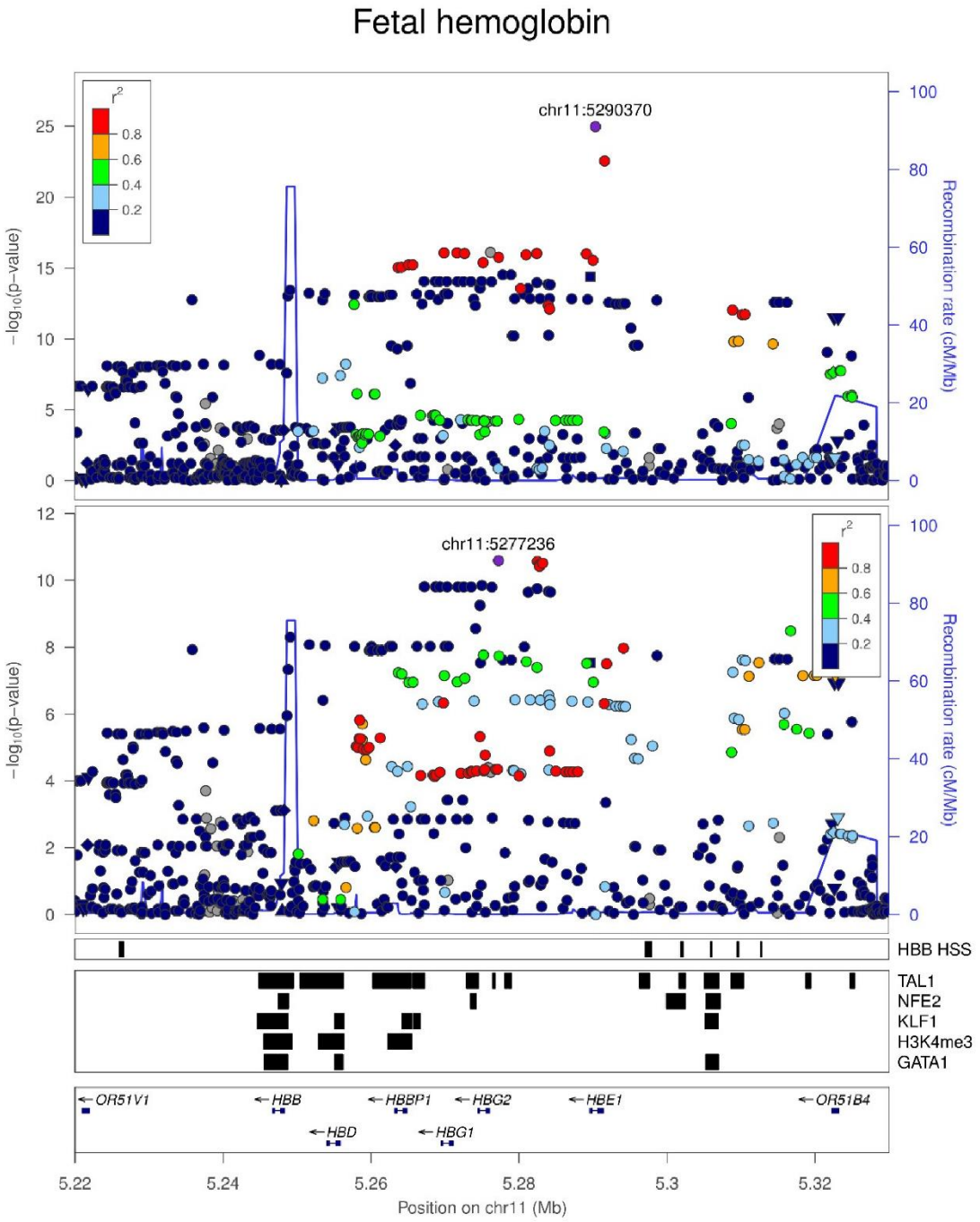
Locus 1



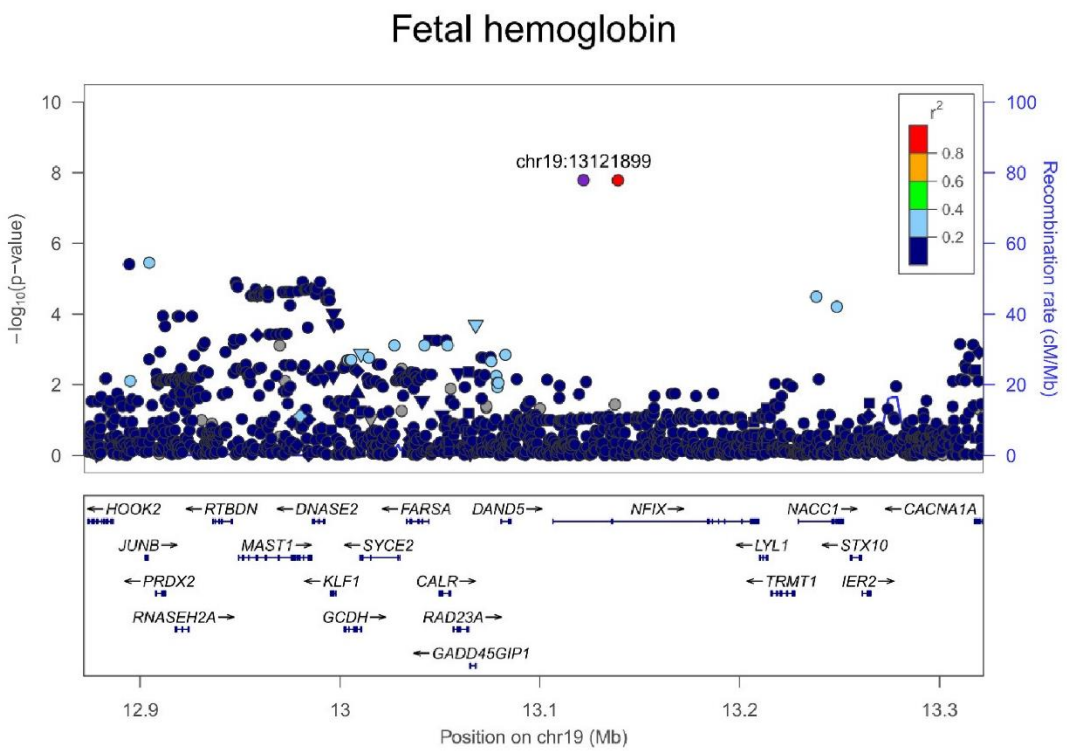
Locus 2


Fetal hemoglobin
chr6:135419018
chr6:135356216
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
HBS1L
MYB
135.35
135.4
135.45
135.5
Position on chr6 (Mb)

Locus 3


Fetal hemoglobin
chr11:5290370
chr11:5277236
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
HBB HSS
TAL1
NFE2
KLF1
H3K4me3
GATA1
OR51V1
HBB
HBBP1
HBG2
HBE1
OR51B4
HBD
HBG1
5.22
5.24
5.26
5.28
5.3
5.32
Position on chr11 (Mb)

Locus 4


Fetal hemoglobin
chr19:13121899
r²
0.8
0.6
0.4
0.2
−log10(p-value)
Recombination rate (cM/Mb)
HOOK2
RTBDN
DNASE2
FARSA
DAND5
NFIX
NACC1
CACNA1A
JUNB
MAST1
SYCE2
LYL1
STX10
PRDX2
KLF1
CALR
TRMT1
IER2
RNASEH2A
GCDH
RAD23A
GADD45GIP1
12.9
13
13.1
13.2
13.3
Position on chr19 (Mb)

SUPPLEMENTARY TABLES

Supplementary Table 1. Characteristics of the studied cohort.
Basic descriptive statistics for the 6,305 individuals with genotyped data and full measurements for all three hemoglobin types, out of the 6,602 genotyped and imputed. Statistics are reported separately for carriers and non-carriers of the $\beta^0 39$ mutation.

All genotyped samples				
Males/ Females	2823 / 3779			
Age mean (sd)	43.6 (17.6)			
	Non-carriers		Carriers	
	N	Median hemoglobin levels (5th percentile - 95th percentile)	N	Median hemoglobin levels (5th percentile - 95th percentile)
Total Hb (g/dl)	5662	14.000 (11.900 - 16.200)	643	12.000 (10.300 - 14.000)
HbA2 (%)	5662	2.700 (2.200 - 3.200)	643	5.800 (5.000 - 6.700)
HbF (%)	5662	0.300 (0 - 0.900)	643	0.900 (0 - 2.400)
HbA1 (g/dl)	5662	13.538 (11.519 - 15.746)	643	11.168 (9.557 - 13.057)
HbA2 (g/dl)	5662	0.382 (0.277 - 0.483)	643	0.699 (0.556 - 0.871)
HbF (g/dl)	5662	0.036 (0 - 0.126)	643	0.107 (0 - 0.275)

Supplementary Table 2. Association results for the $\beta^0$39 and –α 3.7 deletion type I.

The table shows the association results for the $\beta^0$39 mutation and –α 3.7 deletion type I, on all three hemoglobin forms. For each mutation, we indicated the chromosome and genomic position (hg19 build), the rs ID when available, the effect allele tested for association (EA) and the other allele at the SNP (OA), the SNP effect allele frequency (EAF) and the regression coefficients.

Mutation	Alleles (EA/OA)	EAF	Hemoglobin form	Effect (StdErr)	p-value
$\beta^0$39 (rs11549407)					
	T/C	0.048	HbA1 (g/dl)	-1.5482 (0.029)	$<1.0x10^{-300}$
			HbA2 (%)	1.9455 (0.034)	$<1.0x10^{-300}$
			HbF g/dl	1.1396 (0.036)	$6.173x10^{-206}$
–α 3.7 deletion type I					
	del/wild type	0.154	HbA1 (g/dl)	-0.2349 (0.026)	$4.578x10^{-19}$
			HbA2 (%)	-0.1877 (0.032)	$4.434x10^{-09}$
			HbF (g/dl)	-0.0710 (0.029)	$1.460x10^{-02}$

Supplementary Table 3. Joint-association analysis results for loci with multiple associated variants.

The table presents the five loci at which different variants were associated at genome-wide significant levels, with effects and p-values corrected for all variants associated at the locus.

Trait	Loci	variant	p-value	Effect (StdErr)
HbA1				
	α-globin gene cluster	rs570013781	1.50×10^{-18}	-0.2025 (0.023)
		chr16:391593	3.28×10^{-12}	-0.4028 (0.058)
HbA2				
	β-globin gene cluster	rs35152987	8.09×10^{-97}	-2.2450 (0.106)
		rs7944544	7.90×10^{-42}	-1.3020 (0.095)
		rs12793110	3.53×10^{-18}	-0.1800 (0.021)
		rs11036338	5.20×10^{-14}	0.1256 (0.017)
		rs7936823	5.00×10^{-13}	0.1117 (0.015)
	α-globin gene cluster	rs141494605	3.70×10^{-43}	-0.3540 (0.026)
		chr16:391593	1.38×10^{-15}	-0.5035 (0.063)
		rs148706947	1.04×10^{-08}	0.2892 (0.051)
HbF				
	BCL11A	rs4671393	3.00×10^{-76}	0.4644 (0.025)
		rs13019832	9.12×10^{-33}	-0.2024 (0.017)
	HBS1L-MYB	rs9399137	7.46×10^{-94}	0.5195 (0.025)
		rs11754265	5.04×10^{-12}	-0.1421 (0.021)
	β-globin gene cluster	rs67385638	4.67×10^{-35}	0.3103 (0.025)
		rs2855122	2.57×10^{-11}	-0.1458 (0.022)

Supplementary Table 4. Validation of top genome-wide findings.

For each SNP, we show the number of heterozygotes and homozygotes for the reference and alternative alleles that were imputed using the Sardinian panel, the method used for validation, the number of these that were genotyped or sequenced along with the genotype mismatch rate, the p-value observed in our primary analysis (as reported in **Table 1**) and the p-value obtained replacing imputed genotypes with those inferred from Sanger sequencing or Taqman assays.

SNP	EAF	N hom.EA/het/hom.OA	Validation Method	Hom EA (Mismatch %)	Het (Mismatch %)	Hom OA (Mismatch %)	Passed Validation	Original p-value	Final p-value
chr12:123681790	0.0105	2/130/6173	Sanger	2 (0%)	64 (6.3%)	6 (0%)	Yes	1.68×10^{-08}	2.10×10^{-08}
rs112233623 [1]	0.1007	63/1158/5084	TaqMan	8 (0%)	244 (3.3%)	1132 (0.8%)	Yes	1.11×10^{-29}	3.28×10^{-30}
rs183437571	0.0101	0/124/6181	Sanger	0 (na)	21 (0%)	17 (0%)	Yes	1.61×10^{-08}	1.79×10^{-08}
rs7936823	0.4660	1339/3187/1779	TaqMan	300 (1.7%)	704 (2.1%)	365 (1.6%)	Yes	5.00×10^{-13}	1.08×10^{-12}
rs17525396 [2]	0.1360	141/1434/4730	Affymetrix 6.0	40 (2.5%)	526 (2.3%)	1784 (0.3%)	Yes	6.79×10^{-18}	6.65×10^{-18}

[1] tag for rs113267280 – $r^2 = 0.986$

[2] tag for rs570013781 – $r^2 = 0.997$

Supplementary Table 5. Functional annotation of signals reported in Table 1.

The table shows the Combined Annotation Dependent Depletion (CADD) score[29] for each lead SNP and indicates the marker with the highest CADD among those in r^2 >0.9 with the lead, when available (using Sardinian linkage disequilibrium information), together with its functional annotation. The CADD score (in Phred scale) reported in the table, is a measure that integrates diverse genome annotations (including function features from the ENCODE Project, conservation metrics such as GERP and phastCons); it can help to quantitatively prioritize variants across a wide range of functional categories (the higher the score the more likely the functionally relevance).

	Lead SNP (chr:position)	rsID from dbsnp142	CADD score	SNP with highest CADD score and r^2>0.9 with lead			
				SNP	score	r^2 with lead	annotation
HbA1 (g/dl)							
locus1	16:149539	rs570013781	1.12	rs17525396	5.29	1	intronic *NPRL3*
	16:391593 (cond.)	-	2.99	-	-	-	-
locus2	X:153762634	rs5030868	18.52	-	-	-	-
locus3	12:123681790	-	0.153	12:123465483	12.05	1	intronic *ARL6IP4*; TFBS peak[1]
HbA2							
locus1 (%)	11:5255582	rs35152987	20.70	-	-	-	-
	11:5251849 (cond.)	rs7944544	1.57	rs35406175	18.82	0.94	missense *HBD*
	11:5231565 (cond.)	rs12793110	1.80	rs35156960	12.51	0.94	intronic AC104389.16
	11:5242698 (cond.)	rs11036338	0.83	rs12362241	11.04	0.97	-
	11:5250168 (cond.)	rs7936823	1.17	rs3813727	4.83	0.93	TFBS peak[1]
locus2 (g/dl)	16:216593	rs141494605	6.47	-	-	-	-
	16:391593 (cond.)	-	2.99	-	-	-	-
	16:342218 (cond.)	rs148706947	3.33	-	-	-	-
locus3 (%)	6:41952511	rs113267280	2.99	rs112233623	6.34	0.99	TFBS peak[1]
locus4 (%)	6:135418916	rs7776054	12.71	rs9399137	16.95	0.97	-
locus5 (%)	20:44547672	rs59329875	0.19	rs1057208	12.85	0.97	upstream *PCIF1*; TFBS peak[1]
locus6 (%)	16:88601281	rs141006889	5.19	-	-	-	-
HbF (g/dl)							
locus1	2:60720951	rs4671393	0.74	rs1427407	22.4	0.95	upstream *AC009970.1*; TFBS peak[1]
	2:60710571 (cond.)	rs13019832	7.26	-	-	-	-
locus2	6:135419018	rs9399137	16.95	-	-	-	-
	6:135356216 (cond.)	rs11754265	4.51	rs1041480	15.42	0.98	intronic *HBS1L*; TFBS peak[1]
locus3	11:5290370	rs67385638	2.15	rs2855036	9.42	0.91	-
	11:5277236 (cond.)	rs2855122	2.35	rs4910546	11.61	0.91	-
locus4	19:13121899	rs183437571	12.90	-	-	-	-

[1] = transcription factor binding site (TFBS) peak is present in annotation when CADD score reported such evidence.

Supplementary Table 6. Association results at SNPs previously associated with hemoglobin or red blood cell indices.

The table describes association results for the five assessed hemoglobin measurements at SNPs previously associated with hemoglobin or red blood cell indices, and included in GWAS Catalog[30]. The GWAS Catalog query was performed in date 01/25/2015 for the following traits: Erythrocytes sedimentation rate, F-cells distribution, Fetal HB levels, HbA2 levels, Hematocrit, Hematological and biochemical parameters, Hematology traits, Hemoglobin, Hemoglobin levels, Mean corpuscular hemoglobin, Mean corpuscular hemoglobin concentration, Mean corpuscular volume, Other erythrocyte phenotypes (RBC/MCHC), Red blood cell counts, Red blood cell traits. For each SNP, we report the position in genomic build 37, the p-value for each of the five hemoglobin measurements, the PubMed ID of the corresponding article, as well as the category, the original p-value and the notes as listed in the GWAS Catalog.

dbSNP ID	chr:position	Traits					SNPs reported in GWAS Catalog			
		HBA1(g/dl)	HBA2(%)	HBA2(gdl)	HBF(%)	HBF(gdl)	PUBMEDID	Disease/Trait	p-value	p-value notes
rs1043879	1:25570081	9.43E-01	8.04E-02	1.42E-01	2.66E-01	3.10E-01	21700265	Erythrocyte sedimentation rate	$2.0x10^{-09}$	
rs10903129	1:25768937	7.88E-01	3.28E-01	3.35E-01	2.96E-01	3.32E-01	21700265	Erythrocyte sedimentation rate	$5.0x10^{-13}$	
rs12034383	1:207803595	7.99E-01	4.70E-01	6.62E-01	8.45E-01	7.19E-01	21700265	Erythrocyte sedimentation rate	$2.0x10^{-28}$	
rs3091242	1:25674785	7.20E-01	2.99E-01	3.14E-01	2.86E-01	2.99E-01	21700265	Erythrocyte sedimentation rate	$2.0x10^{-13}$	
rs7527798	1:207872290	8.55E-01	6.74E-01	7.94E-01	8.85E-01	9.90E-01	21700265	Erythrocyte sedimentation rate	$2.0x10^{-09}$	
rs10017284	4:121513215	6.35E-01	1.11E-01	9.92E-02	6.45E-01	5.53E-01	21326311	F-cell distribution	$6.0x10^{-06}$	
rs11968814	6:71776971	NA	NA	NA	NA	NA	21326311	F-cell distribution	$9.0x10^{-07}$	
rs12073837	1:221010205	1.49E-01	3.08E-01	9.26E-02	9.02E-02	9.27E-02	21326311	F-cell distribution	$2.0x10^{-06}$	
rs12559632	X:22160992	2.56E-01	2.37E-01	2.17E-01	4.79E-01	5.38E-01	21326311	F-cell distribution	$3.0x10^{-06}$	
rs1318772	5:112723567	1.55E-01	3.15E-01	1.80E-01	9.73E-01	9.86E-01	21326311	F-cell distribution	$1.0x10^{-06}$	
rs1427407	2:60718043	3.80E-01	7.20E-01	9.96E-01	**1.43E-126**	**5.60E-129**	17767159	F-cell distribution	$6.0x10^{-31}$	
rs17135859	5:112996654	4.68E-01	1.83E-01	6.47E-01	4.43E-01	3.54E-01	21326311	F-cell distribution	$8.0x10^{-06}$	
rs1845344	4:120962385	4.93E-01	5.38E-01	6.69E-01	5.57E-01	6.06E-01	21326311	F-cell distribution	$7.0x10^{-06}$	
rs2577720	2:23548614	4.63E-01	4.20E-01	3.31E-01	8.23E-01	7.65E-01	21326311	F-cell distribution	$8.0x10^{-06}$	
rs3800569	7:138411425	8.35E-02	1.87E-01	6.62E-02	9.60E-01	8.36E-01	21326311	F-cell distribution	$6.0x10^{-06}$	
rs11886868	2:60720246	4.72E-01	7.07E-01	9.19E-01	**1.15E-110**	**1.71E-112**	18245381	Fetal hemoglobin levels	$7.0x10^{-35}$	
rs4910742	11:5306509	8.22E-02	4.60E-01	6.52E-01	**4.26E-05**	**9.19E-05**	18245381	Fetal hemoglobin levels	$1.0x10^{-21}$	
rs5006884	11:5373251	1.51E-01	**2.13E-03**	6.96E-02	**2.41E-02**	**1.95E-02**	20018918	Fetal hemoglobin levels	$3.0x10^{-08}$	(African)
rs766432	2:60719970	2.97E-01	7.07E-01	8.99E-01	**3.96E-127**	**1.83E-129**	20018918	Fetal hemoglobin levels	$1.0x10^{-29}$	(African)
rs7950726	11:5225447	1.58E-01	**7.46E-27**	**2.06E-14**	**4.08E-07**	**5.71E-07**	23043469	HbA2 levels	$1.0x10^{-11}$	
rs10224002	7:151415041	1.50E-01	4.39E-01	9.21E-01	2.86E-01	2.52E-01	19862010	Hematocrit	$6.0x10^{-15}$	
rs11065987	12:112072424	**3.57E-03**	1.18E-01	**1.87E-02**	**4.20E-02**	8.01E-02	19862010	Hematocrit	$1.0x10^{-12}$	
rs16926246	10:71093392	**1.16E-07**	2.10E-01	**8.98E-04**	5.16E-01	7.65E-01	19862010	Hematocrit	$1.0x10^{-13}$	
rs7385804	7:100235970	**4.12E-02**	8.27E-01	1.06E-01	2.72E-01	3.12E-01	19862010	Hematocrit	$4.0x10^{-10}$	

rs9483788	6:135435501	4.42E-01	**1.06E-15**	**1.61E-10**	**8.62E-72**	**1.41E-73**	19862010	Hematocrit	$3.0x10^{-15}$	
rs10168349	2:46360907	**1.15E-02**	8.31E-01	1.23E-01	3.51E-01	5.07E-01	20139978	Hematological and biochemical traits	$5.0x10^{-07}$	(Ht)
rs10495928	2:46353166	**1.12E-02**	7.73E-01	1.11E-01	4.00E-01	5.65E-01	20139978	Hematological and biochemical traits	$5.0x10^{-06}$	(Hb)
rs10518733	15:53940307	5.49E-01	7.51E-01	6.20E-01	6.66E-01	6.58E-01	20139978	Hematological and biochemical traits	$2.0x10^{-08}$	(sCr)
rs1077834	15:58723479	2.52E-01	1.25E-01	4.54E-01	8.64E-01	9.17E-01	20139978	Hematological and biochemical traits	$1.0x10^{-14}$	(HDL)
rs11709625	3:66823157	2.16E-01	9.78E-01	4.91E-01	3.22E-01	3.00E-01	20139978	Hematological and biochemical traits	$1.0x10^{-10}$	(BUN)
rs1260326	2:27730940	6.83E-01	8.65E-01	5.96E-01	7.67E-01	7.09E-01	20139978	Hematological and biochemical traits	$4.0x10^{-09}$	(ALB)
rs12678919	8:19844222	9.34E-01	4.66E-01	5.97E-01	5.91E-01	6.26E-01	20139978	Hematological and biochemical traits	$9.0x10^{-06}$	(HDL)
rs2242420	1:21904529	6.95E-01	5.59E-01	6.42E-01	6.37E-01	6.37E-01	20139978	Hematological and biochemical traits	$5.0x10^{-13}$	(ALP)
rs2280401	19:50000009	3.23E-01	5.92E-02	2.99E-01	5.57E-02	**4.27E-02**	20139978	Hematological and biochemical traits	$3.0x10^{-08}$	(TP)
rs2896019	22:44333694	2.47E-01	7.24E-01	6.25E-01	2.60E-01	2.86E-01	20139978	Hematological and biochemical traits	$2.0x10^{-12}$	(ALT)
rs3764261	16:56993324	4.36E-01	9.84E-01	7.07E-01	9.67E-01	8.75E-01	20139978	Hematological and biochemical traits	$5.0x10^{-29}$	(HDL)
rs3782886	12:112110489	NA	NA	NA	NA	NA	20139978	Hematological and biochemical traits	$5.0x10^{-09}$	(ALT)
rs4273077	17:16849139	6.67E-01	4.87E-01	2.30E-01	5.25E-01	4.71E-01	20139978	Hematological and biochemical traits	$3.0x10^{-10}$	(TP)
rs4890568	18:43231622	4.30E-01	6.85E-01	3.45E-01	4.85E-01	3.88E-01	20139978	Hematological and biochemical traits	$2.0x10^{-10}$	(BUN)
rs495828	9:136154867	**9.47E-03**	6.69E-01	4.22E-01	6.80E-01	8.12E-01	20139978	Hematological and biochemical traits	$4.0x10^{-59}$	(ALP)
rs5751902	22:24996630	1.88E-01	1.67E-01	7.74E-01	6.33E-01	5.95E-01	20139978	Hematological and biochemical traits	$8.0x10^{-20}$	(GGT)
rs5756504	22:37467270	9.90E-02	7.93E-01	1.73E-01	3.28E-01	2.71E-01	20139978	Hematological and biochemical traits	$2.0x10^{-10}$	(Hb)
rs671	12:112241766	NA	NA	NA	NA	NA	20139978	Hematological and biochemical traits	$5.0x10^{-09}$	(GGT)
rs6911965	6:24480295	3.36E-01	4.04E-01	8.32E-01	4.41E-01	5.16E-01	20139978	Hematological and biochemical traits	$2.0x10^{-11}$	(ALP)
rs7136716	12:7691134	5.91E-01	7.15E-01	9.12E-01	3.26E-01	2.55E-01	20139978	Hematological and biochemical traits	$3.0x10^{-26}$	(CK)
rs7173947	15:95270467	8.61E-02	5.36E-01	2.41E-01	8.38E-01	7.74E-01	20139978	Hematological and biochemical traits	$3.0x10^{-08}$	(ALP)
rs7350481	11:116586283	2.44E-01	9.52E-02	4.31E-01	7.31E-01	8.75E-01	20139978	Hematological and biochemical traits	$9.0x10^{-10}$	(HDL)
rs9820070	3:187687074	9.51E-01	9.52E-01	9.07E-01	5.07E-02	**4.29E-02**	20139978	Hematological and biochemical traits	$1.0x10^{-11}$	(BUN)
rs11066301	12:112871372	5.95E-02	8.74E-02	**3.59E-02**	1.84E-01	2.65E-01	19820697	Hematological parameters	$8.0x10^{-12}$	(PLT)
rs11970772	6:41925290	4.53E-01	2.44E-01	2.84E-01	6.07E-01	6.12E-01	19820697	Hematological parameters	$7.0x10^{-19}$	(MCV)
rs17609240	17:38110689	3.11E-01	4.69E-01	2.81E-01	3.18E-01	3.11E-01	19820697	Hematological parameters	$9.0x10^{-09}$	(WBC)
rs2227139	6:32413459	7.74E-01	4.94E-01	5.56E-01	1.30E-01	1.62E-01	19820697	Hematological parameters	$1.0x10^{-07}$	(WBC)
rs385893	9:4763176	1.98E-01	5.45E-01	2.62E-01	8.82E-01	9.76E-01	19820697	Hematological parameters	$9.0x10^{-17}$	(PLT)
rs4947019	6:110090049	8.22E-01	9.07E-02	1.14E-01	8.04E-01	7.33E-01	19820697	Hematological parameters	$8.0x10^{-06}$	(MCV)
rs5756506	22:37467392	1.03E-01	7.54E-01	1.65E-01	2.97E-01	2.45E-01	19820697	Hematological parameters	$1.0x10^{-09}$	(MCH)
rs9402686	6:135427817	4.30E-01	**1.56E-17**	**8.79E-12**	**1.47E-83**	**9.62E-86**	19820697	Hematological parameters	$7.0x10^{-42}$	(MCV)
rs9609565	22:32867528	2.71E-01	3.90E-01	2.99E-01	1.07E-01	1.11E-01	19820697	Hematological parameters	$4.0x10^{-10}$	(MCV)
rs1211375	16:240280	**1.11E-06**	**3.06E-13**	**2.37E-17**	2.52E-01	1.71E-01	23263863	Hematology traits	$7.0x10^{-14}$	(RBC)
rs174548	11:61571348	7.56E-01	7.33E-01	7.77E-01	9.51E-01	9.58E-01	23303382	Hematology traits	$4.0x10^{-08}$	(ALB/GLB)
rs2213169	11:5303063	NA	NA	NA	NA	NA	23263863	Hematology traits	$5.0x10^{-11}$	(AA, HCT)
rs2340727	1:161946727	5.49E-02	1.80E-01	8.67E-01	5.95E-01	6.88E-01	23263863	Hematology traits	$2.0x10^{-23}$	(WBC)
rs3130544	6:31058340	4.86E-01	3.88E-01	7.91E-01	6.70E-01	7.27E-01	23263863	Hematology traits	$5.0x10^{-07}$	(WBC)

rs3761944	1:230388989	8.64E-02	4.72E-01	1.29E-01	8.85E-01	9.58E-01	23303382	Hematology traits	$8.0x10^{-06}$	(ALB/GLB)
rs389884	6:31940897	3.05E-01	8.19E-01	8.73E-01	1.61E-01	2.30E-01	23263863	Hematology traits	$2.0x10^{-08}$	(EA, WBC)
rs4561508	17:16848750	6.76E-01	4.80E-01	2.21E-01	5.50E-01	4.95E-01	23303382	Hematology traits	$8.0x10^{-24}$	(ALB/GLB)
rs4657616	1:158971086	3.67E-01	8.39E-01	5.75E-01	9.92E-01	9.38E-01	23263863	Hematology traits	$5.0x10^{-47}$	(AA, WBC)
rs5987027	X:154014107	NA	NA	NA	NA	NA	23263863	Hematology traits	$1.0x10^{-11}$	(AA, RBC)
rs5987027	X:154014107	NA	NA	NA	NA	NA	23263863	Hematology traits	$1.0x10^{-11}$	(AA, RBC)
rs6477998	9:115997249	7.96E-01	5.07E-01	8.01E-01	8.19E-01	8.77E-01	23303382	Hematology traits	$8.0x10^{-06}$	(ALB/GLB)
rs6927090	6:252145	3.83E-01	3.64E-01	2.95E-01	8.18E-01	8.98E-01	23303382	Hematology traits	$4.0x10^{-06}$	(ALB/GLB)
rs7203560	16:184390	NA	NA	NA	NA	NA	23263863	Hematology traits	$2.0x10^{-08}$	(AA, HGB)
rs855791	22:37462936	**1.36E-02**	**3.76E-02**	**1.43E-03**	**2.88E-02**	**1.54E-02**	23263863	Hematology traits	$3.0x10^{-08}$	(HGB)
rs9399137	6:135419018	1.58E-01	**4.76E-19**	**7.67E-12**	**4.07E-91**	**1.10E-93**	23263863	Hematology traits	$3.0x10^{-06}$	(RBC)
rs6013509	20:51318351	8.51E-01	2.37E-01	4.07E-01	4.45E-01	4.80E-01	19862010	Hemoglobin	$1.0x10^{-10}$	
rs10758658	9:4856877	2.34E-01	4.39E-01	7.06E-01	4.97E-01	5.19E-01	19862010	Mean corpuscular hemoglobin	$2.0x10^{-14}$	
rs10815094	9:4845520	4.46E-01	3.66E-01	5.36E-01	4.03E-01	4.03E-01	23263863	Mean corpuscular hemoglobin	$5.0x10^{-06}$	
rs11085824	19:13001547	5.04E-01	3.72E-01	7.53E-01	**2.08E-03**	**3.36E-03**	19862010	Mean corpuscular hemoglobin	$1.0x10^{-11}$	
rs11915082	3:195809139	8.32E-01	7.65E-01	9.31E-01	**2.05E-02**	**1.66E-02**	19862010	Mean corpuscular hemoglobin	$8.0x10^{-13}$	
rs11966072	6:109634828	**3.83E-02**	4.52E-01	1.49E-01	2.87E-01	3.01E-01	20139978	Mean corpuscular hemoglobin	$1.0x10^{-08}$	
rs12127588	1:198595506	4.38E-01	7.21E-01	5.11E-01	1.98E-01	1.97E-01	20139978	Mean corpuscular hemoglobin	$7.0x10^{-10}$	
rs218237	4:55394172	8.94E-02	2.37E-01	8.29E-02	6.19E-02	8.88E-02	20139978	Mean corpuscular hemoglobin	$3.0x10^{-25}$	
rs2279434	10:45955064	1.10E-01	8.65E-02	**3.53E-02**	4.43E-01	3.34E-01	20139978	Mean corpuscular hemoglobin	$4.0x10^{-12}$	
rs2858942	16:225653	3.93E-01	2.13E-01	5.03E-01	9.15E-01	8.56E-01	20139978	Mean corpuscular hemoglobin	$3.0x10^{-09}$	
rs470119	22:50966914	9.57E-01	**1.65E-02**	**4.80E-02**	7.37E-01	8.34E-01	20139978	Mean corpuscular hemoglobin	$4.0x10^{-08}$	
rs4916483	3:195907653	8.45E-02	5.84E-01	2.51E-01	9.96E-01	8.75E-01	20139978	Mean corpuscular hemoglobin	$4.0x10^{-11}$	
rs628751	6:139838419	8.31E-02	**1.50E-03**	**8.29E-04**	7.53E-01	7.68E-01	19862010	Mean corpuscular hemoglobin	$1.0x10^{-17}$	
rs632057	6:139834012	4.30E-01	**9.80E-04**	**2.60E-03**	5.70E-01	5.81E-01	20139978	Mean corpuscular hemoglobin	$1.0x10^{-09}$	
rs6494537	15:66051345	3.30E-01	1.14E-01	4.53E-01	3.42E-01	3.70E-01	20139978	Mean corpuscular hemoglobin	$3.0x10^{-09}$	
rs668459	6:139835689	3.29E-01	**8.07E-04**	**1.65E-03**	4.33E-01	4.26E-01	23263863	Mean corpuscular hemoglobin	$9.0x10^{-09}$	
rs7085433	10:51593354	4.67E-01	2.57E-01	6.41E-01	7.90E-01	7.88E-01	20139978	Mean corpuscular hemoglobin	$6.0x10^{-10}$	
rs7775698	6:135418635	2.47E-01	**6.25E-19**	**3.53E-12**	**5.74E-88**	**8.84E-91**	23263863	Mean corpuscular hemoglobin	$4.0x10^{-13}$	(EA)
rs9349205	6:41925159	3.62E-01	**8.49E-09**	**1.77E-06**	5.25E-02	**4.41E-02**	19862010	Mean corpuscular hemoglobin	$8.0x10^{-20}$	
rs1541252	1:203651927	**2.61E-03**	8.13E-01	2.76E-01	5.62E-01	6.65E-01	23263863	Mean corpuscular hemoglobin concentration	$9.0x10^{-08}$	
rs2266928	16:580124	**2.20E-06**	**1.65E-10**	**3.26E-14**	6.98E-01	6.24E-01	23263863	Mean corpuscular hemoglobin concentration	$3.0x10^{-11}$	
rs4580814	5:1113244	1.53E-01	6.97E-01	3.90E-01	2.81E-01	2.99E-01	20139978	Mean corpuscular hemoglobin concentration	$5.0x10^{-10}$	
rs6684514	1:156255456	6.73E-01	7.88E-01	9.96E-01	3.28E-01	3.12E-01	20139978	Mean corpuscular hemoglobin concentration	$3.0x10^{-09}$	
rs8176746	9:136131322	1.17E-01	1.30E-01	7.55E-02	5.62E-01	4.88E-01	20139978	Mean corpuscular hemoglobin concentration	$4.0x10^{-08}$	
rs837763	16:88853729	6.89E-02	7.49E-01	2.14E-01	6.27E-01	6.46E-01	20139978	Mean corpuscular hemoglobin concentration	$4.0x10^{-13}$	
rs11204538	1:248046272	6.89E-02	9.07E-01	3.36E-01	4.23E-01	3.82E-01	20139978	Mean corpuscular volume	$2.0x10^{-08}$	
rs11239550	10:46024729	2.30E-01	7.85E-01	7.28E-01	4.95E-01	5.93E-01	19862010	Mean corpuscular volume	$1.0x10^{-10}$	

rs12718597	7:50428428	8.85E-01	4.93E-01	4.37E-01	**1.38E-02**	**1.76E-02**	19862010	Mean corpuscular volume	$5.0x10^{-13}$	
rs131794	22:50971752	7.48E-01	**6.93E-04**	**6.05E-03**	6.82E-01	6.16E-01	19862010	Mean corpuscular volume	$1.0x10^{-15}$	
rs172629	4:55407762	7.76E-02	2.48E-01	8.02E-02	5.79E-02	8.52E-02	20139978	Mean corpuscular volume	$1.0x10^{-27}$	
rs198846	6:26107463	**8.11E-03**	**1.24E-02**	**7.22E-04**	8.55E-01	9.48E-01	20139978	Mean corpuscular volume	$2.0x10^{-06}$	
rs2236496	9:4844265	4.36E-01	3.54E-01	5.33E-01	4.37E-01	4.35E-01	23263863	Mean corpuscular volume	$2.0x10^{-07}$	
rs2540917	2:60608759	9.44E-01	7.92E-01	9.56E-01	**1.49E-15**	**2.09E-15**	19862010	Mean corpuscular volume	$1.0x10^{-14}$	
rs3218097	6:41905275	4.77E-01	**5.32E-07**	**5.88E-05**	7.00E-02	5.94E-02	23263863	Mean corpuscular volume	$9.0x10^{-08}$	
rs4466998	14:65475540	4.38E-01	1.45E-01	4.19E-01	4.50E-01	5.17E-01	19862010	Mean corpuscular volume	$5.0x10^{-08}$	
rs4821112	22:21964761	5.79E-01	8.09E-01	9.41E-01	4.14E-01	4.29E-01	20139978	Mean corpuscular volume	$1.0x10^{-08}$	
rs6092477	20:55991695	3.61E-01	9.75E-01	7.41E-01	**7.64E-03**	**9.23E-03**	20139978	Mean corpuscular volume	$1.0x10^{-08}$	
rs643381	6:139839423	8.31E-02	**1.50E-03**	**8.29E-04**	7.54E-01	7.68E-01	19862010	Mean corpuscular volume	$5.0x10^{-25}$	
rs7189020	16:304803	9.35E-02	**9.32E-03**	**2.00E-03**	2.59E-01	2.43E-01	19862010	Mean corpuscular volume	$2.0x10^{-12}$	
rs7255045	19:12932269	3.36E-01	9.31E-02	4.63E-01	**4.51E-04**	**6.41E-04**	19862010	Mean corpuscular volume	$2.0x10^{-12}$	
rs7786877	7:100214015	1.95E-01	7.17E-01	2.06E-01	9.84E-01	9.74E-01	19862010	Mean corpuscular volume	$3.0x10^{-11}$	
rs7843479	8:21820813	7.31E-01	5.00E-01	6.95E-01	5.44E-01	4.86E-01	20139978	Mean corpuscular volume	$3.0x10^{-08}$	
rs9374080	6:109616420	5.48E-02	1.32E-01	**3.50E-02**	4.62E-01	4.94E-01	19862010	Mean corpuscular volume	$4.0x10^{-10}$	
rs9859260	3:195800547	6.29E-01	8.93E-01	8.15E-01	**1.63E-02**	**1.55E-02**	19862010	Mean corpuscular volume	$8.0x10^{-14}$	
rs2075671	7:100345106	9.59E-01	4.67E-01	7.86E-01	7.74E-01	8.01E-01	19862010	Other erythrocyte phenotypes	$1.0x10^{-09}$	(RBC)
rs857721	1:158612548	4.73E-01	4.70E-01	7.48E-01	8.21E-01	8.20E-01	19862010	Other erythrocyte phenotypes	$1.0x10^{-10}$	(MCHC)
rs9373124	6:135423209	7.80E-01	**2.92E-17**	**2.64E-12**	**3.55E-70**	**1.93E-72**	19862010	Other erythrocyte phenotypes	$7.0x10^{-14}$	(MCHC)
rs11611647	12:4333919	5.75E-01	9.36E-01	9.13E-01	6.35E-01	6.96E-01	20139978	Red blood cell count	$6.0x10^{-09}$	
rs2736100	5:1286516	7.93E-01	4.20E-01	5.37E-01	3.47E-01	3.45E-01	20139978	Red blood cell count	$3.0x10^{-08}$	
rs1008084	6:109626965	5.29E-02	1.25E-01	**3.25E-02**	5.15E-01	5.47E-01	23222517	Red blood cell traits	$6.0x10^{-26}$	(EA, MCH)
rs10159477	10:71099888	**3.93E-07**	5.02E-01	**6.74E-03**	3.10E-01	4.92E-01	23222517	Red blood cell traits	$4.0x10^{-20}$	(EA, Hgb)
rs10207392	2:111849659	7.90E-01	5.19E-01	4.41E-01	4.09E-01	4.13E-01	23222517	Red blood cell traits	$4.0x10^{-11}$	(EA, MCV)
rs10445033	16:88840462	**1.14E-02**	8.20E-01	2.27E-01	6.17E-01	5.84E-01	23222517	Red blood cell traits	$2.0x10^{-22}$	(EA, MCHC)
rs10480300	7:151406005	8.69E-02	6.52E-01	6.56E-01	5.47E-01	4.97E-01	23222517	Red blood cell traits	$8.0x10^{-15}$	(EA, Hgb)
rs1050828	X:153764217	NA	NA	NA	NA	NA	23696099	Red blood cell traits	$4.0x10^{-13}$	(RBC count)
rs10849023	12:4332478	2.50E-01	9.74E-01	6.69E-01	4.57E-01	4.74E-01	23222517	Red blood cell traits	$8.0x10^{-12}$	(EA, MCH)
rs11072566	15:76293971	**6.69E-03**	6.60E-01	1.23E-01	4.16E-01	4.64E-01	23222517	Red blood cell traits	$3.0x10^{-10}$	(EA, Hgb)
rs11104870	12:88829294	9.38E-01	6.82E-02	1.48E-01	6.06E-01	5.72E-01	23222517	Red blood cell traits	$6.0x10^{-11}$	(EA, RBCC
rs11190134	10:101282200	6.37E-01	7.01E-01	7.26E-01	9.58E-01	9.69E-01	23222517	Red blood cell traits	$1.0x10^{-10}$	(EA, MCH)
rs1122794	16:309155	8.26E-01	2.02E-01	3.84E-01	2.52E-01	2.97E-01	23446634	Red blood cell traits	$1.0x10^{-08}$	(MCHC)
rs11248850	16:163598	**5.04E-03**	**2.25E-09**	**1.33E-09**	3.87E-01	3.41E-01	23222517	Red blood cell traits	$6.0x10^{-23}$	(EA, MCH)
rs11627546	14:70365924	7.62E-01	4.96E-01	9.88E-01	8.06E-01	8.13E-01	23222517	Red blood cell traits	$1.0x10^{-09}$	(EA, MCV)
rs11717368	3:195834357	8.86E-01	4.09E-01	5.98E-01	**4.74E-02**	**4.82E-02**	23222517	Red blood cell traits	$7.0x10^{-19}$	(EA, MCH)
rs12530845	7:135329978	9.65E-01	5.75E-02	1.17E-01	8.03E-01	8.20E-01	23222517	Red blood cell traits	$3.0x10^{-08}$	(MCH)
rs12718598	7:50428445	4.89E-01	**2.50E-03**	**1.49E-02**	**1.81E-02**	**2.76E-02**	23222517	Red blood cell traits	$2.0x10^{-13}$	(EA, MCV)

rs13008603	2:46355848	**2.94E-02**	4.55E-01	5.89E-01	2.07E-01	1.79E-01	23446634	Red blood cell traits	$4.0x10^{-09}$	(Ht, AA)
rs13335629	16:310380	NA	NA	NA	NA	NA	23446634	Red blood cell traits	$3.0x10^{-23}$	(Hgb, AA)
rs13339636	16:298588	NA	NA	NA	NA	NA	23446634	Red blood cell traits	$2.0x10^{-34}$	(MCH, AA)
rs140522	22:50971266	9.00E-01	**7.32E-03**	**3.64E-02**	8.25E-01	7.72E-01	23222517	Red blood cell traits	$5.0x10^{-23}$	(EA, MCV)
rs1408272	6:25842951	6.70E-01	3.18E-01	4.43E-01	9.75E-01	9.61E-01	23222517	Red blood cell traits	$5.0x10^{-67}$	(EA, MCH)
rs1532085	15:58683366	5.84E-01	9.17E-02	8.55E-02	9.87E-01	9.56E-01	23222517	Red blood cell traits	$7.0x10^{-11}$	(Hgb)
rs17342717	6:25821770	5.40E-01	**4.45E-02**	1.04E-01	6.14E-01	6.79E-01	20927387	Red blood cell traits	$5.0x10^{-08}$	(MCH)
rs17616316	14:103822762	3.24E-01	6.73E-01	9.09E-01	1.55E-01	1.45E-01	23222517	Red blood cell traits	$8.0x10^{-11}$	(MCH)
rs1800562	6:26093141	9.59E-01	2.47E-01	2.63E-01	6.00E-01	6.84E-01	23446634	Red blood cell traits	$4.0x10^{-06}$	(Hgb, AA)
rs2032314	21:35354523	1.73E-01	1.36E-01	8.99E-01	6.65E-01	8.58E-01	23222517	Red blood cell traits	$8.0x10^{-10}$	(PCV)
rs2075672	7:100240296	**3.79E-02**	8.73E-01	1.14E-01	2.67E-01	3.07E-01	23222517	Red blood cell traits	$2.0x10^{-20}$	(EA, RBCC)
rs218238	4:55395024	6.54E-02	5.53E-01	2.58E-01	**2.03E-02**	**2.93E-02**	23222517	Red blood cell traits	$3.0x10^{-39}$	(EA, RBCC)
rs2413450	22:37470224	**1.31E-02**	**3.17E-02**	**1.56E-03**	**1.26E-02**	**6.38E-03**	23935956	Red blood cell traits	$4.0x10^{-07}$	(MCH)
rs243070	2:60620286	6.83E-01	3.47E-01	6.44E-01	**1.12E-08**	**1.53E-08**	23222517	Red blood cell traits	$4.0x10^{-13}$	(EA, MCV)
rs2572207	15:66070693	1.67E-01	8.76E-01	4.62E-01	7.05E-01	7.66E-01	23222517	Red blood cell traits	$3.0x10^{-09}$	(EA, MCV)
rs3184504	12:111884608	**5.55E-03**	2.12E-01	**4.58E-02**	**2.91E-02**	5.56E-02	23222517	Red blood cell traits	$4.0x10^{-19}$	(EA, Hgb)
rs3811444	1:248039451	8.72E-01	9.44E-01	9.47E-01	8.16E-01	6.49E-01	23222517	Red blood cell traits	$5.0x10^{-10}$	(EA, RBCC)
rs3892630	19:33181484	5.52E-01	3.12E-01	3.59E-01	4.25E-01	3.93E-01	23222517	Red blood cell traits	$1.0x10^{-10}$	(EA, MCV)
rs3916164	1:40069939	7.83E-01	8.51E-01	9.69E-01	7.16E-01	7.05E-01	23222517	Red blood cell traits	$3.0x10^{-10}$	(MCH)
rs4820268	22:37469591	**1.34E-02**	**4.87E-02**	**2.58E-03**	**1.46E-02**	**7.53E-03**	20927387	Red blood cell traits	$1.0x10^{-12}$	(MCHC)
rs4895441	6:135426573	5.83E-01	**4.73E-18**	**2.08E-12**	**3.33E-82**	**1.34E-84**	23935956	Red blood cell traits	$3.0x10^{-06}$	(MCV)
rs4953318	2:46355051	**2.01E-02**	7.38E-01	1.02E-01	7.88E-01	9.79E-01	23222517	Red blood cell traits	$3.0x10^{-19}$	(EA, PCV)
rs5749446	22:32880585	1.16E-01	9.07E-02	6.50E-02	2.99E-01	3.42E-01	23222517	Red blood cell traits	$3.0x10^{-13}$	(EA, MCH)
rs5754217	22:21939675	6.82E-01	8.85E-01	9.39E-01	3.22E-01	3.43E-01	23222517	Red blood cell traits	$9.0x10^{-10}$	(EA, MCV)
rs579459	9:136154168	**1.58E-02**	7.09E-01	4.78E-01	5.64E-01	6.79E-01	23222517	Red blood cell traits	$9.0x10^{-18}$	(EA, RBCC)
rs590856	6:139844429	2.14E-01	**1.32E-04**	**3.17E-04**	5.07E-01	5.21E-01	23222517	Red blood cell traits	$5.0x10^{-36}$	(EA, MCV)
rs6569992	6:135452152	3.11E-01	**1.18E-12**	**2.92E-08**	**2.89E-56**	**4.85E-58**	20927387	Red blood cell traits	$6.0x10^{-09}$	(RBC count)
rs6776003	3:141266493	9.02E-01	3.22E-01	2.96E-01	9.23E-01	9.13E-01	23222517	Red blood cell traits	$4.0x10^{-11}$	(EA, MCV)
rs7120391	11:5230907	NA	NA	NA	NA	NA	23696099	Red blood cell traits	$5.0x10^{-09}$	(MCHC)
rs7155454	14:65502239	2.57E-01	1.11E-01	4.86E-01	4.16E-01	4.95E-01	23222517	Red blood cell traits	$2.0x10^{-12}$	(EA, MCH)
rs7312105	12:2523355	**4.35E-02**	6.69E-01	2.19E-01	2.87E-01	3.22E-01	23222517	Red blood cell traits	$3.0x10^{-09}$	(EA, PCV)
rs737092	20:55990405	5.97E-01	2.58E-01	3.36E-01	1.28E-01	1.63E-01	23222517	Red blood cell traits	$4.0x10^{-13}$	(EA, MCV)
rs741702	19:13024250	8.31E-01	4.25E-01	6.77E-01	**4.28E-02**	5.68E-02	23222517	Red blood cell traits	$8.0x10^{-20}$	(EA, MCH)
rs7529925	1:199007208	4.90E-01	**3.23E-02**	**2.19E-02**	**2.20E-03**	**1.51E-03**	23222517	Red blood cell traits	$8.0x10^{-09}$	(EA, RBCC)
rs762516	X:153764663	NA	NA	NA	NA	NA	23446634	Red blood cell traits	$2.0x10^{-16}$	(Ht, AA)
rs7776054	6:135418916	2.47E-01	**3.71E-19**	**2.52E-12**	**9.14E-88**	**1.38E-90**	23935956	Red blood cell traits	$4.0x10^{-06}$	(MCH)
rs857684	1:158575729	4.43E-01	3.86E-01	2.15E-01	4.86E-01	4.13E-01	23222517	Red blood cell traits	$4.0x10^{-16}$	(EA, MCHC)
rs901683	10:45966422	5.45E-02	1.65E-01	**4.99E-02**	9.13E-01	7.46E-01	23222517	Red blood cell traits	$2.0x10^{-16}$	(EA, MCV)

rs9310736	3:24350811	5.68E-01	**3.32E-02**	5.52E-02	1.93E-01	1.64E-01	23222517	Red blood cell traits	$6.0x10^{-16}$	(EA, MCV)
rs9349204	6:41914378	6.65E-01	**1.34E-06**	**1.22E-04**	1.06E-01	8.75E-02	23222517	Red blood cell traits	$2.0x10^{-40}$	(EA, MCV)
rs9386791	6:109608497	**4.66E-02**	2.85E-01	8.42E-02	4.41E-01	4.56E-01	23446634	Red blood cell traits	$1.0x10^{-08}$	(MCH, AA)
rs9389269	6:135427159	5.03E-01	**2.04E-17**	**7.32E-12**	**2.23E-83**	**1.17E-85**	23222517	Red blood cell traits	$3.0x10^{-19}$	(EA, MCV)
rs9494145	6:135432552	5.43E-01	**5.67E-16**	**7.74E-11**	**3.94E-72**	**3.81E-74**	20927387	Red blood cell traits	$3.0x10^{-15}$	(MCV)
rs9924561	16:314780	NA	NA	NA	NA	NA	23696099	Red blood cell traits	$5.0x10^{-29}$	(MCV)

Supplementary Table 7. Genome-wide significant signals using 1000 Genomes phase III as reference panel.
The table shows the association results at the genome-wide significant loci (all results are corrected for β^0 mutations observed in the *HBB* gene, see **Online Methods**). At each locus, we indicate the chromosome and genomic position (hg19 build), rs ID when available, effect allele tested for association (EA) and the other allele at the SNP (OA), the SNP effect allele frequency (EAF), the regression coefficients and imputation accuracy (RSQR).

SNP	rsID (dbsnp142)	chr:pos	Nearest Gene	EA / OA	EAF	p-value	Effect (StdErr)	RSQR	Type
HbA1 (g/dl)									
locus1 [1]	rs570013781	16:149539	*NPRL3*	A/G	0.136	$5.87x10^{-18}$	0.200 (0.023)	0.886	intron
	rs72759456	16:960792 (cond.)	*LMF1*	T/C	0.019	$1.76x10^{-8}$	-0.260 (0.046)	0.976	intron
locus2	rs5030868	X:153762634	*G6PD*	A/G	0.085	$2.78x10^{-11}$	-0.1256 (0.019)	Genotyped	exon
locus3 [2]	rs142385463	12:124632597 [2]	*ZNF664-FAM101A*	A/G	0.010	$1.87x10^{-08}$	-0.680 (0.121)	0.657	intron
HbA2									
locus1 (%)	rs35152987	11:5255582	*HBD*	A/C	0.004	$3.43x10^{-67}$	-2.483 (0.142)	0.987	missense
	rs56778108	11:5252628 (cond.)	*HBB-HBD*	G/T	0.006	$4.33x10^{-39}$	-1.27 (0.096)	0.872	intergenic
	rs12799332	11:5232243 (cond.)	*OR51V1-HBB*	G/C	0.181	$2.77x10^{-37}$	-0.241 (0.019)	0.999	intergenic
	rs11036338	11:5242698 (cond.)	*OR51V1-HBB*	C/G	0.376	$7.19x10^{-14}$	0.127 (0.017)	0.969	intergenic
	rs7936823	11:5250168 (cond.)	*HBB-HBD*	G/A	0.435	$1.89x10^{-10}$	0.111 (0.017)	0.751	intergenic
locus2 [1] (g/dl)	rs141494605	16:216593	*HBM*	C/T		$3.95x10^{-35}$	0.308 (0.025)	0.878	synonymous
	rs72759455	16:960158 (cond.)	*LMF1*	T/C	0.019	$1.27x10^{-08}$	-0.290 (0.051)	0.972	intron
locus3 (%)	rs13205177	6:41749829	*PRICKLE4*	A/G	0.126	$8.91x10^{-28}$	0.335 (0.031)	0.930	intron
locus4 (%)	rs7776054	6:135418916	*HBS1L-MYB*	G/A	0.210	$1.60x10^{-20}$	0.231 (0.025)	Genotyped	intergenic
locus5 (%)	rs73307913	20:44548301	*PLTP-PCIF1*	C/T	0.135	$1.83x10^{-09}$	-0.180 (0.030)	1.000	intergenic
locus6 (%)	rs141006889	16:88601281	*FOG1*	G/A	0.007	$5.33x10^{-09}$	-0.507 (0.087)	Genotyped	missense
HbF (g/dl)									
locus1	rs1427407	2:60718043	*BCL11A*	T/G	0.143	$2.37x10^{-120}$	0.662 (0.028)	Genotyped	intron
	rs7599488	2:60718347 (cond.)	*BCL11A*	T/C	0.376	$2.42x10^{-32}$	0.204 (0.017)	0.980	intron
locus2	rs9399137	6:135419018	*HBS1L-MYB*	C/T	0.205	$4.88x10^{-83}$	0.485 (0.025)	Genotyped	intergenic
	rs1854865	6:135360300 (cond.)	*HBS1L*	G/C	0.373	$1.10x10^{-11}$	-0.141 (0.021)	0.979	intron
locus3	rs67385638	11:5290370	*HBE1*	G/C	0.236	$1.96x10^{-24}$	0.242 (0.024)	0.988	intron
	rs1143541	11:5269330 (cond.)	*HBBP1-HBG1*	A/T	0.023	$1.48x10^{-11}$	0.403 (0.060)	0.865	intergenic
locus4 [3]	rs10031426 [3]	4:181734569	*LINC01098-LINC00290*	G/A	0.362	$1.78x10^{-08}$	-0.130 (0.023)	0.800	intergenic

[1] = association results locally corrected for the –α 3.7 deletion type I (NG_000006.1:g.34164_37967del3804) (see **Online Methods**).
[2] = rs142385463 and the novel variant identified using the SardiNIA cohort (chr12:123681790) represent the same signal: corrected p-value for 12:123681790 is indeed 0.03 when using rs142385463 as a covariate (see Locus 3 on **Supplementary Figure 3**).
[3] = association at SNP rs10031426 was not confirmed when imputing using the SardiNIA reference panel ($p=1.76x10^{-06}$), whereas it provided a greater imputation quality than the 1000Genomes reference panel (RSQR = 0.97 versus 0.80).

Supplementary Table 8. Evidence of relationships between suggestive and genome-wide significant loci.

The table shows results from GRAIL analysis[31] carried out for variants associated with p-values between $1.00x10^{-04}$ and $5.00x10^{-08}$, searching for their relationships with genome-wide associated loci in 3 different databases (see **Online Methods**). Bolded genes occur in more than one dataset report. Only multiple hypothesis-corrected p-values below 0.05 are reported.

Database	Gene	Corrected p-value	Associated traits
Pubmed (2006)			
	NFE2	$6.93x10^{-14}$	HbA2
	HBM	$1.73x10^{-10}$	HbA2
	HBB	$2.90x10^{-08}$	HbF
	SPTB	$1.45x10^{-07}$	HbA1
	FOG1	$5.10x10^{-03}$	HbA2
	KLF2	$5.49x10^{-03}$	HbA1
	MYB	$5.82x10^{-03}$	HbF
	ANK1	$5.83x10^{-03}$	HbA1
	PLA2G4E	$1.33x10^{-02}$	HbA1
	RGS7	$2.04x10^{-02}$	HbF
	ART4	$3.59x10^{-02}$	HbA2
	EVI2B	$4.36x10^{-02}$	HbA1
Gene Ontology[32]			
	ADGB	$7.36x10^{-13}$	HbA2
	HBD	$2.61x10^{-08}$	HbF
	***HBM**, HBQ1, HBZ*	$8.07x10^{-08}$	HbA2
	CYBA	$3.87x10^{-05}$	HbA2
	FADS2	$3.87x10^{-03}$	HbA2
	PSME4	$4.23x10^{-03}$	HbF
	RGS7	$1.12x10^{-02}$	HbF
	CARD11	$1.46x10^{-02}$	HbA2
	STARD13	$1.49x10^{-02}$	HbF
Gene Expression Atlas[33]			
	ANK1	$2.04x10^{-06}$	HbA1
	GLRX5	$9.98x10^{-06}$	HbA2
	SPTB	$3.04x10^{-05}$	HbA1
	CYBRD1	$9.22x10^{-05}$	HbF
	EPB42	$1.04x10^{-04}$	HbA1
	HBA1	$1.40x10^{-04}$	HbA2
	HBB	$2.11x10^{-04}$	HbF
	NFE2	$2.98x10^{-04}$	HbA2
	XPO7	$3.52x10^{-03}$	HbA2
	TFDP2	$3.91x10^{-02}$	HbA2
	RANBP3	$4.26x10^{-02}$	HbF
	RSAD2	$4.39x10^{-02}$	HbA1

Supplementary Table 9. Markers associated with suggestive p-values to a single trait and demonstrating genome-wide significant pleiotropic effects considering between-trait combined p-values.

The table shows markers with a between-trait combined p-value $<5x10^{-8}$. For each SNP, we report chromosome and position in build 37, the corresponding rs ID when available, the effect allele and the other, the effect allele frequency, the imputation accuracy, the effect size and its standard deviation and p-value for each hemoglobin, the between traits combined p-values, and the nearest gene with the indication of the SNP location. Pleiotropy of signals that are already genome-wide significant with one trait is reported in **Table 1**.

chromosome: position	dbSNP 142	EA/OA	EAF	RSQR	HbA1(g/dl)		HbA2(g/dl)		Combined p-value	Gene (SNP location)
					Effect (StdErr)	p-value	Effect (StdErr)	p-value		
12:125860716	rs534082394	A/G	0.006	0.959	-0.443 (0.083)	$9.57x10^{-08}$	-0.390 (0.092)	$2.51x10^{-05}$	$1.18x10^{-11}$	*TMEM132B* (intronic)
7:42298663	rs73098313	A/G	0.199	0.985	-0.068 (0.016)	$2.53x10^{-05}$	-0.090 (0.018)	$6.02x10^{-07}$	$1.05x10^{-10}$	GLI3 / *LOC101928795* (intergenic)
4:75545480	rs13137956	G/A	0.018	0.948	0.226 (0.048)	$2.57x10^{-06}$	0.230 (0.054)	$1.73x10^{-05}$	$1.94x10^{-10}$	*AREGB / BTC* (intergenic)
13:111100266	rs4132608	A/G	0.140	Genotyped	-0.078 (0.019)	$3.16x10^{-05}$	-0.096 (0.021)	$3.36x10^{-06}$	$5.79x10^{-10}$	*COL4A2* (intronic)
5:180097967	rs72821020	A/T	0.081	0.960	0.099 (0.024)	$3.84x10^{-05}$	0.120 (0.026)	$6.23x10^{-06}$	$1.21x10^{-09}$	*LOC100420514* / *FLT4* (intergenic)
1:158763983	rs857854	T/C	0.479	Genotyped	0.057 (0.013)	$1.21x10^{-05}$	0.058 (0.014)	$5.67x10^{-05}$	$2.85x10^{-09}$	*OR2AQ1P* / *OR6N2* (intergenic)
7:138935064	rs541069630	G/C	0.0098	0.984	-0.278 (0.067)	$3.35x10^{-05}$	-0.310 (0.074)	$3.08x10^{-05}$	$4.44x10^{-09}$	*UBN2* (intronic)
5:153856095	rs149968351	T/C	0.021	0.992	-0.175 (0.045)	$9.59x10^{-05}$	-0.219 (0.050)	$1.10x10^{-05}$	$5.37x10^{-09}$	*HAND1* (intronic)
4:117943991	na	A/G	0.020	0.963	-0.189 (0.046)	$4.28x10^{-05}$	-0.214 (0.052)	$3.54x10^{-05}$	$6.57x10^{-09}$	*TRAM1L1* / *MIR1973* (intergenic)
8:144627241	rs12543282	T/C	0.271	0.980	-0.062 (0.015)	$3.20x10^{-05}$	-0.067 (0.016)	$5.08x10^{-05}$	$6.81x10^{-09}$	*ZC3H3* (upstream, 3621bp)
6:158635446	rs36109109	T/G	0.095	0.994	0.090 (0.022)	$4.46x10^{-05}$	0.097 (0.024)	$7.50x10^{-05}$	$1.36x10^{-08}$	*TULP4* / *GTF2H5* (intergenic)
9:3676290	rs55798071	A/T	0.052	0.909	-0.124 (0.030)	$4.09x10^{-05}$	-0.131 (0.033)	$9.26x10^{-05}$	$1.51x10^{-08}$	*LOC101929302* (intronic)
5:111884585	rs6890241	T/C	0.208	0.998	-0.062 (0.016)	$8.53x10^{-05}$	-0.072 (0.018)	$4.99x10^{-05}$	$1.84x10^{-08}$	*LOC100286958* / *APC* (intergenic)
7:124913422	rs139891547	C/T	0.038	0.952	0.139 (0.035)	$5.90x10^{-05}$	0.153 (0.039)	$8.66x10^{-05}$	$2.08x10^{-08}$	*LOC101928283* (intronic)

Supplementary Table 10. Summary of Extended Association for $\beta^0$39.

The table describes phenotypes that show association at Bonferroni level with the $\beta^0$39 mutation. For each trait we show summary statistics for the distribution in the full cohort, as well as the association coefficients (effect size in standard deviation units and its standard error) for the T allele, and the p-value.

Trait	Trait distribution		Association with $\beta^0$39 (rs11549407)	
	Mean	Std. Dev.	Effect (StdErr)	p-value
Total Cholesterol (mg/dl)	208.4	42.3	-0.49 (0.050)	$7x10^{-22}$
LDL-Cholesterol (mg/dl)	126.6	35.4	-0.47 (0.051)	$1x10^{-20}$
Bilirubin, total(mg/dl)	0.69	0.39	0.39 (0.047)	$5x10^{-17}$
Bilirubin, fractioned (mg/dl)	0.13	0.08	0.29 (0.046)	$1x10^{-10}$
White Blood Cell Count (10^6/mmc)	6.674	1.680	0.24 (0.047)	$3x10^{-7}$
Neutrophil Count (10^6/mmc)	3.817	1.296	0.23 (0.048)	$1x10^{-6}$
Transferrin Levels (mg/dl)	307.4	62.9	-0.19 (0.047)	$6x10^{-5}$
Serum Fibrinogen(mg/dl)	329.2	67.6	-0.18 (0.045)	$7x10^{-5}$
HDL Cholesterol (mg/dl)	64.2	14.9	-0.18 (0.045)	$9x10^{-5}$
Platelets Count (10^3/mmc)	243.6	60.1	0.18 (0.047)	$9x10^{-5}$

Supplementary Table 11. Effects of genome-wide significant loci in males and females.

The table shows association in males and females separately, for the 23 lead variants. Last column shows p-value for heterogeneity of effects. None of the SNPs passed Bonferroni correction for significance in heterogeneity (0.002 = 0.05/23).

		Males			Females			Heterogeneity
SNPtype	EA/OA	Effect	StdErr	p-value	Effect	StdErr	p-value	p-value
HBA1								
$\beta^0$39 (rs11549407)	T/C	-1.769	0.052	$6.21x10^{-249}$	-1.642	0.049	$6.09x10^{-242}$	0.071
–α 3.7 deletion type I	del/wild-type	-0.317	0.038	$8.61x10^{-17}$	-0.281	0.036	$9.99x10^{-15}$	0.493
rs570013781	A/G	-0.326	0.041	$2.13x10^{-15}$	-0.248	0.038	$5.58x10^{-11}$	0.162
chr16:391593	T/C	-0.700	0.109	$1.15x10^{-10}$	-0.458	0.093	$9.63x10^{-7}$	0.092
rs5030868	A/G	-0.152	0.028	$5.72x10^{-8}$	-0.170	0.038	$6.46x10^{-6}$	0.700
chr12:123681790	A/C	-0.488	0.107	$4.71x10^{-6}$	-0.478	0.107	$8.81x10^{-6}$	0.947
HBA2								
$\beta^0$39 (rs11549407)	T/C	1.942	0.050	$<1x10^{-300}$	1.945	0.046	$<1x10^{-300}$	1.000
rs35152987	A/C	-2.107	0.150	$8.15x10^{-45}$	-2.136	0.150	$5.75x10^{-46}$	0.887
rs7944544	T/G	-1.325	0.151	$2.05x10^{-18}$	-1.256	0.120	$8.77x10^{-26}$	0.717
rs12793110	T/C	-0.222	0.028	$1.14x10^{-15}$	-0.272	0.024	$5.65x10^{-31}$	0.169
rs11036338	C/G	0.153	0.024	$7.37x10^{-11}$	0.120	0.021	$1.20x10^{-8}$	0.296
rs7936823	G/A	0.083	0.022	$1.79x10^{-4}$	0.129	0.019	$1.32x10^{-11}$	0.113
–α 3.7 deletion type I	del/wild-type	-0.215	0.037	$6.17x10^{-9}$	-0.222	0.034	$7.46x10^{-11}$	0.889
rs141494605	C/T	-0.423	0.038	$1.02x10^{-28}$	-0.340	0.036	$4.41x10^{-21}$	0.113
chr16:391593	T/C	-0.726	0.104	$2.59x10^{-12}$	-0.520	0.088	$3.78x10^{-9}$	0.131
rs148706947	T/C	0.472	0.077	$8.01x10^{-10}$	0.257	0.070	$2.63x10^{-4}$	0.039
rs113267280	G/T	0.268	0.035	$2.35x10^{-14}$	0.271	0.033	$1.33x10^{-16}$	0.950
rs7776054	G/A	0.211	0.026	$1.22x10^{-15}$	0.148	0.024	$6.00x10^{-10}$	0.077
rs59329875	C/T	-0.135	0.033	$3.69x10^{-5}$	-0.123	0.028	$1.07x10^{-5}$	0.780
rs141006889	G/A	-0.528	0.133	$7.60x10^{-05}$	-0.453	0.124	$2.5x10^{-04}$	0.680
HbF								
$\beta^0$39 (rs11549407)	T/C	1.137	0.052	$8.99x10^{-107}$	1.159	0.048	$5.42x10^{-127}$	0.777
rs4671393	A/G	0.507	0.032	$8.46x10^{-56}$	0.626	0.028	$6.34x10^{-110}$	0.005
rs13019832	A/G	-0.179	0.023	$1.45x10^{-14}$	-0.222	0.020	$8.27x10^{-28}$	0.163
rs9399137	C/T	0.373	0.027	$5.42x10^{-44}$	0.440	0.025	$6.09x10^{-72}$	0.065
rs11754265	C/G	-0.146	0.028	$2.69x10^{-7}$	-0.145	0.025	$1.01x10^{-8}$	0.979
rs67385638	G/C	0.216	0.027	$9.44x10^{-16}$	0.202	0.024	$1.35x10^{-17}$	0.696
rs2855122	C/T	-0.110	0.031	$3.28x10^{-4}$	-0.123	0.027	$6.97x10^{-6}$	0.752
rs183437571	T/C	0.555	0.121	$4.32x10^{-6}$	0.397	0.102	$1.05x10^{-4}$	0.318